AMERICAN JOURNAL OF INDUSTRIAL MEDICINE 40:512–522 (2001)

Test-Retest Reliability of an Icon/Calendar-Based Questionnaire Used to Assess Occupational History

Lawrence S. Engel, PhD,[1,2]* Matthew C. Keifer, MD, MPH,[2,3] Mary Lou Thompson, PhD,[4] and Shelia H. Zahm, ScD[5]

Background *Self-reported work histories can be critical for both epidemiologic and clinical purposes. However, the complexity of some work histories, such as those of migrant farm workers, can hamper recall, resulting in inaccurate reporting. Memory aids may reduce such error. This study assesses the reliability of work histories collected using such aids in the form of an icon/calendar-based questionnaire.*

Methods *Thirty-one males engaged in farmwork and other manual labor for a median 28 years (range: 10–64) were interviewed twice, 8–14 months apart, about their lifetime employment. In each interview, subjects were asked about important life events, which were recorded with icons on a calendar. They were then asked to recount their work history, including for each job the tasks, crops or products handled, starting and ending dates, and location. This information was recorded, job-by-job, on the calendar with icons.*

Results *Interquestionnaire agreement of cumulative reported employment duration (as measured by the correlation coefficient) was moderate to high across all time periods for certain crops (e.g., $r = 0.69$–0.92 for apple-related work), by location (e.g., $r = 0.76$–0.95 for Washington State), and for agricultural work in general ($r = 0.67$–0.94), but was lower for specific tasks. Agreement of job counts was high for total work history for certain crops (e.g., $r = 0.93$ for apple-related work), by location (e.g., $r = 0.90$ for Washington State), and for agricultural work in general ($r = 0.89$), but paradoxically decreased with proximity to the interview date. Agreement of both measures tended to be highest for those tasks and crops in which subjects reported spending the most time. Categorization of subjects into tertiles on the basis of either cumulative duration or counts produced results similar to those observed for job counts.*

Conclusions *The icon-calendar questionnaire is an effective tool for estimating cumulative duration of certain work categorizations among subjects with complex work histories.* Am. J. Ind. Med. 40:512–522, 2001. *Published 2001 Wiley-Liss, Inc.*†

KEY WORDS: icon; calendar; questionnaire; occupational history; recall; memory aids; farmwork

[1]Department of Epidemiology, University of Washington, Seattle, Washington
[2]Pacific Northwest Agricultural Safety and Health Center (PNASH), University of Washington, Seattle, Washington
[3]Occupational Medicine Program, Departments of Medicine and Environmental Health, University of Washington, Seattle, Washington
[4]Department of Biostatistics, University of Washington, Seattle, Washington
[5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland

Contract grant sponsor: National Cancer Institute; Contract grant number: QTR50104; Contract grant sponsor: National Institute for Environmental Health Sciences Training; Contract grant number: ES07262; Contract grant sponsor: National Institute for Occupational Safety and Health; Contract grant number: U07/CCU012926-02; Contract grant sponsor: Worker's Compensation Fund of the State of Washington.

*Correspondence to: Lawrence Engel, Occupational Epidemiology Branch, National Cancer Institute, 6120 Executive Blvd., Room 8113, Bethesda, MD 20892-7240.
E-mail: engell@mail.nih.gov

Accepted 27 April 2001

INTRODUCTION

Ascertainment of a person's work history is a critical component in studies of health effects of occupational exposures as well as in clinical diagnosis and treatment of illnesses related to those exposures. Self-report is often the only means for determining a person's work history and thereby estimating occupational exposures. However, migrant laborers and others with transient employment tend to have complex work histories often involving a large number of tasks and/or employers each year. Such persons frequently have difficulty in accurately recalling their work histories, thus complicating health research and clinical decision-making. This is particularly problematic for long-term recall [Checkoway et al., 1989].

We are aware of only a few studies that have examined the validity or reliability of self-reported work histories [Baumgarten et al., 1983; Rosenberg et al., 1987; Stewart et al., 1987; Bond et al., 1988; Bourbonnais et al., 1988; Brisson et al., 1991; Rosenberg, 1993]. These studies have generally found the work histories to be reasonably accurate when compared to company or government records. However, unlike migrant laborers, the subjects in almost all of these studies tended to be stable, long-term employees of a single employer, with relatively few job changes within that employer. This fact, which made the validation studies feasible, limits the interpretation of these results in regard to migrant laborers.

Complex exposures which are difficult to recall and for which there are no ''gold standards'' (i.e., objective and accurate records) are found in other areas of epidemiologic research. In particular, this issue has been extensively addressed by researchers studying the relationship between diet and disease, since obtaining accurate diet histories from subjects is of paramount importance for estimating exposures, but also presents some of the greatest challenges and limitations of such research [Little et al., 1984; Stewart et al., 1987; Willett, 1990a; Rimm et al., 1992; Sempos, 1992; Jarvinen et al., 1993; Block, 1994; Kohlmeier, 1995; Decarli et al., 1996]. Most studies have reported fair to good agreement between nutrient intakes measured either by two questionnaires administered at different times (reliability) or by a questionnaire and another dietary measurement instrument (validity), although agreement has been very low for some nutrients [Willett, 1990b].

The present study examines the reliability of work histories collected twice from the same subjects with the use of an ''icon-calendar questionnaire.'' This questionnaire consisted of (1) a life events calendar similar to those used in some studies of oral contraceptive use and (2) icons representing various life events and jobs. The first provided chronological ''anchors,'' or reference points, around which a subject might more easily recall his or her work history. The second, consisting of stickers portraying such things as cars, flags, babies, fruits, and animals, was used to build an easily interpretable pictorial representation of a subject's life/work history (see Fig. 1). This was particularly useful for illiterate or semiliterate subjects. Reliability was examined for the overall work histories as well as for specific components of those histories that might be of interest to researchers (e.g., jobs with potential pesticide exposure). Different time periods were also considered to see if reliability improved with proximity to the time of interview and because researchers will generally be interested in different time periods when examining different exposure-disease associations.

This study was inspired by a follow-up study of neurological function among farmworkers by some of the present authors (M.C.K. and L.S.E.) [Engel et al., 1998]. In that study, since cumulative lifetime pesticide exposures were of interest, subjects were asked to provide detailed lifetime occupational histories. However, in that study it soon became apparent that a typical farmworker's work history was too complex—in terms of number of jobs, number of employers, and work locations in a given year and over a life time—to capture via ''traditional''questionnaire methods. Subjects had difficulty remembering not only what jobs they had performed in a given year but also what jobs they had already listed for the interviewer, resulting in incomplete and questionable occupational histories. The use of written cues was precluded by the illiteracy or semiliteracy of much of this population.

The need for a more appropriate data collection tool led to the development of an icon-calendar questionnaire for the second round of that study. Results of a comparison of work histories collected via the two methods in that study are presented in an accompanying paper [Engel et al., 2001 (this issue)].

MATERIALS AND METHODS

Subjects

Forty male farmworkers from the Yakima area of central Washington State were recruited and interviewed between May and October, 1996. Participation was restricted to persons who, based on initial screening, were likely to be in the area the following summer for the follow-up interview. Due to joint participation in a concurrent subject-surrogate reliability study of the icon/calendar questionnaire, 32 of these subjects were also required to have spouses/partners. Methods and results of the spousal interviews are presented in an accompanying paper [Colt et al., 2001]. Thirty-four subjects (85.0%) were recruited through an area church, 5 (12.5%) through a local farmworker's union, and 1 (2.5%) through a public health organization. Recruitment and interview were conducted in Spanish. Informed consent was obtained from all participants

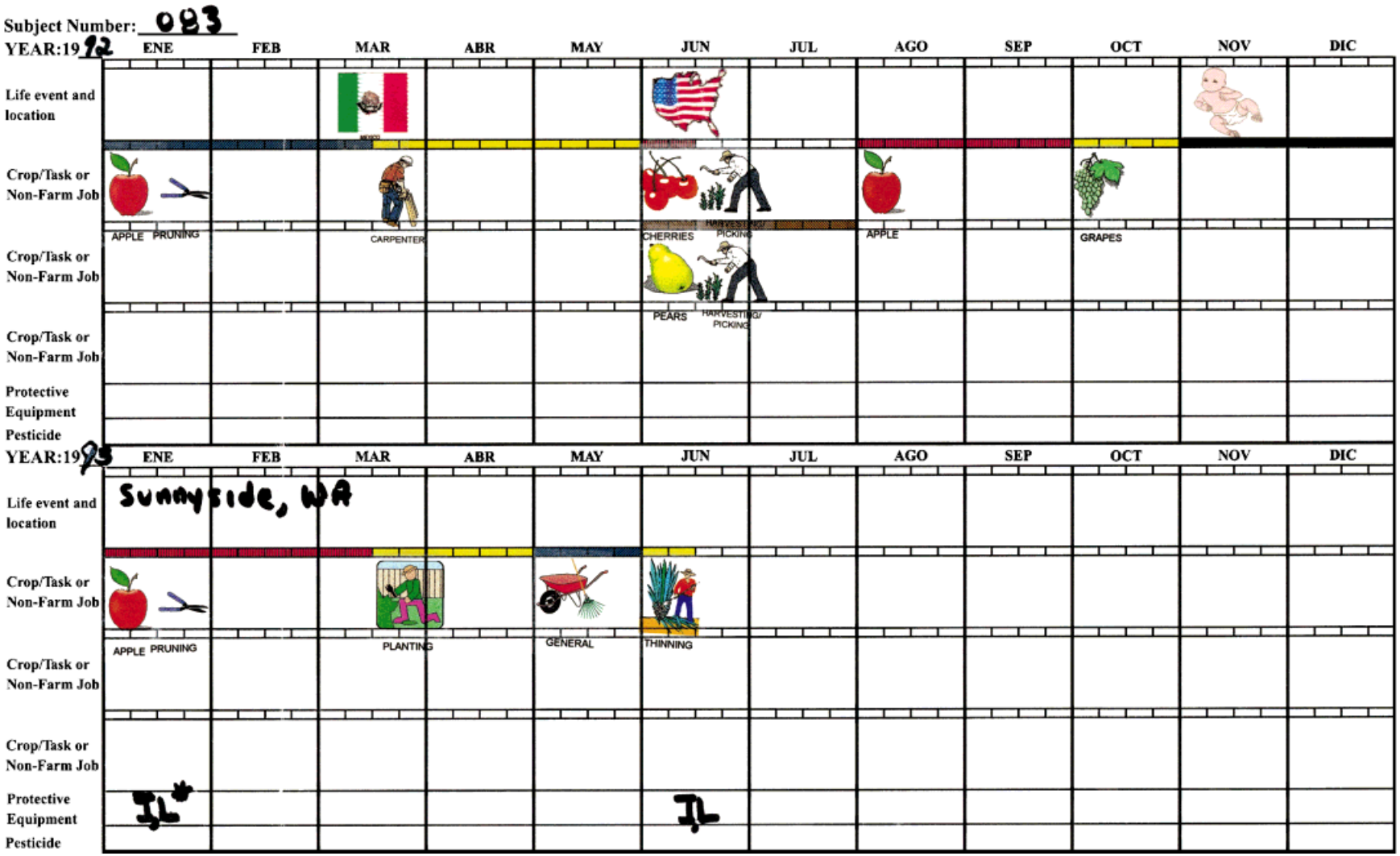

FIGURE 1. Section of a completed life events calendar.

