Maternal Recall of Breastfeeding Duration by Elderly Women

Joanne H. E. Promislow¹, Beth C. Gladen², and Dale P. Sandler¹

¹ Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC.
² Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Received for publication May 17, 2004; accepted for publication September 8, 2004.

Studies of long-term effects of breastfeeding on the health of both infants and mothers often rely on maternal recall of breastfeeding duration after several decades. The authors evaluated this recall by 140 college-educated, US women 69–79 years of age who breastfed a child in 1940–1956 and recorded the duration both prospectively in a diary for the Menstruation and Reproductive History Study and retrospectively in a questionnaire in 1990–1991. Mean prospective breastfeeding duration was 5.6 months (range, 1–12 months). Mean reporting difference, questionnaire minus diary duration, was 0.0 months, with a standard deviation of 2.7 months; women who recorded short durations tended to overreport, while women who noted long durations underreported. The weighted kappa statistic for reporting agreement was 0.55 (95% confidence interval: 0.42, 0.67), with better recall observed for women in the youngest quintile at recall, firstborns, and infants with more siblings. Ever having breastfed was recalled by 94% of women. For categories of 1–3, 4–6, 7–9, and 10–12 months, recalled breastfeeding duration was correctly classified by 54% of women and was classified within ±1 category by 89%.

The observed misclassification, if nondifferential with respect to outcome, would appreciably attenuate estimates of dose-response associations between breastfeeding duration and later health.

Abbreviation: CI, confidence interval.

The protection that breastfeeding affords against numerous infant illnesses has been well documented (1). Infant nutrition may also have long-term effects. The influence of duration of breastfeeding on adult intelligence (2, 3), obesity (4, 5), serum cholesterol (6, 7), diabetes risk (8), and blood pressure (9, 10) has been the focus of recent studies. Lactation also may influence disease risk for mothers, even decades later. Numerous studies have investigated the influence of duration of lactation on risk of breast (11–15) and ovarian (16) cancers and of osteoporosis (17–19).

In such studies, breastfeeding history is often assessed by maternal recall, providing the potential for exposure misclassification and biased measures of association. The validity of maternal recall of breastfeeding history has been evaluated in a handful of studies (20–28), but most have focused on relatively short-term (≤10 years) recall. The longest known recall period previously evaluated was 20–22 years (20).

The purpose of this study was to assess maternal recall of the duration of a breastfeeding event for elderly US women who breastfed a child and reported the duration both prospectively in a diary and retrospectively in a questionnaire administered 34–50 years later. We evaluated the variation in recall with several maternal and infant characteristics, including maternal age at time of recall, years since event, family size, and infant sex and birth order. In this paper, we discuss the effect of the observed misclassification on measures of association and the implications for interpreting results from studies with long-term retrospective assessment of breastfeeding duration.

MATERIALS AND METHODS

Participants and protocol

Prospective diary data on breastfeeding duration were obtained from the Menstruation and Reproductive History...
Study begun in 1934 by Alan Treloar at the University of Minnesota (29). Between 1934 and 1939, 1,807 women younger than age 25 years, primarily White college students, enrolled in this study. Participants prospectively noted menstrual bleeding start and stop dates on the front of annual calendar cards and were asked to record any information concerning events that might influence menstruation, such as lactation, on the back of the cards. Although there were no specific questions about lactation on the calendar card, 173 women prospectively recorded the duration of 310 breastfeeding events. The exclusivity of breastfeeding was not noted.

In 1990, a follow-up study designed to link menstrual characteristics with subsequent health was initiated. The 1,134 women who contributed at least 5 years of menstrual data were eligible for this study. Information necessary for tracing was available for 997, and 943 (94.6 percent) women were successfully located (30). Between 1990 and 1991, a self-administered questionnaire was completed by 716 participants and 158 proxy respondents. For each of their livebirths, participant respondents were asked the following questions pertaining to breastfeeding: “Did you breastfeed?” and “If yes, how many months?” Proxy respondents were not asked for breastfeeding information and consequently were not included in the present analysis.

Of the 716 participants who completed a questionnaire, 407 reported in either the questionnaire or the diary that they breastfed a child, including 147 of the 173 women who prospectively recorded a breastfeeding duration in their diaries. Diary breastfeeding records were matched with questionnaire breastfeeding records by breastfeeding start dates (diary data) and livebirth dates (questionnaire data). Seven women could not be included in the analysis because they did not specify on the questionnaire whether they had breastfed, reported breastfeeding but not the duration, or provided an estimated diary breastfeeding stop date, leaving 140 women for whom 259 matched breastfeeding events occurred between 1940 and 1966. The 259 infants for whom there were prospective records of breastfeeding duration represent just 25 percent of the 1,034 infants reported on the questionnaire to have been breastfed, strongly suggesting that prospective breastfeeding records in the Menstruation and Reproductive History Study are incomplete. Of the women in the final sample, 59 percent kept diary records through menopause.

Statistical analyses

Since both the durations and reporting errors across multiple pregnancies within a woman are likely to be correlated, analyses were restricted to a single breastfeeding event per woman. Analyses were initially performed by using the woman’s first matched breastfeeding event, which took place from 1940 to 1956; all results presented in this paper, unless otherwise noted, derive from these analyses. For comparison, however, the analyses were subsequently repeated by using the women’s last, instead of first, matched breastfeeding event. Of the 140 women in this study, matched diary and questionnaire breastfeeding data for more than one child were available for 72.

Agreement between recalled and prospectively recorded breastfeeding duration was evaluated descriptively with scatter plots and cross-tabulations. Pairwise differences (mean, standard deviation) and the weighted kappa statistic were also computed to evaluate overall agreement and the influence of maternal and infant characteristics. The weighted kappa statistic was calculated by using the Fleiss-Cohen squared error weighting system in which the agreement weight for cell $(i,j)$ is given by $w_{ij} = 1 - (i - j)^2/(c - 1)^2$, with $c$ representing the number of categories. The weighted kappa statistic approximates the intraclass correlation coefficient with this weighting system (31). The values for continuous breastfeeding duration obtained from the diary data were rounded to the nearest month for calculation of the weighted kappa statistic and in all cross-tabulations. Recall of lifetime breastfeeding duration was evaluated in a similar manner for the 45 women who prospectively recorded breastfeeding durations for either each of their livebirths or each of the livebirths that they reported breastfeeding on the questionnaire.

Because the diary data did not differentiate between infants who were not breastfed and infants who were breastfed but for whom that information was not recorded, we could not analyze the specificity with which not breastfeeding is recalled accurately. However, the impact of the observed misclassification error on measures of association between breastfeeding duration and later health was examined by calculating the misclassification-adjusted relative risk estimates for an example dose-response association and reasonable values for both the prevalence of not breastfeeding and the specificity with which not breastfeeding is recalled.

RESULTS

There were 226 women who reported a breastfeeding duration on the questionnaire but not in their diaries, despite having been actively keeping prospective records at the time. Compared with these women, the 140 participants in the recall analysis who did prospectively record breastfeeding duration were similar in mean age at follow-up (73.3 years for both groups), had a slightly large mean number of livebirths (3.8 vs. 3.2; $p = 0.002$), and were more likely to correctly recall the date of their first livebirth (98 percent vs. 88 percent; $p = 0.003$).

Women for whom there was information in both sources reported breastfeeding durations of 1–12 months in the diary data and 0–12 months on the questionnaire. Eight women, who prospectively recorded breastfeeding durations of 2–8 months, retrospectively reported that they had not breastfed the specified child, giving a sensitivity of 94 percent for recall of having breastfed. Six of these women reported little or no breastfeeding for any of their children on the questionnaire; the other two could possibly have been misreporting which of their children they breastfed. Digit preference at 9 and 12 months was evident in the retrospective data, with women reporting these values 1.8 and 5.0 times as frequently, respectively, in the questionnaire data than in the diary data.

