Journal of Epidemiology and Community Health 1996;50:674–680

# Reliability of data from proxy respondents in an international case–control study of cardiovascular disease and oral contraceptives

N R Poulter, C L Chang, T M M Farley, M G Marmot on behalf of the World Health Organization Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception

**Abstract**

***Study objectives*** **– To evaluate the reliability of data supplied in a case–control study by proxy respondents for cases who were too ill to do so themselves.**

***Design*** **–A hospital based, case–control study of the current use of oral contraceptives (OC) and cardiovascular diseases. Data from "true" controls matched to a subset of cases were compared with those supplied by proxy respondents about the true controls.**

***Setting*** **– Hospitals in 21 centres from Africa, Asia, Europe, and Latin America.**

***Patients and participants*** **– For a subset of cases, 403 pairs of controls – one "true" and one proxy – were interviewed. "True" controls were matched by age, place, and time of admission and were admitted with 1 of 27 permissible diagnoses not associated with OC use. Proxy controls were either relatives or friends of true controls.**

***Main results*** **– Levels of concordance between data from proxy and true controls were high for most variables regarding recent events, including current OC use, but were greatly diminished when detailed information, particularly from the past, was required. Husbands were usually the best proxy, although this was question-specific. The sensitivity and specificity of proxy responses were 93% (95% confidence intervals: 77%, 99%) and 100% (98%, 100%) respectively, for current use of OC. Assuming the misclassification of current OC use by proxy cases is similar to that produced by proxy controls, the estimated impact of using proxy data on risk estimates associated with current OC use was to bias the overall estimate of risk of stroke by less than 3% and the risks of both acute myocardial infarction and venous thromboembolism by less than 1%.**

***Conclusions*** **– Friends or relatives, and particularly husbands, provided reliable information when used as proxy respondents for young women. The estimated impact of misclassification by proxy respondents on overall risk estimates in the WHO collaborative study was less than that which would have arisen if information from proxy respondents had not been used.**

**Department of Epidemiology and Public Health, University College London Medical School, 1–19 Torrington Place, London WC1E 6BT**
N R Poulter
C L Chang
M G Marmot

**UNDP/UNFPA/WHO/ World Bank Special Programme of Research Development & Research Training in Human Reproduction, World Health Organization, Geneva 27, Switzerland**
T M M Farley

Correspondence to:
Dr N R Poulter.

Accepted for publication April 1996

(*J Epidemiol Community Health* 1996;50:674–680)

In epidemiological research based on interview data, it is sometimes necessary to use proxy respondents to provide information on subjects who, either through death, inability to communicate, reduced cooperation, or impaired cognition are unable to provide information themselves. It has been reported that almost 1 in 10 aetiological studies of non-infectious diseases published in the *American Journal of Epidemiology* between 1980 and 1985 reported using proxy respondents to some extent.[1] Inability to use proxy respondents when genuine subjects are unable or unavailable to provide information impedes the use of certain types of study design. Exclusion of subjects because they require a proxy respondent results in a reduced sample size and the evaluation of data based on a subsample which may not be representative. However, data from proxy respondents are likely to be less reliable than those provided by true subjects.

A large, multi-centre, international case–control study (WHO collaborative study)[2] designed to evaluate the association between current use of oral contraceptives (OC) and three cardiovascular diseases includes patients, a significant proportion of whom were unable, for various reasons, to provide responses. In the interests of statistical power and generalisibility of results it was decided, a priori, to include these subjects, and hence data from proxy respondents had to be used.

To our knowledge, only two previously reported studies[3,4] have evaluated the reliability of data obtained from proxy respondents on current use of OC, although several studies have evaluated data on many of the other variables incorporated in the WHO study. The reliability of data from proxy respondents may well be specific with regard to time, exposure, population, and instrument,[1] and hence there is a need to evaluate the data acquired from proxy respondents in the WHO study. This study was designed to allow an evaluation of the reliability of the proxy-derived data, as recommended in a recent review of the use of proxy respondents.[1]

## Methods

The WHO collaborative study is a hospital based, case–control study undertaken in 17 countries. It was designed to evaluate the risk of developing either a venous thromboembolic event, a myocardial infarction, or a stroke in

*Table 1a Distributions of reported oral contraceptive (OC) use by true/proxy respondents. Figures in parentheses are the quadratic disagreement weights used to calculate weighted Kappa statistics*

| Proxy respondent | True controls | | | |
|---|---|---|---|---|
| | Never users | Past OC users | Current OC users | Don't know |
| Never users | 212 (0) | 30 (1) | 2 (4) | 0 (0.5) |
| Past OC users | 3 (1) | 97 (0) | 0 (1) | 0 (0.5) |
| Current OC users | 0 (4) | 1 (1) | 26 (0) | 0 (0.5) |
| Don't know | 20 (0.5) | 6 (0.5) | 6 (0.5) | 0 (0) |

*Table 1b Data reported in table 1a collapsed into a 2 × 2 table – ie current oral contraceptive (OC) use and non-current OC use (never/past). Don't know responses are omitted*

| Proxy respondent | True controls | |
|---|---|---|
| | Current | Non-current |
| Current | 26 | 1 |
| Non-current | 2 | 342 |
| Total | 28 | 343 |

Sensitivity = 26/28 = 0.929 (95% CI: 0.765, 0.991). Specificity = 342/343 = 0.997 (95% CI: 0.984, 1.000).

association with the current use of OC. Women aged 20–44 years admitted to hospital with a first time diagnosis of any of the three study diseases not associated with pregnancy or surgical operations and who did not die within 24 hours of admission were study cases. Controls were women matched according to hospital and time of admission and five year age band, admitted with one of a series of specified diagnoses believed to be not associated with OC use.

