IMPACT OF SMOKER STATUS ON MORTALITY RISKS

by Chessman Wekwete

Presented at the Actuarial Society of South Africa’s 2011 Convention
8–9 November 2011, Sandton Convention Centre, Johannesburg

ABSTRACT
This paper provides results on the experience of a block of underwritten, assured lives in South Africa, with respect to differences between the profiles and the mortality risk of smokers and non-smokers. Conclusions as well as outstanding issues are presented in the exploration of questions relating to the setting of mortality rates for pricing and for other purposes. Comparisons are shown with the results of assured lives and general populations in both South Africa and the United Kingdom (UK).

CONTACT DETAILS
Chessman Wekwete, Senior Pricing and Research Actuary, Hannover Life Reassurance Africa Limited, P.O. Box 85321, Emmarentia 2029, South Africa
Email: chessman@hannover-re.co.za
Tel: +27(0)11 481 6750
1. INTRODUCTION
Smoker status is an established rating factor which is used in the pricing of life assurance risks. Along with age, gender, socio-economic class and duration since underwriting, smoker status is an important indicator of the magnitude of the risk when pricing death, dread disease as well as disability covers.

A number of questions arise in life assurance pricing that relate to the risk presented by smoking. Using mortality as an example, it may be required to know the following:
— Is there a difference between the mortality rates of smokers and non-smokers?
— If this difference exists, what is the magnitude?
— Does the level of the difference depend on policyholder characteristics such as age, gender and socio-economic class?
— Does the impact of smoking differ by product type e.g. ‘accidental death’ cover only or ‘all cause mortality’ covers?
— What assumptions should be used to obtain ‘aggregate’ rates from separate non-smoker and smoker rates?

The last question in the above list requires an understanding of the prevalence of smoking in groups of people. The remainder of the questions relate to the relative risk of smoking and the impact on the insurance risk of the policyholder. Similar questions arise in the pricing of critical illness and disability products.

The use of an ‘aggregate’ rate is required when an insurer is unable or unwilling to charge different rates for smokers and non-smokers. All else being equal, this implies that non-smokers, who most likely present a lower mortality risk, pay the same premium as smokers (i.e. cross subsidies exist). If the insurer has reliable ‘smoker’ and ‘non-smoker’ mortality rates, then it is important to use the correct prevalence assumptions when combining these two sets of rates to obtain a single, aggregate rate.

Differentiated pricing, using smoker status, occurs when an insurance company charges different rates for smokers and non-smokers. Each group of lives is charged separately for their insurance costs and there is no cross-subsidy between these groups. In this case, developing a good understanding of the relative risk of smoking and the impact on insurance risks is more important than knowing and understanding the prevalence patterns.

Section 2 provides a summary of the various data sets and literature results that is used in the paper. It outlines some of the similarities and differences observed between the data sets that must be borne in mind when interpreting the results.

The prevalence of smoking from the Hannover Life Re Africa (HLRA) data is analysed for a number of different sub-populations in the data, in Section 3. The prevalence results are compared to the results from the other data sets.

Section 4 contains the analysis of the impact of smoker status on the cause of claims observed in the HLRA data. This provides insight into factors that influence how smoker status impacts on the mortality risk.
Section 5 considers the crude mortality rates of smokers and non-smokers. The significance of the impact of smoking on these rates is assessed using an analysis of the ratio of these rates and the associated confidence intervals. Generalised linear models (GLM) are employed to isolate the influence of smoking from that of confounding risk factors using the HLRA data. The analysis is done for ‘all cause’ mortality as well as for ‘natural cause’ and ‘accidental cause’ mortality separately.

2. DATA AND LITERATURE SOURCES

2.1 Hannover Life Re Africa data (HLRA)

This data relates to medically underwritten lives in South Africa that are reassured by Hannover Life Re Africa during the period from January 2005 to December 2009. The underlying policies are mortality policies and includes a combination of ‘new generation’ (risk only) and ‘old generation’ (risk and investment) policies. The policy profile is skewed towards wealthier policyholders with less than 10% of the data relating to the least and second least wealthy socio-economic classes. The data relates to different life assurance companies and no individual company contributes more than 35% to the total data set.