Considerable recall error existed over the range of reported breastfeeding durations (figure 1). The degree of under- and overreporting was very similar overall, with 36 percent and 37 percent of participants doing each, respectively. However, as might be expected, women who prospectively recorded breastfeeding durations of only 1 or 2 months were more likely to recall longer than shorter durations at follow-up. Conversely, women who noted diary durations of ≥9 months were more likely to underreport. Perfect agreement between recorded and recalled months of breastfeeding was obtained for 26 percent of women; 55 percent correctly recalled the duration within 1 month and 71 percent within 2 months.

In both the diary and the questionnaire data, mean breastfeeding duration was 5.6 months (table 1). The reporting difference, questionnaire minus diary duration, had a mean of 0.0 months (95 percent confidence interval (CI): –0.4, 0.5) and a standard deviation of 2.7 months, indicating substantial misclassification but no overall recall bias. The overall weighted kappa statistic for agreement between diary and questionnaire breastfeeding duration was 0.55 (95 percent CI: 0.42, 0.67). The Pearson correlation coefficient was also 0.55.

For the extended recall of 34–50 years, we found no association between time since the lactation event and recall accuracy (table 1). However, recall did appear better for women who were younger at recall, as well as for children from larger families; weighted kappa statistics were 0.33, 0.53, and 0.73 (p = 0.05) for children whose mothers had 1–2, 3–4, and 5–12 livebirths, respectively. The set of first matched breastfeeding events spans a birth order range of 1–7. Recall was better for children born first and, conversely, was worse for children born last.

When the above analyses were repeated by using women’s last matched breastfeeding event instead of their first, similar results were obtained, except that the slightly improved recall seen for female versus male children (table 1) was no longer observed. Overall, recall was slightly less good for the last events, with a weighted kappa statistic of 0.49 (95 percent CI: 0.34, 0.63).

Breastfeeding duration is often categorized. For commonly used categories of 0, 1–3, 4–6, 7–9, and 10–12 months, recalled breastfeeding duration was correctly classified by 54 percent of women and misclassified by one category by 35 percent of them (table 2). Misclassification percentages were higher for women in the shortest (1–3 months) and longest (10–12 months) prospective breastfeeding categories. A common alternative categorization scheme of 0, 1–2, 3–5, 6–8, 9–11, and 12 months was slightly more prone to misclassification, with recalled breastfeeding duration being correctly classified by only 47 percent of women when this scheme was used.

To illustrate the effect of misclassification on analyses of breastfeeding duration and later health, we calculated the relative risks that would be obtained based on the misclassification observed in table 2 for two breastfeeding prevalence scenarios and two estimates of the specificity with which not breastfeeding is recalled. We assumed that true relative risks were 0.80, 0.60, and 0.40 for breastfeeding durations of ≤3, >3–6, and >6 months versus never breastfed, respectively. For a given breastfeeding duration, misclassification was assumed to be independent of health outcome.

We first used a never-breastfed prevalence of 12 percent. This value was derived by applying the birth-order-specific values for ever breastfeeding rates from the questionnaire data to the birth-order distribution in the first matched event sample. We then repeated the calculations for a much higher never-breastfed prevalence of 50 percent. In both examples, the relative prevalences chosen for the breastfeeding catego-
ries of ≤3, >3–6, and >6 months are those observed for the participants included in the recall analyses.

We calculated the estimated relative risks for two values of the specificity with which not breastfeeding is correctly recalled at follow-up, 100 percent and 75 percent. For the latter example, the distribution of recalled breastfeeding durations among women who did not breastfeed was based on values reported in a Brazilian study with recall at a mean

### TABLE 1. Agreement between breastfeeding duration reported prospectively in a diary in 1940–1956 by Menstruation and Reproductive History Study participants and retrospectively on a questionnaire in 1990–1991, according to maternal and infant characteristics, United States

