Attrition Bias in Rheumatoid Arthritis Databanks: A Case Study of 6346 Patients in 11 Databanks and 65,649 Administrations of the Health Assessment Questionnaire

ESWAR KRISHNAN, KIRSTEN MURTAGH, BONNIE BRUCE, DENISE CLINE, GURKIRPAL SINGH, and JAMES F. FRIES

ABSTRACT. ***Objective.*** Patient dropout (attrition) can bias and threaten validity of databank-based studies. Although there are several databanks of rheumatoid arthritis (RA) in operation, this phenomenon has not been well studied.

Methods. We studied the attrition patterns of patients with RA in 11 long-running databanks where patients were followed using semiannual Health Assessment Questionnaires. Attrition rates were calculated as the proportion of living patients who were in active followup at the cutoff date. Mantel-Haenszel methods and Weibull regression were used to model the relationship between attrition and age, sex, race, education, disease duration, functional disability, and other characteristics.

Results. Overall, 6346 patients with RA were recruited into the study cohorts and followed for 32,823 person-years with 65,649 observations. The crude attrition rate was 3.8% per cycle. Rates were lowest in community-based databanks. Smaller size of the centers, inner-city location, and university clinic settings were associated with worse attrition. In multivariable analyses, younger age, lower levels of education, and non-Caucasian race predicted attrition. Level of disability and disease duration were not associated with attrition.

Conclusion. In terms of person-years of followup and observation-points, this may be the largest study on attrition to date. While it is possible to have very high overall retention rates, certain types of databanks (smaller, inner-city-based, and university-based) are more likely to be biased due to selective retention of older, more educated Caucasian patients. (J Rheumatol 2004;31:1320–6)

Key Indexing Terms:
PATIENT DROPOUTS ARAMIS DISABLED PERSONS RHEUMATOID ARTHRITIS
COHORT STUDIES ATTRITION HEALTH ASSESSMENT QUESTIONNAIRE

Rheumatoid arthritis (RA) is a chronic disease with various phases of disease damage. To observe all disease stages and the time course of outcomes in RA, a long followup time is required. Observational studies in the form of chronic disease databanks have become an important tool in our understanding of this disease and treatments. Several limitations of randomized controlled trials such as short duration of observation and strict selection criteria make it necessary to perform longterm followup studies to evaluate longterm effectiveness of therapeutic interventions[1-4]. Since the patient groups in such studies are less selected than clinical trials, the results from these are thought to be more applicable to real-world situations. With time, however, there is loss of study subjects, i.e., attrition. If the dropouts (attrition) occur randomly, it will not affect the validity of the study as the characteristics of patients remaining under observation will not be systematically different from those dropping out. On the other hand, if patients with, say, higher levels of functional disability preferentially drop out after a short observation period due to physical limitations or psychosocial effects of the disease, then those remaining in the cohort are likely on average to be healthier than those who drop out. Inferences drawn from such cohorts can be seriously biased[5]. Thus the problem of attrition can seriously jeopardize the validity of research findings from databanks.

Since there is no reliable way to test the randomness of attrition in a particular study, empiric and simulation studies play an important role in the study of this phenomenon. The literature on attrition derives from large epidemiological surveys that are very dissimilar in terms of their objectives, methodologies, and outcome assessment from the chronic

From the Division of Rheumatology, Department of Medicine, Stanford University, Palo Alto, California, and Clinical Research Center of Reading, West Reading, Pennsylvania, USA.

Supported in part by grant AR43584 from the National Institutes of Health to ARAMIS (Arthritis, Rheumatism and Aging Medical Information System).

