Regulatory Toxicology and Pharmacology 72 (2015) 49–57

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

# Is the shape of the decline in risk following quitting smoking similar for squamous cell carcinoma and adenocarcinoma of the lung? A quantitative review using the negative exponential model

John S. Fry, Peter N. Lee *, Barbara A. Forey, Katharine J. Coombs

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

ARTICLE INFO

*Article history:*
Received 29 October 2014
Available online 20 February 2015

*Keywords:*
Smoking
Lung cancer
Histological type
Cessation of smoking
Statistical models
Negative exponential model
Meta-analysis
Review

ABSTRACT

One possible contributor to the reported rise in the ratio of adenocarcinoma to squamous cell carcinoma of the lung may be differences in the pattern of decline in risk following quitting for the two lung cancer types. Earlier, using data from 85 studies comparing overall lung cancer risks in current smokers, quitters (by time quit) and never smokers, we fitted the negative exponential model, deriving an estimate of 9.93 years for the half-life – the time when the excess risk for quitters compared to never smokers becomes half that for continuing smokers. Here we applied the same techniques to data from 16 studies providing RRs specific for lung cancer type. From the 13 studies where the half-life was estimable for each type, we derived estimates of 11.68 (95% CI 10.22–13.34) for squamous cell carcinoma and 14.45 (11.92–17.52) for adenocarcinoma. The ratio of the half-lives was estimated as 1.32 (95% CI 1.20–1.46, $p < 0.001$). The slower decline in quitters for adenocarcinoma, evident in subgroups by sex, age and other factors, may be one of the factors contributing to the reported rise in the ratio of adenocarcinoma to squamous cell carcinoma. Others include changes in the diagnosis and classification of lung cancer.


## 1. Introduction

The US Surgeon General (2014) claimed that the risk of developing lung adenocarcinoma from smoking has increased since the 1960s, and concluded that this was because of changes in the design and composition of cigarettes since the 1950s. As support for their claims, they argued that though a big shift from squamous cell carcinoma (SqC) to adenocarcinoma (AdC) occurred in smokers, there had been no overall change in risk of all lung cancer or of adenocarcinoma in never smokers. They also argued that the shift is not explained by changes in diagnostic accuracy or in the intensity or duration of smoking. Though the discussion of this paper does include some comment on these claims, it is not the intention here to examine them here in full. Rather, attention is restricted to one possible contributor to a shift in the relative distribution of the main histological types of lung cancer. This relates to the fact that in the USA and many other countries many smokers, particularly men, have quit, so that, over time, the proportion of long-term quitters has increased. A difference in the pattern of decline in risk following quitting for SqC and AdC might therefore help to explain a change over time in the relative distribution of these two lung cancer types. The main objective of this paper, therefore, is to investigate this possibility.

In an earlier paper (Fry et al., 2013), using a database of epidemiological studies of at least 100 cases of lung cancer, and 106 blocks of relative risks (RRs) from 85 studies comparing current smokers, former smokers (by time quit) and never smokers, we estimated the half-life (H, time in years where the excess risk becomes half that for a continuing smoker) for each block by fitting the negative exponential model. In that paper, which concerned overall lung cancer risk regardless of type, we investigated model fit, studied heterogeneity in H, and conducted sensitivity analyses allowing for reverse causation, either by ignoring short-term quitters (S1) or considering them smokers (S2). Model fit was found to be poor ignoring reverse causation, but much improved for both sensitivity analyses. Overall estimates of *H* were similar for the main and sensitivity analyses, being estimated as 9.93 (95% CI 9.31–10.60) for the best-fitting sensitivity analysis (S1), though varying by sex (females 7.92, males 10.71), and age (<50 years 6.98, 70+ years 12.99).

*Abbreviations:* AdC, adenocarcinoma; CI, confidence interval; H, half-life; K1, Kreyberg type 1; K2, Kreyberg type 2; RR, relative risk; SqC, squamous cell carcinoma; S1, sensitivity analysis 1; S2, sensitivity analysis 2.

* Corresponding author at: P N Lee Statistics and Computing Ltd, 17 Cedar Road, Sutton, Surrey SM2 5DA, United Kingdom. Fax: +44 (0) 2086422135.

*E-mail addresses:* JohnFry@pnlee.co.uk (J.S. Fry), PeterLee@pnlee.co.uk (P.N. Lee), BarbaraForey@pnlee.co.uk (B.A. Forey), Katharine@thecoombses.co.uk (K.J. Coombs).

http://dx.doi.org/10.1016/j.yrtph.2015.02.010
0273-2300/

Here we apply exactly the same techniques to derive and compare estimates of $H$ based on the more limited data that are available separately for SqC and AdC.

## 2. Methods

### 2.1. Study identification, data selection and blocks

The available data were taken from the IESLC database, described fully earlier (Lee et al., 2012a), and the additional literature searching used for the paper on lung cancer and quitting (Fry et al., 2013). The results considered concerned blocks of RRs for SqC, or exceptionally for Kreyberg type I (KI) lung cancer, and for AdC (or KII), each block consisting of RR and 95% confidence interval (CI) estimates, expressed relative to never smokers, for former smokers (by time quit) and for current smokers. On occasion, RRs were provided relative to current smokers. Where possible, blocks were sex-specific, with RRs adjusted for covariates being preferred to unadjusted estimates. For each block, the additional data recorded included study type, sex, location, year of publication, age range, definition of product smoked, and definition of never smoker, as described earlier (Fry et al., 2013). For each RR in each block, the range of each quitting period was also recorded.