The study questionnaire included questions on age, parity, smoking habit, alcohol consumption, marital status, educational attainment, family history of cardiovascular disease, and medical and contraceptive history. The latter section included types of contraceptives used, brands and periods of OC use, and, when applicable, reasons for stopping. Further details of the WHO study are provided elsewhere.[2]

Proxy cases were used if the true case was too ill to complete the questionnaire, or if the patient died 24 hours or more after hospital admission but before the questionnaire had been administered, or if speech or cognitive function were insufficient to allow an interview.

While in most circumstances the ideal proxy respondent would be the husband or partner, occasionally such a person did not exist or was unavailable. In this situation, the most suitable available relative or friend was interviewed. During the pilot phase and the main WHO study, which lasted four years, data on 10% (506) of cases were collected from proxy respondents in this way.

To evaluate the reliability of the proxy data for cases, proxy data were collected for the controls recruited for those cases who required a proxy respondent. Data were also collected in the usual way from controls (true controls) for these cases. Data from these two sets of controls ("proxy" and "true") were then compared. Where possible, the relationship (ie husband, friend, sister, etc) between the proxy and true control was the same as that between the proxy and true case with whom the control was matched.

These procedures were maintained throughout the pilot phase of the study and during the first two years of the main study. During that time, 403 pairs of proxy and true controls were interviewed and were available for this evaluation.

The same questionnaire was administered to proxy cases and controls as to true cases and controls except for an additional explanatory paragraph which was read to the proxy respondents.

The concordance between the information acquired from true controls and their proxy respondents was evaluated using the Kappa statistic[5] for dichotomous and nominal data. Weighted Kappa statistics using quadratic disagreement weights were used for ordinal data (see example in table 1a). These give maximum score for full agreement, minimum score for classifying a current user as a never user (or vice versa) with an intermediate score for partial disagreement or a "don't know" response. The Kappa statistics describe the percentage agreement above that expected by chance alone and are very sensitive to the prevalence of the variable being investigated and the selection of the weights used. They range from 1 (perfect agreement) through 0 (level of agreement expected by chance alone) to negative values (less agreement than expected by chance). Although high levels of exact agreement may exist, the Kappa value may be low if the agreement expected by chance is high.

The sensitivity and specificity of the data obtained from proxy respondents regarding OC exposure (current user, non-user, don't know) were calculated from the pairs of proxy and true controls and applied for each disease to the proportion of cases that required proxy respondents to estimate the impact of the exposure misclassification on the reported odds ratio and sample size (see Appendix).

The intra-class correlation coefficient was used to evaluate agreement on continuous variables.[6] This measure adjusts correlation for systematic bias, being large when "differences among subjects account for a large proportion of the variance relative to the error variance".[1]

## Results

During this evaluation, 403 pairs of "true" and "proxy" controls were interviewed, of whom 385 (95.5%) were matched with stroke cases, nine with pulmonary embolus cases, and nine with acute myocardial infarction cases. Husbands or partners (242) were the most common category of proxy respondent interviewed, with mothers (61) and sisters (60) being the next most common.

Tables 1a and 1b compare responses of true and proxy respondents regarding the key exposure variable in the main study (current, past, and never use of OC). Comparing current/non-current OC user status as reported by true and proxy controls, the sensitivity and specificity of the proxy derived responses were 92.9% (95% CI: 76.5, 99.1) and 99.7% (95% CI: 98.4, 100.0) respectively.

*Table 2a Agreement between true controls and their proxy respondents in relation to contraceptive usage – discrete variables*

| Variables | Husband* (n=242) | | Mother (n=61) | | Sister (n=60) | | All proxy categories (n=403) | |
|---|---|---|---|---|---|---|---|---|
| | % Exact agreement | Kappa (95% CI) | % Exact agreement | Kappa (95% CI) | % Exact agreement | Kappa (95% CI)% | Exact agreement | Kappa (95% CI) |
| Oral contraceptive use† | 90.9 | 0.90 (0.87, 0.94) | 78.7 | 0.64 (0.39, 0.89) | 81.7 | 0.79 (0.67, 0.91) | 83.1 | 0.82 (0.77, 0.87) |
| Injectable contraceptive use† | 94.2 | 0.90 (0.85, 0.95) | 90.2 | 0.49 (−0.08, 1.00) | 78.3 | 0.19 (−0.22, 0.60) | 87.6 | 0.75 (0.67, 0.83) |
| Intrauterine device use† | 92.1 | 0.92 (0.87, 0.97) | 93.4 | 0.88 (0.71, 1.00) | 91.7 | 0.85 (0.67, 1.00) | 89.8 | 0.88 (0.85, 0.93) |
| Oral contraceptive brand (current use)‡ | 75.0 | 0.72 (0.50, 0.93) | 0 | — | 25.0 | 0.14 (−0.34, 0.63) | 47.1 | 0.41 (0.23, 0.60) |
| Oral contraceptive brand (most recent used by past users)‡ | 44.2 | 0.39 (0.28, 0.50) | 14.3 | 0.11 (−0.16, 0.35) | 36.8 | 0.30 (0.06, 0.54) | 38.8 | 0.34 (0.25, 0.43) |