E. Krishnan, MD, MPhil, Division of Rheumatology, Department of Medicine, Stanford University, and Clinical Research Center of Reading; K. Murtagh, MA; B. Bruce, DrPH; D. Cline, BA; G. Singh, MD; J.F. Fries, MD, Division of Rheumatology, Department of Medicine, Stanford

Address reprint requests to Dr. E. Krishnan, Clinical Research Center of Reading, 401 Buttonwood Street, West Reading, PA 19611.
E-mail: iyer@stanford.edu

Submitted August 19, 2003; revision accepted January 29, 2004.

disease databank model being followed in clinic-based studies[2]. The results from such studies on attrition may not necessarily be applicable to RA databanks. Even though there are several databanks operating in different parts of the world, little empiric research has been done on potential biases caused by attrition, and only one recent study has directly addressed this vital issue[6]. We believe it is important to formally study the phenomenon of attrition and analyze its correlates, so that ongoing and future databanks of RA could be alerted to the magnitude and influence of attrition bias. Further, such formal analyses will also provide clues on preventing biases by over-recruiting those subgroups of patients with high attrition rates.

In this study we analyze data from a long-running multicenter followup study of RA in North America, focusing on the following questions: How do people who drop out of longterm followup cohort studies systematically differ from those who continue in the cohort? Are they more disabled? If so, what is the effect on studies done on the remaining subjects? What could be done to prevent biases arising from such dropouts?

MATERIALS AND METHODS

Data collection. The Arthritis, Rheumatism, and Aging Medical Information System (ARAMIS) is a chronic disease databank system consisting of parallel, followup, clinical, and outcome data of patients with osteoarthritis and RA. Data are collected from 8 diverse locations in the United States and Canada. ARAMIS is based at Stanford University and was enabled under the US Arthritis Act of 1974; it has been obtaining clinical and patient self-reported data on disease, disability, death, drug use, and dollar cost for over 25 years. ARAMIS comprises 11 databanks. These databanks comprise data for patients with RA from private rheumatology practices, from community-based practices, and from university clinics. We present the data in an anonymous form using center numbers instead of names.

Followup. After giving informed consent, patients fill in a questionnaire at the time of their entry into the cohort. They are not promised any monetary or other benefit. Subsequently, the Health Assessment Questionnaires (HAQ) are mailed in 2 semiannual "waves" — the first week of January and July of each year. ARAMIS uses standardized procedures designed to optimize patient retention. Vigorous followup is done for nonresponders within an initial 2-week period. A timed sequence of reminders is carried out including postcards, telephone calls, and additional questionnaire mailings. Patient death reports are sometimes obtained by questionnaires returned from the deceased patient's family, but more often from annual queries to the National Death Index[7]. Procedures such as telephone interviews for non-English-speaking patients, illiterate participants, and patients who are physically unable to complete questionnaires for any reason are employed to ensure patient retention. Periodic newsletters are used to maintain patient interest, small gifts of appreciation (e.g., a book of stamps or telephone card worth $5) are provided, and during the holiday season cards of appreciation are sent to study patients. The practice of responding individually to those patients who write in a question or a comment in the remark section adds a personal touch. The outcome assessor at each site makes up to 5 attempts to establish contact by mail and telephone. Detailed description of the databanks and the followup methodology has been published[8].

Outcome measurement. The HAQ collects information on the vital status, health care utilization, drug use, and outcome variables such as disability, pain, etc. The core measurement scale in the HAQ is the HAQ Disability Index (HAQ-DI), which measures physical function. This index measures the inability to perform activities of daily living on a scale of 0–3, where 0 indicates no disability and 3 indicates the worst disability. The questionnaire consists of 20 items that cover the 8 categories of functioning: dressing, arising, eating, walking, hygiene, reaching, gripping, and other usual activities. The HAQ can be viewed and downloaded from http://aramis.stanford.edu

Attrition: etymology and definition. Webster's Dictionary of American English defines attrition as "a reduction in the numbers, size or strength" or "a gradual reduction in the workforce resulting usually from retirements, resignations and deaths"[9]. The term attrition has been used in econometrics and social sciences, where the word "attrite" is often used as a verb. In epidemiology the term attrition has been used less often than the word "dropout" to denote loss of subjects from a study. An excellent review of definitions of attrition is available[10]. The term "retention rate" refers to the proportion of living patients who continue on followup.

In the ARAMIS databanks patients who stopped returning completed questionnaires could be classified into one of 3 categories: (1) Those who did not respond to questionnaires because they were deceased; (2) those who voluntarily discontinued participation; (3) those who were lost to followup but were not known to be deceased. For the purpose of this analysis we defined attrition as noncompletion of the last HAQ mailed to the patient at the cutoff date, the 38th mailing cycle in 1999. We had dead/alive information on all patients at the cutoff date. Individuals who completed the questionnaires erratically over time are not considered dropouts as long as they were known to be alive at the cutoff date and completed the latest questionnaire at that time.

