Published in final edited form as:

Ann Epidemiol. 2015 October ; 25(10): 794–796. doi:10.1016/j.annepidem.2015.07.007.

Comparison of web versus interview participants in a case-control study

Mahsa M Yazdy^a and Martha M Werler^a

^aSlone Epidemiology Center at Boston University, 1010 Commonwealth Ave, Boston, Massachusetts

Abstract

Purpose—To evaluate whether selection bias might be introduced due to differential self-selection into an internet-based case-control study.

Methods—We compared sociodemographic factors and behavioral factors for cases and controls who opted for their first entrée into the study be via the internet (Web group) versus the telephone (Phone group), using data from a study on clubfoot.

Results—Overall, 200 of 1,825 mothers were in the Web group and were more likely to be non-Hispanic White, non-smokers, more educated, wealthier, and took 7 medications/vitamins in pregnancy. These differences were greater for case mothers, introducing the potential for selection bias. Maternal smoking is an established risk factor for clubfoot, and was confirmed in the Phone group (OR=1.61) but was not observed in the Web group (OR=1.15).

Conclusion—Our results suggest there is potential for selection bias if enrollment is solely internet-based for a case-control study.

Keywords

case-control study; selection bias; internet enrollment

Introduction

With the continuing advances in technology and the increasing accessibility of the internet, conducting epidemiologic research online has become more appealing. Performing research online offers many advantages, including the elimination of data entry errors, the ability to incorporate visual and audio aids, the opportunity for immediate validation checks, and the possibility of cost and time savings. [1, 2] Additionally, the internet allows for access to unique populations and can be a useful tool in recruiting participants when the outcomes under study are rare (e.g., birth defects).

Corresponding author (Present address): Mahsa Yazdy, Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 401 Park Drive, Suite 301 West, Boston, MA 02215, Telephone: 617-525-0399, mahsa.yazdy@channing.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The use of the internet for data collection in cohort studies has been studied widely, [3-9] but to our knowledge there are no studies examining its use and the potential for bias in a traditional case-control setting. Persons who self-select to participate in web-based study may well be different than the background population, but such selection doesn't necessarily introduce a bias. In a cohort study, self-selection is less likely to produce biased risk estimated since the health outcomes have yet to occur when participants enroll in the study. [10] Self-selection in a case-control study, on the other hand, is more likely to produce a bias due to participation being associated with both exposure and outcome.[10]

The objective of this study was to compare socio-demographic factors and behavioral factors between cases and controls that opted for their first entrée into a study to be via the internet versus the telephone, using data from a population based case-control study. In particular, we assessed whether relying exclusively on internet based participants for a case-control study could lead to biased results.

Materials and methods

We used data from a population-based case-control study conducted at the Slone Epidemiology Center at Boston University, which has been described elsewhere.[11] The objective of the study was to identify risk factors for clubfoot, a congenital anomaly of the foot. Briefly, cases and controls were ascertained from birth defects registries in Massachusetts, North Carolina, and New York; controls were selected from the same birth population as cases. The mothers were interviewed by telephone within 1 year of delivery on illnesses, medication and vitamin use, reproductive history, demographic information, diet, and behaviors during pregnancy (e.g., alcohol and cigarette use).

Prior to the telephone interview, mothers were sent an introductory packet explaining the objective of the study, along with consent forms, and other study related information. Beginning in June of 2009, we included a card that gave mothers the option of completing the diet portion of the questionnaire online. The card had the web address along with a personal log-in. After a mother completed the online diet questionnaire, she was contacted by a study nurse and the rest of the interview was completed over the telephone. For mothers who choose not to go online, the entire interview, including the diet portion, was conducted on the telephone by the nurse interviewers.

We restricted the data to mothers ascertained after June 1, 2009, the date the online questionnaire was initiated. Two groups were created: the first consisted of participants who opted for their first entrée into the study to be via the internet (Web group) and the second group was participants who opted to enter via the telephone (Phone group) (Figure 1). The Web group served as a proxy for women who would have enrolled in the study via the internet, while the Phone group served as a proxy for women who would have enrolled through traditional recruitment methods. We compared socio-demographic factors and reported medication/ vitamin use in pregnancy, as reported in the telephone interview, between participants who entered the study via the internet and those who entered through traditional recruitment methods, according to case-control status. Additionally, we examined how the use of only the web enrolled women would affect the relationship between clubfoot

and smoking (a consistently identified risk factor for clubfoot). Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using logistic regression and all ORs were adjusted for maternal race, education, and study center.