* Husband = husband/partner. † Possible responses: current/past/never/don't know: weighted Kappa values.[7] ‡ 17 possible brands recorded: unweighted Kappa values.

*Table 2b Agreement between true controls and their proxy respondents in relation to contraceptive usage – continuous variables*

| Variables | Husband* | | | | Mother | | | | Sister | | | | All proxy categories | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n_1/n_2$ | Mean** difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC |
| Duration of current oral contraceptive use (current users) [months] | 18/20 | −0.4 | 2.2 | 0.98‡ | 2/6 | −25.0 | 17.0 | 0.52 | 2/4 | −43.0 | 9.0 | 0.03 | 22/34 | −6.5 | 3.7 | 0.94‡ |
| Duration of last oral contraceptive used (past-users) [months] | 66/94 | −3.6 | 2.7 | 0.79‡ | 2/7 | 42.0 | 42.0 | 0.35 | 7/19 | 4.3 | 1.6 | 0.89† | 77/133 | −1.4 | 2.2 | 0.77‡ |

ICC = intra-class correlation.
* Husband = husband/partner. † $p<0.001$. ‡ $p<0.0001$. $n_1$ = number of true/proxy pairs in analyses. $n_2$ = possible number of true/proxy pairs. ** True control minus proxy respondent.

Table 2a includes the concordance of data relating to method of contraception and, where applicable, brand of OC used, obtained from all control pairs and also for the three major categories of proxy pairs recruited (husbands, mothers, and sisters). In general, husbands provided more reliable information than other proxy categories. The higher Kappa values for husbands partly reflects the lower rates of "don't know" responses for some of these questions compared with the rates among all other categories of proxy controls. While levels of agreement, especially with husbands, regarding method of contraception were good, agreement as to the brand of current and past OC used was less impressive. These latter comparisons could only be made between true controls who were current or past OC users and their proxy respondents, and hence the numbers eligible for comparison are reduced. The data suggest that sisters and mothers, but not husbands, were better informed about the brands of OC used in the past than those currently in use. Of the 34 true controls who were current OC users, 20, 6, and 4 had husband, mother, and sister proxy respondents, of whom 20, 4, and 2 respectively accurately reported this. Only 14 husbands and 1 sister reported the brand of OC in current use. Of the 134 true controls who were past OC users 95, 7, and 19 had husband, mother and sister proxy controls of whom 80, 2, and 13 respectively, accurately reported this. Forty three husbands, 1 mother, and 7 sisters reported the brand of OC last used by these controls.

The calculation of the validity of responses regarding continuous variables using the intra-class correlation coefficient (tables 2b and 3b)

*Table 3a Agreement between true controls and their proxy respondentsin relation to discrete lifestyle variables*

| Variables | Husband* (n=242) | | Mother (n=61) | | Sister (n=60) | | All proxy categories (n=403) | |
|---|---|---|---|---|---|---|---|---|
| | % Exact agreement | Kappa (95% CI) | % Exact agreement | Kappa (95% CI) | % Exact agreement | Kappa (95% CI) | % Exact agreement | Kappa (95% CI) |
| Current alcohol drinker (yes/no) | 94.2 | 0.88 (0.82, 0.94) | 88.5 | 0.72 (0.52, 0.93) | 95.0 | 0.90 (0.79, 1.00) | 92.6 | 0.85 (0.80, 0.90) |
| Smoking (never/ex/current) | 94.6 | 0.92 (0.88, 0.97) | 95.1 | 0.84 (0.63, 1.00) | 100.0 | 1.00 (1.00, 1.00) | 94.8 | 0.90 (0.85, 0.95) |
| Marital status (married or stable union/separated or divorced/single)† | 98.8 | 0.94 (0.86, 1.00) | 98.4 | 0.97 (0.91, 1.00) | 90.0 | 0.85 (0.74, 0.96) | 96.8 | 0.94 (0.90, 0.97) |
| Educational attainment (nil/primary/secondary/ technical/university) | 91.7 | 0.95 (0.92, 0.97) | 85.2 | 0.90 (0.82, 0.98) | 93.3 | 0.93 (0.87, 1.00) | 90.6 | 0.93 (0.91, 0.96) |

* Husband = husband/partner. † Unweighted Kappa values.

