

How rapidly does the excess risk of lung cancer decline following quitting smoking? A quantitative review using the negative exponential model

Additional file 1 – Multistage

John S Fry, Peter N Lee^a, Barbara A Forey and Katharine Coombs

P N Lee Statistics and Computing Ltd, Sutton, Surrey, United Kingdom

^aCorresponding author

**Modelling the effect of quitting smoking
on the risk of lung cancer**

**The negative exponential model as an
approximation to the multistage model**

**Author : P.N. Lee
Date : 5th April 2013**

EXECUTIVE SUMMARY

The negative exponential model is a simple tool for summarizing published epidemiological data on quitting smoking and lung cancer using a single parameter, the half-life. Estimates of the half-life can be derived from blocks of data comparing risk in never smokers, current smokers, and quitters by grouped time of quit. These estimates can be combined by meta-analysis, and heterogeneity can be assessed by meta-regression. While goodness of fit to the model can be investigated by comparison of observed and fitted numbers of lung cancers in the smoking groups, its adequacy can be further assessed by comparing its predictions with those of the multistage model, used on a number of occasions to describe features of the relationship between smoking and lung cancer. This document reports the results of such a comparison.

We consider the scenario where smokers start at the age of S_1 and continue smoking at the same rate to age S_2 , and then either continue or quit. Using a plausible form of the multistage model involving five stages, where smoking affects the first and penultimate stages of the multistage process, and where the later stage effect is twice that of the early stage effect, and assuming that the relative risk from smoking is 10, we showed that the shapes of the decline in excess risk for the same half-life were quite similar for the negative exponential and multistage models. This was particularly true up to the half-life, the subsequent decline predicted by the negative exponential model being somewhat more rapid than that predicted by the multistage model. The negative exponential model also fitted the predictions quite well for alternative forms of the multistage model, in which we varied S_1 , S_2 , the number of stages, the relative risk from smoking and the relative effects on the first and penultimate stages. The estimated half-life for the multistage model varied little according to the assumed value of the relative risk from smoking or the assumed relative effects on the first and penultimate stages. The half-life tended to increase as the assumed duration of smoking in the population increased.

Since the multistage model predicts that the absolute risk of lung cancer remains relatively constant for a period after quitting, and since the declines in excess risk are similar for the two models, the predictions of the negative exponential model are not inconsistent with the observed approximate “freezing” of the absolute risk of lung cancer in quitters.

Especially as it is possible to test goodness-of-fit directly for the negative exponential model, and to test for variations in half-life estimates by study characteristics (e.g. studies of younger populations with shorter smoking durations might be expected to have shorter half-life estimates than studies of older populations with longer durations), the results reported here tend to support the use of the negative exponential model for summarizing published results from exponential studies of quitting and lung cancer.

Index

	<u>Page</u>
<u>Text</u>	
1. Introduction	1
2. Scenario of concern, terminology and notation	2
3. The negative exponential model	6
4. The multistage model	7
5. Comparing the predictions of the negative exponential model with those of the multistage model	11
<u>References</u>	14
<u>Tables</u>	
1. Rise in current smoker risk with age for a multistage model with 5 stages, with $d_1 = 3$, $d_4 = 6$ and $S_1 = 20$	15
2. Effect of varying the parameters of the multistage model on H (the time to half-life) and the goodness-of-fit of the negative exponential model to the predictions of the multistage model	16
<u>Figures</u>	
1. Scaled absolute risks based on multistage and negative exponential models	17
2. Relative excess risks in quitters (compared to current smokers)	

1. Introduction

There is abundant evidence that smoking is associated with an increased risk of lung cancer (Lee et al., 2012) and that this excess risk is reduced in quitters (see e.g. International Agency for Research on Cancer, 2007). However, despite extensive data which demonstrate an increasing benefit of quitting with increasing time quit, no simple model has been proposed which succinctly describes the time pattern following quitting. To be useful in practice it has to be (i) capable of being fitted reliably to data which are often presented for a very limited number of periods of quitting (e.g. 1-9, 10-19 and 20+ years), (ii) fit observed time patterns adequately, and ideally (iii) only involve a single parameter – thus allowing meta-analysis and tests of between-study heterogeneity using meta-regression techniques.

