Dose-response relationship of lung cancer to amount smoked, duration and age starting

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

Published in World Journal of Meta-Analysis 2013 August 26; 1(2): 57-77

Additional file 1 – Goodness of Fit

1. *Introduction*

Additional file 1 describes the method used to estimate a fitted table of pseudo-numbers based on the observed table and the fitted relative risk (RR) estimates. The method differs for prospective studies, where only fitted numbers of cases are derived, and case-control studies, where fitted numbers of both cases and controls are derived. It also describes the tests of goodness-of-fit to the models used.

2. *Prospective studies*

With two exposed levels, we have an observed table of pseudo-numbers:

Level	Cases	At risk
Baseline	A_0	N_0
Low exposure	A_1	N_1
High exposure	A_2	N_2
Total	A_S	N_S

We have fitted a set of RRs: R_0, R_1, R_2 (where $R_0 = 1$)

We wish to derive a set of fitted cases: F_0, F_1, F_2

We have the following formulae:

$F_0 + F_1 + F_2 = A_S$ (marginal totals stay the same) (1)

$R_1 = (F_1N_0)/(F_0N_1)$ (2)

$R_2 = (F_2N_0)/(F_0N_2)$ (3)

From (2) $F_1 = F_0N_1R_1/N_0$ (4)

From (3) $F_2 = F_0N_2R_2/N_0$ (5)

From (1,4,5) $F_0 + \dfrac{F_0N_1R_1}{N_0} + \dfrac{F_0N_2R_2}{N_0} = A_S$ (6)

so $F_0N_0 + F_0N_1R_1 + F_0N_2R_2 = A_SN_0$ (7)

or $\quad F_0 = (A_S N_0 R_0) / \sum_{i=0}^{2} (N_i R_i)$ (8)

From (4,8) $\quad F_1 = (A_S N_1 R_1) / \sum_{i=0}^{2} (N_i R_i)$ (9)

From (5,8) $\quad F_2 = (A_5 N_2 R_2) / \sum_{i=0}^{2} (N_i R_i)$ (10)

This allows derivation of fitted values and is clearly generalizable to multiple exposure levels (k).

A chisquared test of goodness-of-fit on k−1 df is then derived in the usual way from the formula:

$$\chi^2 = \sum_{i=0}^{k} (A_i - F_i)^2 / F_i$$

3. *Case-control studies*

Here the observed table of pseudo-numbers is:

Level	Cases	Controls	Total
Baseline	A_0	B_0	C_0
Level 1	A_1	B_1	C_1
Level 2	A_2	B_2	C_2
Total	A_S	B_S	C_S

The expected table of fitted numbers is:

Level	Cases	Controls
Baseline	F_0	G_0
Level 1	F_1	G_1
Level 2	F_2	G_2

We have fitted RRs: R_0, R_1, R_2 (where $R_0 = 1$)

We can write down the following formulae based on the marginal totals and the RRs:

$$F_0 + G_0 = C_0 \quad (11)$$

$$F_1 + G_1 = C_1 \quad (12)$$

$$F_2 + G_2 = C_2 \quad (13)$$

$$F_0 + F_1 + F_2 = A_S \quad (14)$$

$$R_1 = F_1G_0/(F_0G_1) \quad (15)$$

$R_2 = F_2G_0/(F_0G_2)$ (16)

From (15) $G_1 = F_1G_0/(F_0R_1)$ (17)

From (16) $G_2 = F_2G_0/(F_0R_2)$ (18)

From (12,15) $F_1 + F_1G_0/(F_0R_1) = C_1$ (19)

or $F_0F_1R_1 + F_1G_0 = C_1F_0R_1$ (20)

From (11) $F_0F_1R_1 + F_1(C_0 - F_0) = C_1F_0R_1$ (21)

or $F_1 = C_1F_0R_1/(F_0R_1 + C_0 - F_0)$ (22)

Similarly $F_2 = C_2F_0R_2/(F_0R_2 + C_0 - F_0)$ (23)

From (14,22,23) $$F_0 + \frac{C_1F_0R_1}{(F_0R_1 + C_0 - F_0)} + \frac{C_2F_0R_2}{(F_0R_2 + C_0 - F_0)} = A_S \quad (24)$$

This is an equation in F_0 only, which can be solved using standard Newton-Raphson methodology.

Formula (22) gives F_1 in terms of F_0, while formula (23) gives F_2 in terms of F_0. Formulae (11,12,13) then give G_i in terms of F_i

This gives the whole table of fitted numbers. A chisquared test of goodness-of-fit on 2k−1 df is then derived using the formula:

$$\chi^2 = \sum_{i=0}^{k}(A_i - F_i)^2/F_i + \sum_{i=0}^{k}(B_i - G_i)^2/G_i$$