### 2.2. Analysis

As before, the first analysis step was to use the method of Hamling et al. (2008) in each block, to estimate the pseudo-table of numbers of cases and of controls/at risk corresponding to the observed RRs and 95% CIs. Blocks involving less than 10 cases in quitters were omitted. Midpoint estimates for quitting periods were derived as described by Lee et al. (2012b). That paper also describes the maximum likelihood methods used to estimate $H$ (in years) and its SE.

For prospective studies, the underlying model fitted to a block was:

$$P_j = A + B\exp(-Ct_j)$$

where $P_j$ is the absolute risk of lung cancer for time quit $t_j$ in group $j$ and $A$, $B$ and $C$ are parameters to be estimated. Here $A$ is the risk in never smokers ($t_j$ = infinite) and $B$ is the excess absolute risk in current smokers ($t_j$ = 0), i.e. the increase in absolute risk compared to never smokers. The term $\exp(-Ct_j)$ models the proportional decline in excess absolute risk for quitters, declining asymptotically from 1 to 0, as time increases. $H$ is estimated by:

$$0.5 = \exp(-CH)$$

or

$$H = (\log_e 2)/C$$

For case-control studies, the model used was:

$$F_j = 1 + B\exp(-Ct_j)$$

where $F_j$ is the RR (compared to never smokers) rather than the absolute risk and $H$ is estimated as before. While $C$, and thus $H$, have the same interpretation as for prospective studies, the interpretation of $B$ is different, being the excess relative risk (relative risk in current smokers minus 1) rather than the excess absolute risk.

For both prospective and case-control studies the method, described fully elsewhere (Lee et al., 2012b), allows estimation of the fitted RRs and numbers of cases and controls/at risk by level corresponding to the observed RRs and (pseudo-) numbers, and testing for goodness-of-fit by an approximate chisquared statistic.

Sensitivity analyses S1 and S2 were conducted as briefly described above. In S1 RRs for quitters relating to the shortest quit time were omitted from each block, while in S2 all RRs for quitters in each block relating to a range of quit times with upper limit at most three years were considered to be current smokers.

Overall estimates of $H$ (with 95% CI) were derived, and sources of heterogeneity studied using inverse-variance weighted regression of log H, with estimates converted back to the original scale. The factors studied were sex, study type, location (continent), publication year, midpoint age of subjects (at baseline for prospective studies) and current smoker RR. Sources of heterogeneity were first studied based on an analysis which included all the available estimates for each lung cancer type. A second analysis was restricted to matched estimates, only using data from blocks which provided valid estimates of $H$ for both lung cancer types, so as to avoid confounding by aspects of study quality and conduct. A final analysis, restricted to the matched estimates, was based on inverse-variance weighted regression of the logarithm of the ratio (H for AdC)/(H for SqC).

## 3. Results

### 3.1. Studies providing data

Table 1 summarizes some details of the 16 studies and 19 blocks. Three studies provided sex-specific data, one data for the sexes combined, and the rest data for a specific sex (nine males and three females). Six of the studies were conducted in Europe, four in North America, three in South America, and three in Asia. Apart from two prospective studies, both in the USA, all the other studies were of case-control design. The earliest study (WYNDE3) was conducted in 1966–1968, with the latest (TSE) in 2004–2006. Most studies started in the 1980s or 1990s. Three studies (IARC, LEITZM, LUBIN2) were clearly larger than the rest, involving about 7000 cases. A further five studies (JEDRYC, KUBIK2, PARK, SPEIZE, TSE) included over 1000 cases, with no study less than 200 cases. Of the 16 studies, 13 provided separate information for SqC and AdC, two information for AdC only, and one, the earliest, for KI and KII.

### 3.2. The data blocks

Tables 2 (SqC) and 3 (AdC) give the RRs for each block. Data are shown first for the group considered current smokers in analysis (with mean quit time 0) and then for the former smokers by increasing time of quit. In a number of the blocks, as indicated in the column "time quit groupings", current smokers are combined with recent quitters, the longest such combination being up to 5 years in JEDRYC and WAKAI. In 14 of the 17 blocks for SqC and 13 of the 19 blocks for AdC, the RRs for quitters (ignoring current smokers) decrease strictly monotonically with increasing time quit, and in all the remaining blocks except one (PARK, AdC) the RR for the longest quit time is less than that for the shortest. Current smoker RRs for SqC are always higher than those for AdC in the corresponding block, with the exception of MATOS, where they are equal. However, for both histological types, there is considerable variation in the current smoking RRs, from 4.17 (PARK) to 101.66 (PEZZOT) for SqC, and from 0.87 (PARK) to 13.01 (KUBIK2) for AdC. For SqC, the RR for the shortest quitting time is usually less than that for current smokers, but there are four blocks (KUBIK2, females, LUBIN2 males, LUBIN2 females, TSE) where it is greater, and one (PARK) where it is equal. For AdC, there are five blocks (BARBON, DESTE5, KUBIK2 females, LUBIN2 males, MATOS) where the RR for the shortest quitting time is greater than that for current smokers.

### 3.3. Fit to the negative exponential model and half-life estimates

Table 4 compares the observed number of cases of SqC and of AdC, summed over blocks for never smokers, for current smokers