Statistical methods. All patients with a diagnosis of RA were included in this analysis, performed using Stata® (Stata Corp., College Station, TX, USA). The observation period for each patient began on the date of completion of the first HAQ and ended on the date of completion of the last HAQ. Observations of patients who died while under followup contributed person-time and were considered censored at the time of completing their last questionnaire. Attrition rates were calculated from the number of "attrited" subjects divided by the person-years of observation and by different levels of explanatory variables of interest (e.g., age group). Confidence intervals for the rate were calculated using the standard quadratic approximation to the Poisson log-likelihood for the log rate parameter provided by Stata[11]. The jackknife procedure[12] was used when the Poisson assumption was not expected to be met, such as clustering of patients within centers. This is a relatively unbiased method of calculating confidence intervals. If there are N centers, samples of the data set are drawn N times, dropping one center each time, and calculating overall rates and their standard errors. The distribution of the collected estimates is subsequently used to calculate the jackknifed confidence intervals. Typically, jackknifed confidence intervals are wider (more conservative) than those obtained without them. After adjusting for confounders where needed, rate ratios were calculated using the Mantel-Haenszel method[11], which also served as a trend test for continuous and stratified explanatory variables. Student t-test and Pearson chi-square tests were used for comparing means and proportions, respectively.

Data were analyzed as a survival model that was set up as follows. Each individual enters the study on the day of completing the first HAQ. If they die before the cutoff date, their observations are censored. If they are known to be alive at the cutoff date but did not respond to the latest set of HAQ, the observations were considered to be "failures" at the date of completion of last HAQ. Thus each individual contributes a number of person-years and either one or no failure for calculating rates. We calculated median followup time using the Kaplan-Meier method. Multivariate survival models for attrition were fitted to adjust for the influence of other variables. Weibull regression[13] was used to model attrition, as exploratory analysis showed that the hazard followed a monotonic increasing pattern. The independent variables were age, sex, race, duration of RA, and number of years of education in addition to baseline HAQ-DI. The age and duration variables were used as time variable. Calendar year of first observation was entered in both the models to adjust for secular trends in patient recruit-

ment. Effect modifiers influencing the relationship between HAQ-DI and attrition risk (i.e., interactions) were explored.

RESULTS

Overall, 6346 patients with RA entered the ARAMIS databanks and completed 65,649 HAQ. The cumulative observation time was 32,823 person-years. Table 1 describes the baseline characteristics of the patients per databank. Patients were predominantly women (75%) and Caucasian (88%). There were few ethnic minorities (African Americans 5%, Native Americans 1%, Asian and Pacific Islander 2%, Latino 3%, Others/unknown 1%).

Patients who dropped out of the study were more likely to be younger, non-Caucasian, male, and less educated, and to have had RA for a shorter duration (Table 2), while patients who maintained participation were less disabled at the last observation. During followup, 1306 patients died, representing a crude mortality rate of 19.9 per 1000 person-years of observation (95% CI 16.9–24.9).

Attrition rates. The overall attrition was 3.8 per 100 living patients per cycle of data collection. The retention curve is plotted in Figure 1. The median number of cycles of data collection was 24. Attrition rates varied widely across the databanks (p value for heterogeneity < 0.05; Table 3). Attrition rates were the lowest in community-recruited patients and highest in inner-city-based university practices (Databanks 4 and 5). Centers with larger numbers of patients in general had fewer dropouts. The number of patients entering ARAMIS varied from calendar year to calendar year, reflecting funding and local leadership changes (Table 4). Attrition rates were higher for patients entering in the more recent years. Attrition rates adjusted for center effects in various strata of sociodemographic variables are shown in Table 5. Attrition rates declined with increasing age (with the exception of the highest age group) and increasing level of education. Non-Caucasians and patients with a shorter disease duration experienced higher attrition rates. Attrition rates did not differ significantly among strata of baseline HAQ-DI as well as last recorded HAQ-DI (Table 6).

