

The Finnish Twin Cohort Study: An Update

Jaakko Kaprio^{1,2,3}

¹*Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland*

²*Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland*

³*Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland*

In 2002 and 2006, review papers have described the Finnish Twin Cohort and studies conducted on these population-based, longitudinal data sets with extensive follow-up data. Three cohorts have been established: the older twin cohort in the 1970s, and the Finntwin12 and Finntwin16 studies initiated in the 1990s. The present review provides an update on the latest data collections conducted since the previous review. These cover the fourth waves of data collection in the older cohort (twins born before 1958) and Finntwin12 (twins born 1983–1987). The fifth wave of data collection in Finntwin16 (twins born 1975–1979) also included assessments of their spouses/partners. An analysis of mortality in the older cohort from 1975 to 2009 indicates that the mortality of adult twins (as individuals) does not differ from the population at large. Based on the cohorts, many sub-studies with more detailed phenotyping and collection of omics data have been conducted or are in progress. We also contribute to numerous national and international collaborations.

■ **Keywords:** twins, families, cohort studies, longitudinal, database

The research conducted based on the Finnish Twin Cohort Study has been extensively reviewed in three articles in this journal (Kaprio, 2006; Kaprio & Koskenvuo, 2002; Kaprio et al., 2002). The purpose of this review is to provide an update on the latest data collections and record linkage studies of the cohorts that form the study that have taken place since the last review in 2006. The value of twin studies has been regularly reviewed, most recently in the prestigious *Nature Reviews Genetics* journal (van Dongen et al., 2012). The review indicates the multiple applications of genetically informative data from large studies of well-characterized (phenotypically and omically) twins and twin families.

There are three main data sets consisting first of the older twin cohort (Kaprio & Koskenvuo, 2002), which has twins born before 1958 and with follow-up starting in 1975. Three waves of questionnaires have been sent to unselected members of the cohort in 1975, 1981, and 1990, respectively. The Finntwin16 study of twins born 1975–1979 started in 1991, with baseline assessments of twins and their parents when the twins were aged 16. Follow-up questionnaires of the twins took place when the twins were aged 17 years, 18.5 years, and as young adults (age 24 on average; Kaprio, 2006). In the Finntwin12 study, baseline assessments took place at age 11–12, with follow-up at ages 14 and 17.5 (Kaprio, 2006; Kaprio et al., 2002). See www.twinstudy.helsinki.fi for details and pdfs of most questionnaires (in Finnish, Swedish, and generally in English too). For further details

you may also e-mail me at Jaakko.kaprio@helsinki.fi to ask about data access. We are open to collaborations within the limits set by Finnish legislation and ethics.

Like many other cohorts, we have numerous national and international collaborations, with hundreds of co-authors from dozens of institutions worldwide. Within Europe we have or are taking part in consortia such as NorTwinCan, GenomEUtwin, Eurodisc, Eurohead, Euroclot, Diogenes, I.Family, ENGAGE, Genodisc, Myoage. Internationally, we have contributed to numerous genome-wide association meta-analyses and mega-analyses. To further the collaboration of twin studies, I have been promoting the International Network of Twin Registries, which operates under the auspices of international society for twin studies (ISTS), the International Society (<http://www.ists.qimr.edu.au/>).

Old Twin Cohort

New Data Collection

A fourth wave was initiated in 2011 with a pilot study to test the questionnaire. Since the 1990 questionnaire, many

RECEIVED 22 September 2012; ACCEPTED 14 October 2012.

ADDRESS FOR CORRESPONDENCE: Jaakko Kaprio, Professor of Genetic Epidemiology, Department of Public Health, PO Box 41 (Mannerheimintie 172), University of Helsinki, FIN-00014 Helsinki, Finland. E-mail: jaakko.kaprio@helsinki.fi

twins have been asked to participate in selected sub-studies that have focused on specific study questions and traits, but no unselected questionnaire survey has been conducted. While many follow-up outcomes can be assessed by medical register linkage studies in Finland, many health-related factors need to be assessed by interview or questionnaires given to the participants. To test the acceptability of what turned out to be an extensive questionnaire, we sent 254 pairs with both members alive and 299 twins from pairs where only one twin was available (due to death, migration, or lack of address of the other twin) a questionnaire in February–March 2011. The response rate was 70% for full pairs, but only 55% for ‘single’ twins. With approval from the ethical committee, we revised the invitation cover letter to the ‘single’ twins to better motivate their participation in the 35-year follow-up.

