J Clin Epidemiol Vol. 46, No. 5, pp. 431–434, 1993
Printed in Great Britain.

0895-4356/93 $6.00 + 0.00

THE EFFECT OF USING DIFFERENT REFERENCE DATES FOR CONTROL EXPOSURE MEASUREMENT ON RELATIVE RISK ESTIMATES IN A CASE–CONTROL STUDY

ALVIN H. RAMPEY JR,[1] KATHLEEN L. IRWIN,[2] MARK W. OBERLE,[2] STEVEN KINCHEN,[2] NANCY C. LEE,[2] ANNE MARSDEN[3] and LUIS ROSERO-BIXBY[4]

[1]Statistical and Mathematical Sciences, Lilly Research Laboratories, A Division of Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, [2]Centers for Disease Control, Atlanta, GA 30333, [3]Market Strategies, Inc., 1000 Town Center, Suite 1600, Southfield, MI 48075 and [4]Universidad de Costa Rica San Jose, Costa Rica

(Received in revised form 14 December 1992)

Abstract—In case–control studies in which case and control enrollment periods are not identical, exposure status for time-dependent variables is often measured relative to a reference date. Using data from a case–control study of the relation between cervical cancer and oral contraceptive (OC) use in which control enrollment began 6 months after the end of case enrollment, we evaluated the effect on odds ratios from using five different reference dates to determine the controls' exposure status. The choice of reference date had little effect on the odds ratios in this study. Reference dates for time-dependent exposure variables should be considered carefully in studies when case and control enrollment periods are not identical.

Case–control study Odds ratios Relative risk

INTRODUCTION

Case–control studies of rare diseases often require long or retrospective case enrollment periods to ensure adequate numbers of participants. In some instances, case and control enrollment periods are not identical. To ensure that cases and controls in studies with non-identical enrollment periods have a similar lifetime opportunity for exposure to time-dependent factors, many investigators assign subjects a reference date and consider only those exposures that occurred before this date.

In our recent study of the relation between oral contraceptives (OCs) and cervical cancer in Costa Rica, case-patients identified through a tumor registry were eligible if diagnosed from January 1982 through March 1984 [1]. Only OC use before the date of the case-patient's diagnostic biopsy was considered as a possible etiologic exposure. The controls were identified through a nationwide household survey from September 1984 through February 1985. A single reference date, 15 February 1983, the midpoint of the case diagnosis period, was assigned to all controls. Only OC use before this reference date was considered for each control. Exposure data was subsequently calculated prior to the reference date for each control subject. This paper summarizes our investigation into the effects of this reference date choice for the controls on the resulting odds ratios and confidence intervals for the relation between OC use and invasive cervical cancer and carcinoma *in situ* of the cervix.

METHODS

Detailed methods of this population-based case–control study have been previously reported [1, 2]. The 876 cases (representing 583

cases of carcinoma *in situ* [CIS] and 293 cases of invasive cervical cancer) enrolled in this study were selected from the Costa Rican National Tumor Registry. These cases were women who were newly diagnosed from 1 January 1982 through 31 March 1984. This period was defined as the case enrollment period. Because cases were interviewed up to several years after their diagnosis, their OC use was calculated relative to their date of diagnosis, not their date of interview. For example, if a woman first used oral contraceptives after her date of diagnosis, she was classified as having never used oral contraceptives. Interviewers attempted to enhance recall by recording important life events and intervals of contraceptive use on a month-by-month life history calendar [3].

The 938 controls enrolled in this study were chosen using a one-time, nationwide household survey conducted from September 1984 through February 1985. Cluster sampling was based on the June 1984 national census sampling frame with an interview completion rate of 92.8%. They were interviewed in person using the same standardized questionnaire as was used for case interviews. Since, by definition, the controls were disease-free, they did not have a date of disease diagnosis and hence did not have an obvious reference date. In the primary analysis [1], all controls were assigned the midpoint of the case enrollment period as their reference date to ensure that their lifetime opportunity for OC use was similar to that of the cases.

To assess the effect that our choice of a reference date for the controls had on the resulting odds ratios, we evaluated four additional choices of reference dates for the control subjects: a different, randomly chosen date from 1 January 1982 through 31 March 1984 (the range of diagnosis dates for the cases) for each control; the date of each control's interview; the endpoint of the case enrollment period (a single date for all controls); and the beginning of the case enrollment period (a single date for all controls). The randomly chosen date was generated using Fishman and Moore's method adapted by SAS [4].