*Table 3b Agreement between true controls and their proxy respondentsin relation to age and anthropometry*

| Variables | Husband* | | | | Mother | | | | Sister | | | | All proxy categories | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n_1/n_2$ | Mean difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC | $n_1/n_2$ | Mean difference | SE | ICC |
| Age (years) | 242/242 | 0.03 | 0.01 | 1.00† | 60/61 | −0.92 | 0.92 | 1.00‡ | 60/60 | 0.07 | 0.08 | 1.00‡ | 398/403 | 0.03 | 0.02 | 1.00‡ |
| Weight (kg) | 149/226 | 0.04 | 0.39 | 0.91† | 29/50 | 0.31 | 0.58 | 0.99† | 21/43 | 0.43 | 0.79 | 0.87† | 212/348 | −0.09 | 0.31 | 0.93† |
| Height (cm) | 146/194 | −0.85 | 0.34 | 0.80† | 28/46 | 0.29 | 0.50 | 0.90† | 21/31 | −0.33 | 0.88 | 0.96† | 208/292 | −0.61 | 0.27 | 0.83† |

ICC = intra-class correlation.
* Husband = husband/partner. $n_1$ = number of true/proxy pairs in analyses. $n_2$ = possible number of true/proxy pairs. † $p<0.001$; ‡ $p<0.0001$.

*Table 4 Agreement between true controls and their proxy respondents in relation to medical history*

| | *Husband* (n=242)* | | *Mother (n=61)* | | *Sister (n=60)* | | *All proxy categories (n=403)* | |
|---|---|---|---|---|---|---|---|---|
| | *% Exact agreement* | *Kappa (95% CI)* | *% Exact agreement* | *Kappa (95% CI)* | *% Exact agreement* | *Kappa (95% CI)* | *% Exact agreement* | *Kappa (95% CI)* |
| High blood pressure | 97.5 | 0.64 (0.35, 0.94) | 98.4 | 0† | 93.3 | 0.68 (0.35, 1.00) | 96.3 | 0.67 (0.50, 0.84) |
| Diabetes | 99.2 | 0.76 (0.37, 1.00) | 100.0 | ND | 98.3 | 0† | 99.3 | 0.87 (0.68, 1.00) |
| Cancer | 99.2 | 0.50 (−0.20, 1.00) | 100.0 | ND | 100.0 | ND | 99.5 | 0.50 (−0.20, 1.00) |
| Other cardiac disease | 99.2 | 0.66 (0.20, 1.00) | 100.0 | ND | 100.0 | ND | 99.3 | 0.57 (0.01, 1.00) |
| Gall bladder disease | 97.5 | 0.64 (0.35, 0.94) | 98.4 | 0.66 (0.00, 1.00) | 98.3 | 0† | 98.0 | 0.68 (0.45, 0.91) |
| Psychiatric illness | 99.2 | 0.85 (0.65, 1.00) | 100.0 | ND | 96.7 | −0.01 (−1.00, 1.00) | 99.0 | 0.78 (0.55, 1.00) |
| Other chronic joint disease | 97.9 | 0.85 (0.73, 0.98) | 100.0 | ND | 98.3 | 0.66 (0.00, 1.00) | 98.3 | 0.82 (0.69, 0.95) |
| Surgical operation | 96.3 | 0.90 (0.83, 0.96) | 95.1 | 0.83 (0.64, 1.00) | 93.3 | 0.81 (0.64, 0.99) | 96.0 | 0.89 (0.83, 0.94) |
| Gynaecological | 94.2 | 0.34 (0.01, 0.68) | 98.4 | 0† | 98.3 | 0.66 (0.00, 1.00) | 94.8 | 0.32 (0.03, 0.61) |
| Varicose veins | 97.5 | 0.69 (0.44, 0.93) | 100.0 | ND | 98.3 | 0† | 97.8 | 0.60 (0.34, 0.86) |

* Husband = husband/partner.
ND = not defined. † Kappa and SE = 0 therefore CI not defined.

cannot incorporate "don't know" responses, and hence numbers of pairs compared are both fewer than those in tables 2a and 3a and vary between questions. Therefore, in addition to the intra-class correlation coefficient (ICC) which evaluates concordance among those who respond to any given question, it is also important to know the frequency of "don't know" responses among the proxy categories. Hence in tables 2b and 3b, the numbers of those who were included in the analyses (excluding "don't know" responses) and the numbers who could have responded to each question are shown. It is clear from tables 2b and 3b that a larger proportion of husbands attempted to answer these questions, and produced more accurate responses when they did so, for the variables shown in table 2b, although not in table 3b

Tables 3a and 3b demonstrate that all categories of proxy controls provided reliable data on reported alcohol intake, smoking habit, marital status, educational attainment, age, height, and weight, although for the last two variables only two thirds of the proxy respondents attempted an estimate. Table 4 demonstrates, by category of proxy, the agreement between proxy and true controls on medical history. Zero values of Kappa were observed despite high levels of exact agreement due to the low prevalence of the conditions. All categories of proxy controls were reliable regarding this type of data, with mothers sometimes providing even more accurate information than husbands and sisters. For certain conditions, Kappa could not be calculated because all true and proxy controls gave the same response.