Mantel-Haenszel analyses. Using the Mantel-Haenszel method we first examined the relationship between attrition and first and last HAQ-DI after adjusting for age, sex, and center effects. Non-Caucasians were at higher risk for attrition even after adjusting for age, sex, and center (RR 1.8, 95% CI 1.6–2.0). Higher level of education was associated with lower attrition rates (RR 0.82, 95% CI 0.77–0.86). Risk of attrition decreased with each quartile of disease duration (RR 0.86, 95% CI 0.83–0.90). We then performed parallel Mantel-Haenszel analyses for calculating the attrition risk for each quartile of baseline and last recorded HAQ-DI using the lowest quartile of each as a reference group after adjusting for age, sex, race, and center effects. In all these analyses there was no significant change in risk of attrition with increasing disability.

Weibull regression models. Next, we fitted Weibull models to investigate the effect of functional disability on the risk of attrition after adjusting for age, sex, race, education, duration of disease, and calendar year of study entry (Table 7). Standard errors are adjusted for clustering on center. The model included 63,784 non-missing observations on 6098 patients with 2333 attrition (model chi-square = 142, 7 degrees of freedom). The number of years of education (6% reduction in risk per year) and age at the time of study entry (1% reduction per year) were independently associated with the risk of attrition. HAQ-DI, sex, and disease duration were not statistically significant predictors of attrition. To investigate any nonlinear relationships of age, disease duration, and HAQ-DI with attrition, we entered their second powers in the model as additional covariates and compared the fit (log-likelihood) of the models by likelihood ratio tests. Addition of these transformed variables did not significantly increase the fit of the models.

Table 1. Baseline characteristics of patients with RA followed in ARAMIS.

Databank	Source of Patients	Subjects, n	Start Year	Median Age, yrs	Men, %	Caucasian, %	Median Years of Education	Mean Baseline HAQ-DI, 0–3	Mean Baseline Disease Duration, yrs
1	P	2446	1981	56	27	92	12	1.04	5.31
2	U	666	1981	55	22	77	13	1.33	11.45
3	U	401	1996	57	26	85	14	1.05	14.07
4	U*	201	1990	58	12	75	12	1.35	14.39
5	U*	186	1990	60	29	88	12	1.01	11.15
6	C	885	1981	59	26	94	11	1.33	16.99
7	C	318	1981	55	17	85	14	1.22	14.54
8	U	477	1988	59	28	92	12	1.52	14.9
9	U	87	1988	57	25	80	12	1.19	11.97
10	U	333	1997	60	18	84	12	1.06	15.82
11	U	346	1981	57	25	89	12	1.22	14.32
Total		6346		57	25	88	13	1.17	11.06

U: university-based clinics; P: private practices; C: community-based. * Inner-city-based centers.

Table 2. Characteristics of patients with RA who continued or dropped out.

Variable	Dropouts, n = 2480	Others, n = 3866	p
Mean age, yrs (SD)	54.5 (13)	58.3 (14)	< 0.001
Men, %	25	22	< 0.001
Non-Caucasians, %	16	9	< 0.001
Mean level of education, yrs (SD)	12 (2.6)	13 (2.7)	< 0.001
Mean duration of disease, yrs (SD)	10.2 (9.8)	12 (10.8)	< 0.001
Mean baseline HAQ-DI (SD)	1.14 (0.77)	1.19 (0.80)	0.01
Mean last recorded HAQ-DI (SD)	1.85 (0.86)	1.26 (0.81)	< 0.001

SD: standard deviation.