The model we have used is the negative exponential model. In this additional file, its predictions are compared to those of the multistage model, often used for detailed analysis of individual subject data on smoking and lung cancer (Brown and Chu, 1987; Lee, 1995; Whittemore, 1988), though the model is too complex to fit to epidemiological data as summarized in typical publications.

2. Scenario of concern, terminology and notation

This report is concerned with a scenario introducing three consecutive time periods, a period ($0 \rightarrow S_1$ years) during which none of the subjects smoke, a period ($S_1 \rightarrow S_2$ years) during which all ever smokers do so at a constant rate, and a period ($S_2 \rightarrow T$ years) during which quitters cease smoking and during which continuing smokers continue to smoke at the same rate.

We are interested in comparing risk at age T between those who continue to smoke, “**Continuers**”, those who stop smoking at time S_2 “**Quitters**”, and those who have never smoked, “**Never smokers**”. We are mainly interested in studying patterns of risk as age increases from S_2 .

In comparing risk between continuers, quitters and never smokers, it is important to have a clear understanding of the various measures of lung cancer risk.

“**Absolute risk**” (AR) is the probability that someone without lung cancer at the beginning of a period will develop lung cancer by the end of the period. This is usually expressed in terms such as risk per 100,000 per year.

“**Relative risk**” (RR) is the absolute risk in one group of subjects divided by that in another comparison group. Here unless otherwise stated, relative risk relates to the comparison group of never smokers. Thus, $RR_C = AR_C/AR_N$ and $RR_Q = AR_Q/AR_N$ where the subscripts C, Q and N refer to the three groups of interest, continuers, quitters and never smokers.

“**Excess risk**” (ER) is equal to relative risk minus 1, and relates to the increase in relative risk associated with the smoking history.

“**Relative excess risk**” (RER) is the ratio of excess risks in different smoking groups, and here is compared to that in continuers. Thus if, at a particular point in time following quitting, the relative risks for continuers, quitters and never smokers are, respectively, 20, 12 and 1, the relative excess risk in quitters compared to continuers is $(12-1) / (20-1) = 11/19 = 0.58$. The time after quitting when the relative excess risk reaches 0.50 is referred to as the “**half-life**” (H) when the negative exponential model is considered.

Note that evidence from numerous epidemiological studies of smoking and lung cancer (Doll and Peto, 1978; International Agency for Research on Cancer, 2007; Lee, 1995; Lee et al., 2012) makes it clear that:

1. The absolute risk in never smokers rises markedly with age.
2. The absolute risk in continuers (who continue to smoke at a constant level) also rises markedly with age.
3. The relative risk of continuing cigarette smokers (to never smokers) varies little by age, though may increase somewhat.
4. In quitters, their relative risk and excess risk, and also their relative excess risk compared to continuing cigarette smokers, declines with increasing time since quit. Eventually their absolute risk approaches that of never smokers.
5. In quitters, their absolute risk stays approximately constant for a time following quitting. It is not true to say that quitting reduces risk unless one makes it clear that one is talking of relative risk, excess risk or relative excess risk.

Below, the various abbreviations defined above are brought together for convenience.

- AR absolute risk
- C subscript used for continuers
- ER excess risk
- H the time to half-life (i.e. the time at which half of the achievable reduction in excess risk occurs)
- N subscript used for never smokers
- Q subscript used for quitters

RER	relative excess risk
RR	relative risk
S_1	age (years) when smoking starts (end of first period)
S_2	age (years) when quitters stop smoking (end of second period)
T	age (years) at the end of the third period
t	time (years) from end of second period (time quit in quitters)

Some additional abbreviations are used in the description of the multistage model (see section 4 for fuller details).

