



Karolinska
Institutet

Karolinska Institutet

<http://openarchive.ki.se>

This is a Peer Reviewed Accepted version of the following article, accepted for publication in *Twin Research and Human Genetics*.

2013-12-20

The Swedish Twin Registry : establishment of a biobank and other recent developments

Magnusson, Patrik K E; Almqvist, Catarina; Rahman, Iffat; Ganna, Andrea; Viktorin, Alexander; Walum, Hasse; Halldner, Linda; Lundström, Sebastian; Ullén, Fredrik; Långström, Niklas; Larsson, Henrik; Nyman, Anastasia; Hellner Gumpert, Clara; Råstam, Maria; Anckarsäter, Henrik; Cnattingius, Sven; Johannesson, Magnus; Ingelsson, Erik; Klareskog, Lars; de Faire, Ulf; Pedersen, Nancy L; Lichtenstein, Paul

Twin Res Hum Genet. 2013 Feb;16(1):317-29.

<http://doi.org/10.1017/thg.2012.104>

<http://hdl.handle.net/10616/41868>

If not otherwise stated by the Publisher's Terms and conditions, the manuscript is deposited under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

The Swedish Twin Registry – establishment of a biobank and other recent developments

Patrik K.E. Magnusson¹, Catarina Almqvist^{1,2}, Iffat Rahman¹, Andrea Ganna¹, Alexander Viktorin¹, Hasse Walum¹, Linda Halldner^{1,3}, Sebastian Lundström^{4,5,6}, Fredrik Ullén⁷, Niklas Långström^{1,5}, Henrik Larsson¹, Anastasia Nyman¹, Clara Hellner Gumpert⁸, Maria Råstam⁹, Henrik Anckarsäter⁴, Sven Cnattingius¹⁰, Magnus Johannesson¹¹, Erik Ingelsson¹, Lars Klareskog¹², Ulf de Faire¹³, Nancy L. Pedersen¹, Paul Lichtenstein¹

¹ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

² Department of Women's and Children's Health and Astrid Lindgren Children's Hospital, Stockholm, Sweden

³ Karolinska Institutet, Center of Neurodevelopmental Disorders (KIND), Stockholm, Sweden

⁴ CELAM (Center for Ethics, Law and Mental Health), University of Gothenburg, Sweden

⁵ Swedish Prison and Probation Service, R&D Unit, Norrköping, Sweden

⁶ Gillberg Neuropsychiatry Center, Institution of Neuroscience and Physiology, University of Gothenburg, Sweden

⁷ Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

⁸ Centre for Psychiatry Research & Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

⁹ Department of Clinical Sciences, Lund University, Lund, Sweden

¹⁰ Unit of Clinical Epidemiology, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

¹¹ Department of Economics, Stockholm School of Economics, Stockholm, Sweden

¹² Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Stockholm, Sweden

¹³ Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Corresponding author: Patrik K.E. Magnusson, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, 17177 Stockholm, Sweden

Email: patrik.magnusson@ki.se

Running title: The Swedish Twin Registry – new developments

Abstract

The Swedish Twin Registry (STR) as of today contains more than 194,000 twins and more than 75,000 pairs have zygosity determined by an intra-pair similarity algorithm, DNA or by being of opposite sex. Of these, approximately 20,000, 25,000 and 30,000 pairs are monozygotic, same-sex dizygotic and opposite-sex dizygotic pairs, respectively. Since the establishment in the late 1950s, the STR has been an important epidemiological resource for the study of genetic and environmental influences on a multitude of traits, behaviors and diseases. Following large investments in the collection of biological specimens in the past 10 years we have now established a Swedish twin biobank with DNA from 45,000 twins and blood serum from 15,000 twins, which effectively has transformed the registry into a powerful resource also for molecular studies. We here describe the main projects within which the new collections of both biological samples as well as phenotypic measures have been collected. Coverage by year of birth, zygosity determination, ethnic heterogeneity, and influences of in vitro fertilization are also described.

Keywords: Population-based, Twin, Registry, Biobank, Sweden,

Ascertainment

The Swedish Twin Registry (STR) obtains information on twin births occurring in Sweden from the National Board of Health and Welfare. STR staff requests the data in batches of consecutive birth-years so that twin identities are known before they turn 9 years of age. Triplets and higher order multiples are not included. Since 2004, we have systematically approached the parents of all identified 9-year old twins with an invitation to participate in the Child and Adolescent Twin Study in Sweden (CATSS) (Anckarsater et al., 2011). Before 2004, the twins were contacted at various ages depending on birth-year cohorts. For same-sexed twins born before 1926 the first contact was made in the beginning of 1960s and for same-sexed twins born 1926-1958 it occurred in the beginning of the 1970s. Opposite-sexed twins born 1906-1925 were contacted for the first time in the GENDER study (Gold et al., 2002) while opposite-sexed twins born 1926-1958 were contacted for the first time during the Screening Across the Lifespan Twin study (SALT) interview 1998-2002 targeting all twins-pairs born 1958 or earlier (Lichtenstein et al., 2002; Pedersen et al., 2002). All twins born 1959-1985 were contacted for the first time in 2005-2006 in the Study of Twin Adults-Genes and Environment (STAGE) study (Furberg et al., 2008), which was the first web-based questionnaire study conducted by the STR. A recruitment scheme similar to the one later used for CATSS was adopted for twins born in 1985 - June 1986 in the project denoted The Swedish Twin study of Child and Adolescent Development (TCHAD) (Lichtenstein et al., 2007). After TCHAD there is a gap in the recruitment contacts for the twins born from July 1986 to 1992, before the CATSS study started systematic recruitment of 9-year old twins. This gap will now be filled in a study denoted Young Adult Twin Study in Sweden (YATSS) initiated in the autumn of 2012.

Phenotypic information on study participants comes from different sources. The STR is updated monthly with regard to current address and vital status. Importantly, medical information comes from regular updates to nationwide total population health registers including the Patient Register, with data from the inpatient and outpatient (from non-private specialized care) settings, the Medical Birth, Prescribed Drug, Cancer, and the Causes of Death Registers. Beyond the recruitment contacts mentioned above, the twins have also been contacted in several additional waves with requests to participate in questionnaire/interview studies covering a broad selection of exposures, behavior and medical information. The history and structure of the STR has been described previously; in 2002 (Lichtenstein, et al., 2002; Pedersen, et al., 2002) and 2006 (Lichtenstein et al., 2006). In Table 1 all STR studies targeting whole Swedish twin cohorts are listed.

Table 1 about here

STR is still one of the world's largest twin resources, currently containing 194,842 twins born between 1886 and 2008. Among these, 75,602 pairs have had their zygosity determined by either an intra-pair similarity algorithm, genotyping or by being of opposite sex. There are currently 19,251 MZ, 25,887 same-sex DZ (SSDZ) and 30,464 opposite-sex DZ (OSDZ) pairs.

Since its establishment in the late 1950s, the STR has been an important epidemiological resource for the study of quantitative genetic aspects of a multitude of traits, phenotypes and disorders. Perhaps the most important development since 2006 is the focus on collecting biological material. These efforts have resulted in a collection of DNA from over 45,000 twins and blood serum from some 15,000 twins, rendering the

STR one of the richest twin resources worldwide also in terms of availability of DNA and serum (van Dongen et al., 2012).

Zygoty determination and validity

For the vast majority of same-sex twins in the STR zygoty has been assigned based on questions about intra-pair physical similarities in childhood; however, DNA-based zygoty testing is used increasingly. The laboratory method of DNA zygoty determination has changed over time from panels of 10 or 13 micro-satellites (Lichtenstein, et al., 2002) to routinely rely on 46 SNP markers. The statistical behavior of the new SNP based method has been thoroughly evaluated under various assumptions on lab-based error rates (Hannelius et al., 2007). Currently, DNA based zygoty is available for 8,788 (13% of all) same-sex twin pairs in the STR. The accuracy of using answers to self-reported similarity questions for determining zygoty has repeatedly been found to be around 98% for the investigated cohorts (Lichtenstein, et al., 2002). A recent independent test of the validity of similarity-based zygoty assignment among the adults in the TwinGene study (see below) was obtained when DNA from 1,487 presumed SSDZ pairs was genotyped by Illumina OmniExpress platform. We found that 38 SSDZ pairs shared all alleles (>99.9%) and therefore were incorrectly classified. This gives a DZ to MZ error rate of 2.56%, corresponding to an accuracy of 97.4% (95%CI: 96.6-98.2%). Since we did not send both members of presumed MZ pairs for genotyping, it was not possible to check the MZ to DZ error rate from GWAS data. However, based on other studies (see for example data from CATSS below) we know it is substantially lower in most instances. Thus, we conclude that the

number of zygosity errors detected by the genome-wide genotyping is in agreement with an overall 98% average accuracy previously reported (Lichtenstein, et al., 2002).