This paper is mainly based on data from this source and it is a particularly useful source since it is available to the author at a granularity that enables multivariable analyses to be done. The data enables the causes of death to be split between natural and accidental causes, which allows for a detailed analysis of the impact of smoking by these major causes. Comparisons are done with the results from the studies described below.

2.2 Continuous Statistical Investigations (CSI data)

The CSI data relates to the mortality study done for South African assured lives for the period from 1999 to 2002. This is the most recent study that is available and it relates to ‘old generation’ policies only. This study has a more balanced socio-economic profile than the HLRA data, with almost 30% of the data relating to the two least wealthy classes.

Calculations that are performed for this paper are based on selected tables of the detailed results that are published with the CSI report (Tables 5 to 8 in the spreadsheets accompanying the report). These tables only relate to mortality policies that do not have any accelerator riders. Data relating to ages of 20 and above is used in this work.

2.3 Continuous Mortality Investigations Bureau (CMIB data)

This data relates to the published results from the 1999 to 2003 experience of the United Kingdom’s life offices’ mortality. The data used in the calculations for this paper relates to both permanent and temporary (term) assured lives. Details on the number of claims and central exposures are published separately for males and females and for smokers and non-smokers, but only for exposure at ultimate durations.
2.4 South African Demographic and Health Survey (SADHS data)
This is a survey conducted during 2003 on 7,756 households across South Africa. Its stated aims include the enabling of tracking of the changes in the health status of the South African population, identification of risk factors and access and utilisation of key health services in the period from 1998 to 2003.

2.5 Health Survey for England (HSE data)
These results are based on a household survey conducted in England between January and December 2006. This is a general population survey that was done on 14,142 adults living in private households. Its stated objectives include estimating the proportion of people in England who have specified health conditions and estimating the prevalence of certain risk factors and combinations of risk factors associated with these conditions.

3. PREVALENCE OF SMOKING
3.1 Overall
HLRA estimates that 25% of the medically underwritten, assured population can be classified as smokers. Table 3.1 shows how the proportions differ by gender and how the HLRA values compare with the information from other published sources.

Table 3.1 Smoking prevalence for assured and general population lives

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASSURED LIVES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Medically underwritten lives in South Africa (Source: HLRA data)</td>
<td>28%</td>
<td>20%</td>
</tr>
<tr>
<td>b) Assured lives in South Africa (Source: CSI data)</td>
<td>34%</td>
<td>19%</td>
</tr>
<tr>
<td>c) Assured lives in the United Kingdom (UK) (Source: CMIB data)</td>
<td>21%</td>
<td>19%</td>
</tr>
<tr>
<td><strong>POPULATION (Above age 20)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) South African population (Source: SADHS data)</td>
<td>35%</td>
<td>10%</td>
</tr>
<tr>
<td>b) United Kingdom (UK) population (Source: HSE data)</td>
<td>24%</td>
<td>21%</td>
</tr>
</tbody>
</table>

3.1.1 Assured lives
The HLRA proportion of smokers for males is lower than that observed in the CSI data. However the proportion of smokers for both these South African assured lives data sets is higher than that observed for the UK assured males, as given in the CMI data. It is noted that when the HLRA data and the CSI data are restricted to the
wealthiest socio-economic class (as shown later) then all three assured lives data sets have a broadly similar smoking prevalence (between 21% and 25%).

HLRA female assured lives have a proportion of smokers that is similar to that observed in the CSI data. Both the HLRA and CSI female assured lives have a similar proportion of smokers to that observed in the UK assured female data.

Comparisons of the prevalence for males and females, in the South African data sets, show that female assured lives have a lower smoking prevalence than male assured lives. While the CMIB data also shows that females have a lower prevalence than males, this difference is not significant.

3.1.2 Comparison to the general population
A comparison of the HLRA male smoking prevalence to the South African male population shows that the HLRA data has a lower smoking prevalence. However the CSI male prevalence is similar to that observed in the general population. In the corresponding UK comparisons the population prevalence is also higher than the assured lives prevalence, but the difference is less marked.

