Appendix B. Derivation of formula for score tests for cohort studies and case-control studies

Cohort study

The score test (*44*) is a commonly used method for assessing trends of risk with dose. In particular, it has been used in this way in assessing trends of cancer risk with dose in various occupational studies (*45-47*). This Appendix outlines its use for this purpose, and how it can be used to assess statistical power. The score is the derivative of the log-likelihood with respect to the dose trend parameter. In particular, if a relative risk model is assumed in which the cancer risk (whether for incidence or mortality) in cell *j* of stratum *i* is given by $p_{ij} \cdot [1+\theta_i \cdot D_{ij}]$, then the log (multinomial) likelihood is given by:

$$L = C + \sum_{i=1}^{S} \left\{ \sum_{j=1}^{K_i} m_{ij} \cdot \ln\left[M_i \cdot p_{ij} \cdot \left(1+\theta_i \cdot D_{ij}\right)\right] - M_i \cdot \ln\left[\sum_{j=1}^{K_i} M_i \cdot p_{ij} \cdot \left(1+\theta_i \cdot D_{ij}\right)\right]\right\} \qquad \text{(B1)}$$

where M_i is the total number of cancer cases or deaths in stratum i, m_{ij} is the observed number of cancer cases or deaths in cell j of stratum i (so that $\sum_{j=1}^{K_i} m_{ij} = M_i$), p_{ij} is the proportion of the population (e.g. proportion of person years of observation) of cell j making up stratum i (so that $\sum_{j=1}^{K_i} p_{ij} = 1$). (This is the likelihood obtained by conditioning on the total number, M_i, of cases in each stratum i.)

If we assume that $\theta_i \equiv \theta$ then:

$$\frac{dL}{d\theta} = \sum_{i=1}^{S} \left\{ \sum_{j=1}^{K_i} \frac{m_{ij} \cdot D_{ij}}{1+\theta \cdot D_{ij}} - M_i \cdot \frac{\sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}}{\sum_{j=1}^{K_i} p_{ij} \cdot \left(1+\theta \cdot D_{ij}\right)} \right\} \qquad \text{(B2)}$$

so that at $\theta = 0$ this reduces to:

$$\left.\frac{dL}{d\theta}\right|_{\theta=0} = \sum_{i=1}^{S}\left\{\sum_{j=1}^{K_i} m_{ij}\cdot D_{ij} - M_i\cdot\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}\right\} \tag{B3}$$

Therefore:

$$E_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = \sum_{i=1}^{S} M_i\cdot\left\{\frac{\sum_{j=1}^{K_i} p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)\cdot D_{ij}}{\sum_{j=1}^{K_i} p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)} - \sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}\right\}$$
$$= \theta\cdot\sum_{i=1}^{S} M_i\cdot\left\{\frac{\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}^{\ 2} - \left[\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}\right]^2}{1+\theta\cdot\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}}\right\} \tag{B4}$$

$$\mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = \mathrm{var}_\theta\left[\sum_{i=1}^{S}\sum_{j=1}^{K_i} m_{ij}\cdot D_{ij}\right] = \sum_{i=1}^{S}\mathrm{var}_\theta\left[\sum_{j=1}^{K_i} m_{ij}\cdot D_{ij}\right]$$
$$= \sum_{i=1}^{S} M_i\cdot\left[\frac{\sum_{j=1}^{K_i} D_{ij}^{\ 2}\cdot p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)}{\sum_{j=1}^{K_i} p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)} - \left\{\frac{\sum_{j=1}^{K_i} D_{ij}\cdot p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)}{\sum_{j=1}^{K_i} p_{ij}\cdot\left(1+\theta\cdot D_{ij}\right)}\right\}^2\right]$$
$$= \sum_{i=1}^{S} M_i\cdot\left[\frac{\left(\sum_{j=1}^{K_i} D_{ij}^{\ 2}\cdot p_{ij} + \theta\cdot\sum_{j=1}^{K_i} D_{ij}^{\ 3}\cdot p_{ij}\right)\cdot\left(1+\theta\cdot\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}\right) - \left(\sum_{j=1}^{K_i} D_{ij}\cdot p_{ij} + \theta\cdot\sum_{j=1}^{K_i} D_{ij}^{\ 2}\cdot p_{ij}\right)^2}{\left(1+\theta\cdot\sum_{j=1}^{K_i} p_{ij}\cdot D_{ij}\right)^2}\right] \tag{B5}$$

(It should be noted that this last expression is not the same as

$\mathrm{var}_{\theta=0}\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = E_{\theta=0}\left[-\left.\frac{d^2L}{d\theta^2}\right|_{\theta=0}\right]$. Note also that this calculation includes the covariance terms induced by the correlations between the multinomial terms in the likelihood.)

