Abstract. The population-based twin registries of Denmark, Finland, Norway and Sweden represent an extraordinary resource for scientific research. Although each register has its own history and composition, they share certain common qualities. All the Nordic countries have a long tradition of population registration, a high standard of living and health-related registers of high quality. The large size of the registers means that they are uniquely placed for representative studies of rare occurrences. Examples of studies that these registers make possible are illustrated with data from the Finnish Twin Cohort, which in its first phase consisted of over 17,000 like-sexed twin pairs born before 1958. It has been recently expanded to include multiple births between 1958 and 1986 (nearly 23,000 sets) and their first-degree relatives. During this period the DZ/MZ ratio decreased, while an increase in MZ twinning rates was observed between 1974 and 1986.

Key words: Twins, Diseases-in-twins, Registries, Longitudinal studies

INTRODUCTION

This paper will review the use of the population-based twin registries, such as those of the Nordic countries, as unique research resources. Instead of going into detail about the history, size, composition and research program of each of the Nordic registers, we want to emphasize what such population-based registries offer in contrast to twin panels recruited as volunteer samples and smaller twin series. Selected illustrations of the...
strengths of such registries will be mostly from the Finnish Twin Cohort Study. A more complete perspective can be found in Schwartz et al [29], who provide a bibliography of twin research emanating from the Nordic countries.

The Nordic registries of Denmark, Finland, Norway and Sweden are all large, population-based twin registers. In all these countries there is a long tradition of population registration going back three to four centuries. Thus local parishes have recorded births, deaths, marriages and migration from the local parish and reported this information to the authorities. Each twin study has compiled a register of twins born in certain years using parish records of births or using centralized population registers based on the parish records. Each twin register has been described in detail elsewhere [3,4,7,10,17,21]. The parish records are family-based, and it is possible to use them to build pedigrees and compile data on first-degree relatives for twin family studies. Such studies have been undertaken in all four twin registries.

**THE FINNISH TWIN COHORT STUDY**

The Finnish Twin Cohort consists of two parts. The older part of the cohort, compiled in 1974, consists of all like-sexed pairs born before 1958 with both twins alive in 1967 [10]. The newer part of the cohort, compiled in 1987, includes all multiple births from 1958 through 1986 that were identifiable from the Central Population Registry (CPR) of Finland.

*The older part of Finnish Twin Cohort* was compiled in 1974 from the CPR using selection procedures described in detail elsewhere [12]. Pairs of persons with the same surname at birth, same date of birth, same place of birth and same sex were chosen. We excluded non-twins, who satisfied these criteria, based on questionnaire data and following inquiries from local parish registers. Thus, the older part of the Finnish Twin Cohort is a nationwide, population-based panel of adult, like-sexed twin pairs, born before 1958 and with both cotwins alive in 1967. The older cohort contains information on 17,000 pairs. Two questionnaire surveys of the older cohort in 1975 and 1981 collected data for genetic epidemiological studies. Nationwide computerized registries yield morbidity and mortality data, including cause-of-death, hospitalizations, disability pensions, and cancer incidence, as well as drugs prescribed free of charge for certain chronic diseases (hypertension, diabetes, coronary heart disease, asthma, epilepsy, etc). The twin cohort has been used in a wide variety of studies, principally of chronic disease and risk factors for disease.

*The younger Finnish Twin Cohort*. The twin cohort was expanded in 1987 to include twins born after 1957, the last birth year of the older cohort. Because the twins were younger, and in nearly all pairs one or both parents would have been alive at the time of transfer of family data to the CPR, compilation of the younger cohort was more efficient than for the old one. Sets of persons born on the same day and sharing the same personal identification number (PIN - a unique 10 digit number composed of a 6-digit date of birth, a 3-digit alphanumeric code and a check digit) of the parent(s) were
chosen. Thus, all types of multiple births from 1958 to 1986 were identified, so that both like-sex and opposite-sex twin pairs, as well as higher-order multiple births (triplets, quadruplets, etc) have been ascertained, a total of 46,059 individuals in 22,898 sets of multiple births (22,650 twin pairs, 236 triplets, 9 quadruplets and 3 quintuplets). Members in the younger twin cohort were then linked to their families by the Central Population Registry to give information on the parents, spouses and offspring.

Ascertainment

Among multiple birth sets with all members alive at the time of compilation, there was a total of 21,958 twin pairs (estimated zygosity distribution: 6,114 MZ, 7,922 DZ and 7,922 opposite-sex pairs), 207 sets of triplets and 7 sets of quadruplets. The distribution of twin pairs by birth year is shown in Fig. 1, the total number of twins declining in the late 1960s and early 1970s. This decline correlates well with the decline in the absolute number of births in Finland (Fig. 2).