South African female assured lives (both HLRA and CSI data sets) have a higher smoking prevalence than females in the general population. The SADHS data shows significant differences in the smoker proportions for different races. Differences between the observed proportions in the assured population and the general population may be explained by the differences in the race profiles of the two groups.

3.2 Age and gender
There are differences observed between the proportion of smokers for different age groups for both males and females in all the data sets.

3.2.1 Males
Figure 1 shows the proportion of smokers by age group for males. The proportion of smokers at younger ages is higher than at the older ages in assured lives in the HLRA data. In both the CSI and the South Africa population data the shape of the prevalence rates by age is different from the HLRA shape. The CSI and SADHS data sets both have peak prevalence in the 40 to 49 age group. This is in contrast with the HLRA data where the peak proportion of smokers is observed in the youngest (20 to 29) age group.

The smoking prevalence by age group for the UK assured lives has a shape which is inverted from the HLRA shape. The CMIB data shows the lowest prevalence in the 20 to 29 age group and highest prevalence at the older ages (60+).

3.2.2 Females
The smoking prevalence by age group for females is shown in Figure 2. The age shape of the prevalence for CSI and SADHS data sets are broadly similar to that observed for males. They, however, have peak prevalence over a broader age band of 40 to 59. An
inverted relationship between the age shapes of the HLRA and the CMIB data sets is also observed for females.

A feature of the South African general population female smoking prevalence is that it consistently has the lowest prevalence of all data sets for every age group. This is in contrast with the male prevalence where the SADHS prevalence data is the highest of all the data sets for all but the youngest age group. Again, this may be explained by differences in the race profiles between the assured and the general population. Another feature of the female smoker patterns by age group is that the overall variation of prevalence is lower than that observed for males.

3.3 Socio-economic class
Smoking prevalence for South African assured lives is also analysed by socio-economic class. Table 3.2 shows the overall prevalence by socio-economic class for the HLRA data and the CSI data for males. Class 1 refers to the wealthiest class while Class 4 relates to the least wealthy class.

The HLRA proportion of smokers, for male lives, in the least wealthy classes is almost double that of the wealthiest classes. It must be noted that in the CSI data, that has greater exposure to Class 3 and Class 4 lives than the HLRA data, the smoking prevalence of Class 3 and 4 lives is very similar to that of Class 2 lives. The proportions of smokers for female lives do not differ much between various socio-economic groups in the HLRA data. Details of the smoker profile by socio-economic class are not available for females in the published CSI data.

Figure 1 Prevalence of smoking by age group for males
3.4 Summary
The above analyses show that the different population subgroups exhibit different smoking patterns by age and gender. It is therefore important to understand the specific profile of the target market when doing a pricing exercise. Further, the commonly used assumption of a constant smoker prevalence of 30% may not always be appropriate, particularly for females or at the older ages.

4. IMPACT OF SMOKER STATUS ON CAUSES OF DEATH
The impact of smoker status on the mortality risk can partly be analysed by considering how the causes of death differ between smokers and non-smokers. Comparison of these differences provides an insight into which causes and risk factors are likely to have an impact on the relationship between smoking and mortality. The results of

---

**Table 3.2 Prevalence of smoking by socio-economic class**

<table>
<thead>
<tr>
<th>Socio-economic class</th>
<th>HLRA</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>24%</td>
<td>25%</td>
</tr>
<tr>
<td>Class 2</td>
<td>34%</td>
<td>33%</td>
</tr>
<tr>
<td>Class 3</td>
<td>40%</td>
<td>36%</td>
</tr>
<tr>
<td>Class 4</td>
<td>44%</td>
<td>34%</td>
</tr>
</tbody>
</table>

---

**Figure 2** Prevalence of smoking by age group for females
such comparisons cannot, on their own, provide evidence of the effect of smoking on mortality risks. An analysis of mortality rates is required to achieve this objective. However, analysis of differences in the causes of death by smoker status assists in identifying the main factors that need to be considered.

The HLRA data was used for the analysis by cause of death. Causes were unavailable or unknown for about 12% of all observed deaths. A greater proportion of unknown causes relates to older ages and to females. For the purposes of the following analyses, all deaths with unknown claim causes are excluded. The purpose of the analysis is to obtain a proportion of claims by various causes and since this is calculated at a gender and age specific level (albeit at a broad age group) it is concluded that excluding unknown causes will not unduly distort the results. A similar conclusion is also assumed for the results of Section 5 where the ratios of the crude mortality rates are considered (as opposed to the absolute level of the mortality rates).