*Table 5 Rates (%) of "don't know" responses to selected questions among proxy cases and proxy controls*

| | *Proxy cases (n=506)* | *Proxy controls (n=403)* | *p* |
|---|---|---|---|
| Oral contraceptive usage | 5.3 | 7.9 | NS |
| Injectable | 5.5 | 7.2 | NS |
| Intra-uterine device | 3.7 | 4.7 | NS |
| Diaphragm | 6.3 | 8.4 | NS |
| History of hypertension | 1.4 | 0.7 | NS |
| History of diabetes | 0.8 | 0.5 | NS |
| History of rheumatic heart disease | 1.2 | 0.0 | <0.05 |
| Smoking | 0.6 | 1.2 | NS |
| Alcohol consumption | 0.2 | 0.2 | NS |
| Educational attainment | 0.8 | 1.2 | NS |
| Family history of acute myocardial infarction | 5.3 | 6.7 | NS |
| Family history of stroke | 5.3 | 6.5 | NS |
| Oral contraceptive type (current) (no)* | 15.2 (79) | 41.2 (34) | <0.01 |
| Oral contraceptive type (past use) (no)* | 27.5 (120) | 33.6 (134) | NS |
| Oral contraceptive duration (current) (no)* | 12.7 (79) | 14.7 (34) | NS |

* Numbers of true cases and controls who were current or past oral contraceptive users.

Table 5 compares the rates of "don't know" responses to several key study variables reported by the 403 proxy controls included in this evaluation study with those reported by the 506 proxies for cases who could not be interviewed in the pilot and four year main study. Proxy cases provided "don't know" responses less frequently than proxy controls on 11 of the 15 variables, significantly so for the brand of OC used by current users. Table 6 compares the distributions of different categories of proxy respondents used for cases and controls and shows that although they appear similar, the overall distribution was statistically significantly different for cases compared with controls ($\chi^2 = 27.1$: $p < 0.0001$) in that more husband and mother and fewer sibling and children proxy controls were used compared with cases.

The levels of agreement between proxy and true controls for the key exposure investigated in the main WHO study (current/past/never OC use) are shown by region in table 7. Sample size limits the ability to detect significant differences among the regions but the table suggests that the optimal category of proxy control may differ from region to region. For example, husbands in Africa and mothers in Latin America may be less suitable as proxy controls than their counterparts in the other regions.

When the same levels of sensitivity and specificity of OC exposure achieved by proxy controls (table lb) are applied to the numbers of cases for whom proxy respondents were required (see Appendix), the odds ratio for stroke was estimated to have changed by 2.9% (table 10). This limited impact was due, in particular, to the high specificity of the proxy responses and is achieved despite a relatively large proportion of proxy cases (16.2%). The

*Table 6 Distribution of categories of proxy respondents used for cases and controls (% of pairs)*

| | *Cases (n=506) No (%)* | *Controls (n=403) No (%)* |
|---|---|---|
| Husband/partner | 273 (54.0) | 242 (60.0) |
| Father | 3 (0.6) | 1 (0.2) |
| Mother | 59 (11.7) | 61 (15.1) |
| Brother | 13 (2.6) | 1 (0.2) |
| Sister | 84 (16.6) | 60 (14.9) |
| Son/daughter | 35 (6.9) | 6 (1.5) |
| Other | 39 (7.7) | 32 (7.9) |
| Total | 506 (100.0) | 403 (100.0) |

*Table 7 Rates of exact agreement on oral contraceptive use status* among proxy and true controls*

| | *Husband†* | *Mother* | *Sister* | *All* |
|---|---|---|---|---|
| Africa | | | | |
| Agreement % | 66.7 | 81.8 | 87.0 | 80.0 |
| No | 3 | 11 | 23 | 45 |
| Weighted Kappa | 0.73 (0.29, 1.00) | 0 (−1.00, 1.00) | 0.56 (0.05, 1.00) | 0.46 (0.06, 0.86) |
| Asia | | | | |
| Agreement % | 93.8 | 90.3 | 77.8 | 88.6 |
| No | 162 | 31 | 18 | 228 |
| Weighted Kappa | 0.93 (0.88, 0.98) | 0.82 (0.57, 1.00) | 0.52 (0.01, 1.00) | 0.87 (0.81, 0.93) |
| Europe | | | | |
| Agreement % | 91.2 | 100 | — | 87.2 |
| No | 34 | 2 | 0 | 39 |
| Weighted Kappa | 0.87 (0.72, 1.00) | 1.00 (1.00, 1.00) | — | 0.74 (0.49, 0.99) |
| Latin America | | | | |
| Agreement % | 81.4 | 52.9 | 78.9 | 69.2 |
| No | 43 | 17 | 19 | 91 |
| Weighted Kappa | 0.75 (0.59, 0.91) | 0.17 (−0.47, 0.81) | 0.64 (0.25, 1.00) | 0.57 (0.40, 0.75) |

* Current/past/never/don't know.
† Husband = husband/partner.

estimated impact of not using any proxy data (and hence treating these cases as missing) was greater (3.9%), though still small. However, not using any proxy data would have required an increase of almost 20% in the sample size to achieve the same power, compared with only a 1% increase in sample size required to offset the rates of "don't know" answers from the proxy cases. For the two other main study end points (acute myocardial infarction and pulmonary embolism/deep vein thrombosis), only 3.3% and 0.9% of cases, respectively, required proxy respondents, so the potential impact on sample size is much smaller. The estimated impact on the odds ratio for acute myocardial infection (5% increase) is greatest since 32.0% of the true cases reported OC use, while none of the nine surrogates reported that the cases were current OC users. Nevertheless, in view of the small numbers involved and the precision of the estimated odds ratios, the use of proxy data has a negligible impact on the results of the study.