<table>
<thead>
<tr>
<th>No. of years since breastfeeding event ended</th>
<th>Questionnaire (mean (SD*))</th>
<th>Diary (mean (SD*))</th>
<th>Difference (months) (mean (SD*))</th>
<th>Weighted kappa statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>All women</td>
<td>140</td>
<td>5.6</td>
<td>5.6</td>
<td>0.0 (2.7)</td>
</tr>
<tr>
<td>33.5–40.0</td>
<td>20</td>
<td>7.0</td>
<td>7.4</td>
<td>−0.4 (2.9)</td>
</tr>
<tr>
<td>40.1–45.0</td>
<td>42</td>
<td>5.6</td>
<td>5.2</td>
<td>0.4 (2.8)</td>
</tr>
<tr>
<td>45.1–49.6</td>
<td>78</td>
<td>5.3</td>
<td>5.4</td>
<td>−0.1 (2.6)</td>
</tr>
<tr>
<td>Maternal age (years) at recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69.6–71.3 (quintile 1)</td>
<td>28</td>
<td>5.8</td>
<td>5.9</td>
<td>−0.1 (2.0)</td>
</tr>
<tr>
<td>71.4–75.1 (quintiles 2–4)</td>
<td>84</td>
<td>5.4</td>
<td>5.4</td>
<td>0.0 (2.8)</td>
</tr>
<tr>
<td>75.2–79.8 (quintile 5)</td>
<td>28</td>
<td>6.1</td>
<td>5.9</td>
<td>0.2 (3.0)</td>
</tr>
<tr>
<td>No. of livebirths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–2</td>
<td>32</td>
<td>5.3</td>
<td>6.1</td>
<td>−0.8 (2.7)</td>
</tr>
<tr>
<td>3–4</td>
<td>77</td>
<td>5.7</td>
<td>5.5</td>
<td>0.2 (2.8)</td>
</tr>
<tr>
<td>5–12</td>
<td>31</td>
<td>5.8</td>
<td>5.4</td>
<td>0.3 (2.3)</td>
</tr>
<tr>
<td>Birth order</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>103</td>
<td>5.6</td>
<td>5.7</td>
<td>−0.1 (2.4)</td>
</tr>
<tr>
<td>2–7</td>
<td>37</td>
<td>5.6</td>
<td>5.3</td>
<td>0.3 (3.3)</td>
</tr>
<tr>
<td>Reverse birth order</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = last born</td>
<td>13</td>
<td>5.0</td>
<td>6.1</td>
<td>−1.1 (3.7)</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>5.8</td>
<td>5.4</td>
<td>0.4 (2.4)</td>
</tr>
<tr>
<td>3–10</td>
<td>96</td>
<td>5.7</td>
<td>5.6</td>
<td>0.0 (2.6)</td>
</tr>
<tr>
<td>Sex of infant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>81</td>
<td>5.7</td>
<td>5.6</td>
<td>0.0 (3.1)</td>
</tr>
<tr>
<td>Female</td>
<td>59</td>
<td>5.6</td>
<td>5.6</td>
<td>0.0 (2.1)</td>
</tr>
</tbody>
</table>

* SD, standard deviation; \( \kappa_w \), weighted kappa statistic calculated by using Fleiss-Cohen weights (approximates the intraclass correlation coefficient); CI, confidence interval.
† \( p \) value associated with chi-square statistic for test of equality of stratum-specific values of the weighted kappa statistic.

### TABLE 2. Breastfeeding duration reported retrospectively on a questionnaire in 1990–1991 according to breastfeeding duration recorded prospectively in a diary in 1940–1956 by Menstruation and Reproductive History Study participants, United States

<table>
<thead>
<tr>
<th>Diary duration (months)</th>
<th>Questionnaire duration (months)</th>
<th>Total</th>
<th>Misclassified (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never 1–3 4–6 7–9 10–12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0 0 0 0 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1–3</td>
<td>3 17 10 3 1</td>
<td>34</td>
<td>50 41 9</td>
</tr>
<tr>
<td>4–6</td>
<td>4 9 32 9 1</td>
<td>55</td>
<td>42 18 24</td>
</tr>
<tr>
<td>7–9</td>
<td>1 5 6 23 6</td>
<td>41</td>
<td>44 15 29</td>
</tr>
<tr>
<td>10–12</td>
<td>0 0 1 6 3</td>
<td>10</td>
<td>70 0 70</td>
</tr>
<tr>
<td>Total</td>
<td>8 31 49 41 11</td>
<td>140</td>
<td>46 21 25</td>
</tr>
</tbody>
</table>
child’s age of 47 months (21). The specificity value of 75 percent is likely to be a low estimate, however. Other studies have reported perfect or near perfect recall of not having breastfed (25, 26, 28).