Therefore the normalized score, given by $Z = \left.\frac{dL}{d\theta}\right|_{\theta=0}\Big/\mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]^{0.5}$, has expectation given by:

$$Z_0 = E_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] \Big/ \mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]^{0.5}$$

$$= \frac{\theta \cdot \sum_{i=1}^{S} M_i \cdot \left\{ \frac{\sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}^2 - \left[\sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}\right]^2}{1+\theta \cdot \sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}} \right\}}{\left[\sum_{i=1}^{S} M_i \cdot \frac{\left(\sum_{j=1}^{K_i} D_{ij}^2 \cdot p_{ij} + \theta \cdot \sum_{j=1}^{K_i} D_{ij}^3 \cdot p_{ij}\right) \cdot \left(1+\theta \cdot \sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}\right) - \left(\sum_{j=1}^{K_i} D_{ij} \cdot p_{ij} + \theta \cdot \sum_{j=1}^{K_i} D_{ij}^2 \cdot p_{ij}\right)^2}{\left(1+\theta \cdot \sum_{j=1}^{K_i} p_{ij} \cdot D_{ij}\right)^2} \right]^{0.5}} \quad \text{(B6)}$$

and has variance 1. For small θ and D_i this varies approximately proportionally to the average dose, and as the square root of the number of cases or deaths, M_i. It is assumed that the normalized score, Z, is approximately Normally distributed, $Z \sim N(Z_0,1)$. Therefore, if the $100 \cdot p$ -centile of the standard Normal distribution is N_p, so that $p = P[N(0,1) \le N_p]$, then:

$$P[Z > N_{1-\alpha}] = P[Z - Z_0 > N_{1-\alpha} - Z_0] = P[N(0,1) > N_{1-\alpha} - Z_0] \quad \text{(B7)}$$

If this is to equal p then:

$$1 - p = P[N(0,1) \le N_{1-\alpha} - Z_0] = P[N(0,1) \le N_{1-p}] \quad \text{(B8)}$$

Therefore it must be that $N_{1-\alpha} - Z_0 = N_{1-p}$, or equivalently that:

$$Z_0 = N_{1-\alpha} - N_{1-p} \quad \text{(B9)}$$

Considering a single stratum, with $M_1 = M$, $K_1 = K$ etc then by (B1) and (B2), in order for the cohort to have power p it must be that:

$$M = \frac{\left(\left(\sum_{j=1}^{K} D_j^{\,2} \cdot p_j + \theta \cdot \sum_{j=1}^{K} D_j^{\,3} \cdot p_j\right) \cdot \left(1 + \theta \cdot \sum_{j=1}^{K} p_j \cdot D_j\right) - \left(\sum_{j=1}^{K} D_j \cdot p_j + \theta \cdot \sum_{j=1}^{K} D_j^{\,2} \cdot p_j\right)^2\right) \cdot \left(N_{1-\alpha} - N_{1-p}\right)^2}{\theta^2 \cdot \left(\sum_{j=1}^{K} p_j \cdot D_j^{\,2} - \left[\sum_{j=1}^{K} p_j \cdot D_j\right]^2\right)^2} \quad \text{(B10)}$$

For small θ and D_i this varies approximately as the inverse of the square of the average dose, and as the inverse of the square of the expected ERR per Sv, θ.

