Cohort Profile: The National Academy of Sciences-National Research Council Twin Registry (NAS-NRC Twin Registry)

Margaret Gatz,1,2* Jennifer R Harris,3 Jaakko Kaprio,4,5 Matt McGue,6 Nicholas L Smith,7,8 Harold Snieder,9 Avron Spiro III10,11 and David A Butler12 for the Institute of Medicine Committee on Twins Studies

1Department of Psychology, University of Southern California, Los Angeles, CA, USA, 2Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, 3Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway, 4Department of Public Health and Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland, 5National Institute for Health and Welfare, Department of Mental Health and Substance Abuse Services, Helsinki, Finland, 6Department of Psychology, University of Minnesota, Minneapolis, MN, USA, 7Department of Epidemiology, University of Washington, Seattle, WA, USA, 8VA Seattle Epidemiologic Research and Information Center, Puget Sound Health Care System, Seattle, WA, USA, 9Department of Epidemiology, University of Groningen, Groningen, The Netherlands, 10Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Jamaica Plain, MA, USA, 11Boston University Schools of Public Health and Medicine, Boston, MA, USA and 12Medical Follow-Up Agency, Institute of Medicine, National Academy of Sciences, Washington, DC, USA

*Corresponding author. Department of Psychology, University of Southern California, 3620 McClintock Avenue, Los Angeles, CA 90089–1061, USA. E-mail: gatz@usc.edu

Abstract

The National Academy of Sciences-National Research Council Twin Registry (NAS-NRC Twin Registry) is a comprehensive registry of White male twin pairs born in the USA between 1917 and 1927, both of the twins having served in the military. The purpose was medical research and ultimately improved clinical care. The cohort was assembled in the early 1960s with identification of approximately 16 000 twin pairs, review of service records, a brief mailed questionnaire assessing zygosity, and a health survey largely comparable to questionnaires used at that time with Scandinavian twin registries. Subsequent large-scale data collection occurred in 1974, 1985 and 1998, repeating the health survey and including information on education, employment history and earnings. Self-reported data have been supplemented with mortality, disability and medical data through record linkage. Potential collaborators should access the study website [http://www.iom.edu/Activities/Veterans/TwinsStudy.aspx] or e-mail the Medical Follow-up Agency at [Twins@nas.edu]. Questionnaire data are being prepared for future archiving with the National Archive of Computerized Data on Aging (NACDA) at the
Why was the cohort set up?

The NAS-NRC Twin Registry was established in the late 1950s within the Medical Follow-up Agency (MFUA), primarily to shed light on the determinants of veterans’ health and to elucidate the role of genetic and environmental influences on disease etiology using a twin design.¹⁻⁴ The mission of MFUA was to use medical records from World War II to advance medical research and improve clinical care. The then director of the National Institutes of Health described the studies within MFUA as offering a unique opportunity “to follow a whole generation of men and trace their life history”.⁵ Original funding was primarily from the Veterans Administration. The National Research Council appointed an advisory committee called the Committee on Veterans Medical Problems (now the Board on the Health of Select Populations) and subsequently a subcommittee on twins (now the Committee on Twins Studies). MFUA is currently within the Institute of Medicine, which is part of the National Academy of Sciences, Washington, DC.

Who is in the cohort?

The registry was launched in 1958–59 through matching birth certificates for White male multiple births (from all states except Arizona, Connecticut, Delaware, Georgia, Maine, Missouri, Utah, Vermont and the city of New Orleans, thus representing 93% of population of the USA at the time). Whereas all multiple births were identified, it was decided that the registry would include only White men (who made up nearly 90% of the WWII-era service members) to create a more homogeneous cohort.³ Veterans Administration (VA) records were used to find twin pairs born 1917 through 1927 where both members served in the Armed Forces.¹⁻²⁴,⁶ These efforts led to identifying 15,924 twin pairs, or 31,848 individuals.

Starting in 1965, the men were mailed a brief questionnaire (Q1) requesting their enrolment in the study and including questions to determine zygosity. Of those who were living, 69.8% responded; 7.9% were deceased.⁶ Questionnaire 2 (Q2), constituting the baseline assessment, was mailed after the man replied to Q1. Q2 was an epidemiological questionnaire largely identical to a survey mailed to the Swedish Twin Registry in 1967.⁷,⁸ Of those who had previously answered Q1, 72.6% responded to Q2; 1.2% were newly deceased.⁶ Military records were reviewed in 1970 at the National Personnel Records Center in St Louis, MO, with the abstracted information archived on microfiche in the MFUA office in Washington, DC. Essentially, 99% of the twin pairs have birth certificates and information extracted from their service records, including induction data. If service record data are used to compare men selected for the cohort who did not participate at baseline with those with baseline data: there is no difference on age at which they entered the service; those who participated at baseline served on average 100 days longer; and officers were more likely to participate than enlisted men. Those who
participated had 0.4 more years of education at the time of enlistment and were taller by 7.5 mm.

Table 1 summarizes the characteristics of the 15,924 pairs in the registry and provides descriptive statistics for the men who participated at each wave.

**How often have they been followed up?**

The entire registry has been followed up by mailed questionnaire (Q) three times—Q3, Q7, Q8—with a fourth mailing comprising the NEO Five Factor Personality Inventory. The other Qs were either administrative mailings or were part of ancillary data collection.

The Q3 questionnaire, focused on schooling and earnings, was sent in 1974 to pairs where both were alive, had valid addresses and had not requested to be removed from the registry (N = 12,640 individuals). In all, 72.8% responded; 0.2% were newly deceased. Responses to Q3 were only coded for pairs where both members responded. The most thorough analysis of representativeness of the sample and of the pairs who responded was carried out at Q3. These investigators reported that little evidence of bias was detected. Q3-responding twins were quite comparable to the entire twin panel: there was slightly better survival starting at age 47 compared with the general US White male population; among responding pairs, rates of receiving disability benefits were similar to those among all World War II veterans; a higher percentage had completed 16 or more years of school compared with all US White male veterans; and twin respondents were more likely to fall into the highest earnings category than were all White male veterans or all US White males.
In 1985, the second epidemiological questionnaire (Q7), similar to the first epidemiological questionnaire (Q2), was mailed to all individuals in the registry. In 1998, the third epidemiological questionnaire (Q8) was mailed to all surviving pairs and to ‘singleton’ twins (i.e. whose co-twin was not available, due to death or lack of a valid address) who had previously completed both Q2 and Q7. A 69% response rate was reported for Q8. In 2005, the NEO Five Factor Personality Inventory was mailed to surviving twins. It measures the personality dimensions of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Questionnaires were returned by 1670 men.

Mortality updates use the Veterans Affairs’ Beneficiary Index and Records Locator System database, the National Death Index and LexisNexis. LexisNexis is a corporation providing an electronic database for legal information and public records. Dates of death are currently available through 2012, and updating cause of death is in progress.

Ancillary data collection, described below, occurred between these waves to target informative subsets of the twin pairs. These ancillary data are not currently available at MFUA.

In 1969–72, funded by the National Heart, Lung and Blood Institute (NHLBI), a sample of 514 twin pairs from the registry living near: Indianapolis; Framingham, MA; or San Francisco, Davis or Los Angeles, CA; were selected for clinical examination. This subset of twins is known as the NHLBI Twin Study. Five longitudinal assessments of this subset took place between 1981 and 2003, including lipid panels, structural MRI during the last three waves and blood samples from 711 men in the third wave for genotyping.

The Duke Twins Study of Memory and Aging was a study of dementia entailing four waves of telephone cognitive screening with the Telephone Interview for Cognitive Status (TICS). The sample comprised all twin pairs from the original registry in which both members were thought to be alive and residing in the USA in 1989. Individuals who screened positive for dementia (N = 550 over the four waves) and their co-twins were given a complete clinical workup, including a neuropsychological battery and collection of blood or buccal samples. The second telephone interview of all individual twins believed to be alive in 1992 included screening not only for dementia, but also for Parkinson disease, cardiovascular disease eye disease and cancer. Based on that screening, participants were identified for further ancillary studies of Parkinson disease and age-related macular degeneration.

What has been measured?

Table 2 summarizes the data that have been collected. Copies of the epidemiological questionnaires may be found on the Institute of Medicine (IOM) Studies of U.S. Veteran Twins website. The main longitudinal data are the repeated epidemiological surveys encompassing height, weight, cardiovascular symptoms, problems breathing, other health conditions, smoking, alcohol use and other health behaviours. There are up to 145 diagnoses reflecting diseases and hospitalizations from 1936 through 1983 coded using the International Classification of Diseases (ICD) based on VA admissions records. Also available at one or more times of measurement are: education; employment history; occupations of father, mother, spouse and twin brother; whether the respondent grew up in an urban or rural area; and contact with his twin brother. From service records, branch, rank and dates of entry and separation were obtained. There are standardized ability test scores including Army General Classification Test (AGCT) / Armed Forces Qualification Test (AFQT) scores for 2438 men and General Classification Test (GCT) scores for 2386 men.

Classification of zygosity, crucially important in twin studies, initially relied on responses to two questions: (i) ‘As children were you and your twin as alike as two peas in a pod or of only ordinary family resemblance?’; and ii) ‘In childhood, did parents, brothers and sisters or teachers have trouble in telling you apart?’ At baseline, responses of both members of a pair were used to classify zygosity and certainty of the classification. When compared with blood typing results for 741 twin pairs, agreement was 95%. More recently, DNA genotyping has become available from various ancillary studies. Accuracy of questionnaire classification of zygosity when validated by DNA markers was 96.8%. The zygosity reported in Table 1 uses DNA if available, and the baseline zygosity classification based on questionnaire responses, blood typing and physical similarity if DNA was not available. To facilitate use of all of the data in biometric modelling, the questionnaire classification of zygosity has been recast as the twin pairs’ probability of being monozygotic, rather than as MZ, DZ and unknown categories, based on a logistic regression model that used available DNA as the gold standard.

What has been found?

The project bibliography currently includes 367 publications addressing cardiovascular risk factors, schizophrenia, air pollution, eye disease, socioeconomic indicators, alcohol use, smoking, obesity and cognitive functioning. A list of publications can be found at the IOM website.

Early publications included methods papers and publications concerning psychopathology that took advantage of linkages with medical records from the armed forces and the VA. For example, in one early analysis of
schizophrenia in the twin panel, the authors found no difference in incidence of schizophrenia among MZ as compared with DZ twins, although onset in the MZ pairs occurred at a younger age. The age-corrected pairwise concordance rate was 15.5% for MZ and 4.4% for DZ twins. In MZ pairs, schizo-affective psychosis had double the pairwise concordance rate compared with schizophrenia, whereas pairwise concordance was similar for schizo-affective psychosis and manic depressive psychosis. Work in this area contributed to formulating the diathesis-stress theory that the occurrence of schizophrenia reflects the interaction of genetic vulnerability with environmental influences (including prenatal environment). The epidemiological questionnaires have been used in multiple studies that have helped to establish the importance of smoking in the population’s health. A 24-year mortality follow-up for smoking-discordant twin pairs identified in Q2 found more than twice the rate of death among the twins who smoked compared with their non-smoking co-twins. For lung cancer, the relative risk was 5.0 in MZ and 11.0 in DZ smoking-discordant pairs. Within the NHLBI subsample, several analyses explored the interaction of measured genes with smoking; for example, whereas smoking was associated with poorer cognitive function, this harmful effect was weaker among APOE epsilon4 carriers than among noncarriers. Smoking was
also shown to be related to more advanced age-related macular degeneration, whereas diet (vitamin D, betaine, methionine) was protective.26

As the twins have grown older, studies of conditions related to ageing have emerged. The Q8 questionnaire was used to derive a trait of ‘wellness’ among the respondents, who were on average nearly 75 years old.10 Wellness entailed the absence of any of the major chronic diseases (heart attack, coronary surgery, stroke, diabetes, prostate cancer). Heritability estimates for wellness were over 50%. Data from the Duke Twins Study have shown lower dementia risk associated with greater cognitive activity in midlife27 and with occupations that require use of data, academic skills or extensive vocational training.28 Exposure to solvents, especially trichloroethylene, ascertained from task-based occupational histories, raises Parkinson disease risk.29 In the NHLBI sub-study, glucose load measured over 30 years previously was associated with risk of death from coronary heart disease, controlling for other vascular risk factors.30 These studies demonstrate the value of the data in connecting exposures over the life cycle to later life morbidity and mortality, while controlling for genetic and environmental factors shared between co-twins.

What are the main strengths and weaknesses?

The NAS-NRC Twin Registry is a unique longitudinal sample, representative of their cohort, the so-called ‘Greatest Generation’.31 Nonetheless, it must be kept in mind that the sample was selected to include only White male pairs in which both members survived and passed a physical examination to enter military. The longitudinal twin design is a strength. Being a twin registry, investigators can study the interplay of environmental with genetic influences on health and socioeconomic conditions. The wealth of data collected from these men over 40 years, including self-assessments as well as record linkage, provides the basis for better understanding the contributions of a lifetime of influences to longevity and health in old age.

The main weakness stems from the lack of infrastructure funding to MFUA for maintaining the registry. Therefore, neither health utilization records nor cause of death have been comprehensively updated for over two decades, although updating of death records is currently in process. Further, the impact of the men’s military experiences has not been fully explored, and a fire in 1973 destroyed original service records for a substantial proportion of the members of the registry, making further data gathering and verification of data problematic. Biological samples, including genotyping, were only collected from men who participated in ancillary projects in which they were evaluated in person—these represent <5% of the cohort. Finally, all of the data from the ancillary projects have not been integrated into the registry, although this too is in process.

How can I access the data?

Information about access to the NAS-NRC Twin Registry can be found on the IOM website, [http://www.iom.edu/Activities/Veterans/TwinsStudy.aspx]. Letters of interest may be sent to David A Butler, Medical Follow-up Agency, NAS-NRC Twin Registry, [Twins@nas.edu]. If obtained from MFUA, there is a one-time fee of $10,000 for an anonymized dataset. To facilitate access, questionnaire data and data from the ancillary studies are being prepared for future archiving with the National Archive of Computerized Data on Aging (NACDA) at the Inter-University Consortium for Political and Social Research (ICPSR), University of Michigan, MI.

Acknowledgements

We would like to thank all of the twins who participated, Harriet Crawford and the other staff of the Medical Follow-up Agency at the Institute of Medicine, and the previous members and leadership of the IOM Committee on Twins Studies.

Funding

The authors acknowledge financial support from the National Institutes of Health: R21 AG039572.

Conflict of interest: None declared.

References