For the child cohorts of CATSS, the intra-pair similarity is reported by parents and includes additional items about hair and eye color. In a recent comparison of similarity-based and DNA-based zygosity assignment in 3,422 same sex pairs, overall similarity-based zygosity was confirmed by DNA in 96.6% (95%CI, 96.0-97.2%). More specifically, 1,664/1,691 (98.4%, 95%CI: 97.8-99.0) MZ and 1,642/1,731 (94.9%, 95%CI: 93.8-95.9%) DZ pairs were correctly classified by the similarity algorithm. The somewhat higher error rate for twin children compared to adults is probably primarily due to bias arising from parental answers about their children (where they usually overemphasize their differences) but may also reflect an increase in discordances with age stemming from genetic determinants of physical appearance not expressed until adulthood.

Coverage of Swedish twin births in the STR

Twins may be missing from the STR for several reasons. For example, they may have been lost in the early compilations from church books, been excluded due to criteria at the time of compilation (for example both have to be alive) or they may have declined participation. One way to investigate the completeness of the STR is to compare the STR with the Multi-Generation Register (MGR) (Statistics-Sweden, 2011). The MGR is a Swedish population registry containing parental information for individuals born 1932 or later, and registered as living in Sweden at any time since 1961. Besides data on parent identity, the information in MGR is limited and no information on zygosity is available. The MGR was constructed from Taxation Office censuses complemented with

information from other sources to increase coverage from earlier years where many individuals are deceased. For parental links to be present, the parents must have been alive and resident in Sweden 1947 when the personal identification number system was introduced (Statistics-Sweden, 2011). We searched the MGR for all multiple-births (for which the majority are twin births). A comparison between registered twin pairs in STR and multiple births in the MGR across birth-years is shown in Figure 1.

Figure 1 about here.

Before 1932 no comparisons with MGR can be made but STR shows a linear increase in the number of twins for each birth year. This increase over time is primarily due to that in the oldest cohorts STR has information only on pairs in which both twins responded to the questionnaire in 1961. As expected, there were more deaths in the older cohorts. After 1932 there are some initial years in which the STR has more twins than the MGR, due to incomplete coverage and censoring from migration or death of the twins or their parents in the MGR. The coverage of index persons with link to mothers in MGR reaches 95% after 10 years (Statistics-Sweden, 2011). The overall correspondence in the total number of twins in the STR and MGR is very good from 1942 and onwards. However, the temporary increase in number of twins missing in STR in the cohorts born 1959-1966 reflects differences in inclusion of deceased individuals in the MGR versus the STR (Medlund et al., 1976). This is also reflected in the absolute number of pairs with known zygosity, which is approximately stable for the entire cohort born 1959-1985 (i.e., the STAGE-responders). The number of twins in the STR compared to the number of twins with known zygosity primarily reflects the response rates to our data collections. The proportions of twins having known zygosity in the STR

are shown in Figure 2. It clearly shows the jumps in zygosity coverage over birth years reflecting the general secular decline in response rates over the different data collections. As already mentioned, for cohorts born 1925 and earlier, we only have information on twins where both responded and thus, almost 100% of the twins have been assigned zygosity. The cohort born 1926-1958 responded to a questionnaire in 1973 and for these birth cohorts we have zygosity info on approximately 90%. In contrast, there are smaller proportions with zygosity for the cohorts 1959-1985, which were compiled in 2005-2006 and suffer from lower response rates. For birth years 1987-1992 and 2005 and onwards, the STR practically only has OSDZ registered. The increase in response rate in the 1986 cohort corresponds to the TCHAD study, which was the first cohort to have zygosity assigned based on parents responses on their children. A similar high proportion of known zygosity can be seen in the cohorts born after 1992 due to the parental responses in the CATSS study. Nevertheless, since 2004 (when the CATSS study was initiated) we notice a constant drop in known zygosity, reflecting a worrying consistent decline in response rates for each year, also seen in other research projects in Sweden. The cohorts born 1986-1992 are about to be contacted in the forthcoming study YATSS. For researchers using data from the STR, knowledge about coverage and zygosity is essential.

Figure 2 about here.

Proportions of twins with one or more parents born abroad

Practically all (>99.9%) of the twins in STR were born in Sweden, but there are twins with one or both parents born abroad. Information on parental birth country was available for 105,678 twins born 1932-2008. The overall percentage of having at least

one parent born abroad is 12.2%. The corresponding figures for having at least one parent born outside Scandinavia is 6.8% and 2.9% for having at least one parent born outside Europe. Figure 3 illustrates the time trends in proportion of twins with one or both parent born outside Sweden, Scandinavia or Europe across birth-year. There has been a steady increase in the proportions of twins with parents born abroad, with less than 3% for years 1932-1940 and more than 20% for years 2001 to 2008. A substantial fraction of this increase can be attributed to immigration from non-European countries in more recent decades; before 1970 there were less than 1% of the twins having non-European ancestry, compared to over 10% after 2000. Thus, the ethnic heterogeneity of the STR depends on birth-cohort with descendants of non-Europeans predominantly among births after 1970. Again, researchers using data from the STR need to be aware of the changing ethnic diversity of the twins, which is especially important in genetic studies, where population stratification may lead to spurious associations (Price et al., 2010).

Figure 3 about here.

Changes in zygosity proportions

Assisted reproduction with in vitro fertilization (IVF) was introduced in Sweden at the beginning of 1980s. Transferring multiple embryos back to the woman has been the common practice in most parts of the world. Such multiple embryo transfer increases the probability of having multiple births, particularly DZ twins. Therefore, the ratio of DZ to MZ twins is likely to have increased since the introduction of IVF. Because systematic changes might introduce bias (obvious examples of potential systematic biases are the socioeconomic status of the parents and number of siblings to the twins)

when the data is used for certain analyses, we wanted to investigate the magnitude of the changes. By utilizing the MGR (with no information on zygosity except that opposite-sex twins are always DZ), we calculated the proportions of opposite-sex twins among all twin births. This provides a way to monitor, in an unbiased manner, changes in the proportion of DZ twins born. Because the ascertainment differs between same- and opposite sex-pairs in the STR, this comparison is less suitable when using STR data for the same purpose. Instead, we calculated the proportions of SSDZ among all same-sex twins with zygosity determined (MZ+SSDZ) in the STR. As evident from Figure 4, in both these comparisons the pattern is similar; first the proportion of DZ twins was rather constant up to 1960 and then decreased to 1980 when it started to rise again. It is likely that the latter increase reflects the introduction of IVF. The causes behind the prior decrease are less obvious but may reflect secular changes in maternal age, parity, smoking or BMI, factors all known to be associated with DZ twinning (Hoekstra et al., 2010; Hoekstra et al., 2008). After a gradual adoption of the single embryo transfer (SET) policy in Sweden (rate of SET was 1.5% in 1991, 30.6% in 2002, 54.3% in 2003 and 67.4% in 2004), a marked decrease of IVF induced multiple births, from 35% in 1991 to less than 5% in 2005 (Karlstrom & Bergh, 2007), has occurred. Effects of this are evident also by the flattened and even declining trend of the proportion of OSDZ in the most recent years (Figure 4).

Figure 4 about here.

Recent large data collection efforts

The breadth of the research activities in the STR is substantial. Below we aim to describe the main data-collection efforts, directed to entire cohorts and with broad

phenotypic scopes. The order of the description is based on birth-year of the target population.

TWINGENE

Between the years 2004 and 2008 we collected blood from 12,614 older twins in the TwinGene project. The primary aim was to systematically enhance the phenotypically very rich SALT study (Lichtenstein, et al., 2002; Pedersen, et al., 2002) (conducted 1998-2002) with biological specimens and thereby transform it into a resource equipped to investigate the molecular influences on a broad spectrum of health related traits and common diseases as well as gene-environment interactions in the etiology of common disorders.

Beginning in 2004, about 200 twin pairs were contacted each month until the data collection was completed in 2008. A total of 22,390 twins were invited. Consent forms and health questionnaires were sent to the subjects along with the invitations. To those who consented, we mailed blood-sampling equipment and asked them to contact their local healthcare facility for sampling and a health checkup. Subjects living in the vicinity of four larger cities were given the option of visiting hospital blood centers in which case the health checkup was omitted.

The study population was restricted to twins participating in SALT. Subjects were excluded from the study if they had previously declined future participation or been enrolled in other recent STR based DNA-sampling projects.

Questionnaire

Participants were asked to fill out a questionnaire about common diseases.