We then calculated odds ratios (OR) and 95% confidence intervals (CI) using each of the four additional reference dates and compared them with the OR and 95% CI calculated using the original reference date, the midpoint of the case enrollment period. For each analysis, we calculated odds ratios using the logistic regression model [4]. The model included as independent variables ever use of OCs and age at the reference date plus the following confounding factors, all of which were determined relative to the assigned reference date: gravidity; number of lifetime sex partners; age at first coitus; history of any sexually transmitted disease or pelvic inflammatory disease; and history of Pap smears before 1982 (the beginning of the case enrollment period). The total number of eligible controls in the final models varied slightly for each analysis for two reasons: first, age at reference date varied according to which reference date was used and only women age 25–58 years at their reference date were included in analyses, and second, the confounding factor age at first coitus was included in the final model and only women who had intercourse at least once before their reference date were included in analyses. The analyses included only those women for whom values for all confounding factors were known. In all analyses, women who had never used OCs served as the reference group.

RESULTS

Very little difference existed in the odds ratios reported for any of the five reference date choices for either invasive cancer or CIS (Table 1). The slight differences in odds ratios were dependent upon the proportion of cases whose date of diagnosis was before or after the chosen reference date for the controls. As the control reference date was moved later in time, the number of controls classified as having used OCs before their reference date increased slightly. Thus, the odds ratios decreased when controls were assigned later reference dates. In this study, the cases' dates of diagnosis were fairly uniformly distributed over the enrollment period such that 48.0% of the cases were enrolled before the midpoint of the case enrollment period. This may explain why use of different reference dates did not appreciably change the odds ratios. The interpretation of the results was unchanged: compared with women who had never used OCs, women who had used OCs had no increase in risk of invasive cervical cancer and only a slightly elevated risk of CIS, regardless of which control reference date was used[1].

Table 1. Risk of invasive cervical cancer and carcinoma *in situ* associated with use of oral contraceptives by reference date choice from the Costa Rica cancer study

Control reference date	Number of cases			Number of controls*			Odds ratio (OR)	95% CI
	Ever used OCs	Never used OCs	Unknown if used OCs	Ever used OCs	Never used OCs	Unknown if used OCs		
Invasive Cervical Cancer								
Random†	48	81	10	292	332	56	0.83	0.51–1.33
Beginning of case enrollment (1/1/82)	48	81	10	273	324	53	0.84	0.52–1.37
Midpoint of case enrollment (2/15/83)	48	81	10	300	331	56	0.79	0.49–1.28
End of case enrollment (3/31/84)	48	81	10	317	327	56	0.71	0.44–1.14
Date of interview	48	81	10	324	324	59	0.64	0.40–1.03
Carcinoma *in situ*								
Random†	256	111	25	292	332	56	1.63	1.19–2.25
Beginning of case enrollment (1/1/82)	256	111	25	273	324	53	1.67	1.21–2.31
Midpoint of case enrollment (2/15/83)	256	111	25	300	331	56	1.59	1.15–2.18
End of case enrollment (3/31/84)	256	111	25	317	327	56	1.43	1.04–1.96
Date of interview	256	111	25	324	324	59	1.29	0.94–1.77

*Total number of controls varies in each analysis because of exclusion of controls with unknown or invalid values for age and confounding factors.
†Randomly assigned date between 1 January 1982 and 31 March 1984.

DISCUSSION

In summary, the use of different reference dates for the controls had little effect on the final risk estimates obtained in this study. However, varying the reference date for the controls could alter conclusions based on statistical significance drawn from a study as evidenced by the fact that for CIS, only one of the five reference date choices yields a 95% CI which includes 1.

Since the cases' dates of diagnosis were fairly uniformly distributed over the 27-month enrollment period, assigning each control a different reference date randomly selected from the beginning to the end of the enrollment period would provide the controls with a lifetime OC exposure opportunity most similar to that of the cases. However, the results obtained were not markedly different from those from any of the other methods. Further, there are practical disadvantages to random assignment of control reference dates. For a given set of 938 reference dates chosen randomly from the case enrollment period, there are $938! = 938 \times 937 \times 936 \times \ldots \times 3 \times 2 \times 1$ ways to assign each of these reference dates to each control. Theoretically, the way in which these dates are assigned may cause the calculated exposure values to change enough to cause the odds ratios to vary, depending on which control is assigned which reference date. Investigating all possible control reference date assignments is necessary to assess the effect of date reassignment on the final estimates, an impractical procedure for a data set this large. Other disadvantages are that this method, when used with logistic regression, is very computer-intensive and the methods are difficult to describe.