The analysis of claim causes are done for males and females separately and by broad age groups.

4.1 Ages below 40
The distribution of deaths by cause for male assured lives younger than age 40 for both smokers and non-smokers are shown in Figure 3.

It is observed that accidental causes of death dominate for both smokers (83% of total deaths) and non-smokers (82% of total deaths). Consequently the proportion of natural deaths is low in both instances. This dominance of accidental causes of death is

![Figure 3 Claim cause proportions for assured males younger than age 40](image-url)
also observed for female lives (especially for female smokers) as seen in Figure 4. The conclusion derived from this is that if accidental causes are such an important cause of death and if smokers have a higher mortality risk than non-smokers then the impact of smoking must, in some way, be manifesting through the accidental death risk.

A feature that is observed for males in Figure 3 is that the proportions of the three broad causes of death are broadly similar for smokers and non-smokers. In other words, a difference in smoker status does not translate into a significant difference in the contribution of various causes of deaths. Therefore, should male smokers have an elevated mortality risk, then smoking (as a cause or as a proxy for other causes) may be associated with an increase of a similar magnitude for both the natural and the accidental risk components of the overall mortality risk.

Figure 4 shows that for females, the smokers have a higher proportion of deaths arising from accidental causes than is the case for non-smokers. This is unlike what is observed for males. Consequently any elevated risk of mortality in female smokers is more likely to be associated with accidental death risk than with natural death risk.

4.2 Ages above 40
Figure 5 shows the proportion of deaths by major cause for males older than age 40. For these ages, natural causes accounts for a higher proportion of deaths than accidental causes, for both smokers and non-smokers. The proportion of natural deaths observed is greater for smokers than for non-smokers. Analysis of the graph for females (not shown) shows that similar features apply to their causes of death.

Figure 4 Claim causes proportions for assured females younger than age 40
Any differences in the mortality rates between smokers and non-smokers are likely to be driven more by natural causes than by accidental causes. Therefore, at these ages, it appears that smoking influences mortality mainly through disease processes.

Conclusions arising out of the analysis above require that the impact of smoking on mortality should be considered separately for 'natural causes' and 'accidental causes'. In addition, the impact that age and/or gender plays in this risk should be investigated.

5. IMPACT OF SMOKER STATUS ON CLAIM RATES
The results of the analysis on the causes of death performed above make it reasonable to conclude that smoking has an impact on both the 'natural causes' mortality risk as well as on the 'accidental causes' mortality risk. The level and nature of these elevated risks are explored in this section by considering the crude mortality rates. This analysis employs statistical methods, including GLM, to achieve an understanding of how the mortality rates are affected by various factors. Such methods are particularly important where the factors interact with each other in how they impact on mortality.

This section illustrates some of the results obtained from these analyses.

5.1 ‘All cause’ mortality
The mortality rate of smokers is estimated at 169% of the non-smokers’ rate, based on the HLRA data. In order to construct a confidence interval around this ratio, it is assumed that the number of deaths observed given a specified exposure, is a Poisson random variable. Therefore the ratio of the smokers’ rate to the non-smokers rates’ is
a ratio of two rates, each with an underlying Poisson random variable. Consequently, using the results by Graham et al. (2003) (see Appendix A) a 95% confidence interval can be constructed around this ratio. This interval is calculated to be [156%, 184%] for the HLRA data. It is concluded that the mortality rate of smokers is significantly higher than that of non-smokers, since this interval is substantially above 100%.

A Poisson GLM is fitted to the data in order to determine whether this difference is entirely due to smoking or due to differences in other features of the profile of the assured lives. Should the other factors play a role, then the GLM enables the impact of smoking to be isolated from the other factors. The fitted model confirms that smokers have a higher mortality rate than non-smokers and that this extra risk depends on age. The model, however, shows no evidence that the extra risk differs by either gender or socio-economic class. The remainder of this section examines why the model does not classify gender as an important factor on how smoking impacts on mortality.