Cardiovascular disease was addressed with questions about the occurrence and year of

onset for angina pectoris; coronary infarct; hypertension; high cholesterol or high triglycerides serum values; claudication (i.e. impairment in walking or pain in legs during walking); venous thrombosis; stroke or blood clot; and TIA (transient ischemic attack). Aortocoronary bypass operations or percutaneous coronary interventions and medications for cardiovascular diseases were also recorded. Questions about diabetes included type of diabetes and diabetes medication, year of onset and whether or not a doctor had given the diagnosis. Questions about pain, migraine or reoccurring headache, knee or hip osteoarthritis and rheumatoid arthritis were also asked together with year of onset. For rheumatoid arthritis, follow up questions about whether or not diagnosis had been established by a medical doctor, if the subject attended regular checkups for the condition and if any medication for arthritis besides pain remedy, were used.

DNA collection and health checkup

Study participants were asked to make an appointment at their local healthcare facility for a morning visit Monday to Thursday and not the day before a national holiday to ensure the sample would reach the Karolinska Institutet (KI) biobank the following morning by overnight mail. The subjects were instructed to fast from 8 pm in the evening the day before the appointed visit. A total of 50 ml of blood was drawn from each subject by venipuncture and tubes with serum and blood for biobanking and separate tubes for clinical chemistry testing were sent to KI. An average of 5 aliquots of 900 µl of serum from each participant was subsequently stored in liquid nitrogen. For back-up purposes, one 7 ml EDTA tube of blood has been stored in -80°C, while a second 7 ml EDTA tube of blood was used for DNA extraction using Puregene extraction

kits (Gentra systems, Minneapolis, USA). Following extraction, the DNA has been stored at -20°C.

Clinical blood chemistry assessments were performed by the Karolinska University Laboratory for the following biomarkers: total cholesterol, triglycerides, HDL cholesterol, LDL cholesterol (by Friedewald formula), C-reactive protein, glucose, apolipoprotein A-I, apolipoprotein B, hemoglobin and hemoglobin A1c. The heritability of these biomarkers in the large and homogenous TwinGene population of elderly Swedish twins has been evaluated (Rahman et al., 2009).

In the simple health checkup, participants were asked to rest for five minutes, then systolic and diastolic blood pressure were taken with the subject sitting upright. After one minute a second blood pressure measurement was performed. The subjects' weight, height, hip and waist circumference were recorded without shoes and in light clothing.

Response rates

Out of the 22,391 contacted twins, 12,614 responded by consenting and donating blood giving an overall individual-wise response rate of 56%. The total number of participating complete pairs was 5,014, giving a pair-wise response rate of 45%.

Response rates and number of pairs and participants by sex and zygosity are specified in Table 2. Age had a pronounced effect on the tendency to participate with a peak in participation for subjects born 1936-1940 and thus likely to consist of recent pensioners. Sex did not affect participation reflected in very similar overall rates among males and females.

Table 2 about here.

Genotyping

Single nucleotide polymorphism- (SNP) based genome-wide genotyping has so far been undertaken in two samples. In 2007, 302 female MZ twin pairs were genotyped using the Illumina 317K SNP chip. In 2011, an additional 9,900 twins (all remaining unique genomes) were genotyped using the Illumina OmniExpress 700K chip, thus constituting the most extensive and expensive genotyping in the STR so far. Research groups use the generated data for genome-wide association studies (GWAS) of quantitative traits and common health-problems while other groups primarily utilize the data as controls in GWAS of rare diseases where cases (non-twins) have been collected separately. The majority of the groups are also participants of consortia to which they contribute data. One important prerequisite for using twins as controls for non-twin cases is that there are no abundant and strong genetic determinants for twinning tagged by the chip used (otherwise such studies would run the risk of confusing twinning and disease variants). The existence of variants predisposing to twinning is a question that deserves a large-scale effort on its own and we are aware of ongoing formation of such collaboration. Meanwhile, and in order to be able to recommend collaborators using the TwinGene genotyping data for control purposes, we performed a preliminary comparison with a non-twin population based Swedish study (a subset of the Swedish controls in the International Schizophrenia Consortium (ISC) (Ripke et al., 2011)) genotyped with the same platform. SNPs passing quality control had very similar allele frequencies in the twins (one randomly selected member per pair, N=6,886) and the non-twins (N=3,729), with no allele frequency difference being larger than 4%. This analysis does not rule out the existence of susceptibility alleles for twinning but it indicates that differences are small and these twins may be used as controls for Swedish case materials.

BIRTH

The pregnancy/birth data available in the Swedish Medical Birth Register (for births in 1973 and later) have been retrospectively extended for older twins in STR by retrieving birth journals from archives for the twins born 1926 to 1972. This data collection is now completed. Birth journals were retrieved from 67,298 (86% of all) twins born 1926-1972. The data generated has been used in several studies on associations between birth-weight and health outcomes in adulthood (Bergvall et al., 2007; Oberg et al., 2011).

SALTY

Screening Across the Lifespan Twin Study: the Younger (SALTY) is a collaborative project initiated in 2007 by researchers in epidemiology, medicine and economics. The data collection consisted of three parts: 1) an extensive self-report paper-questionnaire; 2) saliva collection for DNA extraction; 3) a request to participate in an internet-based investigation that included questionnaires on musical experience, tendency to experience psychological flow and creative achievement as well as tests of cognitive and motor performance.

The target population was the younger part of the SALT cohort born between 1943 and 1958. Reasons for the focus on this particular age group was that we did not contact all the eligible twins in these birth years in the previous TwinGene study due to suboptimal participation rates and lack of funding. These twins belong to the active workforce and we estimated that an important factor for the low participation in TwinGene was limited time for visits to a health care facility during the weekdays to

donate blood. Instead, we hoped to be able to collect DNA from these subjects with the saliva DNA kits that are administered in SALTY.

The first requests for participation in the SALTY study were sent out in early 2009 and the data collection was completed in the summer of 2010 when a total of 24,916 twins had been contacted.

Response rates

All twins were asked to fill out the questionnaire while only twins from whom DNA was lacking were asked to provide saliva. In this survey, 11,372 respondents (subject-wise response rate 46%) gave informed consent to have their data stored and analyzed. Among these, the median birth-year was 1950 and 54.3% were females. The total number of complete (answers obtained from both twins) responding pairs was 3,496, distributed as 1,157 MZ, 1,220 SSDZ pairs, 1,109 OSDZ pairs and 10 where zygosity could not be determined. For numbers of responding subjects and complete pairs by sex and zygosity, see Table 3. It is clear that the tendency to participate varies somewhat by zygosity and sex. Among invited complete pairs, we received responses from both members of the pair in 44% of the female MZ while the corresponding fraction for male DZ pairs was 24% (Table 3).

For the collection of saliva DNA, Oragene DNA OG-500 kits were used. Among all invited to the SALTY study, we received 6,501 saliva samples. Together with blood DNA already donated by 5,272 of the SALTY participants in the TwinGene study, we have saliva or blood DNA available for 11,773 of the subjects invited to the SALTY study (47%). Among SALTY questionnaire respondents, saliva or blood DNA was available for 10,048 (88%). DNA from saliva has been extracted using the Chemagic STAR

instrument from Hamilton Robotics with magnet bead purification kits from Chemagen, dissolved in 500 µL buffer and stored at -20 C.

Table 3 about here.

Paper questionnaire

The SALTY paper questionnaire covered the following areas: 1) Hearing and tinnitus problems, each measured regarding severity on a 4-level Likert-Type scale; 2) Cardiovascular disease: same questions as used in the TwinGene study (see above); 3) Urinary dysfunction (Altman et al., 2011); 4) Perinatal depression (Meltzer-Brody et al., 2011); 5) A comprehensive battery of economical and political preferences (Benjamin et al., 2012; Cesarini et al., 2011); 6) Rheumatoid arthritis with questions on lifetime joint pain, osteoarthritis/arthritis, whether the arthritis diagnosis was made by a doctor, monitoring and treatment; 7) Occupational strain with questions on physical workload, problems of pain in the neck, shoulders, lower back and elbows. A similar battery was used in the STAGE study (Nyman et al., 2009); 8) Scoliosis was measured by asking for lifetime physician-diagnosed scoliosis and treatment; 9) Self-rated health (EQ5D) (EuroQol, 1990); 10) Extent of psychosis-related experiences were recorded through a subset (9 out of 20) of questions on psychotic-like symptoms included in the CAPE (Community Assessment of Psychic Experiences) questionnaire (Hedelin et al., 2010); 11) Mobile phone use (Schuz et al., 2011); 12) Partner attachment and jealousy, through items collected from the Sociosexual Orientation Inventory (Simpson & Gangestad, 1991), the Partner Bonding Scale (Walum et al., 2008) and various jealousy questionnaires (Harris, 2003).