Figure 1 and Table 1 illustrate these formulae with calculations of Monte Carlo simulations of power for a cohort having the natural background dose distribution for the GB population, as described in Appendix A. Tables A2-A3 give the dose distributions assumed. The Monte Carlo simulations are performed by sampling within each stratum the number of cases m_{ij} from the multinomial distribution with probabilities:

$$\frac{p_{ij} \cdot \left(1 + \theta \cdot D_{ij}\right)}{\sum_{j=1}^{K_i} p_{ij} \cdot \left(1 + \theta \cdot D_{ij}\right)} \quad \text{(B11)}$$

The normalized score statistic for simulation s, $\left.\frac{dL}{d\theta}\right|_{\theta=0}$, for this simulated set of cases is then estimated using formula (B3) and the normalized score, $Z = \left.\frac{dL}{d\theta}\right|_{\theta=0} \Big/ \mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]^{0.5}$, then derived using this and formula (B5).

Case-control study

We assume that there are S cases, and that for case i there are K_i controls. Assume that each case or control can lie in one of $N_D + 1$ dose groups, $0,1,...,N_D$ with associated doses $D_0 < D_1 < ... < D_{N_D}$, with probabilities for stratum (case-control set) i of $\left(p_{d0i}\right)_{d=0}^{N_D}$ for the controls and $\left(p_{d1i}\right)_{d=0}^{N_D}$ for the cases. Assume that there are K_i controls in stratum (case-control set) i. Initially we assume simply that the odds ratio associated with dose group d in stratum i is given by $\lambda_d = \frac{p_{d1i} / p_{01i}}{p_{d0i} / p_{00i}} \equiv \lambda_d(\theta)$. This implies that $p_{d1i} = \frac{p_{01i}}{p_{00i}} p_{d0i} \lambda_d$, and since $\sum_{d=0}^{N_D} p_{d1i} = 1$ therefore $\frac{p_{01i}}{p_{00i}} = \frac{1}{\sum_{d=0}^{N_D} p_{d0i} \lambda_d}$. If $(n_{d0i})_{d=0}^{N_D}$ are the number of matched controls and $(n_{d1i})_{d=0}^{N_D}$ the number of cases in the various dose groups for (case-control set) i (so that $\sum_{d=0}^{N_D} n_{d0i} = K_i$, $\sum_{d=0}^{N_D} n_{d1i} = 1$) the retrospective likelihood is given by:

$$l = \prod_{i=1}^{S} \left[\prod_{d=0}^{N_D} p_{d1i}^{\,n_{d1i}} \prod_{d=0}^{N_D} p_{d0i}^{\,n_{d0i}} \frac{K_i!}{\prod_{d=0}^{N_D} n_{d0i}!} \right] = \left[\prod_{d=0}^{N_D} \lambda_d^{\sum_{i=1}^{S} n_{d1i}} \right] \prod_{i=1}^{S} \left[\frac{1}{\sum_{d=0}^{N_D} p_{d0i} \lambda_d} \prod_{d=0}^{N_D} p_{d0i}^{\,n_{d0i}+n_{d1i}} \frac{K_i!}{\prod_{d=0}^{N_D} n_{d0i}!} \right] \tag{B12}$$

This is maximized in $(p_{d0i})_{d=0}^{N_D}$ by:

$$p_{d0i} = \frac{[n_{d0i} + n_{d1i}] / \lambda_d}{\sum_{d'=0}^{N_D} [n_{d'0i} + n_{d'1i}] / \lambda_{d'}} \tag{B13}$$

so that the profile log-likelihood becomes:

$$L = \ln[l] = C - \sum_{d=0}^{N_D} \left[\sum_{i=1}^{S} n_{d0i} \right] \ln[\lambda_d] - \sum_{i=1}^{S} K_i \ln \left(\sum_{d=0}^{N_D} [n_{d0i} + n_{d1i}] / \lambda_d \right) \tag{B14}$$

Therefore the score statistic is:

$$S(\theta) = \frac{\partial L}{\partial \theta} = -\sum_{d=0}^{N_D} \left[\sum_{i=1}^{S} n_{d0i} \right] \frac{1}{\lambda_d} \frac{\partial \lambda_d}{\partial \theta} + \sum_{i=1}^{S} K_i \frac{\sum_{d=0}^{N_D} \frac{[n_{d0i} + n_{d1i}]}{\lambda_d^{\,2}} \frac{\partial \lambda_d}{\partial \theta}}{\sum_{d=0}^{N_D} \frac{[n_{d0i} + n_{d1i}]}{\lambda_d}} \tag{B15}$$

In particular, if we assume that the odds ratio has the form $\lambda_d(\theta) = 1 + \theta D_d$, so that θ is the excess odds ratio (EOR) per Sv. Then

$$\left.\frac{dL}{d\theta}\right|_{\theta=0} = S(0) = -\sum_{d=0}^{N_D}\left[\sum_{i=1}^{S} n_{d0i}\right]D_d + \sum_{i=1}^{S}\frac{K_i}{K_i+1}\sum_{d=0}^{N_D}[n_{d0i}+n_{d1i}]D_d$$
$$= -\sum_{i=1}^{S}\frac{1}{K_i+1}\sum_{d=0}^{N_D} n_{d0i}D_d + \sum_{i=1}^{S}\frac{K_i}{K_i+1}\sum_{d=0}^{N_D} n_{d1i}D_d \qquad \text{(B16)}$$

Therefore:

$$E_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = \sum_{i=1}^{S}\frac{K_i}{K_i+1}\sum_{d=0}^{N_D}[p_{d1i}-p_{d0i}]D_d$$
$$= \sum_{i=1}^{S}\frac{K_i}{K_i+1}\sum_{d=0}^{N_D}\left[\frac{p_{d0i}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0i}\lambda_{d'}} - p_{d0i}\right]D_d \qquad \text{(B17)}$$

and

$$\text{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = \sum_{i=1}^{S}\frac{K_i}{[K_i+1]^2}\left[\sum_{d=0}^{N_D} D_d{}^2 p_{d0i}(1-p_{d0i}) - 2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d0i} p_{d'0i}\right]$$
$$+\sum_{i=1}^{S}\left(\frac{K_i}{K_i+1}\right)^2\left[\sum_{d=0}^{N_D} D_d{}^2 p_{d1i}(1-p_{d1i}) - 2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d1i} p_{d'1i}\right]$$
$$= \sum_{i=1}^{S}\frac{K_i}{[K_i+1]^2}\left[\sum_{d=0}^{N_D} D_d{}^2 p_{d0i}(1-p_{d0i}) - 2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d0i} p_{d'0i}\right]$$
$$+\sum_{i=1}^{S}\left(\frac{K_i}{K_i+1}\right)^2\left[\sum_{d=0}^{N_D} D_d{}^2 \frac{p_{d0i}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0i}\lambda_{d'}}\left(1-\frac{p_{d0i}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0i}\lambda_{d'}}\right) - 2\sum_{0\le d<d'\le N_D} D_d D_{d'}\frac{p_{d0i}p_{d'0i}\lambda_d\lambda_{d'}}{\left(\sum_{d''=0}^{N_D} p_{d''0i}\lambda_{d''}\right)^2}\right]$$
(B18)

Assuming that $p_{d0i} \equiv p_{d0}$, $K_i \equiv K$ then these reduce to:

$$E_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right] = \frac{SK}{K+1}\sum_{d=0}^{N_D}[p_{d1}-p_{d0}]D_d = \frac{SK}{K+1}\sum_{d=0}^{N_D}\left[\frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}} - p_{d0}\right]D_d \qquad \text{(B17')}$$

and

$$
\begin{aligned}
&\mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]=\frac{SK}{[K+1]^2}\left[\sum_{d=0}^{N_D} {D_d}^2 p_{d0}(1-p_{d0})-2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d0} p_{d'0}\right]\\
&+S\left(\frac{K}{K+1}\right)^2\left[\sum_{d=0}^{N_D} {D_d}^2 p_{d1}(1-p_{d1})-2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d1} p_{d'1}\right]\\
&=\frac{SK}{[K+1]^2}\left[\sum_{d=0}^{N_D} {D_d}^2 p_{d0}(1-p_{d0})-2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d0} p_{d'0}\right]\\
&+S\left(\frac{K}{K+1}\right)^2\left[\sum_{d=0}^{N_D} {D_d}^2 \frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}}\left(1-\frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}}\right)-2\sum_{0\le d<d'\le N_D} D_d D_{d'} \frac{p_{d0}p_{d'0}\lambda_d\lambda_{d'}}{\left(\sum_{d''=0}^{N_D} p_{d''0}\lambda_{d''}\right)^2}\right]
\end{aligned}
\tag{B18'}
$$