From October 2011 to June 2012, we approached all available twins in the same-sexed twin cohort irrespective of their zygosity or earlier response status. Overall, we sent the questionnaire, either in Finnish or Swedish (Finland being a bilingual country) to 11,738 twins born 1945–1957, originally identified in 1974 when the cohort was established. Their vital status was updated in 2011 from the Finnish national population register and all twins known to be alive with an address in Finland formed the target population for this survey. We obtained 8,501 responses, a response rate of 72%. Data entry and cleaning are in progress, so the numbers may still change somewhat. The questionnaire was mailed, rather than administered by Internet, given the age of the subjects and the fact that the three prior questionnaires had also been paper ones. Therefore, a change of mode of administration was not warranted in order to retain comparability across waves.

The content of the questionnaires was a balance between retention of items used in the earlier questionnaires in order to get 2, 3, or 4 measures over time and a need to assess age-specific outcomes of interest, such as intentions to retire. Despite the development of better measures for some topics, we retained the original questions (such as used in 1975 and 1981) wherever possible. A total of 241 items were included in the questionnaires, covering domains such as twin pair relationships, general health and risk factor measurements (blood pressure, diabetes, cholesterol), standard symptoms (chronic bronchitis, dyspnea), musculoskeletal pain, bruxism, history of common diseases, and use of various medications including hormone replacement therapy for women. We asked six questions on sleep and seven items on work and retirement. A detailed smoking history was taken again, supplemented by nicotine dependence scales (Fagerström Test for Nicotine Dependence and the Wisconsin Inventory of Smoking Dependence Motives) and exposure to second-hand smoke over their lifetime. Quantity, frequency, and problem measures of alcohol use and abuse were asked, including a screen for alcohol problems. Coffee and tea use were queried. Physical activity and sedentariness were asked

using 10 questions. Life events and traumatic experiences in childhood and adulthood were assessed. Close social relationships and social support were asked using seven items. Depression was assessed using the CES-D scale (The Center for Epidemiologic Studies Depression Scale) and the short extraversion and neuroticism scales from 1975 and 1981 were repeated. Finally, a history of weight, weight changes, and dieting were included. The invitation letter included a tape measure, and the respondents were asked to record their waist circumference; preliminary data indicates that 98% of subjects reported this.

Register Linkage and Follow-Up

Register linkages to various medical registers have also taken place. Cancer incidence to the end of 2010 is currently updated, and we are participating in NorTwinCan, which is the Nordic twin cancer study, a 4-country study with almost 40,000 incident cancers in 300,000 twins from the Nordic twin cohorts. First results were presented at the Twin Congress in Florence, but no papers have yet been published.

Disability pension data linkage has led to publications on risk factors, genetic and environmental, for becoming incapacitated to work. These include an estimate of the genetic contribution to disability pension risk, which was larger in younger age groups (Harkonmäki et al., 2008). In addition, the disability pension data has been the outcome for many analyses linking our early questionnaire data to the long-term risk of reduced work capacity. The analyses have been epidemiological, utilizing the large sample size and long follow-up period, as well as using the discordant pair design to test causality of the epidemiological association (Ropponen et al., 2012). Vital status and cause of death are regularly updated. Genome-wide genetic variation data using the Illumina 670k array is available for over 1,500 cohort members, mainly from the Nicotine Addiction Genetics sub-sample of heavy smokers (Broms et al., 2012). Other currently active sub-sample studies of interest include the Finnish Twin Study of Ageing (<https://www.jyu.fi/sport/laitokset/tutkimusyksikot/sgt/en/Research/fitsa>) of elderly female twins. The TwinSpine study has assessed disc degeneration and back disorders among 600 male twins using magnetic resonance imaging (MRI) assessments across three waves in the past 15 years. This study has won numerous awards, including the American Academy of Orthopaedic Surgeons’ 2008 Kappa Delta Elizabeth Winston Lanier Award (Battié et al., 2009).