k	number of stages in the multistage process
n	number of cells at risk
i	subscript used for specific stage
p_i	transition probability for stage i
a_i	background rates of cell transformation (“transformation probabilities”)

- b_i transition probabilities when smoking ($b_i > a_i$ for stages affected by smoking, $b_i = a_i$ for unaffected stages)
- c_i transition probabilities for tobacco affected stages in relation to background ($c_i = b_i/a_i$, $c_i > 1$ for stages affected by smoking, $c_i = 1$ for unaffected stages)
- d_i increase in transition probabilities for affected stages relative to background ($d_i = c_i - 1$, $d_i > 0$ for stages affected by smoking, $d_i = 0$ for unaffected stages)
- y the relative stage effect, the ratio of the promoting to the initiating effect (taken as d_4/d_i in this work)

3. The negative exponential model

The negative exponential model proposed does not attempt to predict absolute risk, but only changes in relative risk, excess risk or relative excess risk following quitting. Its predictions do not depend upon the age at which smoking started (S_1) or stopped (S_2) but only relate to the time since stopping (t). Formally the relative excess risks in the third period are estimated by the equations

$$RER_C(t) = 1 \quad \text{and} \quad (1)$$

$$RER_Q(t) = \exp(-t(\log_e 2)/H) \quad (2)$$

where t is the time since quit.

$$\text{At } t = 0, \quad RER_Q(t) = 1$$

$$\text{At } t = H, \quad RER_Q(t) = 0.5$$

$$\text{At } t = zH, \quad RER_Q(t) = (0.5)^z$$

As t tends to infinity, $RER_Q(t)$ tends to zero

If one can assume that the relative risk of a continuer is constant over time the equation

$$ER_Q(t) = (RR_C - 1) \exp(-t(\log_e 2)/H) \quad (3)$$

will apply.

4. The multistage model

In the multistage model (see Lee, 1995 for a detailed discussion), a tissue for an individual is assumed to consist of a very large number, n , of identical cells. They all start in an untransformed stage (i.e. at stage 0) and cancer occurs when the first cell undergoes each of k successive independent transformations (i.e. reaches stage k). The transformations are (or may be) different, and agents affecting risk of cancer (such as smoking) may affect only one or two of them.

Providing that the transition probabilities ($p_1, p_2 \dots p_k$) are small and constant throughout life, the incidence rate, I_T , of cancer at age T will be given by the simple formula

$$I_T = BT^{k-1} \quad (4)$$

where B is a constant equal to $np_1p_2 \dots p_k / (k-1)!$

The incidence rate can also be calculated, again assuming that transition probabilities are small, for situations where the p_i are time-dependent. Thus for $k = 5$, the relevant formula is

$$I_T = np_5 \int_0^T p_4 \int_0^{t_4} p_3 \int_0^{t_3} p_2 \int_0^{t_2} p_1 dt_1 dt_2 dt_3 dt_4 dt_5 \quad (5)$$

Although it is theoretically possible to take account of any form of functional dependence of the transition probabilities on age, the most common use of the multistage model has been where transition probabilities are either unaffected by exposure and take “background” values which are invariant of age, or are affected by exposure, taking an increased value when exposure occurs.

In the work carried out here, attention is restricted following precedent (Brown and Chu, 1987; Lee, 1995) to forms of the multistage model in which only and first the penultimate stages are affected by smoking. Evidence that the decline in risk following quitting is reasonably rapid suggests a late stage (“promoting”) effect, though it is clear that the decline is not an instantaneous drop, as would occur if smoking affected the final stage. Evidence of a strong role of duration of smoking suggests an early stage (“initiating”) effect is also present.

For a five stage model, let us define the background rates of cell transformation (“transformation probabilities”) as a_1, a_2, a_3, a_4 and a_5 . While smoking, let us assume they become b_1, a_2, a_3, b_4 and a_5 .

We further define $c_i = b_i/a_i$ and $d_i = c_i - 1$, where d_i are the increases in transition probabilities relative to background, and $y = d_4/d_1$ as the relative stage effect, the ratio of the promoting to the initiating effect.

The formulae for the incidence rate can be simplified in our scenario. For a k stage process we first convert absolute risk to scaled absolute risk by dividing through by $a_1 a_2 \dots a_{k-1}$. The following equations can then be used to estimate the scaled absolute risks of time T .

Never smokers

$$AS_N = T^{k-1} \tag{6}$$

Continuers

$$AS_C = T^{k-1} + d_1(T - S_1)^{k-1} + d_{k-1}(T^{k-1} - S_1^{k-1}) + d_1 d_{k-1}(T - S_1)^{k-1} \tag{7}$$

Quitters

$$AS_Q = T^{k-1} + d_1((T - S_1)^{k-1} - (T - S_2)^{k-1}) + d_{k-1}(S_2^{k-1} - S_1^{k-1}) + .. \tag{8}$$

$$d_1 d_k (S_2 - S_1)^{k-1}$$

These formulae can then be used to calculate relative risks by dividing through by T^{k-1} .

It should be noted that formula (7) does not imply that the relative risk for current smokers is constant over time. As illustrated in Table 1, where k is set at 5, S₁ at 20, d₁ at 3 and d₂ at 6, the relative risk varies from 9.57 at age 50 to 13.62 at age 80.

In the following, where we define the relative risk from smoking (RR_c), this relates to the relative risk at time S₂. Given the relative stage effect $y = d_{k-1}/d_1$ and given RR_c, S₁ and S₂, d₁ can readily be determined from the quadratic equation

$$S_2^{k-1} + d_1(S_2 - S_1)^{k-1} + d_1y(S_2^{k-1} - S_1^{k-1}) + d_1^2y(S_2 - S_1)^{k-1} = RR_c S_2^{k-1} \quad (9)$$

Given d₁, and hence d_{k-1} = d₁y, the scaled absolute risk in continuers and quitters at time T can then be derived.

For the purpose of examining the effect of changes in the various parameters on the adequacy of the negative exponential model fit to the multistage model, a standard model was defined as follows:

age at starting to smoke	S ₁ = 20 years
age at quitting	S ₂ = 50 years
number of stages	k = 5
relative stage effect	y = 2
relative risk from smoking	RR _c = 10

Variants of the model were studied as follows:

age of starting to smoke	S ₁ = 15, 20, 25 years
age of quitting	S ₂ = 40, 50, 60 years
number of stages	k = 4, 5, 6
relative stage effect	y = 0.5, 1, 2
relative risk from smoking	R = 5, 10, 20

5. Comparing the predictions of the negative exponential model with those of the multistage model

Based on the main multistage model, Figure 1, plots the scaled absolute risks for never smokers, continuers and quitters from time S_2 in yearly increments of time up to 30 years after quitting, while Figure 2 plots the corresponding relative excess risks in quitters. Figure 1 shows that while the absolute risks for never smokers and continuers increase steadily with time, those for quitters are relatively unchanged. The Figures also show the half-life estimated from the multistage model, and superimposed on it the corresponding negative exponential curve with that half-life. As can be seen, the fit of the negative exponential model seems very good visually, up to the half-life, though there is some visible divergence subsequently, in that the negative exponential predictions decline somewhat more rapidly than those of the multistage.

As an index of goodness-of-fit, a statistic was calculated based on the area between the two curves from zero up to the half-life. The time to half-life is divided into 20 periods, the difference is then divided by the time to half-life to give relative scaling between various models being compared. Finally, the sum is subtracted from 1, so that values close to 1 represent a good fit. In Figure 2, the value of this statistic is 0.9753.

Table 2 shows how H and the goodness-of-fit statistic vary according to the different multistage parameters. Various conclusions can be drawn from this:

Variation in number of stages: H declines with increasing k . Inasmuch as if the multistage model applies to lung cancer, it will apply with a specific number of stages, it is not important that H varies with k .