![Fig. 1. Number of twin pairs in the younger twin cohort by birth year. Estimated numbers of MZ and DZ pairs with both cotwins alive in 1987.](https://www.cambridge.org/core)
DZ pairs born in each year, but also by mortality rates, which if different in MZs and DZs, will change the DZ/MZ ratio as well.

In the younger twin cohort file, there were 940 recorded deaths in 692 twin sets and 32 higher order multiple births. There was a total of 1,488 persons in these death affected sets. We used these data to divide the younger twin cohort into two parts concerning completeness of ascertainment. We examined the number of deaths recorded for each birth cohort and the distribution of years of death. For birth cohorts between 1958 and 1973, the annual number of deaths was under 30 equally distributed over time. From 1974 onwards there were more than 30 (range 31-74) deaths recorded in each birth year cohort with 80-95% of deaths occurring in the same year as the year of birth (Fig. 3). Thus infant mortality in multiple births has evidently been adequately recorded only from 1974 onwards in the CPR.

Also, the total number of multiple births derived from the CPR data was compared with official statistics on multiple births. We compared the number of multiple birth sets identified from CPR files, excluding those born abroad (N = 701, 3.1% of all sets), with the number reported by the Central Statistical Office (CSO) in its annual Yearbooks:

<table>
<thead>
<tr>
<th>Year of birth</th>
<th>No. of multiple births</th>
<th>Coverage (CPR/CSO)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPR</td>
<td>CSO</td>
</tr>
<tr>
<td>1958-1973</td>
<td>13422</td>
<td>16156</td>
</tr>
<tr>
<td>1974-1986</td>
<td>8775</td>
<td>9095</td>
</tr>
<tr>
<td>1958-1986</td>
<td>22197</td>
<td>25251</td>
</tr>
</tbody>
</table>
Fig. 3. Number of deaths in the younger twin cohort data set. Proportion of deaths occurring in year of birth and later by birth year.

Fig. 4 shows the yearly numbers of multiple births from these two sources of data. The differences between the 1987 CPR data and the number of multiple births annually recorded at the CSO are due to several features in these registers. The CSO records all births regardless of outcome. If one or both twins were stillborn or died shortly after birth, a PIN might not be assigned and such pairs would be missing from the CPR. Loss of only one twin would result in a pairwise deletion. Other sources of differences could be incomplete reporting of family data to the CPR before the early 1970s, migration (most of which has been to and back from Sweden), clerical or computer errors.

As the twin data set seemed to represent well all twin births from 1974 onwards based on the two pieces of evidence presented above, we have examined mortality rates and twinning rates by birth year and zygosity between 1974 and 1986. Fig. 5 shows that the mortality of MZ pairs is consistently greater than that of DZ pairs in each birth year, provided the mortality of like-sexed and opposite-sexed DZ pairs does not differ.

We then computed twinning rates per 1000 births for all pairs, including pairs with one or both twin dead, by birth year between 1974 and 1986. There was no significant variation in the DZ rates over time (Fig. 6), but there was a significant increase (regression of rate on year, \( p = 0.01 \)) in MZ twinning rates between 1974 and 1986.

**Twin families**

For the twins and their parents basic demographic information is present in the data files compiled at the CPR for the formation of the younger cohort. These include name and
Fig. 4. Total number of multiple births recorded in the younger twin cohort (compiled from the Central Population Registry CPR) and by the Central Statistical Office (CSO) of Finland. No. of multiple births by birth year.

Fig. 5. Cumulative mortality rates in twin pairs by birth year.

current address, vital status (died, emigrated, or temporarily abroad), marital status, and linkage data to the family members.

To illustrate the data already available to characterize the twins and their parents, we use data from the cohorts born 1973-75. There was a total of 1,940 pairs of twins identified through the CPR for these years, of which 1,833 were such that both members
of the twin pair were alive at the time of compilation of the data file in 1987. In addition, there were 20 higher-order multiple births (17 triplets, two quadruplets and one set of quintuplets). All mothers (except one) of the twins and other multiple births had data on their vital status. There were 19 maternal deaths. Nine of the mothers were adoptive mothers, and no information on the biological mother was available. In addition, 19 had a missing address or unknown place of residence, and 104 (5.4%) were permanently or temporarily resident outside Finland. Thus, 1,808 (93.2%) of the twins’ biological mothers were alive and resident in Finland with a known address. For 1,702 (94.1%) of these mothers, both cotwins were, themselves, alive with known address in 1987.