## Discussion

Overall, the levels of concordance between data from proxy and true controls were high for most recent issues, and husbands appeared to be the best proxy for several critical variables, in keeping with the results of previous studies.[7] However, for variables such as smoking habit and body weight, other proxy controls (sisters and mothers respectively) were more reliable, highlighting the fact that the optimal proxy tends to be question-specific and hence it may be inappropriate to expect similar levels of agreement from proxy data in different studies. As might be expected, responses requiring more detail, such as the duration or brand of OC use, were less accurately reported.

Evidence for the reliability of data derived from proxy respondents is limited, and particularly regarding OC use[1] because only two studies,[3,4] involving 138 and 99 pairs of subjects respectively, have evaluated such information. In the first of these studies only current OC use was evaluated, and in the second only husbands were involved. This report is the first to include data from proxy respondents on several aspects of OC use from a number of different types of proxy respondents from a wide range of countries.

The purpose of evaluating the data produced by proxy controls is to extrapolate the findings to the proxy data which had to be collected on 10% of the cases recruited in the main WHO study.[2] This design, however, cannot measure or account for differential recall of episodes of OC use between proxy cases and controls, because cases had more serious illnesses than controls. Ideally, therefore, this evaluation of proxy data would also have compared data from true and proxy cases. However, because the questions relating to OC use were only a small part of the total questionnaire, important recall bias on this key exposure was less likely. Data in table 5 suggest that important differences in recall bias between proxy cases and controls were not apparent, at least with regard to the method of contraception used. However, the large and highly significant differences between "don't know" response rates among case control proxy respondents regarding the brand of OC used is difficult to explain. It may be that proxy cases, as a result of the death or the seriousness of the true case's illness, are more likely to take over the possessions of true cases and hence have direct access to the information necessary to identify the OC in current use. A second possible explanation for obtaining different responses from proxy cases and controls is that the relationship between true and proxy cases are different from those between true and proxy controls. Although table 6 shows significant differences in the relationship of the proxy repondents to their respective case or control, these are unlikely to affect the accuracy of the responses since all categories of proxy controls had high levels of agreement.

Assuming the misclassification of OC exposure among proxy controls was the same as for proxy cases, the estimated impact on risk estimates of developing any of the three cardiovascular events studied in association with OC use was small, due to the high sensitivity and, more importantly, specificity of the proxy responses. The impact of misclassification is less than would result if those cases who could not supply information were omitted. While a

certain number of proxy cases must be expected not to know certain information about the cases on whom they are providing information, the loss of power due to these "don't know" responses is small compared with the estimated loss due to not using proxy cases.

While reservations concerning the misuse and misinterpretation of the Kappa statistic have been expressed,[8] this method was preferred for dichotomous or ordinal data because of its ability to allow for concordance expected by chance.[5 9 10] Unfortunately, the Kappa value is greatly affected by the prevalence of the variable under investigation,[10] which in part explains the different Kappa values shown in the tables for the same degree of exact agreement. This property of the Kappa statistic limits comparability of results from one study to another. Although the $\chi^2$ test and Pearson correlation coefficients relate associations in the context of chance, neither test differentiates between increased disagreement or increased agreement. Correlation analyses were only used for continuous variables and then only as intraclass coefficients[6] since this method is unable to detect important systematic bias as a result of using one category of respondent which may be present, even when the correlation is strong.

Although reliability is a prerequisite for validity, at least at the individual level, the data presented above do not necessarily imply validity. The data should also be viewed in the light of how "reliable" data might be if true cases and controls were interviewed twice and comparisons made between the two interviews.

In conclusion, this study demonstrates that information on many variables, including current OC use, obtained from proxy controls was reliable. It seems reasonable to assume therefore, that for this variable, the use of data from proxy cases is sufficiently reliable. For OC use and several other variables, husbands were the most reliable category of proxy. However, the optimal category of proxy appears question-specific and the reliability of proxy data from any source was greatly diminished when detailed information, particularly from the past, was required.