Variation in S_1 and S_2 : H declines somewhat with increasing S_1 and increases somewhat with increasing S_2 . This is consistent with saying that effects of quitting are more rapid for smokers who have smoked for a small proportion of their lives. Since, in many studies, subjects will have started to smoke at about the same time on average, the facts that H declines with S_1

may not be important. One might, however, expect estimates of H to be lower in populations with shorter durations of smoking, such as younger populations.

Variation in y: H is little affected by varying y.

Variation in R: H declines slightly with increasing R, i.e. effects of quitting are proportionally somewhat more rapid for heavy smokers or in countries where relative risks from smoking are higher.

Goodness of fit: The statistic is relatively similar for all the models tested.

While it might be possible to find a single model that fits the predictions of the multistage model, slightly better than the negative exponential, the negative exponential still seems a good approximation. Given data from epidemiological studies on quitting are typically only presented for a small number of quit periods, and it is clearly not possible to reliably estimate multiple parameters from a more complex model, the negative exponential model has a number of advantages. Principally, these are that:

- (i) it is simple,
- (ii) it can be fitted to data from a study using available techniques (Lee et al., 2012b) and
- (iii) it is dependent on only a single parameter, H, so its estimates can easily be subject to meta-analysis and meta-regression.

Note also that the negative exponential model can also be extended to situations where, instead of quitting, subjects reduce the number of cigarettes smoked or switch to a lower risk product. Thus the relative excess risk in the third period (during which smokers have reduced or switched) may be written as

$$\text{RER}(t) = F + (1 - F) \exp(-t(\log_e 2)/H)$$

where F is 1 for continuers, 0 for quitters and an intermediate value for reducers or switchers. Additional analyses varying F (results not shown) showed that estimated half-lives for reducing exposure are (under the multistage model) virtually identical to those for quitting.

References

- Brown, C.C., Chu, K.C., 1987. Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. *J. Chronic Dis.* 40(Suppl 2), 171S-179S.
- Doll, R., Peto, R., 1978. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J. Epidemiol. Community Health* 32, 303-313.
- International Agency for Research on Cancer, 2007. IARC Handbooks of Cancer Prevention, Tobacco Control, Vol 11: Reversal of risk after quitting smoking. IARC, Lyon, France.
- Lee, P.N., 1995. Studying the relationship of smoking to lung cancer using the multistage model of carcinogenesis. A review. P N Lee Statistics and Computing Ltd, Sutton, Surrey. www.pnlee.co.uk/Reports.htm [Download LEE1995D]
- Lee, P.N., Forey, B.A., Coombs, K.J., 2012. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 12, 385.
- Whittemore, A.S., 1988. Effect of cigarette smoking in epidemiological studies of lung cancer. *Stat. Med.* 7, 223-238.

Table 1
Rise in current smoker relative risk with age for a
multistage model with 5 stages, with $d_1 = 3$,
 $d_4 = 6$ and $S_1 = 20$

<u>Age</u>	<u>Current smoker RR</u>
40	7.94
45	8.77
50	9.57
55	10.34
60	11.07
65	11.77
70	12.43
75	13.04
80	13.62

Table 2
Effect of varying the parameters of the multistage model on H
(the time to half-life) and the goodness-of-fit of the negative
exponential model to the predictions of the multistage model

Model	Model parameters					Estimate of half-life	Goodness of fit
	k	S ₁	S ₂	y	R		
1	5	20	50	2	10	8.45	0.9753
2	4					11.31	0.9746
1	5					8.45	0.9753
3	6					6.79	0.9772
4	5	15				9.17	0.9736
1		20				8.45	0.9753
5		25				7.89	0.9772
6		20	40			6.32	0.9715
1			50			8.45	0.9753
7			60			10.69	0.9784
8			50	0.5		8.69	0.9530
9				1		8.43	0.9699
1				2		8.45	0.9753
10					5	9.24	0.9737
1					10	8.45	0.9753
11					20	7.75	0.9763

Notes: Results for main model (1) repeated when varying each model parameter in order to allow ready comparison of results when varying a specific parameter. If the value of a model parameter is not shown for a model, it is assumed to be the same as that for the main model.