All data on the father were missing for 74 twin pairs; 50 of the fathers had died and 9 were adoptive (nonbiological) fathers. In addition 43 fathers had no address and 76 were permanently or temporarily abroad. Thus, 1,706 (87.9%) fathers of the twins were alive and resident in Finland with a known address. For 1,605 (94.1%) of these fathers, both cotwins in a pair were alive and at a known address in 1987. In the Nordic registries that have started from birth records, the characteristics of twin pairs, with both cotwins surviving, can be compared with those in which either one or both cotwins have died or emigrated.

**ZYGOSITY DETERMINATION**

Accurate zygosity determination is essential in twin studies. When studying very large numbers of twins, genetic markers are not of use for logistical and financial reasons. Therefore, the Nordic registries have developed accurate zygosity determination...
methods based on questionnaire data [2]. A combination of deterministic and stochastic approaches permits the classification of over 95% of adult pairs [20,27,28]. The applicability of questionnaire methods of zygosity determination among adolescents and younger children has not been assessed in the Nordic registries.

REPRESENTATIVENESS OF THE FINNISH TWIN COHORT

The Nordic twin registries are able to assess the representativeness of the twin pairs in manners that other studies are unable to do. Being based on population register data, the twin samples can be compared to the entire population for characteristics recorded in the population register such as marital status, place of residence, occupation. Even more fundamentally, we can assess the basic demographic structure of the twin sample. The proportion of each gender and of each age-group can be compared with expected distributions.

Thus, for the older part of the Finnish Twin Cohort, the gender distribution of the respondent twin pairs is very similar to that found in the population, which is very different from Anglo-Saxon volunteer samples. Volunteer samples have an over-representation of females, as well documented by David Lykken and his colleagues as the rule of “two thirds” [19]. While we found a tendency for females to be more active in responding in some studies on subsamples in the Finnish Cohort, this has not been universally so. For example, in a study on smoking-discordant twins invited to participate in a clinical assessment of CHD risk factors and atherosclerosis, men had a higher participation rate than women [6].

![Graph showing smoking rates in twins and singletons.](https://www.cambridge.org/core)
The Finnish twin cohort has fewer elderly twin pairs than expected; this follows from the requirement that both twins have been alive in 1967, when the older cohort was formed. This is the major difference between the twin cohort and the general population in terms of common demographic characteristics. In addition, the twin cohort has slightly more unmarried twins and slightly more twins living in rural areas than the population at large [12].

Health care in Finland is a mixture of public and private care with nearly all hospital care provided by the local communities. The government covers a substantial proportion of costs of health care. For administrative reasons several centralized, computerized medical registries exist, which, with administrative permission, are also accessible by the research community. These permit follow-up of mortality and cause of death with little loss to follow-up. For certain diseases such as cancer, there also exist incidence registries. Follow-up studies of exposure discordant twins [14] have used these sources of data. Koskenvuo et al [16] found that ischemic heart disease mortality rates among men under 60 years of age and responding to the 1981 twin questionnaire were not significantly different from the population rates during three years of follow-up.

In specific studies, we can compare the distribution of traits in twins with the distribution in singletons to assess whether the trait is as common in twins as in singletons. The information on singletons can be from separate studies that are judged comparable or, as in the case shown here, from including singletons in the twin study. Besides the more than 25,000 twins that replied to the 1975 questionnaire of the Finnish Twin Cohort, over 3,000 singletons were also surveyed. Thus, we can compare the responses to an identical mailed questionnaire from twins and singletons. We see that the proportion of current smokers among men is very similar in twins and singletons for nearly all age-groups (Fig. 7). Also, the age-adjusted proportion of smokers was very similar to that found in concurrent interview studies. Such comparisons are very important in defining the reference population to which the estimates from twin studies can be related.

**SIZE OF NORDIC REGISTRIES**

By virtue of their large size, with many thousands of twin pairs in each twin registry, the Nordic registries are able to analyze effects that smaller studies might only hint at. Exclusion of a genetic effect by MZ to DZ comparison requires a substantially larger number of twin pairs than showing the existence of a significant genetic effect; the size of the confidence interval of the heritability estimate is dependent on sample size.