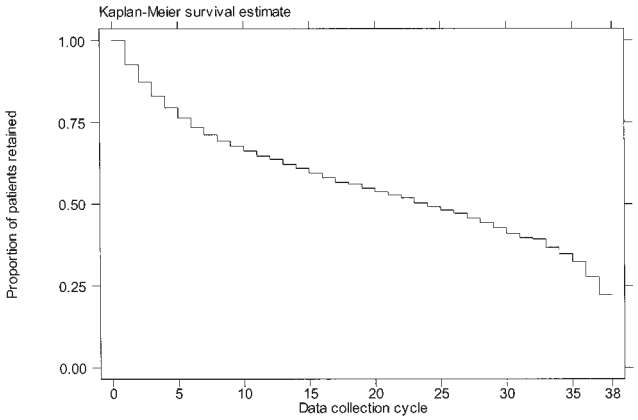

Figure 1. Kaplan-Meier curve showing the pattern of attrition from the ARAMIS databanks. Each data collection cycle is spaced apart by 6 months. Observations of patients who die are censored at the cycle of last observation.

DISCUSSION

We have described patient attrition from the ARAMIS databanks, one of the earliest multicenter cohort studies of RA. To our knowledge this is the largest study of attrition reported in the rheumatology literature. By sheer number of observation-points and person-years of followup, this may also be the largest study to date on attrition in any cohort study. The biggest strength of this study is the potential generalizability of our findings to other databank model observational studies. ARAMIS collects data from general community, inner-city hospital, university, and private practice settings. Thus the databanks of ARAMIS in a sense portray a microcosm of data sources in North America.

Our results confirm that it is possible to retain a large proportion of patients with RA in studies spanning decades with attrition rates less than 4% per cycle of data collection. We were able to locate only one other English language study focused on attrition in longterm studies of RA[6]. Our attrition rates of 3.8% compared well with their results (cycle-to-cycle attrition rate of 4.6%) even though the study populations, design, and followup strategies were different. Comparing to large general population-based cohorts, we note that our cycle-to-cycle attrition rates were comparable to the National Health and Nutrition Examination Survey Epidemiologic followup study (NHANES/NHEFS)[14], and were better than the 17% reported for the Black Women's Health Study (BWHS)[15]. Using a different, more restrictive definition of attrition than ours (i.e., patients specifically requesting to leave the followup rather than not just failing to follow up), other disease-specific cohorts, for example the Multi-Center AIDS Cohort Study[16], have reported a much lower cycle-to-cycle attrition of 1.5%. Recalculating

Table 3. Crude attrition rates in various ARAMIS databanks sorted in ascending order.

Databank	Source of Patients	Subjects, n	Cumulative No. of HAQ	Dropouts, n	Attrition Rate*	95% CI
7	Community	318	6,908	108	1.56	1.56–1.89
6	Community	885	12,597	358	2.84	2.70–3.39
8	University	477	4,242	140	3.3	2.85–4.04
1	Private practice	2446	25,916	961	3.7	3.47–4.28
2	University	666	7,829	304	3.88	5.59–7.28
11	University	346	2,404	127	5.28	5.46–7.29
9	University	87	920	60	6.52	4.80–6.56
3	University	401	1,722	116	6.73	7.39–10.18
10	University	333	1,042	90	8.63	6.64–7.88
5	University**	186	797	72	9.03	10.34–8.54
4	University**	201	1,091	144	13.2	1.56–1.89
Overall			65,469	2480	3.81	3.7–3.96

* Percentage of living patients lost per questionnaire cycle. ** Inner-city-based centers.

Table 4. Length of followup and attrition rates of patients by calendar year of entering the study.

Calendar Year	No. of Patients	No. of Dropouts	Crude Attrition Rate* (95% CI)	Mean followup, yrs (SD)
< 1982	1142	399	1.97 (1.80–2.11)	10.3 (6.44)
1983–5	785	305	3.19 (2.88–3.52)	7.75 (6.1)
1985–7	295	127	3.52 (2.99–5.14)	7.49 (5.43)
1987–9	932	351	3.68 (3.30–4.10)	4.95 (3.76)
1989–90	418	221	5.69 (4.94–6.54)	4.53 (3.96)
1991–2	353	184	6.14 (5.26–7.17)	4.25 (3.46)
1993–4	366	158	5.04 (4.20–6.06)	3.87 (2.71)
1995–6	298	149	6.23 (4.96–7.83)	2.85 (1.9)
> 1997–8	1574	570	5.93 (5.20–6.77)	2.41 (1.18)
Overall	6346	2464	3.8 (3.7–3.9)	5.4 (5.17)

* Percentage of living patients per questionnaire cycle.