Figure 1.
Scaled absolute risks based on multistage and negative exponential models

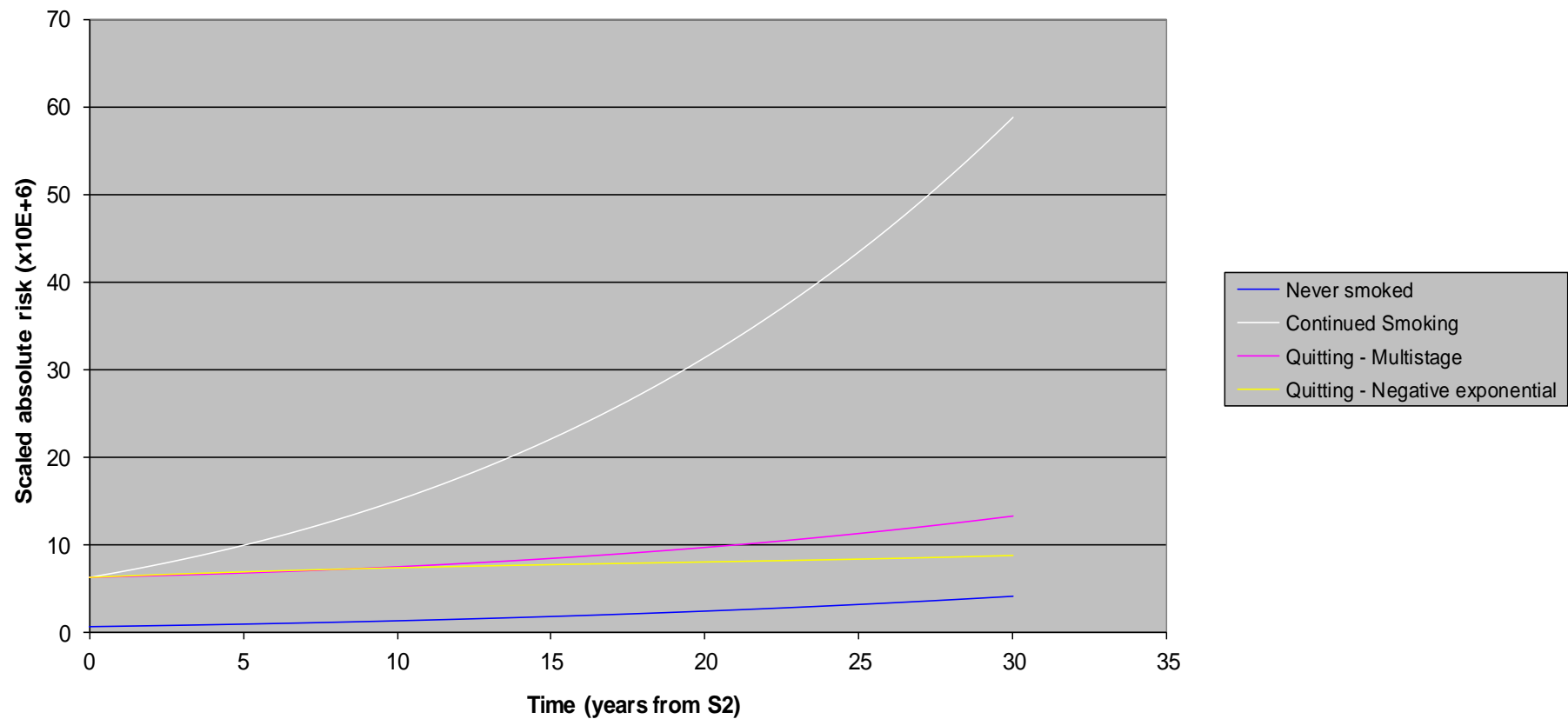


Figure 2.
Relative excess risks in quitters (compared to current smokers)
based on multistage and negative exponential models