To assess test-retest variability, a request to fill out an identical questionnaire at second time point was sent out to a random set of 800 SALTY participants with a bonus of lottery tickets for those willing to fill out the questionnaire again. This was performed after finishing the primary data collection, so the time between the two test occasions varied between 2 and 18 months. We obtained a retest from 495 (62%) participants. The test-retest variability has not been investigated for all items, but for 9 scales on economic and political preferences the average test-retest correlation was 0.78 (range, 0.53-0.89).

SALTY web-based extension

A separate invitation to participate in a web-based test was included in the postal shipment of the paper questionnaire. This part was aimed at measuring creative achievement, proneness for psychological flow experiences, cognitive ability and stability in motor tasks. It contained questions about creative achievement in different domains (music, visual arts, theater, literature, dance, science, invention) and flow proneness (Mosing et al., 2012; Ullén et al., 2012). The web application also included a battery of tests for reaction time and temporal accuracy (Holm et al., 2011), as well as cognition (Wiener Matrizen test, WMT). The WMT is a 24-item test which is similar in construction to the Raven test, with which it is highly correlated ($r = .92$) (Formann & Piswanger, 1979). A total of 3,070 twins aged between 50 and 67 (mean 58.9) choose to participate in this web-based extension of the SALTY study.

STAGE DNA COLLECTION

All participants of the phenotypically extensive STAGE study (Furberg, et al., 2008) (conducted in 2005-2006) were approached again (in 2009-2010) with a request to

donate saliva for DNA extraction (no additional questionnaire or request for other measures was included in this DNA collection effort). The same procedure of collection as used in the SALTY study (Oragene DNA OG-500 kits) and the same extraction method as used in the TwinGene study (Puregene extraction kits) was adopted. A total of 10,282 STAGE participants provided DNA, corresponding to a subject-wise donation rate of 48%.

CATSS EXTENSION STUDIES

Building on the recruitment study of 9-year-old twins (CATSS) that has been ongoing since 2004, several extensions and follow-ups have started. The CATSS is comprehensively described, including detailed descriptions of instruments and measures at various ages, in a recent overview article (Anckarsater, et al., 2011) and will also be briefly described here. Although CATSS collects information on a wide range of health issues such as inflammatory and respiratory diseases common in childhood, the phenotypic focus is on neurodevelopmental problems (NDP).

Screening for neurodevelopmental problems

An inventory called the “Autism – Tics, ADHD and other Comorbidities” (A-TAC) was especially designed for use in the initial telephone interview of the parents (Hansson et al., 2005; Larson et al., 2010). A-TAC allows for a systematic screening for autism spectrum disorder, attention deficit/hyperactivity disorder (ADHD), tic disorders, developmental coordination disorder, learning disorders, oppositional defiant disorder, conduct disorder, obsessive–compulsive disorder, and/or eating problems. The overall response rate of the initial telephone interview was 80%.

Test-retest reliability of the telephone interview was investigated from 200 interviews performed a second time concerning 218 boys (55%) and 182 girls (45%). The administration of the second interview varied from one to four months (mean 70 days) after the first. The test-retest correlations have up to now only been investigated for the psychiatric problems; test-retest reliability range from >0.8 for autism spectrum and ADHD, over >0.60 for most other neuropsychiatry to 0.4-0.5 for anxiety, compulsion, and feeding problems.

Follow-up of A-TAC positive families

Families in which one or both of the twins were found positive for any of the NDPs screened for and randomly selected control families were administered paper questionnaires containing several instruments providing background information relevant to mental health. Although data collection is ongoing, by January 1, 2010 the overall response rate for the questionnaire phase was 1,065/1,756 mailed questionnaires (60.6%).

A clinical validation of the 9-year-old children positive for any of the NDPs screened for in the telephone interview is ongoing among families living in the regions of Skåne, Västra Götaland, or Stockholm (about 50% of the Swedish population) to confirm or dismiss the diagnoses. The study also invites 5% of the screen-negative twin pairs. Furthermore, information on previous contacts with health care services is collected. The study is designed to identify factors predicting contact with the health care services, especially the Child and Adolescent Mental Health Services, and will also provide a clinical validation of the A-TAC inventory in a population-based sample. The study started during 2011 at three study sites and is planned to continue at least until 2013.

Follow-up at older ages

When the twins participating in CATSS reach 15 years of age, the families are contacted again. The major aim of this follow-up is to study the effect of NDPs on primarily four outcomes: social marginalization, substance abuse, criminality, and the onset of new mental health problems. Also, a follow-up on somatic diseases from the interview at 9 years is done. Parents and twins fill out separate questionnaires. This part of the study started in the winter 2009–2010 with the cohorts born in 1994 and 1995. The number of individuals responding so far is 4,102. In addition, in a subsample, families with same-sex twins, one or both of whom that were positive in the A_TAC screening at age 9 or selected as random controls for the questionnaire, were invited to a clinical examination at age 15. So far this activity has been limited three sites: Stockholm (about 50 pairs yearly), Malmö (about 20 pairs yearly), and Göteborg (about 30 pairs yearly) during years 2008-2010. This clinical examination is focused on psychosocial marginalization, includes a comprehensive psychiatric work-up according to all five axes of the DSM-IV-TR and covers twins born between 1993 and 1995. The total number of twins examined so far is 452.

When the twins reach 18 years of age they are invited, as are their parents, to answer a web-based questionnaire. The focus of this study is to provide longitudinal information on neuropsychiatric conditions with regards to criminality, substance abuse and psychosocial marginalization and to detect possible initiation of psychosis and other mental health problems. The study currently includes information from roughly 3,500 individuals.

To follow up previous CATSS studies on the association between growth and subsequent disease (Lundholm et al., 2010; Ortqvist et al., 2009), growth charts with

longitudinal records of weight and height were collected from the child health care centers (ages 0-5) and school health care (ages 6-18). In the 1994 and 1995 birth cohorts, approximately 81% of the twins have consented and between 55% (child health care centers) and 82% (school health care) of records have been retrieved.

DNA collection

Since 2008, saliva for DNA extraction has been collected directly after the initial telephone interview at age 9. Oragene DNA OG-500 and Puregene extraction kit (Gentra systems, Minneapolis, USA) have been used for collection and DNA extraction, respectively. Cohorts born 2003-2007 have recently been re-contacted to complement the CATSS biobank, which presently includes samples from more than 12,400 individuals. For twins contacted for clinical examination at age 15, additional biomaterial is retrieved through donations of capillary blood and from both parents by saliva.

STOPPA

Based on the CATSS study ascertainment, the Swedish Twin study On Perinatal characteristics to Prevent Asthma (STOPPA) addresses if MZ twins, discordant and concordant for asthma, differ from each other in objective markers of lung function, serology, genetic variation, early exposures or epigenetic changes. From answers in the CATSS study and an asthma and/or wheezing algorithm, 9-14 year old MZ twins discordant or concordant for asthma are invited to participate in a clinical examination. A control group of concordant healthy MZ twin pairs and DZ twins is also included. Background characteristics, such as history of asthma, BMI, physical activity, quality of

life, tobacco usage, medication and socioeconomic factors, along with questions screening for neuropsychiatric disorders and personality (anxiety) are collected. We also collect information on perinatal characteristics from the Medical Birth Register. A clinical examination is performed to test lung function and capacity (spirometry and exhaled nitric oxide) and blood is collected for serology and DNA. We will make comparisons of the effect of environmental factors on outcomes across discordant, concordant and a control group of unaffected twin pairs. Genome-wide methylation will be investigated in blood DNA. Possible candidate methods considered now include next generation sequencing and methylation chips.

The data collection is ongoing. To date, 355 twin pairs (710 individual twins) from Stockholm and Göteborg have been invited; 131 twin pairs (262 twins), 37%, have participated in both the questionnaire and clinical examination while 206 twins (28%) have declined. Currently, additional twin pairs in Umeå, Karlstad, Linköping and Malmö/Lund are invited to participate in the STOPPA study.

PETSS

The initial data collection in the first Swedish twin study of preschool children has recently been finished. Parents of all twins born in Sweden between January 2004 and May 2005 were identified by a request to the National Board of Health and Welfare and contacted as part of the Preschool Twin Study in Sweden (PETSS) one month prior to the twins 5th birthday. PETSS was designed to address how genes and environments in early childhood contribute to the development of behavioral problems. Questionnaires to parents and teachers were used to measure conduct problems, ADHD, aggression, and temperament in the twins. The parents also responded to questions regarding their

own behavioral problems and different family characteristics (e.g. parenting). A special feature of PETSS is that a web-based platform for the distribution of a computer-based neuropsychological test battery (e.g. working memory and response inhibition) has been used. Questionnaires were sent to parents and preschool teachers of 1,261 twin pairs. Non-responders were approached with up to three reminders. Parents were approached separately, resulting in 828 (65%) responses from the mothers and 698 (55%) responses from fathers. Mother ratings or father ratings were available for 879 twin pairs. The response rate among the preschool teachers was 54% (n=686).

Zygoty was determined by fitting a 2-class (i.e. MZ or DZ) latent class model (Heath et al., 2003) in Mplus Version 4.1 (Muthén & Muthén, 2006) to standard physical similarity questions (Lichtenstein, et al., 2006). Latent class models were fitted separately for mother and father reports. Zygoty was scored as unknown for 25 twin pairs due to contradictions between the mother reports and father reports (20 twin pairs) or due to low predicted probabilities of class membership (5 twin pairs were assigned as MZ or DZ with a probability lower than 0.95). The final sample consisted of 1,708 twins (854 twin pairs) with zygoty assigned; of which 284 were MZ male twins, 246 were DZ male twins, 292 were MZ female twins, 238 were female DZ twins and 648 were OSDZ twins.

Because PETSS is an exemption from the usual ascertainment scheme (i.e. having the recruitment contact to STR at age 9, adopted since 2004), PETSS data are not part of the STR database at this point. Thus, similarity-based zygoty data from PETSS have not been included or used in the overall description of coverage and zygoty in STR (Figure 1 and 2).

Discussion

Since it was established in the late 1950s the STR has been utilized to contribute to our understanding of genetic and environmental influences on disease and other traits of interest. The main contributions has been through quantitative twin modeling of trait variances and the use of co-twin control designs, requiring collection of phenotypic and exposure data in large quantities and of high quality, which is still ongoing in several projects on twins of various ages. However, the major development of the STR during the past 10 years has been the collection of biological samples and the creation of a biobank, at present covering a total of 45,000 twins. As intended, the effort has led to initiation of various molecular studies, so far dominated by studies based on genome-wide genotyping. The genotypes have been used for studies on quantitative traits, behavior, common health-problems and disease.

The ethnic composition of the populations from which the study material is drawn is important in most genetic studies. The STR is population-based and it is likely that the ethnic structure in the STR mirrors Swedish born singletons. As expected, we find that the vast majority of the twins have both parents born in Sweden. However, the proportion of twins being second generation immigrants has increased from less than 5% among twins born before 1945 to over 20% today. The proportion of twins having parents born outside Europe has increased, predominantly after 1970. Hence, for genetic association studies, researchers therefore ought to be cautious about influences from genetic stratification in studies utilizing the younger (e.g. CATSS) compared to the older (e.g. TwinGene) cohorts. Nevertheless, there are marked gradients of stratification also among the Swedish citizens considered native. Particularly the difference between the northernmost counties and the rest of Sweden is pronounced, with larger genetic

differences observed between them than between Southern Swedes and the HapMap CEU samples (Humphreys et al., 2011).

The CATSS ascertainment scheme introduced in 2004 with systematic contacts made with parents to all twins the year they turn 9 years, has led to a longitudinal approach following the development of twin children through adolescence and into adulthood. CATSS and its follow-up studies cover practically all aspects of neurodevelopmental problems and their potential medical and psychosocial consequences. The major health problem of asthma in young individuals is also under intense study within the STOPPA study. Until recently, the STR had not been used for studies with pre-school children; the PETSS study is the first to do so and will therefore be highly interesting to follow for potential associations between measures at pre-school age to later health and behavioral problems in school and young adulthood.

MZ twins discordant for highly heritable diseases have puzzled researchers over the years. Following developments in biotechnology, with the new possibilities to perform systematic large-scale investigations of DNA primary sequence, DNA modifications, RNA, metabolites and proteins levels, the interest in using such twins for molecular epidemiology has increased. The rationale being that identified molecular differences may reflect causal mechanisms involving environmental exposures and/or constitute a direct cause of the phenotypic difference. Mutations occurring after the split of the initial intact embryo, and epigenetic (e.g., DNA methylation, histone acetylation) differences have been put forward as promising candidate mechanisms and projects based on discordant MZ twins are ongoing in STR and other twin register resources around the globe (van Dongen, et al., 2012).

Related to the introduction of IVF, twin births have increased in numbers. The practice to transfer multiple embryos (in order to increase the probability of delivery

per attempt) varies between countries as a function of how the IVF treatment is usually financed. In Sweden, where government funding for multiple IVF-cycles has existed for many years as well as a recommendation from the National Board of Health and Welfare of only one embryo transfer, the influence of IVF on twinning rates has probably been smaller as compared to many other countries not providing such funding or similar recommendations. Yet we have observed a change in the ratio of DZ to MZ twins coinciding with the introduction of IVF in Sweden, although the magnitude is not more pronounced than in a prior decreasing trend observed between 1960 and 1980.

Applications to use STR data for research purposes are open to Swedish as well as international collaborators. Information on application procedures and policies is available in both Swedish and English through our Internet site:

www.tvillingregistret.se.

The prospects of moving the biomedical field forward rely on several critical factors. STR works towards improving some of them, in particular the breadth and depth of phenotypic and genetic data available. Together with improvements in study design, molecular, statistical and computational tools, we expect the STR to continue to contribute to our common strive towards a better understanding of life.

Acknowledgements

The Ministry for Higher Education financially supports the STR. For the various projects described here financial support has been provided by: TwinGene; the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF), the Heart and Lung foundation no. 20070481. STOPPA; the Strategic Research Program in Epidemiology at Karolinska Institutet, the Swedish Research Council (grant number 2011-3060), the Swedish Asthma and Allergy Association and the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet. CATSS; Support was provided by the Swedish Council for Working Life and Social Research, the Swedish Research Council, Systembolaget, the National Board of Forensic Medicine, the Swedish Prison and Probation Service, Bank of Sweden Tercentenary Foundation, the Söderström-Königska foundation, and the Karolinska Institutet Center of Neurodevelopmental Disorders (KIND). BIRTH; Supported by the Swedish Council for Working Life and Social Research (2004-0174 and 2007-0231), the Swedish Research Council (K2006-71X-14676-04-2 and 2008-54X20638-01-3), the Swedish Cancer Society (4594-B01-01XAC and 4594-B04-04XAB), and the European Union-funded Network of Excellence Lifespan (FP6 036894).

References

- Altman, D., Lundholm, C., Milsom, I., Peeker, R., Fall, M., Iliadou, A. N., & Pedersen, N. L. (2011). The genetic and environmental contribution to the occurrence of bladder pain syndrome: an empirical approach in a nationwide population sample. *Eur Urol*, *59*(2), 280-285.
- Anckarsater, H., Lundstrom, S., Kollberg, L., Kerekes, N., Palm, C., Carlstrom, E., Langstrom, N., Magnusson, P. K., Halldner, L., Bolte, S., Gillberg, C., Gumpert, C., Rastam, M., & Lichtenstein, P. (2011). The Child and Adolescent Twin Study in Sweden (CATSS). *Twin Res Hum Genet*, *14*(6), 495-508.
- Benjamin, D. J., Cesarini, D., van der Loos, M. J., Dawes, C. T., Koellinger, P. D., Magnusson, P. K., Chabris, C. F., Conley, D., Laibson, D., Johannesson, M., & Visscher, P. M. (2012). The genetic architecture of economic and political preferences. *Proc Natl Acad Sci U S A*, *109*(21), 8026-8031.
- Bergvall, N., Iliadou, A., Johansson, S., de Faire, U., Kramer, M. S., Pawitan, Y., Pedersen, N. L., Lichtenstein, P., & Cnattingius, S. (2007). Genetic and shared environmental factors do not confound the association between birth weight and hypertension: a study among Swedish twins. *Circulation*, *115*(23), 2931-2938.
- Cederlöf, R. (1966). The Twin Method in Epidemiological Studies on Chronic Disease. *Dissertation, Karolinska Institutet*.
- Cesarini, D., Johannesson, M., Magnusson, P. K., & Wallace, B. (2011). The Behavioral Genetics of Behavioral Anomalies. *Management Science* (10.1287/mnsc.1110.1329).
- EuroQol. (1990). EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy*, *16*(3), 199-208.
- Formann, A.K., & Piswanger, K. (1979). *WMT - Wiener Matrizentest. Manual*. Weinheim: Beltz Test Gesellschaft.
- Furberg, H., Lichtenstein, P., Pedersen, N. L., Thornton, L., Bulik, C. M., Lerman, C., & Sullivan, P. F. (2008). The STAGE cohort: a prospective study of tobacco use among Swedish twins. *Nicotine Tob Res*, *10*(12), 1727-1735.
- Gold, C. H., Malmberg, B., McClearn, G. E., Pedersen, N. L., & Berg, S. (2002). Gender and health: a study of older unlike-sex twins. *J Gerontol B Psychol Sci Soc Sci*, *57*(3), S168-176.
- Hannelius, U., Gherman, L., Makela, V. V., Lindstedt, A., Zucchelli, M., Lagerberg, C., Tybring, G., Kere, J., & Lindgren, C. M. (2007). Large-scale zygosity testing using single nucleotide polymorphisms. *Twin Res Hum Genet*, *10*(4), 604-625.
- Hansson, S. L., Svanstrom Rojvall, A., Rastam, M., Gillberg, C., & Anckarsater, H. (2005). Psychiatric telephone interview with parents for screening of childhood autism - tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): preliminary reliability and validity. *Br J Psychiatry*, *187*, 262-267.

- Harris, C.R. (2003). Factors Associated with Jealousy Over Real and Imagined Infidelity: an Examination of the Social-Cognitive and Evolutionary Psychology Perspectives. *Psychology of Women Quarterly*, 4, 319-329.
- Heath, A. C., Nyholt, D. R., Neuman, R., Madden, P. A., Bucholz, K. K., Todd, R. D., Nelson, E. C., Montgomery, G. W., & Martin, N. G. (2003). Zygosity diagnosis in the absence of genotypic data: an approach using latent class analysis. *Twin Res*, 6(1), 22-26.
- Hedelin, M., Lof, M., Olsson, M., Lewander, T., Nilsson, B., Hultman, C. M., & Weiderpass, E. (2010). Dietary intake of fish, omega-3, omega-6 polyunsaturated fatty acids and vitamin D and the prevalence of psychotic-like symptoms in a cohort of 33,000 women from the general population. *BMC Psychiatry*, 10, 38.
- Hoekstra, C., Willemsen, G., van Beijsterveldt, C. E., Lambalk, C. B., Montgomery, G. W., & Boomsma, D. I. (2010). Body composition, smoking, and spontaneous dizygotic twinning. *Fertil Steril*, 93(3), 885-893.
- Hoekstra, C., Zhao, Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I., & Montgomery, G. W. (2008). Dizygotic twinning. *Hum Reprod Update*, 14(1), 37-47.
- Holm, L., Ullen, F., & Madison, G. (2011). Intelligence and temporal accuracy of behaviour: unique and shared associations with reaction time and motor timing. *Exp Brain Res*, 214(2), 175-183.
- Humphreys, K., Grankvist, A., Leu, M., Hall, P., Liu, J., Ripatti, S., Rehnstrom, K., Groop, L., Klareskog, L., Ding, B., Gronberg, H., Xu, J., Pedersen, N. L., Lichtenstein, P., Mattingsdal, M., Andreassen, O. A., O'Dushlaine, C., Purcell, S. M., Sklar, P., Sullivan, P. F., Hultman, C. M., Palmgren, J., & Magnusson, P. K. (2011). The genetic structure of the Swedish population. *PLoS One*, 6(8), e22547.
- Karlstrom, P. O., & Bergh, C. (2007). Reducing the number of embryos transferred in Sweden-impact on delivery and multiple birth rates. *Hum Reprod*, 22(8), 2202-2207.
- Larson, T., Anckarsater, H., Gillberg, C., Stahlberg, O., Carlstrom, E., Kadesjo, B., Rastam, M., & Lichtenstein, P. (2010). The autism--tics, AD/HD and other comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. *BMC Psychiatry*, 10, 1.
- Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., & Pedersen, N. L. (2002). The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med*, 252(3), 184-205.
- Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlstrom, E., Bjork, C., Svartengren, M., Wolk, A., Klareskog, L., de Faire, U., Schalling, M., Palmgren, J., & Pedersen, N. L. (2006). The Swedish Twin Registry in the third millennium: an update. *Twin Res Hum Genet*, 9(6), 875-882.
- Lichtenstein, P., Tuvblad, C., Larsson, H., & Carlstrom, E. (2007). The Swedish Twin study of CHild and Adolescent Development: the TCHAD-study. *Twin Res Hum Genet*, 10(1), 67-73.

- Lundholm, C., Ortqvist, A. K., Lichtenstein, P., Cnattingius, S., & Almqvist, C. (2010). Impaired fetal growth decreases the risk of childhood atopic eczema: a Swedish twin study. *Clin Exp Allergy*, *40*(7), 1044-1053.
- Medlund, P., Cederlof, R., Floderus-Myrhed, B., Friberg, L., & Sorensen, S. (1976). A new Swedish twin registry containing environmental and medical base line data from about 14,000 same-sexed pairs born 1926-58. *Acta Med Scand Suppl*, *600*, 1-111.
- Meltzer-Brody, S., Zerwas, S., Leserman, J., Holle, A. V., Regis, T., & Bulik, C. (2011). Eating disorders and trauma history in women with perinatal depression. *J Womens Health (Larchmt)*, *20*(6), 863-870.
- Mosing, M.A., Magnusson, P.K, Pedersen, N.L. , Nakamura, J. , Madison, G. , & Ullén, F. (2012). Heritability of proneness for psychological flow experiences. *Personality and Individual Differences*, *53*(5), 699-704.
- Muthén, L.K., & Muthén, B.O. . (2006). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Nyman, T., Mulder, M., Iliadou, A., Svartengren, M., & Wiktorin, C. (2009). Physical workload, low back pain and neck-shoulder pain: a Swedish twin study. *Occup Environ Med*, *66*(6), 395-401.
- Oberg, S., Cnattingius, S., Sandin, S., Lichtenstein, P., & Iliadou, A. N. (2011). Birth weight predicts risk of cardiovascular disease within dizygotic but not monozygotic twin pairs: a large population-based co-twin-control study. *Circulation*, *123*(24), 2792-2798.
- Ortqvist, A. K., Lundholm, C., Carlstrom, E., Lichtenstein, P., Cnattingius, S., & Almqvist, C. (2009). Familial factors do not confound the association between birth weight and childhood asthma. *Pediatrics*, *124*(4), e737-743.
- Pedersen, N. L., Lichtenstein, P., & Svedberg, P. (2002). The Swedish Twin Registry in the third millennium. *Twin Res*, *5*(5), 427-432.
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*, *11*(7), 459-463.
- Rahman, I., Bennet, A. M., Pedersen, N. L., de Faire, U., Svensson, P., & Magnusson, P. K. (2009). Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet*, *12*(3), 286-294.
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H., Pato, C. N., Malhotra, A. K., Purcell, S., Dudbridge, F., Neale, B. M., Rossin, L., Visscher, P. M., Posthuma, D., Ruderfer, D. M., Fanous, A., Stefansson, H., Steinberg, S., Mowry, B. J., Golimbet, V., De Hert, M., Jonsson, E. G., Bitter, I., Pietilainen, O. P., Collier, D. A., Tosato, S., Agartz, I., Albus, M., Alexander, M., Amdur, R. L., Amin, F., Bass, N., Bergen, S. E., Black, D. W., Borglum, A. D., Brown, M. A., Bruggeman, R., Buccola, N. G., Byerley, W. F., Cahn, W.,

- Cantor, R. M., Carr, V. J., Catts, S. V., Choudhury, K., Cloninger, C. R., Cormican, P., Craddock, N., Danoy, P. A., Datta, S., de Haan, L., Demontis, D., Dikeos, D., Djurovic, S., Donnelly, P., Donohoe, G., Duong, L., Dwyer, S., Fink-Jensen, A., Freedman, R., Freimer, N. B., Friedl, M., Georgieva, L., Giegling, I., Gill, M., Glenthøj, B., Godard, S., Hamshere, M., Hansen, M., Hansen, T., Hartmann, A. M., Henskens, F. A., Hougaard, D. M., Hultman, C. M., Ingason, A., Jablensky, A. V., Jakobsen, K. D., Jay, M., Jurgens, G., Kahn, R. S., Keller, M. C., Kenis, G., Kenny, E., Kim, Y., Kirov, G. K., Konnerth, H., Konte, B., Krabbendam, L., Krasucki, R., Lasseter, V. K., Laurent, C., Lawrence, J., Lencz, T., Lerer, F. B., Liang, K. Y., Lichtenstein, P., Lieberman, J. A., Linszen, D. H., Lonnqvist, J., Loughland, C. M., Maclean, A. W., Maher, B. S., Maier, W., Mallet, J., Malloy, P., Mattheisen, M., Mattingsdal, M., McGhee, K. A., McGrath, J. J., McIntosh, A., McLean, D. E., McQuillin, A., Melle, I., Michie, P. T., Milanova, V., Morris, D. W., Mors, O., Mortensen, P. B., Moskvina, V., Muglia, P., Myin-Germeys, I., Nertney, D. A., Nestadt, G., Nielsen, J., Nikolov, I., Nordentoft, M., Norton, N., Nothen, M. M., O'Dushlaine, C. T., Olincy, A., Olsen, L., O'Neill, F. A., Orntoft, T. F., Owen, M. J., Pantelis, C., Papadimitriou, G., Pato, M. T., Peltonen, L., Petursson, H., Pickard, B., Pimm, J., Pulver, A. E., Puri, V., Quested, D., Quinn, E. M., Rasmussen, H. B., Rethelyi, J. M., Ribble, R., Rietschel, M., Riley, B. P., Ruggeri, M., Schall, U., Schulze, T. G., Schwab, S. G., Scott, R. J., Shi, J., Sigurdsson, E., Silverman, J. M., Spencer, C. C., Stefansson, K., Strange, A., Strengman, E., Stroup, T. S., Suvisaari, J., Terenius, L., Thirumalai, S., Thygesen, J. H., Timm, S., Toncheva, D., van den Oord, E., van Os, J., van Winkel, R., Veldink, J., Walsh, D., Wang, A. G., Wiersma, D., Wildenauer, D. B., Williams, H. J., Williams, N. M., Wormley, B., Zammit, S., Sullivan, P. F., O'Donovan, M. C., Daly, M. J., & Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*, *43*(10), 969-976.
- Schuz, J., Elliott, P., Auvinen, A., Kromhout, H., Poulsen, A. H., Johansen, C., Olsen, J. H., Hillert, L., Feychting, M., Fremling, K., Toledano, M., Heinavaara, S., Slottje, P., Vermeulen, R., & Ahlbom, A. (2011). An international prospective cohort study of mobile phone users and health (Cosmos): design considerations and enrolment. *Cancer Epidemiol*, *35*(1), 37-43.
- Simpson, J. A., & Gangestad, S. W. (1991). Individual differences in sociosexuality: evidence for convergent and discriminant validity. *J Pers Soc Psychol*, *60*(6), 870-883.
- Statistics-Sweden. (2011). Multi-generation register 2010: a description of contents and quality. Background facts on population and welfare statistics 1654-4331; 2011:2.
- Ullén, F., de Manzano, Ö., Almeida, R., Magnusson, P.K., Pedersen, N.L., Nakamura, J., Csikszentmihályi, M., & Madison, G. (2012). Proneness for psychological flow in everyday life: Associations with personality and intelligence. *Personality and Individual Differences*, *52*(2), 167-172.
- 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. *Nat Rev Genet*.
- Walum, H., Westberg, L., Henningson, S., Neiderhiser, J. M., Reiss, D., Igl, W., Ganiban, J. M., Spotts, E. L., Pedersen, N. L., Eriksson, E., & Lichtenstein, P. (2008). Genetic

variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. *Proc Natl Acad Sci U S A*, 105(37), 14153-14156.

Tables

Table 1. Studies conducted in STR targeting whole Swedish twin cohorts.

Study name (Year conducted)	Birth-years included	N (respondents)	References
Q61 (1961)	1886-1925 ^a	21,870	(Cederlöf, 1966; Lichtenstein, et al., 2002)
Q63 (1963)	1886-1925 ^a	18,616	(Lichtenstein, et al., 2002)
Q67 (1967)	1886-1925 ^a	16,779	(Lichtenstein, et al., 2002)
Q70 (1970)	1886-1925 ^a	8,763	(Lichtenstein, et al., 2002)
Q73 (1973)	1926-1958 ^a	30,279	(Lichtenstein, et al., 2002; Medlund, et al., 1976)
SALT (1998-2002)	1911-1958	44,919	(Lichtenstein, et al., 2002)
TwinGene (2004-2008)	1911-1958	12,614	(Lichtenstein, et al., 2006; Rahman, et al., 2009), This paper
SALTY (2009-2010)	1943-1958	11,372	This paper
STAGE (2005-2006)	1959-1985	25,364	(Lichtenstein, et al., 2006)
TCHAD (1994, 1999, 2002, 2005)	1985-1986	1,480	(Lichtenstein, et al., 2007)
CATSS (2004--)	1993--	22,004	(Anckarsater, et al., 2011)
YATSS (started)	1986-1992	-	

^a Restricted to same sex pairs only.

Table 2. Number of twins consenting and donating blood to the TwinGene study and participation rates over zygosity and sex.

Zygosity	MZ		SSDZ		OSDZ	UNKNOWN		TOTAL
Sex	Male	Female	Male	Female		Male	Female	
Complete pairs ^a	587	838	786	1078	1722	3 (7%)	0	5014
	(52%)	(56%)	(43%)	(44%)	(41%)		(0%)	(45%)
Twin individuals ^b	1376	1831	2087	2643	4645	23	9	12614
	(61%)	(61%)	(57%)	(54%)	(55%)	(25%)	(13%)	(56%)

^a Number of pairs in which both twins consented and donated blood, percentage shows complete pair participation rate. ^b Sum of individuals consenting and donating blood in complete and incomplete pairs, percentage shows individual-wise participation rate.

Table 3. Number of twins participating in the SALTY paper questionnaire study and participation rates over zygosity and sex.

Zygosity	MZ		SSDZ		OSDZ	UNKNOWN		TOTAL
Sex	Male	Female	Male	Female		Male	Female	Total
Complete pairs ^a	465	692	502	718	1109	9	1 (2%)	3496
	(35%)	(44%)	(24%)	(34%)	(26%)	(10%)		(30%)
Twin individuals ^b	1347	1722	1902	2199	4115	59	28	11372
	(47%)	(52%)	(42%)	(49%)	(44%)	(26%)	(19%)	(46%)

^a Number of participating complete pairs, percentage shows complete pair participation rate ^b Sum of participating individuals in complete and incomplete pairs, percentage shows individual-wise participation rate.

Figure legends

Figure 1. Number of twins pairs in STR (solid thick line), number of twin pairs in STR with zygosity determined (solid thin line) and the number of twins pairs in the MGR (dotted line), by birth-year.

Figure 2. The proportion of twin pairs in the STR with zygosity determined.

Figure 3. Proportion of twins with at least one: non-Swedish (solid thick line), non-Scandinavian (dotted line) or non-European (solid thin line), born parent, by birth year.

Figure 4. Proportion of DZ twin births by birth-year: The solid line is based on STR data only and shows the proportion of SSDZ out of all same-sex twins in STR. The lack of data between 1986 and 1992 is due to the gap in the ascertainment contacts of twins to the STR for these years. The dotted line is based on MGR data only and shows proportion of OSDZ out of all multiple births in the MGR.

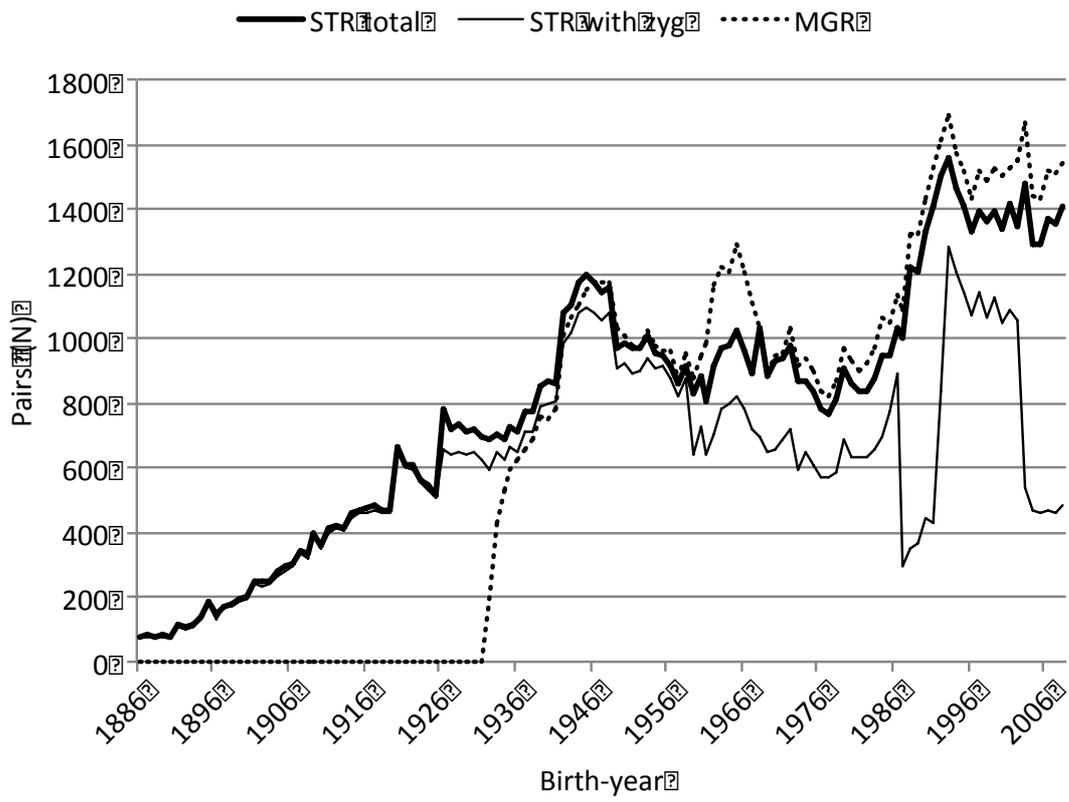


Figure 1.

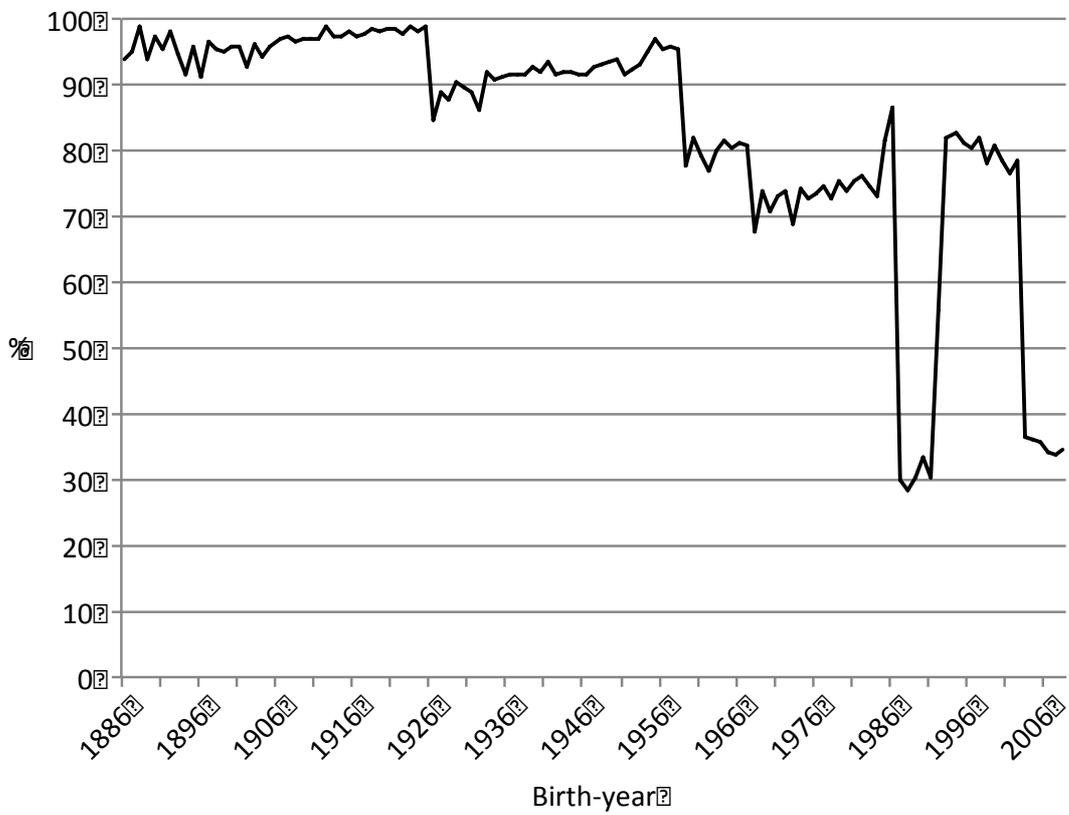


Figure 2.

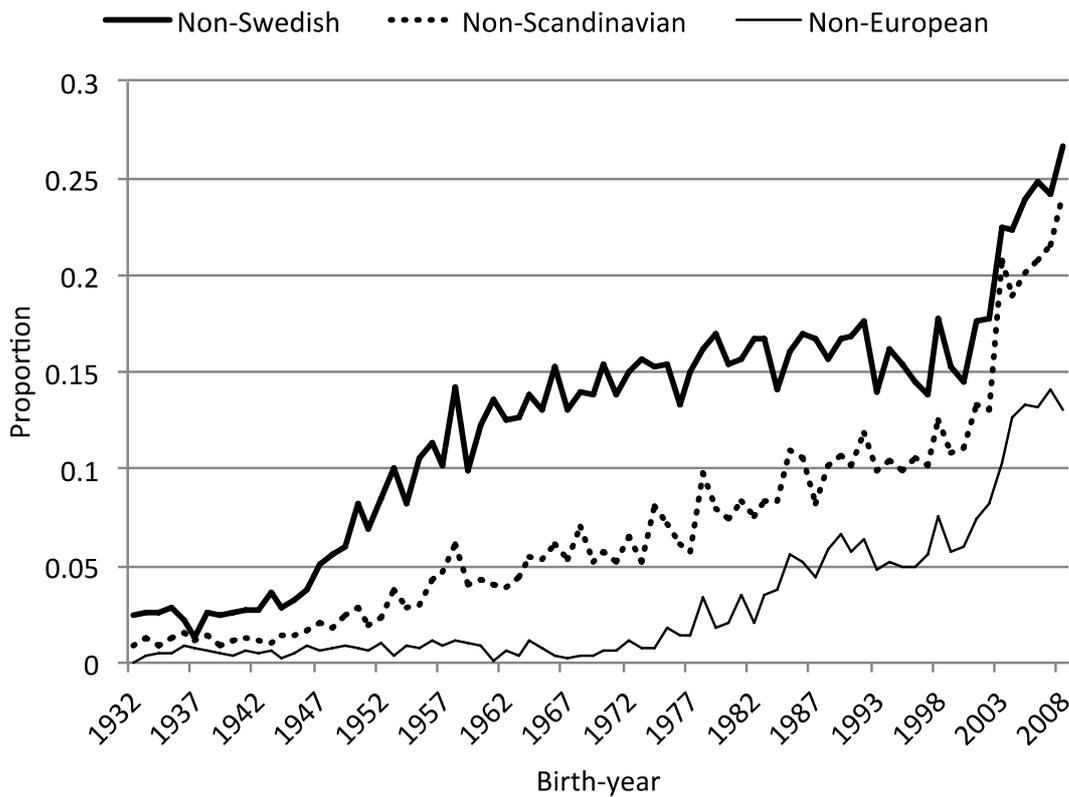


Figure 3.

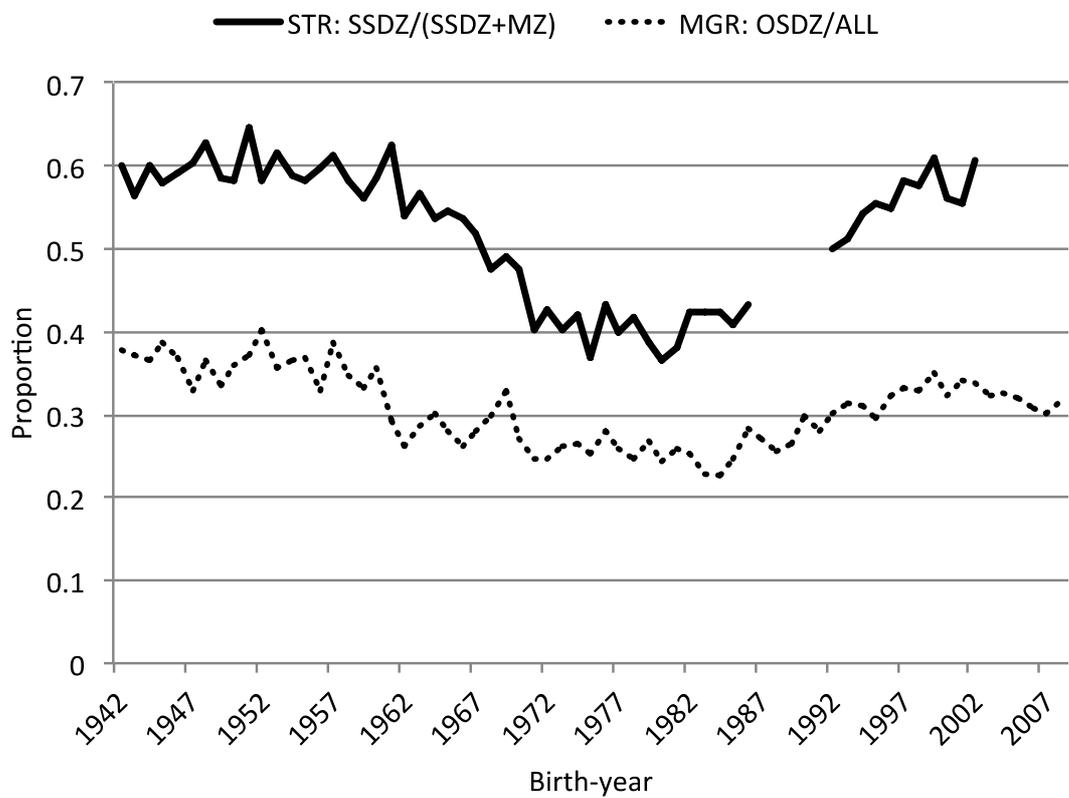


Figure 4.