In this study, using each control's interview date as her reference date also has drawbacks. The cases were diagnosed between 1 January 1982 and 31 March 1984 and their exposures were truncated at the date of diagnosis; however, the controls were interviewed between 13 September 1984 and 28 January 1985. Because all of the controls were interviewed several months after the end of the case enrollment period, using their interview date as the reference date allows the controls, as a group, to have been at risk of OC exposure for a longer period of time than the cases, assuming OC availability was constant over time. Increasing the opportunity for exposure in the control group decreased the odds ratio for both the *in situ* and invasive analyses.

The other three choices assigned a single reference date to all controls, either the beginning (1 January 1982), the midpoint (15 February 1983), or the end (31 March 1984) of the case enrollment period. Using the beginning date gives an estimate that is greater than the odds ratio based on the midpoint of case enrollment because it reduces the proportion of ex-

posed controls. Similarly, use of the end date gives an estimate that is less than the odds ratio based on the midpoint of case enrollment because it increases the proportion of exposed controls. Matching cases and controls on the month and year of birth and then using the corresponding case's diagnosis date as the control's reference date is the best way to ensure that the controls have a lifetime opportunity for OC exposure equal to that of the cases. However, even with a large number of controls, it is nearly impossible to obtain such a match *post hoc*. Given the limitations of these other methods, using the midpoint of the case enrollment period seems to be the best choice, on practical grounds. Furthermore, in this study, results obtained using the midpoint of the case enrollment period easily approximate the results obtained using the randomly chosen date. Indeed, the differences among the odds ratios for the five possible references date choices are minimal.

Finally, it is likely that the effect of changing the reference date for controls is dependent upon the magnitude of the relative risk. The effect may also be related to sample size, the length of the enrollment period, and the uniformity of the distribution of case diagnosis dates over the enrollment period. Even though changing the reference date of the controls had little effect on the odds ratios for this data set, this effect may not be small for other studies.

In conclusion, in case–control studies where case and control enrollment periods are not identical, the choice of a control reference date should be carefully considered. An investigation into how sensitive the results are to the choice of the reference date is warranted. If varying the reference date for the controls substantially affects the conclusions of the study, then an appropriate justification for the particular reference date chosen and a description of how alternative choices affect the conclusions should be provided.

Acknowledgements—From the Costa Rican Demographic Association and the Centers for Disease Control. *Principal Investigators*: Dr Luis Rosero-Bixby, Dr Mark W. Oberle. *Project Coordinators*: Carmen Grimaldo, Martin Fallas, Dr Daisy Fernández. *Data Managers*: Anne S. Whatley, Hernan Caamano, Elizabeth Z. Rovira, Steven Kinchen, Dr A. H. Rampey Jr. *Project Associates*: Dr Oscar Fallas, Dr Nancy C. Lee, Dr Kathleen L. Irwin, Dr Judith A. Fortney, Dr Gary S. Grubb, Michele G. Bonhomme. *Project Consultants*: Dr Raimundo Riggioni, Miguel Gómez, Phyllis A. Wingo, George L. Rubin, Dr Howard W. Ory, Dr Peter M. Layde, Jacquelyn Arthur, Dr Emilia Leon. *Costa Rican National Tumor Registry*: Georgina Muñoz de Brenes. *Laboratory Consultants*: Dr Mary E. Guinan, Jorge Ramírez, Dr Sandra Larson, Dr Andre J. Nahmias, Dr Julius Schachter. *Pathology Consultants*: Dr Saeed Mekbel, Dr Jorge Salas Cordero, Dr León Tropper.

Partial support was provided by Family Health International with funds from the United States Agency for International Development (USAID). The opinions expressed do not necessarily represent those of USAID.

REFERENCES

1. Irwin KL, Rosero-Bixby L, Oberle MW *et al.* Oral contraceptives and cervical cancer risk in Costa Rica. **JAMA** 1988; 259: 59–64 (*Erratum*, **JAMA** 1988; 259: 1182).
2. Lee NC, Rosero-Bixby L, Oberle MW *et al.* A case–control study of breast cancer and hormonal contraception in Costa Rica. **J. Natl Cancer Inst** 1987; 79: 1247–1254.
3. Rosero-Bixby L, Oberle MW. Fertility change in Costa Rica 1960–84: analysis of retrospective lifetime reproductive histories. **J Biosoc Sci** 1989; 21: 419–432.
4. SAS Institute Inc. **SAS User's Guide: Basics, Version 5 Edition.** Cary, NC: SAS Institute Inc; 1985.