Many references, including Gorman and Read (2007) and Lew and Garfinkel (1987), conclude that smoking has a greater impact on the mortality of males and that smoking accounts for a substantial part of the gender gap in mortality rates. However, the question explored by the GLM fitted above is very specific and can be phrased as: For fully underwritten assured lives where the effect of age, gender, socio-economic class, select duration and smoker status is allowed for in the mortality rates, is the residual (excess) risk due smoking dependant on gender?

It is to this question that the model finds no evidence of a gender differential in the excess mortality risk due to smoker status. It is felt that this is a reasonable conclusion in light of the following observations.

The ratios of the smoker mortality rates to the non-smoker rates by age group for the HLRA data are shown in Figure 6 for both males and females. It can be seen from the graph that there are no significant differences between the male and female ratios, particularly for the ages 30 to 59 (which is where the greatest exposure lies). A 95% confidence band constructed around the male ratios would encompass the female ratios. The volatility of the female ratios is likely to be due to the paucity of the data for females at the youngest and oldest age groups.

The corresponding ratios observed for the CMIB data are provided in Figure 7. The ratios for males and females diverge more than for the HLRA data but show a generally similar pattern. Based on an analysis of the 95% confidence band around the male ratios it is also not possible to conclude that the ratios for females are different to those for males. It is interesting to note, based on the Lew and Garfinkel (1987) data, for lives in good health, that the ratios of the smoker mortality rates to the non-smoker rates for females also lie within the Graham et al. (2003) 95% confidence limits of the male ratios.

A comparison of the ratios for the CSI data, also shown in Figure 7, leads to the conclusion that the ratios by gender for this data set are significantly different. What cannot be determined from this analysis is whether this difference is due to gender
Figure 6 Ratio of smokers to non-smokers ‘all cause’ mortality rates by gender and age group for HLRA data

Figure 7 Ratio of smokers to non-smokers ‘all cause’ mortality rates by gender and age group for other data sets
only or whether it is also due to other differences in the profiles of the sub-populations. Fitting a GLM to this data would provide insight into the source of the difference.

5.2 Accidental causes
The accidental death rates for smokers and non-smokers are calculated separately for males and females. The overall ratios of the smoker rates to the non-smoker rates for accidental death claims are shown in Figure 8. In all cases, the ratio is greater than 100% indicating that the smoker rates are higher than the non-smoker rates. For both males and females combined, the estimate of this ratio is 158% with a 95% confidence interval of [140%; 178%]. It could also be interpreted from Figure 8 that the impact of smoking on accidental death rates is greater for females than males.

GLM modelling confirms that smokers have an ‘accidental’ mortality risk that is higher than for non-smokers. It is estimated that the elevation of risk is of magnitude 148% of the risk for non-smokers with 95% confidence interval of [133%; 165%], based on the data set used. The ratio of the crude rates of 158% estimated above is higher than the value of 148% estimated by the GLM. This is due to the fact that the crude estimate does not allow for any differences between the profile of smokers and non-smokers or of that between males and females as done by the GLM.

The GLM result, however, does not support the hypothesis that the ratio (i.e. the elevated accidental risk of death for smokers) is different between males and females. Further, there is no evidence from the model results to conclude that this elevation of risk depends on age. It is worth noting here that, despite these results, the substantial

Figure 8 Ratio of the smoker rates to the non-smoker rates for accidental deaths
difference in the ratios of mortality observed for ages below 30 in the HLRA data (see Figure 6) is also observed in the CMIB data (see Figure 7). A GLM analysis of the CMIB data could potentially resolve this anomaly.

5.3 Natural causes
The results of the GLM analysis of ‘natural cause’ mortality risk also show an elevated risk for smokers relative to non-smokers. However, unlike the ‘accidental cause’ risk, the risk elevation differs with age but does not depend on gender or socio-economic status. The modelling was inconclusive with respect to the level and nature of any dependence of the smoker risk on duration since underwriting.

5.3.1 Gender
Figure 9 shows the ratio of smokers ‘natural causes’ mortality rates to those for non-smokers, based on the HLRA data. This is shown separately for males and females. This indicates that the risk for smokers is generally higher than non-smokers.