(or from a parent or guardian of minors). The study protocol was approved by the University of Washington's and National Cancer Institute's Human Subjects Committees.

All subjects in Round 1 were asked to provide an address and phone number where they expected to be the following summer and also the address and phone number of a contact person. Follow-up of all subjects was attempted beginning the following spring. For all subjects not reachable at the address or phone number provided in Round 1, the contact person was approached for information concerning the subject. Thirty-two subjects (80%) (excluding surrogates) were located and reinterviewed in July, 1997.

In the first round, all work histories were collected in an office with minimal distractions via interviewer-administered questionnaire. In the second round, 29 interviews (90.6%) were conducted in the same location and 3 (9.4%) were conducted elsewhere (i.e., either in the subject's home or in another quiet location). The same two interviewers were used in both rounds and each interviewed the same subjects in both rounds. The interviewers were trained to administer the questionnaires and were instructed to conduct the interviews in the same manner in each round. No information from the first interview was made available during the second interview. Interviewers were told not to incorporate into the second interview any subject-specific information learned during the first interview. Each interview lasted approximately 1 h.

The questionnaire, which was translated initially from English into Spanish, was back translated by another translator to verify the initial translation. Any necessary corrections were made at that time.

Interviews

The method of interview using the icon-calendar questionnaire is described in the accompanying paper by Engel et al. [2001]. Three aspects of that method were changed for the present study based on experiences in that earlier study. First, the small toys and objects used as icons in the original study were replaced by illustrated stickers, creating a permanent record of the life events and work history (Fig. 1). Second, subjects were asked about their work histories starting from their first job and moving forward in time, in contrast to the original study that worked backwards from the present. Third, crop icons, when appropriate, were used together with job icons.

Data Analysis

Only those subjects who were interviewed twice were included in analyses. We examined agreement between the two interviews for both the number (count) as well as cumulative duration of various job categorizations. Three statistics were used to assess agreement: the correlation between reports at the two interviews, their median difference (to assess bias), and the interquartile range (IQR) of the differences in reporting (to assess variability in response). A measure of correlation alone does not allow one to assess bias and variability [Bland and Altman, 1986]. We chose to use the median and IQR rather than the mean and standard deviation as measures of bias and variability since they are less sensitive to outliers in the data. For normally distributed data, the IQR corresponds to an interval of ± 0.67 SD about the mean. For the same reason, we used Spearman's correlation coefficient, measuring significance at the 5% two-sided level.

In addition to comparing job counts and durations across both interviews we also examined degree of temporal overlap for the same/similar jobs (e.g., jobs reported in the same state, of the same task, or in the same crop) in the two interviews. This was done by calculating the amount of overlapping time for the same/similar jobs reported in both interviews as a percentage of duration of those jobs (e.g., percent of overlapping time of an apple-picking job reported in the same year in both interviews).

The above analyses were also performed with subjects stratified by age ($<$median vs. $\geq$median [median $=$ 36 years]), reported current alcohol use (light or moderate vs. heavy consumption [where heavy consumption was defined as either (a) six or more drinks per drinking session or (b) three or more drinks per session and three or more sessions per week]), reported alcohol consumption on day of either interview (any vs. none), years of education ($<$median vs. $\geq$median [median $=$ 6 years]), and interviewer. Additional analyses were performed by grouping subjects into tertiles on the basis of either job count or cumulative duration and examining agreement of these ''exposure'' groups via the weighted and unweighted kappa statistics. We focused on apple and cherry-related jobs, as well as harvesting, thinning, and pesticide application based on either the high percentage of subjects or time engaged in that work or the greater potential for pesticide exposure in that work.

To assess the viability of reconstructing potential exposure to different pesticides used in different time periods and also to make results from this study comparable with results from our other icon/calendar research, work histories were divided into four nonoverlapping time periods for analysis: before 1/1/86, 1/1/86 to 12/31/90, 1/1/91 to 12/31/95, and 1/1/96 to first interview date. When calculating job counts, a job was assigned to one of these time periods if any portion of that job occurred within that period. If a job spanned more than one period, it was counted separately for each. The sum of all period counts frequently exceeded the total count since a job that spanned multiple periods was counted once for each period in which it fell but only once for the total.

For analyses in which the starting date of a subject's work history mattered (e.g., temporal agreement of reported

jobs in the earliest time period), if the reported starting date differed between the two interviews, the earlier date was used for both interviews and the time between the earlier and later starting dates was treated as ''missing'' in the later-starting interview (resulting in a more conservative estimate of agreement between interviews). Total duration of work history was calculated as the number of months between the start of the first reported job and the first interview date. For all analyses, data in the second interview were right-censored at the date of the first interview. Data were analyzed with Visual FoxPro (Microsoft Inc., Redmond, WA) and SPSS for Windows (SPSS Inc., Chicago, Illinois).