Therefore the normalized score, given by $Z=\left.\frac{dL}{d\theta}\right|_{\theta=0}\Big/\mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]^{0.5}$, has expectation given

by:

$$Z_0 = E_\theta \left[\left. \frac{dL}{d\theta} \right|_{\theta=0} \right] \Big/ \mathrm{var}_\theta \left[\left. \frac{dL}{d\theta} \right|_{\theta=0} \right]^{0.5}$$

$$= \frac{[SK]^{0.5} \sum_{d=0}^{N_D} \left[\frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}} - p_{d0} \right] D_d}{\left[\begin{array}{l} \sum_{d=0}^{N_D} D_d{}^2 p_{d0}(1-p_{d0}) - 2 \sum_{0 \le d < d' \le N_D} D_d D_{d'} p_{d0} p_{d'0} \\ +K \left[\sum_{d=0}^{N_D} D_d{}^2 \frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}} \left(1 - \frac{p_{d0}\lambda_d}{\sum_{d'=0}^{N_D} p_{d'0}\lambda_{d'}} \right) - 2 \sum_{0 \le d < d' \le N_D} D_d D_{d'} \frac{p_{d0} p_{d'0} \lambda_d \lambda_{d'}}{\left(\sum_{d''=0}^{N_D} p_{d''0}\lambda_{d''} \right)^2} \right] \end{array} \right]^{0.5}}$$

$$= \frac{[SK]^{0.5} \frac{\sum_{d'=0}^{N_D} p_{d'0} D_{d'}{}^2 - \left(\sum_{d'=0}^{N_D} p_{d'0} D_{d'} \right)^2}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})}}{\left[\begin{array}{l} \sum_{d=0}^{N_D} D_d{}^2 p_{d0}(1-p_{d0}) - 2 \sum_{0 \le d < d' \le N_D} D_d D_{d'} p_{d0} p_{d'0} \\ +K \left[\begin{array}{l} \sum_{d=0}^{N_D} D_d{}^2 \frac{p_{d0}(1+\theta D_d)}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})} \left(1 - \frac{p_{d0}(1+\theta D_d)}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})} \right) \\ -2 \sum_{0 \le d < d' \le N_D} D_d D_{d'} \frac{p_{d0} p_{d'0} (1+\theta D_d)(1+\theta D_{d'})}{\left(\sum_{d''=0}^{N_D} p_{d''0}(1+\theta D_{d''}) \right)^2} \end{array} \right] \end{array} \right]^{0.5}} \quad \text{(B19)}$$

For small θ and D_i this varies approximately proportionally to the average dose, and as the square root of the number of cases or deaths, S . As above, for the score statistic to have power $100p\%$ to detect an excess with type I error α it must be that $Z_0 = N_{1-\alpha} - N_{1-p}$. This implies that:

$$S=\frac{\left(N_{1-\alpha}-N_{1-p}\right)^2\left[\begin{array}{l}\sum_{d=0}^{N_D} {D_d}^2 p_{d0}(1-p_{d0})-2\sum_{0\le d<d'\le N_D} D_d D_{d'} p_{d0} p_{d'0} \\ +K\left[\begin{array}{l}\sum_{d=0}^{N_D} {D_d}^2 \frac{p_{d0}(1+\theta D_d)}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})}\left(1-\frac{p_{d0}(1+\theta D_d)}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})}\right) \\ -2\sum_{0\le d<d'\le N_D} D_d D_{d'} \frac{p_{d0}p_{d'0}(1+\theta D_d)(1+\theta D_{d'})}{\left(\sum_{d''=0}^{N_D} p_{d''0}(1+\theta D_{d''})\right)^2}\end{array}\right]\end{array}\right]}{K\theta^2\left[\frac{\sum_{d'=0}^{N_D} p_{d'0}{D_{d'}}^2-\left(\sum_{d'=0}^{N_D} p_{d'0}D_{d'}\right)^2}{\sum_{d'=0}^{N_D} p_{d'0}(1+\theta D_{d'})}\right]^2} \qquad \text{(B20)}$$

For small θ and D_d this varies approximately as the inverse of the square of the average dose, and as the inverse of the square of the expected EOR per Sv, θ.