The ratios in Figure 9 show that the impact is generally lower for females than for males and modelling results estimate that the extra risk for female smokers should be about 85% of the extra risk for male smokers. However and crucially, the model suggests that this difference is not statistically significant. As a result, it would be concluded that there is no evidence, based on the data analysed, to support the difference in the ratios for males and non-smokers observed in Figure 9. There was a detailed exploration of the profiles of the assured lives to determine the factors that

![Figure 9 Ratio of smoker rates to non-smoker rates by gender for natural deaths](image-url)
lead to the observation of the difference in Figure 9. The failure to identify such factors lead to the conclusion that the gender difference in the relative risk may exist but that it is not large enough to satisfy the measure of statistical significance.

5.3.2 Age
The use of the GLM allows us to interpret the results in Figure 9 in light of other factors like socio-economic class and duration since underwriting. Modelling confirms that an elevated risk of mortality due to smoking increases with increasing age. As an example, the ‘natural causes’ risk of a smoker aged 30 is approximately 150% that of a non-smoker of the same age. For lives aged 60, the ratio would be in excess of 250%.

While age is one factor where there is significant evidence of influence on the smoker risk, there is still a need to understand the level and nature of the influence more clearly. The fitted GLM could not inform whether the level tapers off at the older ages or whether it continues to increase. This needs to be explored in other data sets.

5.3.3 Other analyses
The analysis of whether socio-economic status or duration since underwriting has an influence on the risk due to smoking did not produce conclusive results. Reasons for this include the results being inconsistent with expectation and the statistical significance being marginal. The use of other data sets for similar analyses would most likely resolve these shortcomings.

6. Summary
The proportion of smokers differs significantly by the age and gender of assured lives. Different groups within the population, as well as assured lives, exhibit their own smoking patterns. It is therefore important to correctly assess the suitability and appropriateness of assumptions used for the prevalence of smoking.

The impact of smoking differs between the risk of accidental deaths and of natural deaths. This was shown for both the claim causes and the death rates. For accidental deaths, the impact of smoking impacts at the same level across all subgroups of the assured lives. This is after allowing for accidental death risk due to other factors like age and gender. The impact on natural deaths increases with increasing age. A GLM analysis of larger data sets is required to assess the age shape of the extra mortality risk that arises due to smoking.

Smoking should be appropriately allowed for in the pricing and valuation assumptions for individual life business. Equally analysis should be done to correctly understand the impact of other factors such as age, gender, duration since underwriting, socio-economic class and occupation class on mortality. The use of multivariable techniques like GLM on data sets like the CSI data and the CMIB data would greatly enrich the understanding of the influence of these important rating factors on mortality and on other insurance risks.
ACKNOWLEDGEMENTS
The author wished to acknowledge the assistance of Petro Appelo (Pricing Actuary, Hannover Life Re Africa) in the preparation of this paper.

REFERENCES
APPENDIX A

CONFIDENCE LIMITS FOR THE RATIO OF TWO RATES

Graham et al. (2003) states that for two independent Poisson variables \( X_0 \) and \( X_1 \) with corresponding parameters \( \gamma_0 \) and \( \gamma_1 \) then for

\[
\rho = \frac{\gamma_0}{\gamma_1}
\]

(the ratio of the two rates) then the estimate of the ratio is:

\[
\frac{x_1 n_0}{n_1 x_0}
\]

with a lower 95% confidence limit of and

\[
\frac{n_0}{n_1} \left\{ \frac{2 x_0 x_1 \cdot 1.96^2 - \sqrt{(1.96^2 (x_0 + x_1) \times (4 x_0 x_1 + 1.96^2 (x_0 + x_1)))}}{2 x_0^2} \right\}
\]

an upper 95% confidence limit of

\[
\frac{n_0}{n_1} \left\{ \frac{2 x_0 x_1 \cdot 1.96^2 + \sqrt{(1.96^2 (x_0 + x_1) \times (4 x_0 x_1 + 1.96^2 (x_0 + x_1)))}}{2 x_0^2} \right\}
\]

where \( x_0 \) and \( x_1 \) are the observed occurrences in samples of size and \( n_0 \) and \( n_1 \) respectively.