SUPPLEMENTARY FILE 1 TO

"Environmental tobacco smoke exposure and lung cancer – a systematic review", Lee PN *et al*

FURTHER DETAILS OF METHODS

Data recorded

The study database contains a record for each study, identified by a 6-character reference code ("Ref") based on the principal author's name. The record includes the following aspects: relevant publications; study title; study design; sexes considered; age range; location; timing and length of follow-up; number of cases and extent of histological confirmation; number of controls or subjects at risk; types of controls and matching factors used in case-control studies; use of proxy respondents, interview setting and response rates; confounding variables considered; availability of results by histological types; availability of results for each ETS index; and study quality, i.e. having or not having defined serious study weaknesses[11].

The RR database holds the detailed results, typically containing multiple records for each study. Each record is linked to the relevant study and refers to a specific RR, describing the comparison made and the results. This record includes the sex, age range, race, lung cancer type, and (for prospective studies) the follow-up period. The ETS exposure of the numerator of the RR is defined by the exposure type (spouse, household etc) and timing of exposure if it occurred in adulthood. Information is recorded about the denominator of the RR to indicate whether subjects had no exposure to any ETS source, or just to the exposure type recorded for the numerator. For dose-related indices, the level of exposure is recorded. The source of the RR is also recorded, as are details on adjustment variables. Results recorded included numbers of cases for the numerator and denominator, and, for unadjusted results, numbers of controls, persons at risk or person-years at risk. The RR itself and its lower and upper 95% confidence

limits are always recorded. These may be as reported, or derived by various means (see below), with the method of derivation noted.

Derivation of RRs

Methods used as required to provide estimates of the RR and CI included the following:

Correction for zero cell

If the 2 x 2 table has a zero cell, 0.5 was added to each cell, and the standard formulae applied.

Combining independent RRs

RRs were combined over *l* strata(e.g. from a 2 x 2 x *l* table) using fixed-effect meta-analysis[13], giving an estimate adjusted for the stratifying variable.

Combining non-independent RRs

The Hamling et al. method[14] was used (e.g. to derive an adjusted RR for those exposed at home from available adjusted RRs for those exposed at home, work or both locations, relative to those not exposed at home, or to combine adjusted RRs for several histological types, each relative to a single control group).

Estimating CI from crude numbers

If an adjusted RR lacked a CI or p-value but the corresponding 2 x 2 table was available, the CI was estimated assuming that the ratio of the upper to lower comfidence limits was the same as for the equivalent unadjusted RR.

Meta-analyses

Analyses conducted

For a given exposure type, a pre-planned set of up to 20 analyses was conducted. Meta-analyses 1 and 2 used the overall data available, while meta-analyses 3 and 4 were separated by region (North America, Europe, Asia or other regions), with meta-analyses 1 and 3 using most-adjusted and 2 and 4 least-adjusted data. Analyses 5-20 were based on most-adjusted data

only and studied variation by the following factors: country within Asia, region within Europe, publication year, number of cases, histological confirmation, study type (prospective or case-control), study control type (prospective, case-control separated by healthy, diseased or both), study quality, number of confounders considered, adjustment for age, adjustment for marital status, dose-response results available, whether the index used was actually the spouse, never smoker definition, interview setting, and proxy use.

The primary index of exposure used was "spousal smoking (or nearest equivalent)" where, for studies which provided no results for spousal exposure, results for household, total or both spousal/home and other exposure were chosen instead. This identified a single exposure definition for each study. For overall lung cancer, the full set of 20 meta-analyses was carried out restricted to females, and unrestricted on sex (i.e including separate RRs for males and females if available, and RRs for sexes combined otherwise). Further meta-analyses for the principal index of exposure corresponded to meta-analyses 1 to 4 only. These included analyses for spousal smoking (or nearest equivalent) for males, spousal smoking (specifically) for females, males and unrestricted on sex, and analyses for spousal smoking (or nearest equivalent) for squamous cell carcinoma and for adenocarcinoma, each for females, males and unrestricted on sex.

Analyses for the other types of exposure were run only for overall lung cancer, without restriction on sex, and were equivalent to meta-analyses 1-4 only. The childhood exposure analyses were run using four alternative indices – (1) most comprehensive index, choosing mother if no more comprehensive index available, (2) mother specifically, (3) father specifically and (4) parents specifically. The household exposure analyses were run using seven alternative indices - (1) most comprehensive index, choosing mother if no more comprehensive index available, (2) as (1) but choosing father, (3) mother specifically, (4) father specifically, (5) excluding ETS from parents and spouse, otherwise most comprehensive index available, (6) excluding ETS

from spouse, otherwise most comprehensive index available, choosing mother if no more comprehensive index available and (7) as (6) but choosing father.

Selecting RRs for the meta-analyses

All meta-analyses are restricted to records with available RR and CI values (i.e. non-missing values). The process of selecting RRs for inclusion in a meta-analysis aimed to include all relevant data and avoid double-counting. For studies with multiple RRs, the one used is determined by a preference order defined for the meta-analysis. For example, for an analysis of exposure from a smoking spouse, one study might provide data relevant to any exposure during the marriage and to current exposure; the order of preference would determine which result to include from this study, whilst allowing either definition from other studies which provided no such choice. Preference orders may be required for exposure status, timing of exposure and the unexposed base. As the definitions of RR available may differ by sex, the RRs chosen for each sex may not necessarily have the same definition. Sexes combined results are only considered where sex-specific results are not available. When multiple preference orders are involved, the sequence of implementation may affect the selection, so preferences for the most important aspects, usually concerning ETS exposure, are implemented first. The preferences used are given in the detailed output for each analysis, made available as Supplementary File 4.

Sofware used

All data entry and meta-analyses were conducted using ROELEE version 3.1 (available from P.N.Lee Statistics and Computing Ltd, 17 Cedar Road, Sutton, Surrey SM2 5DA, UK). Some derivation of RRs and CIs were conducted using Quattro Pro 9 or Excel 2003.

References

See main report.