Assume now that there are M groups of cases, with S_i cases in each ($1 \le i \le M$), with K controls each. Assume that in case group i each case or control can lie in one of $N_{D,i}+1$ dose groups, $0,1,\ldots,N_{D,i}$ with associated doses $D_{i0} < D_{i1} < \ldots < D_{iN_{D,i}}$, with probabilities for stratum i of $\left(p_{id0}\right)_{d=0}^{N_{D,i}}$ for the controls and $\left(p_{id1}\right)_{d=0}^{N_{D,i}}$ for the cases; we assume that the odds ratio associated with dose group d in stratum i is given by $\lambda_{id}=\frac{p_{id1}/p_{i01}}{p_{id0}/p_{i00}}\equiv\lambda_{id}(\theta)$. Then with trivial changes (B17') and (B18') become:

$$E_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]=\frac{K}{K+1}\sum_{i=1}^{M}S_i\sum_{d=0}^{N_{D,i}}[p_{id1}-p_{id0}]D_d=\frac{K}{K+1}\sum_{i=1}^{M}S_i\sum_{d=0}^{N_D}\left[\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_D} p_{id'0}\lambda_{id'}}-p_{id0}\right]D_{id}$$

(B17'')

and

$$\mathrm{var}_\theta\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]=\frac{K}{[K+1]^2}\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2 p_{id0}(1-p_{id0})-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}p_{id0}p_{id'0}\right]$$

$$+\left(\frac{K}{K+1}\right)^2\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2 p_{id1}(1-p_{id1})-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}p_{id1}p_{id'1}\right]$$

$$=\frac{K}{[K+1]^2}\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2 p_{id0}(1-p_{id0})-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}p_{id0}p_{id'0}\right]$$

$$+\left(\frac{K}{K+1}\right)^2\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_{D,i}}p_{id'0}\lambda_{id'}}\left(1-\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_D}p_{id'0}\lambda_{id'}}\right)-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}\frac{p_{id0}p_{id'0}\lambda_{id}\lambda_{id'}}{\left(\sum_{d''=0}^{N_{D,i}}p_{id''0}\lambda_{id''}\right)^2}\right]$$

(B18'')

Therefore the normalized score statistic has expectation given by:

$$Z_0=\frac{K^{0.5}\sum_{i=1}^{M}S_i\sum_{d=0}^{N_D}\left[\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_D}p_{id'0}\lambda_{id'}}-p_{id0}\right]D_{id}}{\left[\begin{array}{l}\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2 p_{id0}(1-p_{id0})-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}p_{id0}p_{id'0}\right]\\+K\sum_{i=1}^{M}S_i\left[\sum_{d=0}^{N_{D,i}}D_{id}{}^2\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_{D,i}}p_{id'0}\lambda_{id'}}\left(1-\frac{p_{id0}\lambda_{id}}{\sum_{d'=0}^{N_D}p_{id'0}\lambda_{id'}}\right)-2\sum_{0\le d<d'\le N_{D,i}}D_{id}D_{id'}\frac{p_{id0}p_{id'0}\lambda_{id}\lambda_{id'}}{\left(\sum_{d''=0}^{N_{D,i}}p_{id''0}\lambda_{id''}\right)^2}\right]\end{array}\right]^{0.5}}$$

(B19')

Figures 1-2 and Tables 1-2 illustrate these formulae with calculations of Monte Carlo simulations of power for a case-control study drawn from a population having the natural background dose distribution for the GB population, as described in Appendix A. Tables A2-A3 give the dose distributions assumed. The Monte Carlo simulations are performed by sampling within each stratum i the indicator for the case n_{d1i} in each case-control set from the multinomial distribution with probabilities:

$$\frac{p_{id0} \cdot \left(1+\theta \cdot D_{id}\right)}{\sum_{d=0}^{N_{D,i}} p_{id0} \cdot \left(1+\theta \cdot D_{id}\right)} \qquad \text{(B21)}$$

The normalized score statistic for simulation s, $\left.\frac{dL}{d\theta}\right|_{\theta=0}$, for this simulated set of cases is then estimated using formula (B16) and the normalized score, $Z = \left.\frac{dL}{d\theta}\right|_{\theta=0} \Big/ \operatorname{var}_{\theta}\left[\left.\frac{dL}{d\theta}\right|_{\theta=0}\right]^{0.5}$, then derived using this and formula (B18’’).