Principal investigators in this study were: Myriam Debert-Ribeiro, Escola Paulista de Medicina, São Paulo, Brazil; Ernesto Medina, Universidad de Chile, Santiago, Jorge Artigas, Escuela de Medicina, Valparaiso, Chile; Shen He, National Research Institute for Family Planning, Beijing, Zhong Yu Hui, Sichuan Family Planning Research Institute, Chengdu, Zhang De-Wei, Zhao Weijin, Shanghai Institute for Planned Parenthood Research, Shanghai, People's Republic of China; Oscar Rojas, Universidad del Valle, Cali, Colombia; NR Poulter, CL Chang, MG Marmot, UCL Medical School, London, Martin Vessey, Department of Public Health and Primary Care, University of Oxford, Oxford, England; Lothar Heinemann, Zentrum für Epidemiologie und Gesundheitsforschung, Berlin, Germany; Stuart Donnan, Suzanne Ho, Chinese University of Hong Kong, Hong Kong; Gyorgy Bartfai, Albert Szent-Gyorgyi Medical University, Szeged, Hungary; J Kisjanto, University of Indonesia, Jakarta, Indonesia; Rainford Wilks, University of the West Indies, Kingston, Jamaica; Reuben Agwanda, Kenya Medical Research Institute, Nairobi, Kenya; Rossario Ruiz, Grupo Interuniversitario Mexicano de Investigacion Epidemiologica en Salud Reproductiva, Durango, Mexico; Mateja Kozuh-Novak, University Institute of Public Health, Ljubljana, Slovenia; TMM Farley, S Holck, O Meirik, World Health Organization, Geneva, Switzerland; Nikorn Dusitsin, Pramuan Virutamasen, Kammant Phanthumchinda, Chulalongkorn Hospital, Suporn Koetsawang, Manee Piya-Anant, Siriraj Hospital, Bangkok, Thailand; Jasenka Demirovic, Karen Belkic, University of Belgrade, Belgrade, Federal Republic of Yugoslavia; WS Mwandila, CM Mutale, University Teaching Hospital, Lusaka, Zambia; Jonathon Matenga, Adrian Wilson, University of Zimbabwe, Harare, Zimbabwe. This study received primary financial support from the UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, and supplemental support from the USA National Institutes of Health Contraceptive and Reproductive Evaluation Branch. We thank J Kelaghan, O Meirik, J Olsen, M Shipley and M Thorogood for their advice in producing the manuscript and Ms Sandra Johnson for its preparation.

## Appendix

### IMPACT OF PROXY DATA ON ESTIMATED ODDS RATIOS AND POWER

The impact of the use of proxy respondents of estimated odds ratios depends on the sensitivity and specificity of these responses, the prevalence of OC use as well as on the proportion of proxy cases required. Suppose the prevalence of OC use reported by true cases is $p_C$, that $r$ is the proportion of proxy cases and that a proportion $p_p$ of these report that the true case was a current OC user. Then the estimated true prevalence of OC use among cases requiring a proxy respondent, $\pi_p$, is given by:

$$\pi_p = [p_p - (1 - \text{Specificity})] / [\text{Sensitivity} - (1 - \text{Specificity})]$$

[see Copeland *et al*[12]]. Reported exposure among all cases including proxy cases, $p_R$, is given by:

$$> -p_R = (1-r)p_C + r\,p_p \text{ and estimated exposure is } p_E = (1-r)\,p_C + r\,\pi_p$$

The estimated impact on the odds ratio of using proxy responses in place of true responses is $p_R/(1-p_R)/[p_E/(1-p_E)]$.

The argument above can be extended to incorporate "don't know" responses from proxy respondents to estimate the true prevalence of OC use among cases. Suppose that proportions $\gamma_1$ and $\gamma_0$ of OC users were reported by their proxies as users or non-users, respectively, and for a proportion $\gamma_U$ the proxy reported that they did not know the OC exposure status (see table 8). Similarly, let $\delta_1$, $\delta_0$ and $\delta_U$ be the corresponding proportions of those who were non-OC users. Estimates of these misclassification rates are available from the comparison of responses between proxy and true controls (table 1b). The EM algorithm[11] is used to estimate the proportion of those reported as unknown who are actually exposed, and the misclassification rates are used to estimate how many of these would be reported as exposed or not exposed, respectively. Then Copeland's argument,[12] conditional on a reported exposure status, is applied to compute the actual exposure rate among all subjects. The new estimate of exposure is used to update the reported exposure among the unknowns,

*Table 8 Actual and reported exposure among subjects*

| *Reported status* | *True status* | | *Observed exposure* |
|---|---|---|---|
| | *Exposed* | *Not exposed* | |
| Exposed | $\gamma_1$ | $\delta_1$ | $n_1$ |
| Not exposed | $\gamma_0$ | $\delta_0$ | $n_0$ |
| Don't know | $\gamma_U$ | $\delta_U$ | $n_U$ |
| Total | 1 | 1 | |

*Table 9 Cases and proxy respondents by diagnosis*

| | True cases | Cases requiring proxies | |
|---|---|---|---|
| | | Reported | Estimated |
| Stroke | | | |
| Exposed | 327 | 63 | 76.9 |
| Not exposed | 1577 | 283 | 294.1 |
| Don't know | 0 | 25 | 0.0 |
| Total | 1904 | 371 | 371.0 |
| Myocardial infarction | | | |
| Exposed | 85 | 0 | 0.0 |
| Not exposed | 181 | 9 | 9.0 |
| Don't know | 0 | 0 | 0.0 |
| Total | 266 | 9 | 9.0 |
| Pulmonary embolus or deep vein thrombosis | | | |
| Exposed | 518 | 2 | 2.1 |
| Not exposed | 688 | 9 | 8.9 |
| Don't know | 0 | 0 | 0.0 |
| Total | 1206 | 11 | 11.0 |

