

Appendix CValidation checks on completeness and consistency of the data1. Study database

Appendix D gives a key to the six character codes used for cards and fields on the 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:

CONFND, OTHRES

or for the following fields:

Card DESCR: TITLE, FTITLE, SSEX, SAGELO, SAGEHI, SRACE, CONT, BEGYR, ENDYR, PUBYR, REFID, OVERL, PRINC

Card DESIGN: POPUL, MEDEXC, OTHEXC, RESPON, CHISMO

Card ASTHMA: LIFAST, CURAST, NTOT

‘Zero’ data not allowed for fields:

Card DESCR: TITLE, FTITLE, SSEX, SRACE, CONT, NAMER, SCAMER, WEUR, EEUR, ASIA, AUSLIA, AFRICA, LOCAT, BEGYR, ENDYR, FINFYR, PUBYR, REFID, ADDREF, OVERL, PRINC

Card DESIGN: STYPE, CONTRL, CONDIS, POPUL, RESPON, CHISMO

Card ASTHMA: DIAGLS, TIMLAS, DESLAS, DIAGCS, TIMCAS, DESCAS, NLAST, NCAST, NTOT

Other checks on card DESCR

If STYPE=2 (prospective), then –

$SAGELO \leq SAGEHI \leq SAGEHF$

Otherwise (CC or CS study) – $SAGELO \leq SAGEHI$ and $SAGEHF$ NA

Fields NAMER, SCAMER, WEUR, EEUR, ASIA, AUSLIA, AFRICA must be NA except for one field, depending on the value of CONT as follows:

1- NAMER, 2- SCAMER, 3- WEUR, 4- EEUR, 5- ASIA, 6- AUSLIA, 7- AFRICA.

USSTAT must be NA unless NAMER=1 or 3, in which case USSTAT must be +ve.

If STYPE=2 (prospective), then –

$BEGYR \leq ENDYR \leq FINFYR$

Otherwise (CC or CS study) – $BEGYR \leq ENDYR$ and $FINFYR$ NA

If OVERL=1, then PRINC must be 1, otherwise a comment must be entered.

Other checks on card DESIGN and MATCH

If STYPE=2,3 (prospective or cross-sectional study), then fields CONTRL, CONDIS, POPCON, MATSEX, MATAGE, MATRACE, MATLOC, MATSES, MATHOS must be NA.

If STYPE=1 (case-control study), then –

Fields CONTRL, MATSEX, MATAGE, MATRACE, MATLOC, MATSES, MATHOS must not be NA.

CONDIS must be NA if and only if CONTRL=1 (i.e. healthy controls)

MATSEX may only be 1 if SSEX=1.

Other checks on card ASTHMA

Fields DIAGLS, TIMLAS, DESLAS must be NA if and only if LIFAST=0.

Fields FIRAST, REPCAS, DIAGCS, TIMCAS, DESCAS must be NA if and only if CURAST=0.

If NTOT is non-missing, then NLAST < NTOT if LIFAST=1, and NCAST < NTOT if CURAST=1

Other checks on card CONFND

TOTCO must equal the sum of the other fields in the card excluding COREJE (or TOTCO may be missing).

If COREJE=1, then a comment must be entered.

COSEX may only be 1 if SSEX=1 (i.e. both sexes in study)

Other checks on card OTHRES

If any field has the value 1, then a comment must be entered.

2. RR database

Appendix F gives a key to the six character codes used for cards and fields on the relative risk 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 card RRADJ,

or for the following fields:

Card RRDEF: NRR, RSEX, RAGELO, RAGEHI, RRACE, RASTIM, ONSET,
 EXPOS, MEASEX, UNEXSO, SOURCE,

Card RRDATA: CA1, CA0, RR, RRL, RRU, DERIVE

Other checks on card RRDEF

If EXPOS = 1, 2, 6 (i.e. parent), then WHOPAR must be +ve; otherwise WHOPAR must be NA.

If EXPOS = 3, 7 (i.e. household), then WHOHOU must be +ve; otherwise WHOHOU must be NA.

If EXPOS = 4, 8 (i.e. total), then WHOTOT must be +ve; otherwise WHOTOT must be NA.

If EXPOS = 1-4 or 6-8 (i.e. not biochemical), then WHESMO and UNEXTI must be +ve; otherwise they must be NA. If UNEXTI=3, then a comment must be entered.

If EXPOS = 5, 9 (i.e. biochemical), then BIOMEA and BIOMAR must be +ve; otherwise they must be NA.

ODDSON must be NA if and only if ONSET=0.

If DOSER=1, 11 (i.e. not standard dose-response category) then EXPLO and EXPHI must be NA

If DOSER is in the range 2-10 (i.e. standard dose-response category) then

MEASEX >0

EXPLO, EXPHI must not be NA

EXPLO ≤ EXPHI

UNEXHI ≤ EXPLO

UNEXHI may be NA if and only if MEASEX=1 (i.e. not dose-response, and not denominator is “low” exposure), except if DOSER=11.

Either both RAGELO = RAGEHI = 99, or $RAGELO \leq RAGEHI < 99$.

Other checks on card RRDATA

$RRL \leq RR \leq RRU$

If CA1, CA0, CO1 and CA0 are all +ve, then RR, RRL, RRU must equal (to 2 decimal places) the relative risk and CI as calculated according to the formula given in §3.4.3; 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 7.

Consistency checks between cards RRDEF and RRADJ

ADSEX may be +ve only if RSEX=1

Consistency checks between cards RRADJ and RRDATA

CO1 and CO0 must be NA if and only if at least one field in card RRADJ is +ve.

Consistency checks between card RRDEF and study database

RSEX may be 2 only if SSEX is 1 or 2

RSEX may be 3 only if SSEX is 1 or 3

RSEX may be 1 only if SSEX is 1

$RAGELO \geq SAGELO$ (except if $RAGELO=RAGEHI=99$)

If STYPE is 1, 3 (case-control or cross-sectional study) –

$RAGEHI \leq SAGEHI$ (except if $RAGELO=RAGEHI=99$)

Must not have both $RAGELO = SAGELO$ and $RAGEHI = SAGEHI$.

If STYPE is 2 (prospective study), similar conditions apply but with SAGEHF instead of SAGEHI.

RRACE must not be same as SRACE

RASTIM may be 1 only if LIFAST=1, and it may be 2 only if CURAST=1.

ONSET may be 1 only if STYPE=2 (prospective study).

Consistency checks between card RRADJ and study database

ADSEX may be 1 only if COSEX is 1.

ADAGE may be +ve only if COAGE is +ve.

ADRACE may be +ve only if CORACE is +ve.

ADOTHR may be +ve only if COMSP, COPS MC or COHSM is +ve.

$ADOTHR+ADOETS \leq TOTCO-COSEX-COAGE-CORACE$, except that ADOTHR may be 20 (meaning +ve but unknown) provided the sum is +ve. If less than (i.e. not equal) then a comment must be entered.

Consistency checks between card RRDATA and study database

$CA1 + CA0 \leq NLAST$ if RASTIM=1, or $CA1 + CA0 \leq NCAST$ if RASTIM=2.

If ONSET=0 (i.e. prevalence analysis) $CA1+CA0+CO1+CO0 \leq NTOT$.

If ONSET=1 (i.e. onset analysis) $CO1+CO0 \leq NTOT$. [This validation requirement was checked individually and waived for RRs 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, ODDSON, SOURCE and comments).