Mortality of Twins Compared to the Population

We examined the mortality of the birth cohorts born 1880–1957 in the like-sexed twin pairs of the older twin cohort, in which both twins were alive and resident in Finland in 1975. The beginning of follow-up was May 1, 1976, and end of follow-up is death, emigration, or alive at end of follow-up in December 31, 2009. National mortality rate by 5-year

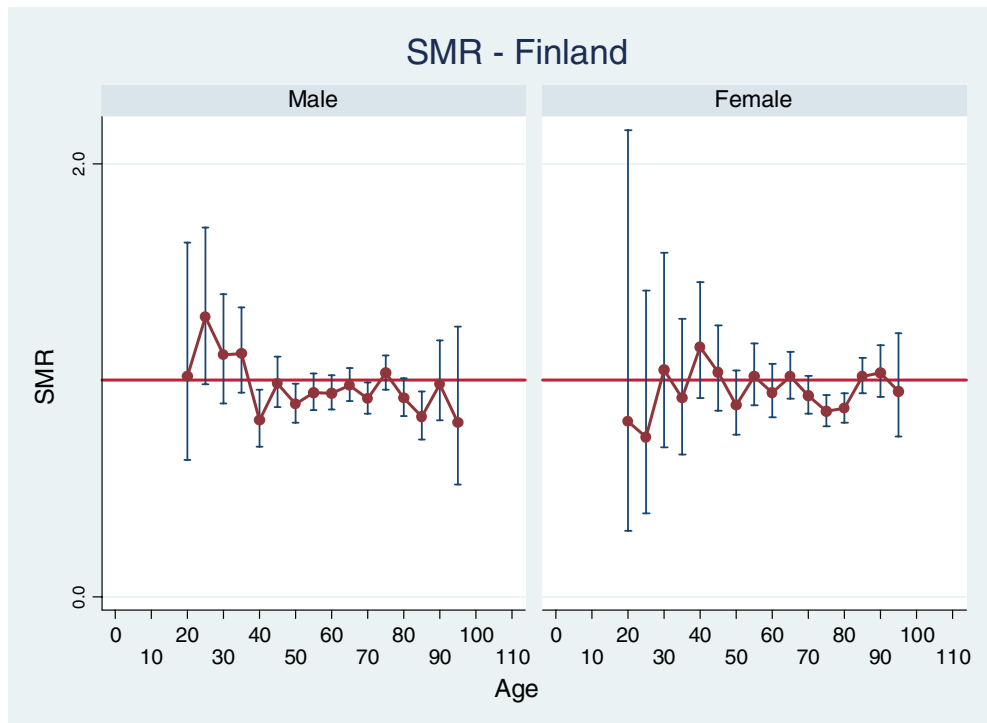


FIGURE 1

(Colour online) Standardized mortality ratios (SMR) of observed to expected mortality among twin individuals in the older twin cohort (with 95% confidence intervals) by sex and 5-year age group (the red line). Expected numbers are based on calendar-specific mortality rates for the Finnish population in the same sex-age groups. The horizontal line represents an SMR of 1 (or 100%), that is, the population rates.

age groups and 5-year calendar periods were obtained from the Human Mortality Database (www.mortality.org). Standardized mortality ratios (SMR) were computed as the ratio of observed to expected (based on person years and population rates). The overall SMR for male twins was 94.95 (95% confidence interval [CI] 92.3 to 97.7, based on 4,752 deaths) compared to 100 for the population, while for female twins it was 94.5 (91.5 to 97.6, 3,714 deaths). Monozygotic (MZ) twins had the lowest SMRs (87.2 for men, 89.5 for women), dizygotic (DZ) twins somewhat higher (94.0 for men and 94.3 for women), while the small number of twins with unknown zygosity had higher than expected mortality (SMR for men was 113, based on 810 deaths and for women the SMR of 108 was based on 507 deaths). The unknown zygosity group includes those who had not replied to our questionnaires with zygosity items, had left the zygosity items unanswered, or whose responses were conflicting such that the zygosity determination algorithm left them unclassified as to being MZ or DZ.