Results

Web participation increased steadily over time in our study, with 5%, 10%, and 16% of mothers completing the online questionnaire in 2009, 2010, and 2011, respectively. After excluding 7 women who completed the web questionnaire but did not complete the telephone interview, there were a total of 200 women (74 cases and 122 controls) in the Web group and 1625 women (542 cases and 1083 controls) in the Phone group. A higher percentage of both case and control mothers in the Web group reported being older, being non-Hispanic White, having >12 years of education, having an annual income \$90,000, being non-smokers, and taking 7 medications/vitamins during pregnancy. These differences were more apparent for cases (Table 1). Among cases, 45% in the Web group were 30 years of age compared to 32% in the Phone group, resulting in a 13% difference between the Web and Phone groups; whereas for controls this difference was only 3% (44% compared to 41%, respectively). Compared to the control Web group, the case Web group captured more women who reported non-Hispanic white race, higher incomes, being non-smokers, and taking 7 medications/ vitamins in pregnancy. For maternal smoking, the adjusted OR was 1.61 (95% CI: 1.25, 2.06) in the Phone group; it was attenuated to 1.15 (95% CI: 0.45, 2.94) in the Web group (Table 2).

Discussion

The results of our study suggest there was self-selection among women whose first entrée was via the internet, which was more apparent among cases. Selection bias resulting from such self-selection was indicated in our analysis of maternal smoking in relation to clubfoot. Previous studies have consistently identified smoking as a risk factor for clubfoot, with ORs ranging from 1.4-2.6.[12-22] We observed a similarly positive association for women who entered the study through the standard introductory letter and telephone interview, whereas very little association was observed for women who elected to enter the study via the internet.

We considered women who completed the diet questionnaire online to be a proxy for participation in an internet-based study, while women who chose not to do the online questionnaire would not enroll in an online study. It is probable that women who are comfortable using the internet for a questionnaire would also enroll in an online study, though we cannot rule out the possibility that our online questionnaire may be an inadequate proxy. For example, participants knew they would be contacted by telephone and may, in fact, have been opting to complete the interview in one session rather than splitting the participation up into two pieces (i.e., a web questionnaire and phone interview). In addition, it is possible that women who chose not to do our questionnaire online would still have chosen to participate in a web-based study. Further, this study was conducted from 2009 to 2012 when internet access was approaching universal in the U.S., but facility and preferences were evolving.

Conclusion

The disparity by case-control status in our study suggests that in a case-control study, there is potential for selection bias if enrollment is internet-based. Despite the opportunity for efficient case and control accrual via the internet, traditional methods may be necessary for the time being.

Acknowledgments

This work was supported by Eunice Kennedy Shriver National Institute for Child Health and Human Development grant RO1-HD051804. We thank Lisa Crowell RN and Mary Beth Pender RN, Interviewers; Michelle Heinz and Eileen Mack, Research Assistants; Michael Bairos, Oleg Starobinets, and Elie Sirota, Database Analysts; and the mothers who participated in the study.

References Cited

1. Rothman KJ, Cann CI, Walker AM. Epidemiology and the internet. *Epidemiology*. 1997; 8(2):123–5. [PubMed: 9229201]
2. Van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: The future in epidemiology? *Am J Epidemiol*. 2010; 172(11):1292–8. [PubMed: 20880962]
3. Huybrechts KF, Mikkelsen EM, Christensen T, Riis AH, Hatch EE, Wise LA, et al. A successful implementation of e-epidemiology: The danish pregnancy planning study 'snart-gravid'. *Eur J Epidemiol*. 2010; 25(5):297–304. [PubMed: 20148289]
4. Smith B, Smith TC, Gray GC, Ryan MA. When epidemiology meets the internet: Web-based surveys in the millennium cohort study. *Am J Epidemiol*. 2007; 166(11):1345–54. [PubMed: 17728269]
5. West R, Gilman A, Coste F, Zhou X, Brouard R, Nonnemaker J, et al. The attempt cohort: A multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. *Addiction*. 2006; 101(9):1352–61. [PubMed: 16911736]
6. Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F. Feasibility of recruiting a birth cohort through the internet: The experience of the ninfea cohort. *Eur J Epidemiol*. 2007; 22(12): 831–7. [PubMed: 17955333]
7. Turner C, Bain C, Schluter PJ, Yorkston E, Bogossian F, McClure R, et al. Cohort profile: The nurses and midwives e-cohort study—a novel electronic longitudinal study. *Int J Epidemiol*. 2009; 38(1):53–60. [PubMed: 18202083]
8. Andreeva VA, Galan P, Julia C, Castetbon K, Kesse-Guyot E, Hercberg S. Assessment of response consistency and respective participant profiles in the internet-based nutrinet-santé cohort. *Am J Epidemiol*. 2014
9. van Gelder M, Pijpe A. E-epidemiology: A comprehensive update. *Cancer*. 2013; 15:0.007.
10. Rothman, K.; Greenland, S.; Lash, T. Validity in epidemiologic studies. In: Rothman, K.; Greenland, S.; Lash, T., editors. *Modern epidemiology*. City: Lippincott Williams & Wilkins; 2008.
11. Werler MM, Yazdy MM, Mitchell AA, Meyer RE, Druschel CM, Anderka M, et al. Descriptive epidemiology of idiopathic clubfoot. *Am J Med Genet A*. 2013; 161(7):1569–78. [PubMed: 23686911]
12. Dickinson KC, Meyer RE, Kotch J. Maternal smoking and the risk for clubfoot in infants. *Birth Defects Res Part A: Clin Mol Teratol*. 2008; 82(2):86–91. [PubMed: 18022868]
13. Kancherla V, Romitti PA, Caspers KM, Puzhankara S, Morcuende JA. Epidemiology of congenital idiopathic talipes equinovarus in iowa, 1997–2005. *Am J Med Genet A*. 2010; 152A(7):1695–700. [PubMed: 20583169]