Table 5. Attrition rates and sociodemographic characteristics adjusted for center effects.

	N	Attrition Rate*	95% CI	p for Trend
Age group				
16–24	77	6.7	3.8–9.9	0.01
25–34	394	4.7	3.9–5.7	
35–44	832	4.2	3.5–4.9	
45–54	1318	3.4	2.7–4.3	
55–64	1745	3.1	2.5–3.7	
65–74	1434	3.8	2.9–5.0	
> 75	546	6.1	4.2–8.2	
Women	4762	3.8	3.0–4.6	0.64
Men	1584	3.7	3.1–4.4	
Race				
Caucasian	5585	3.5	2.9–4.1	< 0.001
Others	761	7.4	4.6–11.6	
Education, yrs				
< 5	37	6.4	2.6–12.2	< 0.001
5–12	3771	4.2	3.2–5.2	
13–16	1315	3.5	2.7–4.4	
> 16	1223	3	2.3–3.9	
Disease duration, yrs				
First quartile (0.30–2.28)	2319	5.6	4.7–6.4	< 0.001
Second quartile (2.28–8.02)	1394	2.9	2.1–3.6	
Third quartile (8.02–16.65)	1259	3.2	2.6–3.7	
Fourth quartile (16.65–38.22)	1135	3	2.5–3.6	

* Percentage of living patients per questionnaire cycle.

Table 6. Functional disability and center-adjusted attrition rates in the ARAMIS databanks.

	Attrition Rate*	95% CI	p for Trend
Baseline HAQ-DI			
First quartile (0.0–0.50)	3.9	3.2–4.7	0.23
Second quartile (0.5–1.13)	3.7	3.0–4.4	
Third quartile (1.13–1.75)	3.7	3.0–4.4	
Fourth quartile (1.75–3)	3.8	2.7–5.0	
Last recorded HAQ-DI			
First quartile (0–0.75)	4.0	3.2–5.1	0.09
Second quartile (0.75–1.50)	3.8	3.1–4.6	
Third quartile (1.50–2.13)	3.8	3.0–4.5	
Fourth quartile (2.13–3.0)	3.6	2.7–4.7	

* Percentage of living patients per questionnaire cycle. HAQ-DI: Health Assessment Questionnaire Disability Index.

Table 7. Multivariate Weibull models of factors associated with attrition.

Characteristic	Hazard Ratio	Standard Error
Older age at entry, yrs*	0.99	0.002
Male sex	1.01	0.05
Increasing formal education, yrs*	0.94	0.009
Duration, yrs	1.00	0.02
Non-Caucasian*	1.64	0.16
Calendar year of entry*	1.07	0.002
Baseline HAQ-DI†	1.01	0.06

* $p < 0.01$. † Health Assessment Questionnaire Disability Index (0–3).

our attrition rates using their definition, we arrived at a rate of attrition of less than 1% per cycle in the ARAMIS databanks. The Women's Interagency HIV Study, using a definition of attrition similar to ours, reported an attrition rate of about 1.8% per year for following HIV infected women followed using questionnaires, gynecological examination, and venipuncture[17]. These studies involved pre-agreed cash incentives (offered as an inducement to joint the study in the first place) ranging from $20 to $50 per visit. We did not promise any benefit and enrollment was achieved based on appeals to altruism. That the ARAMIS program was funded by the National Institutes of Health and not by the pharmaceutical industry may have increased our credibility. Our small occasional gifts were intended to be unsolicited tokens of gratitude rather than a contractual obligation. We believe that attrition is lowest when subjects are volunteers and the subject matter of the research is personally relevant, as in the case of patients with RA[18].

Attrition rates in ARAMIS databanks varied considerably from databank to databank, suggesting that there was a center effect. Interestingly, the lowest attrition rates were found in the community-based centers, suggesting that patients recruited from the community rather than from outpatient clinics are more likely to be retained. This could indicate that over the long term, subject commitment to the study rather than loyalty to a particular clinic is more important in subject retention. Within the university clinic-based databanks there also was considerable heterogeneity in attrition rates, inner-city-based databanks doing relatively poorly. We believe that some of the center differences also reflect commitment and availability of protected time for the staff at the center.

Our finding that younger age predicted increased attrition confirms the findings from the NHANES-I Epidemiologic Follow-up Study[19], the National Institutes of Mental Health Epidemiologic Catchment Area program (ECA)[20], the Black Women's Health Study (BWHS)[15], and the Netherlands Mental Health Survey and Incidence Study (NEMESIS)[21]. One reason for this could be that younger people tend to move and change address more often than older persons. For example, in the BWHS at 3-year followup 56% of the study subjects (especially the younger ones) had changed address at least once[15]. Another explanation could be that younger subjects in general are less committed to contributing to medical research.

Our study does not agree with the finding of the smaller study by Reisine, *et al* that men tend to have lower retention rates[6]. Although non-Caucasian patients in general have been reported to have a higher attrition rate than Caucasians, there is considerable heterogeneity within the non-Caucasian group in terms of attrition risk. For example, in the ECA program study[20], Hispanics were found to have higher attrition rates compared to African Americans, who in turn had higher rates than Caucasians. In our study we had relatively few non-Caucasians, with considerable ethnic heterogeneity among minorities (African Americans, Hispanics, Puerto Ricans, Native Americans, Vietnamese, Chinese, etc), and therefore did not attempt to analyze attrition rates for each ethnic group. Our analysis predicts that retaining ethnic minority patients with RA is likely to be more difficult. Lower levels of education, a powerful surrogate variable for poorer socioeconomic status, were associated with higher attrition risk even after adjusting for other predictors. This explains why inner-city-based databanks have a lower retention rate.

In RA, damage to joints, especially hand joints, accrues over time and leads to deformities that lead to premature work disability[22]. In our analysis, increasing disease duration was associated with decreased attrition in the univariate analysis, but had no effect in the multivariate models. This may be explained by the fact that we utilize telephone interviews for those patients unable to fill in the questionnaire, ensuring better followup than with mailed questionnaires. We have also reported previously that the most severely disabled patients in ARAMIS are more likely to die than those who are less disabled[23].

In some studies poorer general health has been associated with increased attrition[24] and in others with decreased attrition[25]. Our data are consistent with findings by Reisine, *et al*[6] that functional disability did not predict attrition. In that study, the disease variables predicting attrition were the number of joint groups with flares, education level, and level of social support. Interestingly, they did not find intercorrelations between disease duration, disease "stage," and number of deformed joints.

Literature from the mental health epidemiology field is inconclusive about whether depression and other illnesses lead to increased attrition. Some reports suggest that psychosocial factors have only a weak to moderate effect on attrition in general[20,21]. Among patients with RA, psychological factors *per se* had no influence on continued study participation[6].

Several caveats to our findings are due. First, the terms "dropout" and "attrition" lack a universally accepted definition. Ideally, the dropout status needs to be verified by interviewing the patient. In real-life clinical research, such

confirmation is not always carried out, given the program priorities and constraints. We believe that the conservative definition of attrition we used is clinically relevant and useful. While it can be argued that more complex definitions, such as missing 2 or more consecutive questionnaires, should be used, we believe that those are unlikely to add to our analyses. Until methodological research establishes a more satisfactory way to define attrition, we recommend that our simple definition be used.

Secondly, loss of study subjects due to death is a major issue, as the incidence of RA increases with age and because RA is associated with increased mortality. In calculating attrition rates, we censored the observations of those who died; in doing so, the calculated rates of attrition may be underestimated. However, mortality itself is an important outcome and is not necessarily a loss to longterm studies.

Our results have implications for our understanding of the validity of followup studies being reported, and for the design of better followup studies in the future. We have provided the evidence that followup studies can potentially be biased due to selective loss of subjects who are younger, are non-Caucasian, and are from poorer socioeconomic background. Strategies like oversampling a high-risk group may easily be used to minimize such bias. Measures for reducing attrition are available and should be utilized[1,10,26,27]. We believe that all databanks should be required to report their attrition pattern so that the results they report can be interpreted correctly. Sensitivity analyses to assess the effect of attrition on inferences about the study subjects also needs to be mandatory[28].

REFERENCES

1. Krishnan E, Fries JF. Measuring effectiveness of drugs in observational databanks: promises and perils. Arthritis Res Ther 2004;6:41-4.
2. Fries JF. The chronic disease data bank: first principles to future directions. J Med Philos 1984;9:161-80.
3. Krishnan E, Singh G, Tugwell P. Long-term observational studies. In: Smolen J, Lipsky P, editors. Targeted therapies in rheumatology. London: Martin Dunitz; 2003.
4. Wolfe F, Lassere M, van der Heijde D, et al. Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. J Rheumatol 1999;26:484-9.
5. Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. Stat Med 2001;20:3715-28.
6. Reisine S, Fifield J, Winkelman DK. Characteristics of rheumatoid arthritis patients: who participates in long-term research and who drops out? Arthritis Care Res 2000;13:3-10.
7. The National Death Index (NDI): National Center for Health Statistics. [Cited March 4, 2004] Available from: http://www.cdc.gov/nchs/r&d/ndi/ndi.htm
8. Singh G. Arthritis, Rheumatism and Aging Medical Information System Post-Marketing Surveillance Program. J Rheumatol 2001;28:1174-9.
9. Daglish GM. Random House Webster's dictionary of American English. New York: Random House; 1997:46.
10. Ribisl K, Walton W, Mowbray C, Luke D, Davidson W, Bootsmiller B. Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: review and recommendations. Evaluation and Program Planning 1996;19:1-25.
11. Clayton D, Hills M. Statistical models in epidemiology. Oxford, UK: Oxford University Press; 1993.
12. Gould W. Jackknife estimation. Stata Technical Bulletin 1995;24:25-9.
13. Peto R, Lee P. Weibull regressions for continuous-carcinogenesis experiments. Biometrics 1973;29:457-70.
14. Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. Am J Epidemiol 2000;152:874-83.
15. Russell C, Palmer JR, Adams-Campbell LL, Rosenberg L. Follow-up of a large cohort of Black women. Am J Epidemiol 2001;154:845-53.
16. Dudley J, Jin S, Hoover D, Metz S, Thackeray R, Chmiel J. The Multicenter AIDS Cohort Study: retention after 9 1/2 years. Am J Epidemiol 1995;142:323-30.
17. Hessol NA, Schneider M, Greenblatt RM, et al. Retention of women enrolled in a prospective study of human immunodeficiency virus infection: impact of race, unstable housing, and use of human immunodeficiency virus therapy. Am J Epidemiol 2001;154:563-73.
18. Streib GF. Participants and drop-outs in a longitudinal study. J Gerontol 1966;21:200-9.
19. Farmer ME, Locke BZ, Moscicki EK, Dannenberg AL, Larson DB, Radloff LS. Physical activity and depressive symptoms: the NHANES I Epidemiologic Follow-up Study. Am J Epidemiol 1988;128:1340-51.
20. Eaton WW, Anthony JC, Tepper S, Dryman A. Psychopathology and attrition in the epidemiologic catchment area surveys. Am J Epidemiol 1992;135:1051-9.
21. de Graaf R, Bijl RV, Smit F, Ravelli A, Vollebergh WA. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Am J Epidemiol 2000;152:1039-47.
22. Sokka T, Kautiainen H, Mottonen T, Hannonen P. Work disability in rheumatoid arthritis 10 years after the diagnosis. J Rheumatol 1999;26:1681-5.
23. Wolfe F, Mitchell DM, Sibley JT, et al. The mortality of rheumatoid arthritis. Arthritis Rheum 1994;37:481-94.
24. Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. Am J Epidemiol 1978;108:367-72.
25. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR Jr. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 1987;126:310-8.
26. Leigh JP, Ward MM, Fries JF. Reducing attrition bias with an instrumental variable in a regression model: results from a panel of rheumatoid arthritis patients. Stat Med 1993;12:1005-18.
27. Kessler RC, Little RJ, Groves RM. Advances in strategies for minimizing and adjusting for survey nonresponse. Epidemiol Rev 1995;17:192-204.
28. Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. Epidemiology 2002;13:347-55.