APPENDIX C
Validation checks on completeness and consistency of the data

1. Study database

'Blank' data (i.e. no data entered) not allowed for any field in any card

'Missing' data not allowed for field: Card DESCR: STYPE

'Not applicable' (NA) data not allowed for any field in the following cards:
RESUL2, RESUL3, CONFN2
or for the following fields:
Card DESCR: TITLE, FTITLE, SSEX, SAGELO, SAGEHI, SRACE, CONT, BEGYR, ENDYR, PUBYR, REFKEY, NLC, NCONTR
Card DESIGN: PROSP, POPUL, PROX, FHIST2, HISTPC
Card RESUL1: REHIST, REEX, RECURR, REEVER, RECIGT, REHR, REPIPE, RECIGR, REOTHR, RENCIG, REMORE

Other checks on card DESCR
If STYPE 2,3,4 (prospective, nestedCC or case-cohort study) –
SAGELO ≤ SAGEHI ≤ SAGEHF
Otherwise (CC study) – SAGELO ≤ SAGEHI and SAGEHF NA

Fields NAMER, SCAMER, WEUR, ASIA, AUSLIA, AFRICA must be NA except for one field, depending on the value of CONT as follows, which must have +ve value:
1- NAMER, 3- SCAMER, 4- WEUR, 5- EEUR, 6- ASIA, 8- AUSLIA, 9- AFRICA.
USSTAT must be NA unless NAMER is 1 or 3, in which case USSTAT must be +ve.
USSR must be NA unless EEUR is 6, in which case USSR must be +ve.

If STYPE is 2,3,4 (prospective, nestedCC or case-cohort study) –
BEGYR ≤ ENDYR ≤ FINFYR
Otherwise (CC study) – BEGYR ≤ ENDYR and FINFYR NA

NLC (number of lung cancer cases) ≥ 100, or missing

Other checks on card DESIGN
If STYPE is 2 (prospective study) –
Fields POPUL, PROXPB, RESRPB must not be NA

Fields CONTRL, CONDIS, PROXCA, PROXCO, PROXDI, IVDI, VITDI, RESRCA, RESRCO, MATSEX, MATAGE, MATRACE, MATOTH, MATSES, MATUR, MATEDU, MATOCC, MATMAR must be NA.

Field PROSP must be 1.

If PROX is 0 then both PROXPB must be 0. If PROX is 1 then PROXPB must be +ve or missing.

If STYPE is 1 (case-control study) –

Fields CONTRL, PROXCA, PROXCO, PROXDI, IVDI, VITDI, RESRCA, RESRCO, MATSEX, MATAGE, MATRACE, MATOTH, MATSES, MATUR, MATEDU, MATOCC, MATMAR must not be NA.

Fields PROXPB, RESRPB must be NA.

Field PROSP must be 0.

CONDIS must be NA if and only if CONTRL is 1, 4, 8 (i.e. no diseased/decedent controls) .

HOSPDI must be NA if and only if CONTRL is 1, 5, 6, 8 (i.e. no hospital controls) .

If PROX is 0 then both PROXCA and PROXCO must be 0. If PROX is 1 then at least one of them must be +ve or missing.

PROXDI must be 0 (or missing) unless PROXCA ≠ PROXCO.

MATSEX can only be 1 if SSEX is 1 (i.e. both sexes in study)

Other study types have same conditions as case-control studies except:

Fields PROXPB, RESRPB must not be NA, and if PROX is 0 then PROXPB must be 0.

And for case-cohort study only: Fields RESRCA,RESRCO,MAT… must be NA

FHIST2 must be 1 (full) if and only if HISTPC is 100.
If FHIST2 is 2 or 7 (partial), then HISTPC must be +ve or missing.
If FHIST2 is 3 or 5 (unchecked) the HISTPC must be 0.
If FHIST2 is 4 or 6 (unknown) the HISTPC must be missing.

Consistency checks between cards DESIGN and RESUL1
If FHIST2 is 3-6 then REHIST must not be 1.

Either REHR is 0 and DEFHR is NA; or REHR is 1 and DEFHR is +ve or missing.