*Table 10 Impact of proxy responses on odds ratio and sample size*

| | No of cases | Exposure (%) | Odds ratio (% change) |
|---|---|---|---|
| Stroke | | | |
| All cases | 2275 | 17.8 | — |
| True cases only | 1904 | 17.2 | −3.93 |
| True cases and proxies | 2250 | 17.3 | −2.85 |
| Myocardial infarction | | | |
| All cases | 275 | 30.9 | — |
| True cases only | 266 | 32.0 | 5.02 |
| True cases and proxies | 275 | 30.9 | 0.05 |
| Pulmonary embolus or deep vein thrombosis | | | |
| All cases | 1217 | 42.7 | — |
| True cases only | 1206 | 43.0 | 0.88 |
| True cases and proxies | 1217 | 42.7 | −0.04 |

and the procedure is repeated until convergence.

Suppose that a proportion $\pi$ of the subjects are in fact exposed, then the proportion whose proxy would respond that they do not know the exposure status, $p_U$, is given by:

$$p_U = \pi\ \gamma_U + (1-\pi)\ \delta_U$$

The proportions of these women who are, and are not, in fact exposed are $\pi\gamma_U/[\pi\gamma_U+(1-\pi)\delta_U]$ and $(1-\pi)\delta_U/[\pi\gamma_U+(1-\pi)\delta_U]$, respectively. Hence if the exposure for $n_U$ women is reported as "don't know", then $n_1' = n_U\pi\gamma_U/[\pi\gamma_U+(1-\pi)\delta_U]$ women are and $n_0' = n_U(1-\pi)\delta_U/[\pi\gamma_U+(1-\pi)\delta_U]$ are not actually exposed. In view of the misclassification rates between the actual and reported exposure, the number who are reported as exposed is $n_1'\gamma + n_0'\delta$, where $\gamma = \gamma_1/(\gamma_1+\gamma_0)$ and $\delta = \delta_1/(\delta_1+\delta_0)$ are the probabilities of being reported as exposed or not, respectively, conditional on the proxy respondent providing an exposure status. Similarly the number who are reported as not exposed is $n_1'(1-\gamma)+n_0'(1-\delta)$. Thus a total of $n_1'' = n_1 + n_1'\gamma + n_0'\delta$ women are reported as exposed and $n_0'' = n_0 + n_1'(1-\gamma) + n_0'(1-\delta)$ as not exposed. The estimated proportion of women reported as exposed is:

$$p' = n_1''/(n_1''+n_0'').$$

Copeland's argument[12] is applied to the reported proportion of exposed, $p'$, using theconditional misclassification rates $\gamma$ and $\delta$, to estimate the actual proportion exposed:

$$\pi' = (p'-\delta)/(\gamma-\delta)$$

where $\gamma$ and $(1-\delta)$ correspond to the sensitivity and specificity, respectively. This value $\pi'$ is used as the revised estimate and the procedure repeated until convergence.

Table 9 gives the numbers of true and proxy cases for each of the three main diseases in the WHO study and uses the information provided by the proxy and true controls to estimate the distribution of exposure among proxy cases. Table 10 shows, for each disease, the number of cases and the rates of exposure based on the computations in table 9. Exposure among all cases corresponds to that which could theoretically have been observed if all cases had provided information. The impact on estimated odds ratios is shown as the percentage change from the theoretical value by using only data from true cases and by using true cases' data supplemented by proxy responses. The increase in sample size shows the number of additional cases that would have to be recruited to the study to have the same power as a study where all cases were able to provide information.


1 Nelson LM, Longstreth W Jr, Koepsell TD, Van Belle G. Proxy respondents in epidemiologic research. *Epidemiol Rev* 1990;**12**:71–86.
2 WHO Collaborative Study of Cardiovascular Disease and SteroidHormone Contraception. A multinational case–control study of cardiovascular disease and steroid hormone contraceptives: description and validation of methods. *J Clin Epidemiol* 1995;**48**:1513–47.
3 Krueger DE, Ellenberg SS, Bloom S, *et al.* Fatal myocardial infarction and the role of OCs. *Am J Epidemiol* 1980;**111**: 655–74.
4 Thorogood M, Vessey M. The reliability of surrogate information about oral contraceptive use, smoking, height and weight collected from men about their wives. *Contraception* 1989;**39**:401–8.
5 Fleiss JL. *Statistical methods for rates and proportions.* New York; John Wiley & Sons 1981:212–36.
6 Kramer MS, Feinstein AR. Clinical biostatistics. IV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981; **29**:111–23.
7 Pickle LW, Brown LM, Blot WJ. Information available from surrogate respondents in case–control interview studies. *Am J Epidemiol* 1983;**118**:99–108.
8 Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;**126**(2):161–9.
9 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–10.
10 Grove WM, Andreason NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis. Theory and practice. *Arch Gen Psychiatry* 1981;**38**:408–13.
11 Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society*, Series B 1977;**39**:1–38.
12 Copeland KT, Checkoway H. McMichael AJ, Holbrook RH Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;**105**:488–95.