In all four scenarios, the dose-response association was attenuated by the observed misclassification of breastfeeding duration (table 3). For a never-breastfed specificity of 100 percent, the misclassification made the shortest breastfeeding category, ≤3 months, appear more protective than it actually was and the longest breastfeeding category, >6 months, appear less protective than it actually was, thereby weakening evidence for a dose-response association. A true relative risk of 2.0 for breastfeeding ≤3 months versus >6 months was attenuated to a value of 1.4 for both prevalence examples. For a never-breastfed specificity of 75 percent, the misclassification of breastfeeding duration made breastfeeding appear less protective than it actually was across all breastfeeding categories. In all examples, however, the misclassification-adjusted relative risk estimates did remain suggestive of a dose-response association.

Evaluation of recall of lifetime breastfeeding duration was possible for 45 women. These women breastfed one to four children (median, two). For this group, mean lifetime breastfeeding duration was 10.4 months in the diary data and 9.9 months in the questionnaire data. The reporting difference mean and standard deviation were −0.5 and 4.7 months, respectively. The weighted kappa statistic for lifetime months of breastfeeding was 0.82 (95 percent CI: 0.72, 0.93). The Pearson correlation coefficient was 0.84.

For comparison with the cohort as a whole, the weighted kappa statistic for the first matched breastfeeding event was 0.66 (95 percent CI: 0.50, 0.82) for this more select group.

**DISCUSSION**

Because few prospective records exist of breastfeeding history from several decades ago, studies of the long-term health effects of breastfeeding duration on infants and their mothers often rely on retrospective data. Long-term (≥18 years) maternal recall of breastfeeding duration has been used in studies of the association of infant feeding with adult obesity (32), adult vascular function (10), and both child-(33, 34) and adult- (8) onset diabetes. Similarly, studies of lactation and maternal disease, for example, breast cancer (11–15), ovarian cancer (16), Hodgkin’s disease (35), and osteoporosis (17–19), often rely on maternal recall of the duration of breastfeeding events that ended over 30 years ago. These studies also frequently include women over 70 years of age at the time of recall.

The accuracy of maternal recall of breastfeeding duration after 20–22 years has been investigated in a cohort of Jerusalem residents (20). Other studies have evaluated only relatively short-term (≤10 years) recall (21–28), and very little information is available from US women (28). To our knowledge, this study is the first to evaluate maternal recall of breastfeeding duration among elderly US women after more than 30 years.
In this cohort of women who breastfed, we observed appreciable misclassification in breastfeeding duration recalled after 34–50 years. The overall weighted kappa statistic of 0.55 suggests moderate agreement (36) between recalled and prospectively recorded breastfeeding duration. However, for comparison, this value is toward the higher end of the range found for nutrients when comparing food frequency questionnaires with diet records.

Nonetheless, under a number of reasonable scenarios, the observed misclassification would appreciably attenuate a true dose-response association between duration of a breastfeeding event and later health. This attenuation diminishes the ability to detect true weak or moderate dose-response associations and could contribute to some of the inconsistent results reported in the literature. The data on recall accuracy presented here could enable evaluation of the potential effect of measurement error on estimates of the effect of breastfeeding duration, as assessed by maternal recall, on later health of the infant (37, 38).

Previous studies of maternal recall of breastfeeding duration have tended to report better agreement than that observed in this cohort. When their children were 4 years of age, 70 percent of mothers in a Brazilian study correctly recalled breastfeeding duration in the 3-month categories that they had initially reported when their children were 11 months of age (21). In an Australian study, 79 percent of mothers correctly recalled duration of breastfeeding within 1 month at a mean child age of 3 years (25). Other studies have reported correlation coefficients of 0.94 and 0.86 for the association between prospectively recorded breastfeeding duration and that recalled by mothers after 8 and 20 years, respectively (20, 22). Median breastfeeding duration was shorter in each of these previous studies than in the current study, and some of the previous studies included a small proportion of women (≤10 percent) who did not breastfeed, which could have affected the measures of agreement and association. However, the substantially longer recall period and older age at time of recall seem probable reasons for the reduced agreement observed in this study.