Either REPIP, REOTHR are both 0 and DEFPIP is NA; or at least one of REPIP, REOTHR is 1 and DEFPIP is +ve or missing.

Either REEX is 0 and DEFEX is NA; or REEX is 1 and DEFEX is +ve or missing.

Other checks on card CONFN2
TOTCO must equal the sum of the other fields (or missing).

2. RR database

'Blank' data (i.e. no data entered) not allowed for any field in any card.

'Missing' data not allowed for any fields on the following cards:
RRDEF, RRADJ
or for the following field:
Card RRDATA: DERIVE

'Not applicable' (NA) data not allowed for any field in the following cards:
RRADJ
or for the following fields:
Card RRDEF: NRR, RSEX, RAGELO, RAGEHI, RRACE, LCTYPE, SMKSTA, PROD, NCIGLO, NCIGHI, DENOM
Card RRDATA: RR, RRL, RRU, DERIVE

Other checks on card RRDEF
If PROD = 2, 4 or 5 (i.e. cigarettes), then CIGTYP must be +ve, otherwise CIGTYP must be NA.

For the following pairs of fields, either both LO = HI = 0, or 0 < LO # HI:
RAGELO and RAGEHI; NCIGLO and NCIGHI.
Similarly for FOLPLO and FOLPHI (except for case-control study they must both be NA – see below).

Other checks on card RRDATA
RRL # RR # RRU
If any two of CA1, CA0, CO1 and CA0 = 0, then RR, RRL, RRU must be missing. If all four of them are +ve, then RR, RRL, RRU must equal (to 2 decimal places) the relative risk and CI as calculated according to the formula given in Section 3.4.5; if three are +ve and one zero, then the calculation will include the correction for zero cells described in that section, and DERIVE must be 14.

Consistency checks between cards RRADJ and RRDATA
CA1, CA0, CO1 and CA0 must be NA if and only if at least one field in RRADJ is +ve.

Consistency checks between card RRDEF and study database
RSEX can be m only if SSEX is m or b
RSEX can be f only if SSEX is f or b
RSEX can be c only if SSEX is b

RAGELO $SAGELO
If STYPE is 1 (case-control study) RAGEHI must be # SAGEHI, and must not have both RAGELO = SAGELO and RAGEHI = SAGEHI.

If STYPE is 2,3,4 (prospective, nested CC or case-cohort study), similar conditions apply but with SAGEHF instead of SAGEHI.

RRACE must not be 1 (all) unless SRACE is 1 (all).
RRACE must not be 2, 4, 5 (white) if SRACE is 3 (black).
RRACE must not be 3 (black) if SRACE is 2 or 5 (white).

If SMKSTA is 1 then REEVER must be 1.
If SMKSTA is 2 then RECURR must be 1.
If SMKSTA is 3 then REEX must be 1.

If CIGTYP is 4 then REHR must be 1.
If NCIGLO >0 then RENCIG must be 1.
If PROD is 7, then REPIPE must be 1.
IF PROD is 8, then RECIGR must be 1.

If STYPE is 2,4 (prospective or case-cohort study) and FOLPHI is >0, then FOLPHI must be # FINFYR-BEGYR+1.
If STYPE is 1,3 (case-control or nested CC study) FOLPLO and FOLPHI must be NA.

Consistency checks between card RRADJ and study database
ADSEX can be 1 only if COSEX is 1.
ADAGE can be 1 only if COAGE is 1.
ADRACE can be 1 only if CORACE is 1.
ADOTHR must be equal to the sum of COTOT-COSEX-COAGE-CORACE, except that ADOTHR may be 20 (meaning +ve but unknown) provided the sum is +ve.

Consistency checks between card RRDATA and study database
CA1 + CA0 must be # NLC.
CO1 + CO0 must be # NCONTR. [This validation requirement was checked individually and waived for RRs from prospective studies where numbers of man-years at risk had been entered.]

Consistency checks between records within each study
NRR is unique, and one record has NRR=1.
Each record has a unique set of values for the fields in cards RRDEF and RRADJ (excluding NRR and comments).