RESULTS

Characteristics of Study Subjects

Thirty-two of the original 40 subjects were located and agreed to participate in a second round of interviews. One subject was subsequently excluded due to his lack of co-operation with the interviewers and consequent concerns about the validity of his self-reported work history. The 31 remaining subjects had a median age at first interview of 36 years, with a median 6 years of education and a median 28 years of employment (Table I). All subjects reported some employment in all time periods in at least one of the interviews. The median time between interviews was 12 months.

Interquestionnaire Agreement

Agreement between the two interviews for the timing of reported life events (''chronological anchors'') was very high, with a typical median difference of only 0.0–0.5 months (Table II). Variability between repeat interviews was quite small for all these measures, the upper end of IQR of differences in reported times being only 5 months. Maximum discrepancies were within 2–3 years, with the exception of one subject who was off by a decade on several reported dates.

TABLE I. Selected Characteristics of Study Subjects

Characteristic	n(%)
Hispanic	31 (100.0)
Male	31 (100.0)
Country of birth:Mexico	31 (100.0)
Characteristic	Median (Range)
Age, years	36 (16–72)
Years of education	6 (0–12)
Years in U.S.A.	16 (1–56)
Duration of work history, years	28 (10–64)
Months between interviews	12 (8–14)

TABLE II. Agreement of Timing of Reported Life Events

Life event	n	Median Δ (IQR) (months)[a]
Birth date	30	0.0 (0.0, 0.0)
First came to U.S.A.	30	0.0 (−1.3, 0.2)
First came to WA	31	0.0 (−1.0, 5.0)
First married	23	0.0 (0.0, 0.0)
First child	20	0.0 (0.0, 0.0)
First US job	31	−0.5 (−2.0, 0.5)
First WA job	31	0.0 (−3.9, 1.6)

[a]Median difference (in months) and the interquartile range of that difference in reported timing of event between the two interviews.

Table III shows summary statistics for job count across the different time periods. For the aggregated data (over entire work history), for instance, the correlation between total number of jobs reported at the two interviews was 0.82, the median number of jobs held by these workers was 168.5, the median difference in job counts across the two interviews was 23, and the lower and upper quartiles of these differences were −33 and 53, respectively. The total work history correlation for apple-related jobs was high, at 0.93, and remained consistently high across time periods except the most recent. The correlation dropped to 0.43 for cherry-related jobs and became less consistent across time periods. Agreement was also less consistent for specific tasks and task-crop combinations (e.g., thinning apples), ranging from 0.04 for applying pesticides to 0.77 for harvesting when examining total work history. Contrary to expectations, agreement tended to decrease in more recent time periods.

The range of job count discrepancies between the two interviews was often fairly wide. Considering all jobs in a subject's work history, the interquartile range of the discrepancies was 86, representing over half of the median job count (168.5 jobs) reported. For some tasks and task-crop combinations, the ratio of the variability to the median count was even greater. This pattern was consistent across time periods.

Table IV provides summary statistics for cumulative job duration across the different time periods. For the job categories we examined, correlations of cumulative job duration showed a similar pattern to those for job count, although they tended to be somewhat higher. For total work history, correlations above 0.90 were observed for total jobs, general agriculture in Washington, and apple-related work. As with job counts, agreement was higher and more consistent for these job categories than for specific tasks and task-crop combinations. For these latter categories, correlations ranged from 0.40 for applying pesticides to 0.59 for harvesting when examining total work history.

TABLE III. Job Counts in Both Questionnaires for Selected Jobs

	Total work history					< 1/1/86					1/1/86–12/31/90				
Job type	**n**	**Correlation**	**Median count**	**Median countΔ**	**IQR of countΔ**	**n**	**Correlation**	**Median count**	**Median countΔ**	**IQR of countΔ**	**n**	**Correlation**	**Median count**	**Median countΔ**	**IQR of countΔ**
All (including unemployment)	31	0.82[a]	168.5	23.0	−33.0, 53.0	31	0.85[a]	101.0	9.0	−15.0, 51.0	31	0.58[a]	36.0	−2.0	−10.0, 7.0
WA agriculture	31	0.89[a]	57.5	1.0	−18.0, 13.0	22	0.92[a]	25.5	−3.0	−8.5, 0.0	27	0.70[a]	23.5	2.0	−5.0, 9.0
Apple-related	30	0.93[a]	21.8	0.0	−4.5, 11.0	21	0.84[a]	12.5	0.0	−3.5, 5.5	24	0.85[a]	12.0	0.0	−4.0, 4.8
Cherry-related	25	0.43[a]	6.5	0.0	−4.0, 1.5	15	0.49	1.5	−2.0	−3.0, 1.0	18	0.35	2.8	0.0	−2.3, 1.0
Harvesting	31	0.77[a]	37.0	6.0	−10.0, 17.0	31	0.73[a]	20.5	2.0	−8.0, 20.0	29	0.54[a]	10.0	1.0	−3.0, 6.0
Thinning	27	0.59[a]	8.0	−1.0	−5.0, 2.0	21	0.64[a]	3.0	−1.0	−2.0, 0.5	17	0.02	3.5	0.0	−2.5, 2.0
Applying pesticides	21	0.04	10.5	−2.0	−19.5, 4.5	15	−0.10	6.0	−4.0	−18.0, 4.0	14	0.01	3.3	−2.0	−5.3, 3.5
Thinning apples	26	0.48[a]	6.3	−1.0	−4.0, 2.0	17	0.37	3.0	−1.0	−2.0, 0.0	17	0.14	3.5	0.0	−2.0, 2.0

	1/1/91–12/31/95					1/1/96–1st interview				
Job type	**n**	**Correlation**	**Median count**	**Median countΔ**	**IQR of countΔ**	**n**	**Correlation**	**Median count**	**Median countΔ**	**IQR of countΔ**
All (including unemployment)	31	0.32	35.0	0.0	−7.0, 4.0	31	0.63[a]	4.5	0.0	−1.0, 1.0
WA agriculture	31	0.48[a]	23.5	2.0	−6.0, 9.0	30	0.75[a]	4.3	0.0	−1.0, 1.0
Apple-related	28	0.84[a]	12.5	0.5	−4.0, 5.0	27	0.69[a]	2.0	0.0	−1.0, 0.0
Cherry-related	21	0.16	2.5	0.0	−2.0, 0.5	11	—	1.0	0.0	0.0, 1.0
Harvesting	29	0.38[a]	8.0	1.0	−3.0, 5.0	18	0.30	1.0	0.5	0.0,1.0
Thinning	23	0.06	3.0	0.0	−2.0, 3.0	15	−0.45	1.0	0.0	0.0, 1.0
Applying pesticides	11	0.05	5.0	0.0	−4.0, 5.0	7	−0.29	2.0	0.0	0.0, 2.0
Thinning apples	23	0.10	2.5	0.0	−1.0, 3.0	15	−0.30	1.0	0.0	0.0, 1.0

n, number of subjects in a category; correlation, Spearman's correlation coefficient for the number of jobs reported in both questionnaires; median count, median number of jobs reported in both questionnaires; median count Δ, median difference in number of jobs reported; IQR of count Δ, lower and upper quartiles of the difference in number of jobs reported.

[a] $P < 0.05$

TABLE IV. Cumulative Duration of Jobs in Both Questionnaires for Selected Jobs

	Total work history					< 1/1/86					1/1/86–12/31/90				
Job type	n	Correlation	Median months	Median months Δ	IQR of months Δ	n	Correlation	Median Months	Median Months Δ	IQR of Months Δ	n	Correlation	Median Months	Median Months Δ	IQR of Months Δ
All (including unemployment)	31	0.97[a]	333.9	0.0	−12.0, 36.0	31	0.97[a]	204.0	0.0	−12.0, 24.0	31	—	60.0	0.0	0.0, 0.0
WA agriculture	31	0.93[a]	97.0	6.2	−15.1, 23.9	22	0.94[a]	31.0	−0.7	−5.1, 10.4	27	0.81[a]	38.6	2.0	−3.9, 9.0
Apple-related	30	0.92[a]	43.7	0.3	−11.4, 26.9	21	0.86[a]	23.6	0.1	−2.2, 15.0	24	0.85[a]	22.5	1.5	−3.2, 8.5
Cherry-related	25	0.45[a]	3.7	0.3	−3.3, 1.7	15	0.56[a]	0.8	−1.1	−3.1, 0.5	18	0.51[a]	1.4	−0.6	−2.3, 0.8
Harvesting	31	0.59[a]	66.2	8.9	−6.0, 38.0	31	0.61[a]	37.6	6.5	−1.9, 21.2	29	0.43[a]	10.2	−0.7	−3.3, 4.0
Thinning	27	0.55[a]	10.0	−2.0	−8.0, 4.6	21	0.67[a]	4.3	−0.6	−3.1, 1.8	17	−0.05	4.4	−0.7	−3.6, 4.7
Applying pesticides	21	0.40	11.1	−1.7	−9.4, 15.2	15	0.03	4.7	−0.4	−7.6, 5.0	14	−0.19	6.3	1.0	−3.7, 11.0
Thinning apples	26	0.52[a]	8.1	−1.7	−6.2, 5.9	17	0.40	4.3	−0.5	−3.9, 2.4	17	0.01	4.4	−0.7	−3.2, 4.7

	1/1/91–12/31/95					1/1/96–1st interview				
Job type	n	Correlation	Median months	Median months Δ	IQR of months Δ	n	Correlation	Median months	Median months Δ	IQR of months Δ
All (including unemployment)	31	—	60.0	0.0	0.0, 0.0	31	1.00[a]	6.7	0.0	0.0, 0.0
WA agriculture	31	0.67[a]	40.4	−0.5	−6.6, 7.6	30	0.79[a]	5.2	0.0	−1.3, 1.5
Apple-related	28	0.81[a]	23.8	−0.5	−3.0, 5.8	27	0.69[a]	3.4	0.0	−1.2, 0.8
Cherry-related	21	0.22	1.5	−0.2	−1.9, 1.6	11	0.35	0.3	0.4	−0.1, 0.5
Harvesting	29	0.56[a]	10.7	1.6	−3.1, 6.6	18	0.38	1.1	0.4	−0.5, 0.9
Thinning	23	−0.10	3.8	−0.8	−4.0, 3.5	15	−0.34	0.8	−0.3	−0.7, 0.9
Applying pesticides	11	0.42	14.0	1.2	−0.9, 12.1	7	0.71	2.9	−0.3	−0.6, 0.0
Thinning apples	23	−0.13	3.8	0.0	−4.0, 3.5	15	−0.34	0.8	−0.3	−0.7, 0.9

n, number of subjects in a category; correlation, Spearman's correlation coefficient for the cumulative duration of jobs reported in both questionnaires; median months, median cumulative job duration reported in both questionnaires; median months Δ, median differences in cumulative job durations reported; IQR of months Δ, lower and upper quartiles of the difference in cumulative job durations reported.

[a] $P < 0.05$.

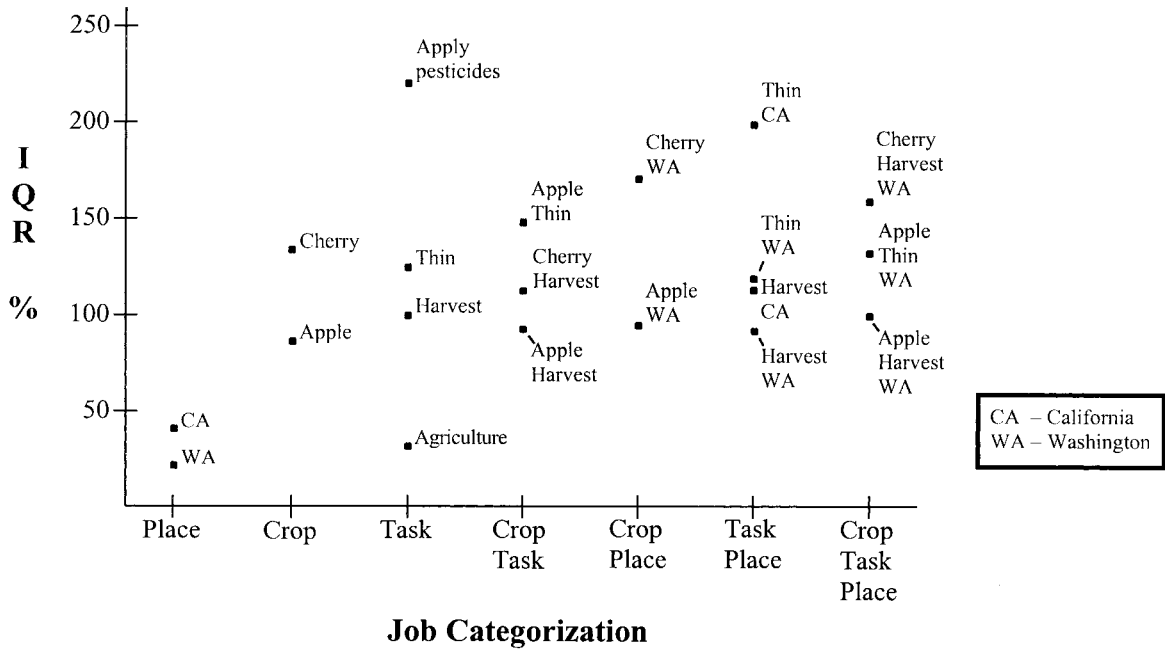

FIGURE 2. IQR % of cumulative job duration for selected crop/task/place combinations.

The ranges of interquestionnaire differences in cumulative job duration also were substantial relative to the median durations, although they tended to be somewhat less pronounced than for job count. Figure 2 illustrates the extent of reporting variability between the two interviews for various job categorizations (e.g., by crop, task, crop-task, etc.) with the interquartile range percent (i.e., the interquartile range as a percentage of the median [IQR %], analogous to a coefficient of variation). IQR % was substantial for almost all job categorizations, with most values falling between about 100% and 150%, although values go as high as 221% for pesticide application jobs. There is less variability for jobs grouped either as agricultural work or by a particular place, with values ranging from 32.8% for agricultural work to 22.9% and 42.3% for employment in Washington and California, respectively.

The amount of time unaccounted for by subjects was very low, with a total work history median of 0.0% (data not shown). For total work history, 13 subjects (41.9%) left some time unaccounted for, with a range of 0.1–3.6% (median = 0.3%). The number of subjects with unexplained time decreased with increasing proximity to the interview date; no subjects had unexplained time in the period 1/1/96 to first interview.

The percentage of temporal overlap (i.e., temporal agreement) over total work history varied from highs of 87% for jobs in the same country and 73% for jobs in the same state to a low 26% for the same task (data not shown). Percent overlap of jobs in the same crop was a modest 47%. Unlike agreement between job counts and between cumulative durations, percent overlap tended to increase in more recent time periods, although it remained modest for both crop and task even in the 1/1/96 to first interview period.

Agreement was also assessed with subjects grouped into exposure tertiles on the basis of both job counts and cumulative durations (data not shown). Weighted kappa statistics tended to be low to moderate for both these measures, although they were generally higher for tertiles based on cumulative job durations than for those based on job counts. Exceptions to this were the broad categories, over entire work history, of cumulative duration of total work and Washington agriculture, with weighted kappas of 0.93 and 0.78, respectively, and of apple-related job count, with a weighted kappa of 0.77. Weighted kappas for cumulative duration across total work history varied from 0.30 for apple thinning to 0.93 for all jobs, with a typical value of 0.4–0.5. As observed for job count and duration as continuous measures, agreement of tertiles tended to be the highest in earlier time periods.

Results did not change appreciably in any of the above analyses when subjects were stratified by age, reported alcohol use, years of education, or interviewer. However, because of the relatively small number of subjects in this study (n = 31), the degree of stratification was limited and the resulting strata were small.

DISCUSSION

The current study was designed to examine reliability of occupational histories collected via an icon/calendar-based questionnaire. Because of the complicated and transient nature of farmwork, we lacked records on a

subject's actual work history with which to validate the self-reported information. However, by interviewing subjects twice in the same manner about one year apart, we attempted to assess the reliability of the occupational history information they provided.

Use of the icon/calendar questionnaire resulted in a very detailed picture of a subject's occupational history. Subjects ''accounted for'' almost all of their working history either as work or unemployment, with very little time left unexplained.

Reliability of both the number of jobs and the cumulative duration of those jobs varied considerably by the type of work being reported. Reliability, judged by correlation between the two interviews, tended to be the highest for general agricultural employment and for certain crops, but decreased for specific tasks and task-crop combinations. Contrary to expectation, it also decreased with proximity to the interview date, although correlations generally remained high across time periods for cumulative duration of general agricultural and certain crop-related employment.

The extent of discrepancy between the two interviews is noteworthy. No bias in either job count or cumulative duration was detected between the first and second interviews; the error appeared to be random. However, for most job categories, the range of error tended to be wide relative to the median value of the measure being examined even when the correlation was high. The impact of this variability will depend on what is being measured and how it is being used. Greater error will generally be more tolerable for stronger exposure-disease associations and for wider exposure distributions among subjects.

Such error in self-reported work histories among farmworkers may be inevitable given the complexity of those work histories. The average subject in this study reported 156 jobs and 15 periods of unemployment over a 28 year work history. The top 25% of subjects by total job count reported an average of 243 jobs in a 42 year work history. It would be surprising *not* to find appreciable error in the self-reporting of such work histories.

The consistent trend of declining agreement with more recent time periods was unexpected and difficult to explain. It was probably due, in part, to the subjects painting their earlier work histories with broader brush strokes. In addition, many subject's earlier jobs, which usually predated their migrant lifestyle, were longer than their later jobs, and were therefore easier to accurately recall. A likely explanation for the generally low agreement in the most recent time period (1/1/96 to first interview) relates to the short duration of this period—approximately 6.7 months. Since the cumulative duration of most jobs was less than 1 month and the questionnaire solicited starting and ending dates in 1-week increments, differential reporting between the two interviews of only 1 week for either date would have had a large impact on observed agreement.

Another possible explanation for this temporal trend in reporting error is that, given the length of the interview, subjects may have become increasingly tired and consequently less careful as the interview progressed, although interviewers reported that subjects remained engaged and cooperative. If interview length was a factor, then better results might be obtained by shortening or splitting up the interview, the most appropriate method depending on the purpose of the data being collected and on logistical considerations. If interview length was not an issue, this implies that the icon-calendar is most appropriate for earlier, rather than more recent, work history. Since our other research into farmworker work history self-reporting indicates that the icon-calendar produces far more information than traditional questionnaire methods do (see accompanying paper by Engel et al.), the icon/calendar would still be advantageous for collecting early work histories.

The few studies we are aware of that have examined validity or reliability of self-reported work histories have generally found them to be reasonably accurate when compared to company, union, or social security records. Rosenberg, [1987; 1993] in two studies of workers who had changed job categories an average of 14 times over 15 years at a capacitor manufacturing plant, observed approximately 75% agreement when comparing self-reported job categories either between repeat interviews or between an interview and company personnel records. This agreement dropped to 60% among workers with the greatest job diversity. Baumgarten et al. [1983] observed 74.0% agreement when names of reported employers were compared to governmental employment records year-by-year over a 13 year period for subjects reporting two or more jobs. Stewart et al. [1987], in a study of shipyard workers, almost half of whom had been employed for five or more years and most of whom had kept the same job title during their employment, found modest agreement for reporting of exact starting and ending years (57% and 53%, respectively) compared to company personnel records. A validation study comparing work area assignments reported by chemical plant workers to company personnel records [Bond et al., 1988] found specificity of almost 100%, but sensitivity of only 48.4%. Sensitivity decreased with decreasing duration of assignment and increasing number of assignments. Brisson et al. [1991], in a study of female garment workers which compared self-reported employer names to public and union records year-to-year for up to 29 years, observed 73% agreement among subjects with two or more jobs. The number of jobs was inversely associated with percentage of agreement.

While differences in study designs, measures, and populations preclude direct comparison of the present study with the above studies, the agreement we observed seems reasonable compared to the agreement reported in those studies given that (1) most of those studies showed an inverse relationship between reliability/validity and number

of jobs and (2) the average number of jobs reported by subjects in the current study was far greater than in any of the above studies.

Exposure assessment poses many of the same challenges for nutritional epidemiologists as it does for occupational epidemiologists studying mobile populations such as farmworkers. In particular, the ''exposures'' (i.e., foods vs. jobs) tend to be highly varied, transient, repetitive, and lacking objective means of measurement (''gold standards''). Most reliability and validity studies of food frequency questionnaires have reported fair to moderate agreement of nutrient intake, with correlations typically ranging from 0.5 to 0.7, although agreement has been quite low for some nutrients [Little et al., 1984; Willett, 1990a, b; Rimm et al., 1992; Sempos, 1992; Jarvinen et al., 1993; Block, 1994; Kohlmeier, 1995; Decarli et al., 1996]. Yet despite its modest reliability, such dietary data has been successfully used in epidemiologic research [Willett, 1990b]. Reliability coefficients for a number of the job categories examined in the current study compare favorably with those observed in studies of diet and disease, suggesting a useful role for the icon-calendar questionnaire in epidemiologic research.

One limitation of this study is the possibility that a subject might have remembered and reported during the second interview the information that he had reported during the first interview, as opposed to the details of his actual work history. This is unlikely since the interviews were conducted about a year apart and since the average subject reported over 150 jobs. Given the complexity of this information, the *reported* work history, as opposed to the *actual* work history, was unlikely to be remembered and repeated in the second interview. For the same reason, it is unlikely that an interviewer would have remembered many details of a subject's work history and used this information to prod the subject. However, to address this possibility, interviewers were instructed not to use any information from the first interview in the second interview.

Given the large number of evaluations made in this study, the problem of multiple comparisons must be kept in mind when interpreting the results. However, the large number of significant correlations observed is unlikely to have arisen by chance alone. In any case, our primary focus was on the magnitude of the correlations, with less concern for their significance.

Since the icon-calendar questionnaire was originally developed for the purpose of exposure reconstruction, it is important to recognize its limitations in that context. We included in our analyses two job categories with typically higher pesticide exposures: pesticide application and apple thinning. Both of these showed only moderate interquestionnaire agreement over total work history, and variability was very high. Our data suggest that an exposure reconstruction based on time employed in various crops and places will be more accurate than one based on time spent in different tasks, but will result in a relatively coarse estimate of exposure.

In spite of its limitations, we feel that the icon-calendar questionnaire provides a truer picture of a farmworker's work history than does a traditional questionnaire. The ''quality'' of a work history is determined by both its accuracy and its completeness. While the results of the present study suggest that work histories collected via the icon-calendar questionnaire are of moderate accuracy, our other research indicates that these work histories are far more complete than those obtained via more traditional questionnaire methods.

We have observed in our previous research that farmworkers respond more favorably to the icon-calendar questionnaire than to the traditional questionnaire, being more patient and cooperative with the interviewers. Farmworkers have a great deal of difficulty, often becoming frustrated and impatient, when asked to recall details of their lifetime employment using traditional questionnaire methods. Interviewers have difficulty in collecting complete and detailed work histories as the subjects increasingly lose focus. On the other hand, most subjects enjoyed seeing their ''lives'' literally drawn before them with the icon-calendar questionnaire, remaining cooperative and engaged. We believe that such a subject is likely to provide better information than one who is confused, bored, or annoyed.

In conclusion, occupational histories of farmworkers obtained via the icon-calendar questionnaire, while very detailed and full, were of moderate accuracy. Given the shortcomings of alternative methods for reconstructing occupational exposures in this population, we feel that the icon-calendar questionnaire can be a valuable tool. However, its greatest strength appears to be in assessing the extent of agricultural employment by the rather broad categories of place and crop.

ACKNOWLEDGMENTS

The authors thank the interviewers Marni Aaron and Emma Moreno. We are also grateful to Nick Heyer for all of his ideas, work, and advice in this project.

REFERENCES

Baumgarten M, Siemiatycki J, Gibbs GW. 1983. Validity of work histories obtained by interview for epidemiologic purposes. Am J Epidemiol 118:583–591.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310.

Block G. 1994. Improving diet methods, improving epidemiologic methods. Ann Epidemiol 4:257–258.

Bond GG, Bodner KM, Sobel W, Shellenberger RJ, Flores GH. 1988. Validation of work histories obtained from interviews. Am J Epidemiol 128:343–351.

Bourbonnais R, Meyer F, Theriault G. 1988. Validity of self reported work history. Br J Ind Med 45:9–32.

Brisson C, V'ezina M, Bernard PM, Gingras S. 1991. Validity of occupational histories obtained by interview with female workers. Am J Ind Med 19:523–530.

Brownson RC, Jackson-Thompson J, Wilkerson JC, Kiani F. 1994. Reliability of information on chronic disease risk factors collected in the Missouri Behavioral Risk Factor Surveillance System. Epidemiology 5:545–549.

Checkoway H, Pearce NE, Crawford-Brown DJ. 1989. Research methods in occupational epidemiology. Oxford, UK: Oxford University Press.

Decarli A, Franceschi S, Ferraroni M, Gnagnarella P, Parpinel MT, La Vecchia C, Negri E, Salvini S, Falcini F, Giacosa A. 1996. Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy. Results for specific nutrients. Ann Epidemiol 6:110–118.

Engel LS, Keifer MC, Checkoway H, Robinson LR, Vaughan TL. 1998. Neurophysiological function in farm workers exposed to organophosphate pesticides. Arch Environ Health 53:7–14.

Engel LS, Keifer MC, Zahm SH. 2001. Comparison of a traditional questionnaire with an icon/calendar-based questionnaire to assess occupational history. Am J Ind Med 40:502–511 (this issue).

Jarvinen R, Seppanen R, Knekt P. 1993. Short-term and long-term reproducibility of dietary history interview data. Int J Epidemiol 22:520–527.

Kohlmeier L. 1995. Future of dietary exposure assessment. Am J Clin Nutr 61:702S–709S.

Little RE, Worthington-Roberts B, Mann SL, Uhl CN. 1984. Test-retest reliability of diet and drinking estimates for pregnancy and post partum. Am J Epidemiol 120:794–797.

Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. 1992. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 135:1114–1126.

Rosenberg CR, Mulvihill MN, Fischbein A, Blum S. 1987. An analysis of the validity of self reported occupational histories using a cohort of workers exposed to PCBs. Br J Ind Med 44:702–710.

Rosenberg CR. 1993. An analysis of the reliability of self reported work histories from a cohort of workers exposed to polychlorinated biphenyls. Br J Ind Med 50:822–826.

Sempos CT. 1992. Invited commentary: some limitations of semi-quantitative food frequency questionnaires. Am J Epidemiol 135: 1127–1136.

Stewart WF, Tonascia JA, Matanoski GM. 1987. The validity of questionnaire-reported work history in live respondents. J Occup Med 29:795–800.

Willett WC. 1990a. Epidemiologic studies of diet and cancer. Prog Clin Biol Res 346:159–168.

Willet WC. 1990b. Nutritional epidemiology. New York: Oxford University Press.