14. Parker SE, Mai CT, Strickland MJ, Olney RS, Rickard R, Marengo L, et al. Multistate study of the epidemiology of clubfoot. *Birth Defects Res Part A: Clin Mol Teratol.* 2009; 85(11):897–904. [PubMed: 19697433]
15. Eeden SKVD, Karagas MR, Daling JR, Vaughan TL. A case-control study of maternal smoking and congenital malformations. *Paediatr Perinat Epidemiol.* 1990; 4(2):147–55. [PubMed: 2362871]
16. Alderman BW, Takahashi ER, LeMier MK. Risk indicators for talipes equinovarus in washington state, 1987-1989. *Epidemiology.* 1991; 2(4):289–92. [PubMed: 1912045]
17. Honein M, Paulozzi L, Watkins M. Maternal smoking and birth defects: Validity of birth certificate data for effect estimation. *Public Health Rep.* 2001; 116(4):327. [PubMed: 12037261]
18. Honein MA, Paulozzi LJ, Moore CA. Family history, maternal smoking, and clubfoot: An indication of a gene-environment interaction. *Am J Epidemiol.* 2000; 152(7):658–65. [PubMed: 11032161]
19. Skelly AC, Holt VL, Mosca VS, Alderman BW. Talipes equinovarus and maternal smoking: A population-based case-control study in washington state. *Teratology.* 2002; 66(2):91–100. [PubMed: 12210013]
20. Pavone V, Bianca S, Grosso G, Pavone P, Mistretta A, Longo MR, et al. Congenital talipes equinovarus: An epidemiological study in sicily. *Acta Orthop.* 2012; 83(3):294–8. [PubMed: 22489891]
21. Dodwell E, Risoe P, Wright J. Factors associated with increased risk of clubfoot: A norwegian national cohort analysis. *Journal of Pediatric Orthopaedics.* Publish Ahead of Print.
22. Leite M, Albieri V, Kjaer SK, Jensen A. Maternal smoking in pregnancy and risk for congenital malformations: Results of a danish register-based cohort study. *Acta Obstet Gynecol Scand.* 2014; 93(8):825–34. [PubMed: 24861914]

Abbreviations

OR	Odds Ratio
CI	Confidence Interval

Figure 1. Flowchart of study recruitment and determination of web and telephone groups, Massachusetts, New York, and North Carolina, 2009–2011

Table 1

Sociodemographic Factors and Pregnancy Behaviors Reported in the Telephone Interview by Cases and Controls, Massachusetts, New York, and North Carolina, 2009–2011.

	Cases		Controls	
	Web group (n=74) %	Phone group (n=542) %	Web group (n=122) %	Phone group (n=1083) %
	%	%	%	%
Maternal Age				
<20	3	10	4	11
20-29	53	58	52	48
30	45	32	44	41
Maternal Race				
White non-Hispanic	93	69	82	64
Black non-Hispanic	4	13	8	18
Hispanic	0	13	5	13
Other/Unknown	3	5	5	6
Education				
<high school	1	14	3	15
high school grad	18	27	9	22
>high school	81	59	88	63
Income				
<\$30,000	12	41	17	33
\$30,000 - \$60,000	22	20	12	21
\$60,000 - \$90,000	20	10	23	12
\$90,000 +	41	16	39	20
Missing/ refused	5	13	9	14
Center				
NC	34	45	39	47
NY	50	37	31	32
MA	16	18	30	21
Smoked after the first lunar month of pregnancy				
No	84	70	88	80
Yes	16	30	12	20
Number of medications or vitamins reported				
0	1	0	0	0
1-3	19	29	16	35
4-6	38	41	48	39
7	42	30	36	26

Table 2

Association Between Smoking and Clubfoot by Participation Group, Massachusetts, New York, and North Carolina, 2009–2011.

	Crude Odds Ratio	95% Confidence Interval	Adjusted Odds Ratio ^a	95% Confidence Interval
All participants	1.67	1.33, 2.10	1.56	1.23, 1.98
Phone group only	1.72	1.36, 2.18	1.61	1.25, 2.06
Web group only	1.38	0.61, 3.14	1.15	0.45, 2.94

^a Adjusted for maternal race, education, and study center