Huttly et al. (21) reported that maternal recall of breastfeeding duration was less accurate after 3 years than after 1 year. In contrast, recall accuracy did not differ over recall periods of 34–50 years in this study, suggesting an earlier plateau in the decline in recall accuracy. However, even over the limited age range of 69–79 years, there was some evidence to suggest that younger women had better recall, consistent with the increasing prevalence of memory problems with age.

To our knowledge, the influence of family size on recall of breastfeeding duration has not been reported previously; however, recall of birth weight has been found to decrease among mothers with five or more children (39). Surprisingly, in this study, agreement was found to improve with increasing family size. This effect could derive from enhanced recall by mothers who maintained similar breastfeeding patterns for multiple children. Prospectively recorded breastfeeding durations for first and second births were positively correlated (Pearson correlation coefficient = 0.52).

In the Brazilian study, better educated women were more likely to recall a longer breastfeeding duration than originally reported (21). In the present study of predominantly college-educated women, no overall bias toward over- or underreporting breastfeeding duration was observed either for the cohort as a whole or for any specific maternal or infant characteristic. Maternal recall of child-rearing practices has been reported to err in the direction of experts’ recommendations (28). While the American Academy of Pediatrics now recommends breastfeeding for ≥12 months (1), expert support for extended breastfeeding was less pronounced both when participants breastfed and at follow-up (40). The tendency for participants who breastfed for 9–12 months to underreport at follow-up may instead reflect societal norms at the time and the perceived social stigma associated with extended breastfeeding (41).

For studies evaluating the effect of lactation on later maternal health, recall of lifetime lactation may be most relevant. The higher kappa value found for recall of lifetime lactation probably reflects improved recall correlation due to both the greater variation between women in lifetime lactation and the increased selectivity of the subset of women for whom this analysis could be performed.

This study has some limitations. Recall of neither non breastfeeding nor exclusive breastfeeding could be evaluated. Additionally, study participants prospectively noted breastfeeding duration without being specifically prompted and might therefore reflect a selective subset of the follow-up cohort. Indeed, compared with women who reported breastfeeding retrospectively but not prospectively, women who did keep prospective breastfeeding records had significantly better recall of the date of their first livebirth, suggesting that the recall observed in this study might be better than average.

The absence of specific breastfeeding questions in the menstrual diaries also raises the possibility that some women may no longer have noted continued lactation once menses resumed. A prospective breastfeeding stop date within 1 month after the date menses resumed was recorded for only 27 of the 82 women with a breastfeeding stop date on (three women) or after the date menses resumed, however. This possible limitation also runs counter to the observation that women who noted longer prospectively recorded breastfeeding durations tended to underreport at follow-up. Study participants were also predominantly college-educated Whites from Minnesota who had demonstrated motivation for record keeping and who breastfed the children included in the analysis between 1940 and 1956, further limiting the generalizability of these results to other populations.

The appreciable misclassification observed in this cohort of older US women suggests that if misclassification is independent of disease status for a given breastfeeding duration, observed dose-response associations between breastfeeding duration and later health are likely to underestimate the true associations and that weak or moderate associations may be difficult to detect. Recall patterns in this cohort also suggest that after several decades, time since event may be a less important predictor of accuracy than age at recall. The better recall of breastfeeding duration observed for firstborns and
children from larger families is intriguing and merits further investigation.

ACKNOWLEDGMENTS

The authors are grateful to the late Dr. Alan Treloar, who conceived of and developed the Menstruation and Reproductive History Study, and to Dr. Ann Voda, former Director of the Tremin Trust Research Study, for her permission to contact women from the study. The authors also thank Drs. Ken Smith, Jan Root, and Elizabeth Whelan for their work in assembling and tracing the cohort and Drs. Donna Baird and Walter Rogan for their comments on the manuscript.

REFERENCES


8:1051–69; discussion 1071–3.