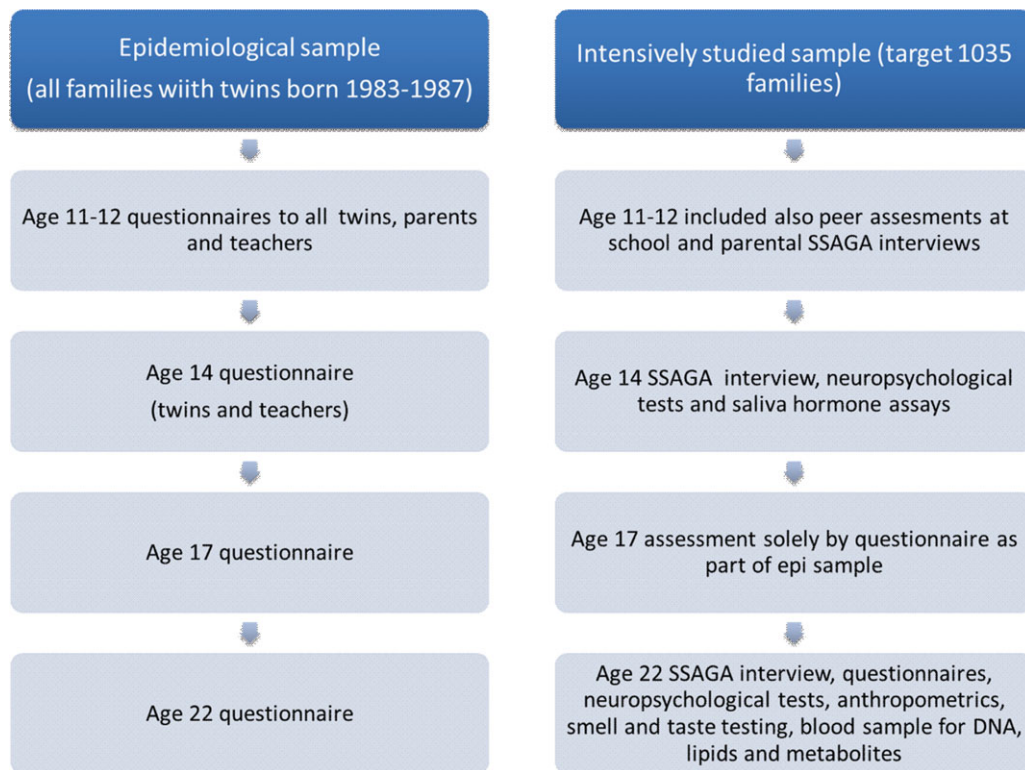
The smaller than expected mortality was primarily due to the lesser mortality in the older birth cohorts. Among twins born in the 1940s and 1950s, the SMRs were very close to one (SMR = 99.9 for men and women combined). Because both twins in a pair had to be alive to be included in the cohort when it was compiled, the oldest ones in particular were probably healthier than average. This ‘healthy pair’ effect

is equivalent to the health worker effect in occupational health studies. The SMRs by age, averaged over the 33 years of follow-up, are shown in Figure 1 separately for men and women. In most age groups the mortality of twins is very similar to that of the population, with some mortality deficit in older age groups. Thus, we find no evidence for excess mortality in twins compared to the population at large.

Finntwin16

The first four waves of data collection have been described earlier (Kaprio, 2006). Based on these data, several sub-studies of selected pairs have been conducted, but this review will not cover those in detail as they are or will be the subject of separate reviews. See Naukkarinen et al. (2012) for a review of the study of weight-discordant MZ pairs. There is an ongoing study of physical activity discordant pairs led by Professor Urho Kujala (University of Jyväskylä), while results from our study of alcohol dependence discordant and concordant pairs have been reported (Koskinen et al., 2011; Latvala et al., 2011).

The fifth wave of data collection of the Finntwin16 cohort was done in May 2010 (a pilot study), October 2010–November 2011 for Finnish-speaking subjects, and in the spring of 2012 for Swedish-speaking subjects. The cohort has over 3,000 twin pairs, and the invitation was sent to all twins living in Finland irrespective of earlier participation.

**FIGURE 2**

(Colour online) Two schematic charts of data collection in the Finntwin12 study (Epidemiological sample and intensive study sample).

Note: In the epidemiological sample 5 birth cohorts of twins born 1983–1987 were assessed at 4 waves of data collection. An intensively studied subset was subject to additional data collections in waves 1, 2, and 4.

Twins who had indicated that they no longer wished to participate were naturally not contacted. Of the 6,132 twins that we contacted, 4,414 participated, a response rate of 72%.

The Internet survey permitted us to tailor questions depending on answers to certain key questions. Thus, questions on their children were asked only for twins who indicated that they had children, items on nicotine dependence were asked only of smokers, and abstainers were not asked about detailed alcohol items. As for the older cohort, the balance of questions was between retaining items from earlier questionnaires to permit longitudinal analyses and including items of relevance to these twins now in their mid-30s. We asked them about their twin relationships, study and work, general health, symptoms, selected specific diseases, and reading disorders. Seven items on weight and weight changes were included. Ten physical activity items were asked. Then they were asked about their partners, whether they had children, and some details about each child and intentions to have more children. Alcohol use and alcohol-related problems, diet, and tobacco use were assessed. A short personality scale was included before asking about satisfaction in various life domains. Waist circumference was recorded using the tape measure supplied in the invitation letter.

After completion of the survey items, some respondents were asked to administer a net-based battery of neuropsychological tests on spatial ability, memory, and reading speed.

Responses to these tests were obtained from 1,391 twins. Because many of the twins also had partners or spouses, we asked the twins to invite their partner to participate. This was done using a fully anonymous code, supplied in a closed envelope to the twin to be passed on to the partner. Partner responses were obtained from 1,948 persons, with spouse pair responses (i.e., both twin and partner replied) from 1,826 dyads. Some of the partners were also asked to take the neuropsychological tests, and 710 did so successfully, with 677 twin-partner dyads.

Finntwin12

We have enrolled 5,600 twins and their families ascertained from nation-wide birth cohorts born 1983–1987. In the first phase, the twins, their parents, and teachers filled out questionnaires when the twins were 11–12 years old, with two follow-ups at ages 14 and 17½ years (Kaprio, 2006; Kaprio et al., 2002). Parents and twins completed questionnaire assessments of their use of alcohol, smoking, lifestyle, and health status; parents and teachers assessed aspects of the twins' behavior. We have achieved response rates of 85–90% across all aspects of the study. Parents also reported on pregnancy (such as infertility treatments) and pregnancy events (such as elevated blood pressure), use of alcohol and

tobacco during pregnancy, and early childhood conditions and child development. From the five consecutive birth cohorts a subset of 1,035 families of twins was formed for the intensive study. All intensive subset parents were interviewed ($N = 1,860$) with a semi-structured psychiatric assessment interview, and at age 14, their twins have been interviewed ($N = 1,852$) with the adolescent version of the same instrument to assess substance use, dependence, and psychiatric morbidity in adolescence. Teachers, parents, and peers assessed the externalizing (inattention, impulsivity, and aggressiveness), internalizing (depression and anxiety), and pro-social behaviors of the twins. Figure 2 summarizes the data collection procedures in Finntwin12.

Wave 4 data collection in 2006–2009 conducted structured psychiatric interviews at an average age of 21.9 (SD 0.8, range 21–26) years, with neuropsychological assessments and collection of DNA and serum samples (for basic biochemistry and metabolomics) from twins of 900+ families most informative for substance abuse/dependence with existing psychiatric interview data (at age 14 for the twins). During this 1-day, in-person assessment in Helsinki, we collected information about substance use, abuse, and dependence (smoking, alcohol, cannabis, and other drug use) based on questionnaires and interviews for DSM-IV diagnoses (alcohol, illicit drugs, depression, eating disorders, suicidal behavior, and anti-social behavior), but also about lifestyle (physical exercise, food habits), chemosensory preferences, mood and regulation of emotions (depression — General Behavior Inventory) and mood (General Health Questionnaire-12), personality (NEO-Five Factor Inventory), schizotypy (Schizotypal Personality Questionnaire), sense of coherence, general health (illnesses, body composition, metabolism, experienced health), cognitive functions (general cognitive capacity, spatial ability, memory, executive functioning, attentiveness, perceptual speed, motor dexterity, and social cognition as measured with ability to recognize emotional facial expressions), and work. The assessments were completed on 1,347 twins (73% of target sample), of which 812 were seen in person (with neuropsychological testing) and the rest by telephone psychiatric interviews, mailed questionnaires, and DNA collection ($n = 1,295$) from blood and/or saliva samples. We have also sent a questionnaire to those not part of the intensive cohort, and this data collection was just completed.

We now have genome-wide genotyping association study data from the wave 4 twins genotyped on the Illumina 670k custom chip at the Wellcome Trust Sanger Institute. In addition, 200+ metabolites, primarily of lipids and their fractions, have been determined from all twins, who provided blood samples ($n = 780$; Kettunen et al., 2012) and cotinine, 3-hydroxycotinine and the nicotine metabolite ratio have been determined for smokers and their co-twins by Professor Rachel Tyndale's lab at the University of Toronto for all occasional and daily smokers. We are currently completing the first epigenetic analyses using targeted gene-

specific methylation assays, as well as genome-wide approaches (such as the Infinium 450k array).

Conclusions

This brief review has focused on data collections done on essentially unselected twin pairs. Over the years, many twins have been invited to participate in selected sub-studies for more intense assessments, but these are not described in detail here. Likewise, there have been more than 300 peer-reviewed publications since 2006. A dozen PhDs have been completed since 2006 under my supervision, and many more as part of other research projects, to which the twin data has contributed at least in part. In 2013, we start a Marie Curie Initial Training Network in Epigenetics (EPI-TRAIN), which will offer training to three PhD students based in Helsinki, and many more through our other European partners. The numerous national and international scientific and researcher training collaborations indicate the continued value of large follow-up studies of twins and their family members. These studies also highlight the representativeness of adult twins, as shown here in our mortality analysis. The newest data collections will increase the informativeness and uniqueness of the cohort data, permitting more detailed analyses of factors predicting changes and stability of health-related behaviors, health status, and other outcomes of interest over decades of follow-up.

Acknowledgments

Studies within the Finnish Twin Cohort have been supported, among others, by the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 213506 and 129680), the Academy of Finland (grants 100499, 205585, 118555, and 141054), Global Research Awards for Nicotine Dependence (GRAND), European Network for Genetic and Genomic Epidemiology (ENGAGE), FP7-HEALTH-F4-2007 (grant 201413), NIH (grants DA12854, AA12502, AA00145, AA09203, AA15416), and the Sanger Institute. The support of the Nordic Cancer Union, The Finnish Cancer Society, the Sigrid Juselius Foundation, and The Jenny and Antti Wihuri Foundation is also gratefully acknowledged. I would like to thank Axel Skytte (Danish Twin Register) and Eero Pukkala (Finnish Cancer Register) for computing the SMR results. I would like to warmly thank the dedicated staff and researchers in our group, the numerous collaborators in Finland and internationally, and finally, a great thanks is due to all the twins and their family members who have participated in these studies over the years.

References

- Battié, M. C., Videman, T., Kaprio, J., Gibbons, L. E., Gill, K., Manninen, H., . . . Peltonen, L. (2009). The Twin Spine Study: Contributions to a changing view of disc degeneration. *Spine Journal*, 9, 47–59.

- Broms, U., Wedenoja, J., Largeau, M. R., Korhonen, T., Pitkäniemi, J., Keskitalo-Vuokko, K., . . . Loukola, A. (2012). Analysis of detailed phenotype profiles reveals CHRNA5-CHRNA3-CHRNA4 gene cluster association with several nicotine dependence traits. *Nicotine & Tobacco Research*, *14*, 720–733.
- Harkonmäki, K., Silventoinen, K., Levälähti, E., Pitkäniemi, J., Huunan-Seppälä, A., Klaukka, T., . . . Kaprio, J. (2008). The genetic liability to disability retirement: A 30-year follow-up study of 24,000 Finnish twins. *PLoS One*, *3*, e3402.
- Kaprio, J. (2006). Twin studies in Finland 2006. *Twin Research & Human Genetics*, *9*, 772–777.
- Kaprio, J., & Koskenvuo, M. (2002). Genetic and environmental factors in complex diseases: The older Finnish Twin Cohort. *Twin Research*, *5*, 358–365.
- Kaprio, J., Pulkkinen, L., & Rose, R. J. (2002). Genetic and environmental factors in health-related behaviors: Studies on Finnish twins and twin families. *Twin Research*, *5*, 66–71.
- Kettunen, J., Tukiainen, T., Sarin, A.-P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L. P., . . . Ripatti, S. (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genetics*, *44*, 269–276.
- Koskinen, S. M., Ahveninen, J., Kujala, T., Kaprio, J., O'Donnell, B. F., Osipova, D., . . . Rose, R. J. (2011). A longitudinal twin study of effects of adolescent alcohol abuse on the neurophysiology of attention and orienting. *Alcoholism, Clinical and Experimental Research*, *35*, 1339–1350.
- Latvala, A., Tuulio-Henriksson, A., Dick, D. M., Vuoksimaa, E., Viken, R. J., Suvisaari, J., . . . Rose, R. J. (2011). Genetic origins of the association between verbal ability and alcohol dependence symptoms in young adulthood. *Psychological Medicine*, *41*, 641–651.
- Naukkarinen, J., Rissanen, A., Kaprio, J., & Pietiläinen, K. H. (2012). Causes and consequences of obesity: The contribution of recent twin studies. *International Journal of Obesity (London)*, *36*, 1017–1024.
- Ropponen, A., Svedberg, P., Huunan-Seppälä, A., Koskenvuo, K., Koskenvuo, M., Alexanderson, K., . . . Kaprio, J. (2012). Personality traits and life dissatisfaction as risk factors for disability pension due to low back diagnoses: A 30-year longitudinal cohort study of Finnish twins. *Journal of Psychosomatic Research*, *73*, 289–294.
- van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, *13*, 640–653.