Only with very large numbers, can one confidently exclude a major genetic effect. In a study of the genetics of myopia, we used data on drivers’ licenses about compulsory use of eye-glasses while driving [31]. We computed probandwise concordance rates for possession of a driver’s license by age, gender and zygosity in Finland based on 3,676 MZ and 8,109 DZ pairs [31]. Apart from older females, there was virtually no difference in concordance by zygosity. With the large numbers involved, one can confidently say that there are no major genetic influences on the possession of a driver’s license. This was of importance for the analysis of the trait in question, ie, myopia assessed by compulsory use of eye-glasses while driving.
Another example of being able to examine age and gender effects is shown in Fig. 8 [11]. Mean intrapair differences in alcohol use are clearly smaller among adult sisters than among adult brothers, and, for men, intrapair differences are age-dependent: young and old men have smaller differences than middle-aged men. Therefore, age effects might or might not be seen depending on the age-range and gender chosen for study.

**VOLUNTEER TWIN SAMPLES VERSUS POPULATION-BASED REGISTRIES**

Thus, population-based large twin registries are a valuable resource, but are they fundamentally different from volunteer samples: do volunteer and population-based samples give similar or different results? Do all the strengths of the Nordic registries then imply any differences in the inferences to be drawn from twin studies. In specific cases, there are differences. As one illustration, Table 2 compares the degree of intrapair resemblance for Neuroticism-scale scores in five studies of MZ twins reared apart (MZA). There are striking differences across samples.

The three volunteer samples [22,30,32] report much greater resemblance for MZAs and yield much higher heritability estimates than do the two samples [18,23] ascertained
Table 2 - Volunteer and population-based studies reporting intraclass correlations for Neuroticism among MZ pairs reared apart (MZA)

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of MZAs</th>
<th>rMZA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newman et al [22]</td>
<td>USA</td>
<td>1937</td>
</tr>
<tr>
<td>Shields [30]</td>
<td>UK</td>
<td>1962</td>
</tr>
<tr>
<td>Tellegen et al [32]</td>
<td>USA</td>
<td>1988</td>
</tr>
<tr>
<td>Langinvainio et al [18]</td>
<td>Finland</td>
<td>1984</td>
</tr>
<tr>
<td>Pedersen et al [23]</td>
<td>Sweden</td>
<td>1988</td>
</tr>
</tbody>
</table>

from Nordic population registries. And, while the volunteer samples give no hint of shared environmental effects, the Nordic results suggest otherwise; eg, Swedish MZ twins, who were reared together, but matched on age, gender, and county of birth to the Swedish MZAs in Table 2, were more alike (r = 0.41, se = 0.074) for N-scale scores. Differences in MZA correlations found in volunteer samples and the Nordic registries may reflect scale differences in assessment of Neuroticism, or differences in ascertainment of the separated twins. Reporting results for the Swedish MZAs, the investigators concluded that volunteer samples which rely on “response to media appeals” or “identification by third parties” may inflate MZA correlations [23, p. 955]. We agree.

We have previously shown that intrapair social contact frequency is a significant modulator of pairwise similarity for personality traits, such as neuroticism, and alcohol use [13,15,25,26]. Concordance for place of residence also affects disease concordance rates [24]. Thus, differences in place of residence such as living in different provinces also reflect differences in the amount of shared environment within twin pairs. It is probable that even among pairs that do not live together, cotwins that live in different provinces see and contact each other less often than those that live in the same province. Romanov et al [24] demonstrate yet another trait, in this case hospitalization for psychosis, for which pairwise similarity is modulated by shared environment. The causal relationship between shared environment and pairwise similarity could not be assessed in that analysis, for it was not known which event preceded the other, migration or hospitalization.

STUDY DESIGNS

There are many study designs that are usable by the Nordic registries. Classical comparisons of MZ and DZ similarity for the estimation of heritability and shared environment can currently be done by sophisticated modelling approaches [8], that can be extended to many different kinds of situations. Käre Berg and his coworkers [1] have advanced the concept of “variability genes” (as opposed to “level genes”) with respect to variability of lipids and other risk factors. They postulate that a person’s total genetic risk for coronary heart disease may depend on his or her combination of “level genes”
and "variability genes". To detect variability genes, intrapair variability of MZ pairs with different genotypes at a single locus is assessed. They suggest that this method of analysis may be the best available for the study of gene-environment interaction [1]. With more specific genetic information becoming available and a changing environment, the study of gene-environment interaction becomes more important.

The Nordic registries are also of special value for studying twin pairs discordant for exposure [5,9,14] or disease [33]. Such twin "case-control" studies are very powerful designs for demonstrating exposures that have occurred a long time ago, or for detecting effects confounded by genetic influences. The Nordic registries are a unique resource that increases in value each year. It is in the interest of the research community to maintain and use these valuable data bases.

Acknowledgments. This paper was made possible by grants from the Academy of Finland, and the US National Institutes of Health (AA-06232 and AA-07611).

REFERENCES


Correspondence. Dr. J. Kaprio, Department of Public Health, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland.