**The Arthritis, Rheumatism and Aging Medical Information System (ARAMIS): Still young at 30 years**

B. Bruce, J.F. Fries

**ABSTRACT**

Chronic diseases such as atherosclerosis, arthritis, diabetes, and cancer are among major public health concerns. To understand their cumulative risk factors and antecedents, a chronic disease databank consisting of time-oriented, multidisciplinary longitudinal data, prospectively collected on consecutive patients and describing their clinical courses, provides a systematic anthology of patient reported outcome (PRO) data. ARAMIS, which began in the mid-1970s, was the first large-scale chronic disease data bank system. Outcomes data are collected using the Health Assessment Questionnaire (HAQ), a well established PRO instrument that collects patient-centered data in the areas of disability, pain and other symptoms, adverse effects of treatment, economic impact, and mortality. More than 900 peer-reviewed studies have emanated from ARAMIS since its inception. In the earlier days, and even today, ARAMIS had to invent its own tools for the study of these new sciences. ARAMIS has made dominant contributions to the understanding of PROs and to helping improve treatment and health outcomes in rheumatoid arthritis (RA), osteoarthritis (OA), scleroderma, lupus, aging, and drug side effects. It continues to traverse terrain with participation in the NIH “Roadmap” project, the Patient Reported Outcome Measurement Information System (PROMIS). PROMIS is designed to provide improved assessment of health status across all chronic illnesses as part of an improved infrastructure for clinical research. As initiator of the rich history of chronic disease data banks with “rolling” consecutive open patient cohorts, ARAMIS has enabled the study of real-world PROs in rheumatology, with a wealth of resultant improved approaches to treatment, outcome, cost effectiveness, and quality of life.

**Birth of the chronic disease databank**

Chronic diseases, such as atherosclerosis, arthritis, diabetes, and cancer, are major public health concerns in developed nations. A collective understanding of the cumulative factors that affect outcome of these illnesses is essential and presents major challenges. A chronic disease databank is comprised of time-oriented, multidisciplinary longitudinal data, collected prospectively with consecutive “rolling” patient cohorts, with ongoing open recruitment and careful quality control. Typically, patients come from varying locations. Data describe their clinical courses and provide a systematic anthology of patient reported outcome (PRO) data that enables rigorous study (1). As a chronic disease data bank matures and patients follow up approaches the average duration of the disease under study, the cumulative impact of the illness can be examined in a manner not possible by any other method.

The principles underlying the chronic disease data bank model are consistent across disease conditions: 1) the purpose of medical care is to improve patient outcomes; 2) outcomes in modern societies are largely linked to chronic illness, which will continue to increase; 3) these outcomes have numerous risk factors or antecedents; 4) these risk factors often precede overt clinical disease by several years; and 5) computerized, longitudinal study is requisite for clinical research. As initiator of the rich history of chronic disease data banks with “rolling” consecutive open patient cohorts, ARAMIS has enabled the study of real-world PROs in rheumatology, with a wealth of resultant improved approaches to treatment, outcome, cost effectiveness, and quality of life.
as the pilot National Arthritis Data Resource intended to establish the science of longitudinal study of patients with chronic disease (1-3). Since then ARAMIS has been federally funded by the NIH for more than 30 years and has become the model for data banks in other chronic diseases, such as cardiology, stroke and coma, AIDS, aging, and gastrointestinal diseases.

The initial ARAMIS concept was that of a computerized medical record room which would objectify data collection with a time-oriented medical record and make clinical research more efficient using computer techniques; this was at a time (1971) when computers were calculators with little capacity for data storage or data communication. Early problems included excessive drop-outs and a sparse matrix of data with considerable missing data. These led to the need to overlay a protocol on top of observational data in order to regularly collect outcome data in patient-relevant terms, to use mail and telephone to maintain patient contact, and to use PROs as dependent variables in studies. These evolved into the Health Assessment Questionnaire (HAQ) protocols, the Background Information Questionnaire, and the 400-page ARAMIS assessment protocols, developments which transformed and greatly enhanced ARAMIS.

ARAMIS initiated the principles of the chronic disease data bank of enrolling consecutive patients and following them for life, during which they are regularly assessed for multiple factors, including demographics, socioeconomic status, the biology of disease, the influence of comorbid conditions, the mechanics and setting of care, specific medical and surgical treatments, and associated costs. The multiple ARA- MIS data sets contain systematically collected longitudinal PRO data that are gathered prospectively at regular intervals and which are rigorously quality controlled by protocol. Data are organized for investigation, observations are standard, and in essence each patient is on a universal, prospective research protocol. PRO data are supplemented by review of all hospitalizations and determination of all deaths through the NDI Plus. In some cohorts, ARAMIS collects data on genotypes, inflammatory markers, chronic disease laboratory-based risk factors, and protein microarrays.

**ARAMIS databank centers**

ARAMIS is characterized by a consortium of data bank centers with a common core support group at Stanford University. The Stanford center provides design and biostatistical support and serves as the data repository. There have been 16 active data bank centers in the United States and Canada which include a selected group of practices and institutions, both public and private. The use of multiple data bank centers with unique and diverse characteristics but common protocols permits comparisons of health outcomes among different settings. These include African-American, Hispanic, American Native, and non-White Hispanic populations, capitatted and fee-for-service organizations, and both inner city and more affluent settings.

**ARAMIS datasets**

ARAMIS contains PRO data on about 14,000 rheumatoid arthritis (RA), osteoarthritis (OA) patients and individuals from normal and healthy aging populations. The characteristics of ARAMIS RA and OA patients are similar to other patient groups reported in the literature relative to age, gender, and disease duration. RA patients are in their mid 50s and are about two-thirds female. OA patients are in their mid 60s and about 75% female. Patients with a wide range of disease severity are included. In addition, ARAMIS has been following the RA National Inception Cohort (NIC), which consists of more than 900 RA patients seen within the first year of disease onset and recruited in the mid-1990s. Patients were recruited from clinical members of the American College of Rheumatology, and data also include a serum and DNA data bank and radiographic data.

**Data collection: The Health Assessment Questionnaire (HAQ)**

The HAQ (4, 5) is the basis for the analytic data used by ARAMIS. HAQ data are collected semi-annually from RA and OA patients and annually from the normal and healthy aging cohorts and the NIC using the scanning-formatted version of the HAQ, the ScanHAQ. The HAQ has been deployed in ARAMIS studies since 1979 and provides up to 24 years of longitudinal PRO data at this time (6).

The HAQ, as generally referred to, covers five generic patient-centered dimensions: disability, pain and other symptoms, adverse effects of treatment, economic impact, and mortality (4, 5). However, it is the short or 2-page HAQ, containing only the HAQ Disability Index (HAQ-DI) and the pain and patient global health visual analog scales, that has received the greatest attention. Three recent reviews each reference over 200 selected publications on the reliability, validity, and application of the HAQ-DI and its pain and global health scales (4, 7, 8). With the exception of the HAQ-DI and pain and global health scales, the remaining components of the HAQ are modified periodically to address issues of contemporary scientific interest. The HAQ-DI and its pain and global health scales have remained essentially unchanged since 1981 (5). The HAQ is among the most cited and used PRO instruments, particularly but not exclusively in the rheumatic disease literature.

At present, more than 500 such studies on the HAQ have been published, and the HAQ-DI has been translated and validated in more than 60 languages. Multiple studies compare the HAQ-DI with other instruments (4, 9). Reliability and validity have been repeatedly documented for all HAQ variables (4, 8-10), and it is among the outcome measures required by the FDA for new drug approval. It has been used in thousands of clinical trials and other research studies, has frequently been used in the clinic, and its normative values are well established (11). In addition, other validated self-assessment instruments have periodically been included in ARAMIS assessments. These include items from the SF-12 (12-14), Lorig’s Self-Efficacy scale (15,16), a patient preference “feeling thermometer” (17, 18), the RADAR (Rapid Assessment of Disease
Activity in Rheumatology) self-administered joint counts, the AIMS (Arthritis Impact Measurement Scales) anxiety and depression scales (19, 20), and the CES-D (Center for Epidemiologic Studies Depression Scale) (21-23).

Data collection protocols
Standardized protocols for patient enrollment, HAQ administration, and data quality control are followed by all data bank centers.

Patient enrollment
RA and OA patients are recruited by the data bank physician or a trained staff member during their first clinic visit as part of their usual care; over 95% accept. All patients provide informed consent and complete an initial questionnaire to establish their demographic profile and drug history. Patients also provide periodic updates on demographic variables such as health behaviors, weight, employment status, health insurance, and living arrangements.

HAQ administration
Rigorous protocols standardize questionnaire administration. Procedures have been established to accommodate non-English speaking Hispanic patients, as well as other non-English speaking patients and those with low levels of literacy. For some data bank centers, questionnaires are administered at clinic visits instead of, or in addition to, mailed questionnaires. If patients prefer, they may complete the questionnaire during a telephone interview. Vigorous follow-up identifies the outcome status of patients and minimizes loss to follow up. Follow-up protocols commence for those patients who do not respond within the initial two-week period after mailing. Non-responders are traced using standardized procedures, which includes search of the National Death Index (NDI Plus). At the end of each questionnaire cycle, all patients are classified according to their study status: ongoing, dead, lost or unable to contact, withdrawn for personal reasons, or administratively withdrawn. The annual patient retention rate averages 98%.

PRO data coordination
The ARAMIS Outcome Assessment core unit at Stanford coordinates the administration of studies and projects among the different population groups. New and ongoing staff undergo thorough training, and there is regular follow-up training for protocol revisions, new data coding manuals, and to insure compliance with study protocols. Outcome assessors at each databank are responsible for questionnaire administration, processing, patient follow up, and their data bank’s management.

Data quality control (QC)
Rigorous protocols function to maximize data quality. Scanning technology is used for QC of data entry, and SAS programs are used to QC data prior to analyses. Uniform outcome assessment scanning and clinical abstracting manuals are used to insure document consistency and quality of project data. Pilot studies of instruments and items test the clarity and consistency of the data, the suitability of abstracting, coding, and entry procedures, and reliability and validity (e.g., test-retest reliability, validation against a gold standard, etc.). Instruments, procedures, and manuals are revised regularly to clarify procedures or coding rules, add new items of interest, update medication lists, etc. All returned questionnaires are checked for completeness, ambiguities, or inconsistencies, and patients are contacted for clarification. Medical records pertaining to numerous types of inpatient hospitalizations, surgeries, emergency room visits, and nursing home care are obtained from providers and the patient, and then are reviewed, coded, and entered.

Data entry
Data entry is performed using scanning technology, which has eliminated hand coding and has reduced data entry time and delays in data entry, while increasing data standardization, quality, entry efficiency and accuracy. The scanning coordinator and staff have been trained in scanner operation, software usage, document preparation, and problem resolution. The scanning software performs coding, range and relationship checks, and provides snippets for review by the operator for ambiguous entries, which are then corrected and exported into SAS for QC. For example, a man who responds to the “women only” questions or patients with total counts not equal to the sum of the component items would be flagged. New master research data sets, devoid of personal health information for maintenance of confidentiality, are then prepared for analyses.

Results
ARAMIS has received more than 30 million dollars in federal funding over the years, and its research productivity deserves examination, both with regard to research volume and the importance of the results generated. Since its inception, more than 900 peer-reviewed studies have emanated from ARAMIS. Detailed listings can be found at ARAMIS.Stanford.edu.

ARAMIS began in an era of as yet under-developed capabilities and methodologies in an area that was conceptually new. Content-independent data banks (with schemas) required conceptual and software development. Computers could not efficiently store and manage large datasets, so the historically very important concepts of a relational database (24), associated data storage techniques (25,26), and the clinical representations of “time-oriented” longitudinal data (24), had to be invented by ARAMIS. Statistically, converting clinical information “at the margin” using stepwise logistic regression needed development (27), as did the medical use of recursive partitioning (28). Classification criteria for rheumatoid arthritis (28), lupus (29), arteritis (30-33), and other rheumatic diseases required more sophisticated methods, and ARAMIS performed these studies. The first example of what is now called “computerized adaptive testing” (CAT), a central tenet of the ARAMIS project, was introduced by ARAMIS three decades ago (34). In all of these centrally important and now familiar advances, ARAMIS was a central and leading force.

The concepts of dominant chronic illness rather than acute outcomes; risk...
Factors for disease onset, disease progression and health rather than medical models; patient-valued outcomes; prevention of disease as well as cure; and morbidity as being even more important than mortality have had many parents and are in the ascendancy. ARAMIS has made two critical contributions to this shift of paradigms. First, as discussed above and elsewhere in this issue, the Health Assessment Questionnaire (HAQ) was introduced in 1979-80, and has had profound effects on how we conceive, value, and measure patient outcomes. The initial HAQ paper (35) is the most cited article in rheumatology of all time. Second, the Compression of Morbidity hypothesis was introduced in 1980, and over 25 years has gone from an “interesting notion” (36) to an established paradigm (37) that now underlies the new science of successful aging and healthy aging initiatives in many settings, including the WHO and U.S. Medicare. The Compression of Morbidity paradigm envisions health futures, now beginning to occur in the U.S. and other developed nations (37,38) where postponement of the onset of morbidity begins to come closer to the age at death, resulting in decreased total lifetime morbidity. The original article (35) has been the most cited article in the history of gerontology.

The ARAMIS themes have generally been based on a series of focused studies around an hypothesis. Early studies in scleroderma developed avascular hypothesis that led to the successful treatment of renal fibrosis, while lupus studies separated some of the subsets of the disease and helped individualize treatment to the subset (39). In RA, the “inversion of the pyramid” movement drew on ARAMIS data for many of its concepts and most of its data (40-42), and ARAMIS has recorded the positive results of this clinical shift (43). The NSAID gastropathy “epidemic” was uncovered by ARAMIS studies, which quantitated its importance, documented risk factors (44-46), proved differences in toxicity between traditional NSAIDS, studied the importance of measures to reduce incidence, and chronicled the reduction in incidence which resulted from these measures (47). In OA, studies indicated the importance of exercise and other behavioral risk factor interventions (48), and the neutral to protective effects of exercise on total joint destruction. In morbidity compression, the ARAMIS longitudinal studies of aging established lower cumulative lifetime disability with good health habits (49, 50) and, together with the national health studies (NLTCs, NHIS) and the intervention trials, it established the proof. In patient education, the ARAMIS data banks have supported and contributed to the landmark advances of Dr. Kate Lorig and her group (51).

ARAMIS today
ARAMIS themes continue to evolve, with emerging interests in areas of ethnic disparities in health, the influence of RA and its treatment on comorbidity and the reciprocal influences of morbidity upon RA outcomes, and new technological approaches to disseminating effective patient education to broader audiences. A set of new projects will allow the simulation of head-to-head comparisons of treatment where none exist, enabling major improvement in clinical decision making.

ARAMIS continues to traverse new terrain with its participation in the NIH “Roadmap” project, the Patient Reported Outcome Measurement Information System (PROMIS). PROMIS is designed to provide improved assessment of health status across all chronic illnesses as part of an improved infrastructure for clinical science (52). As envisioned, PROMIS will develop large banks of patient value-based items and will improve these items with regard to verbal clarity and comprehension issues, face validity, patient relevance, uniqueness, comprehensiveness and, finally, psychometric properties such as degree of difficulty and fit with a particular content area through the mechanisms of Item Response Theory and Computerized Adaptive Testing. ARAMIS will compare effectiveness of new approaches with that of more traditional PRO approaches (such as the SF-36 or the HAQ-DI) by clinical trials, with particular attention dedicated to the ability of each approach to detect clinically important change. This approach should permit the blending of generic and disease-specific approaches to health status assessment, which will yield improved methods for PRO research and improve consensus in the field. Sample size requirements for clinical trials will be reduced substantially with improved precision of outcome assessment.

The world outside the research laboratory is more complex than that within, and chronic disease data banks such as ARAMIS have led to fundamental advancements in the longitudinal study of chronic disease outcomes and how to improve upon them. Chronic disease data banks have enabled the study of real-world PROs that are not otherwise obtainable. This has resulted in a wealth of significant new information regarding treatment, outcome, cost effectiveness, and quality of life. ARAMIS has made major contributions to theory, methods, constructs, and clinical research and continues to do so.

References


