

Article

The Swedish Twin Registry: Content and Management as a Research Infrastructure

Ulrika Zagai, Paul Lichtenstein, Nancy L. Pedersen and Patrik K. E. Magnusson

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

Abstract

The Swedish Twin Registry functions as research infrastructure containing information on 216,258 twins born between 1886 and 2015, of whom 86,199 pairs have zygosity determined by DNA, an intrapair similarity algorithm, or being of opposite sex. In essence, practically all twins alive and currently 9 years or older have been invited for participation and donation of DNA on which genomewide single nucleotide polymorphisms array genotyping has been performed. Content, management and alternatives for future improvements are discussed.

Keywords: Twin; registry; biobank; population-based; Sweden; genetics; environment

(Received 19 May 2019; accepted 31 May 2019; First Published online 21 November 2019)

The Swedish Twin Registry (STR) was established in the late 1950s to investigate the role of environmental factors such as smoking and alcohol on disease. Since then, the STR has developed into an infrastructure of broad utility, with the mission to provide a longitudinal resource for epidemiological and molecular studies of twins.

Content of the STR

The last overview article about the STR was published in TRHG in 2013 (Magnusson et al., 2013). In this update, we provide a summary of the major, so-called base-data collections, structured according to the source of data: (1) Self- or parental responses to questionnaires/interviews, (2) biobanking, (3) genotyping, (4) biomarkers, (5) health check-ups and (6) national register linkages.

Stemming from the base-data, there have been several ancillary studies directed toward specific phenotypic or geographic selections, conducted by various users over the years. Sometimes, the data collected in such studies have been fed back into the STR database and may, therefore, be open for secondary use. However, because of their heterogeneity and incomplete coverage, we do not describe them further here. Thus, the description is limited to include only *nationwide efforts to collect data with a broad scope*.

Self- or Parental Responses

Between 1960 and 2005, the STR was almost exclusively a collection of phenotypic and exposure data and most resources were put into obtaining responses to questionnaires or interviews. Several large efforts to gather this type of data have been made throughout the history of the STR, each directed toward a

specific range of birth years (Table 1). Here, we describe these efforts in the chronological order of when data collection was conducted (finished).

Q61, 63, 67, 70 — The first compilation (same-sex twins born in 1886–1925). During the end of the 1950s, all Swedish parishes were contacted and asked to deliver details of all church book records of same-sex multiple births between 1886 and 1925. The twin births were followed through to 1947 when the system of unique personal numbers for each citizen living in Sweden was introduced. Because the initial aims focused on the health effects of smoking and alcohol, the birth years were restricted to 1886–1925, which generated a suitable age span at risk, of ~35–75 years, at the beginning of the 1960s (Cederlöf, 1966). All identified twins received the so-called ‘green questionnaire’ in 1960, but unfortunately only the answers from pairs in which both twins responded were saved ($N = 21,990$). The same selection of twins received follow-up questionnaires in 1963 (red questionnaire) and 1967 (blue questionnaire). In 1970, a complementary questionnaire was sent to twins who did not respond or who had missing information about alcohol or tobacco in 1967. The requested information in all these questionnaires was about similarity (for zygosity), demographic, medical and lifestyle aspects (Lichtenstein et al., 2002).

Q73 — The second compilation (same-sex twins born in 1926–1958). A new twin cohort was compiled in 1973 by using national registrations of births from 1926 to 1967. Identified same-sex twin-pairs born in 1926–1958 in which both were alive and living in Sweden were sent a questionnaire with content similar to that given to the older cohorts. Individual responses were obtained from 36,536 participants (Lichtenstein et al., 2002).

Author for correspondence: Patrik K. E. Magnusson, Email: patrik.magnusson@ki.se
Cite this article: Zagai U, Lichtenstein P, Pedersen NL, and Magnusson PKE. (2019) The Swedish Twin Registry: Content and Management as a Research Infrastructure. *Twin Research and Human Genetics* 22: 672–680, <https://doi.org/10.1017/thg.2019.99>

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<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 unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Table 1. Questionnaire or interview data in nationwide STR base-studies

Acronym	Years	Birth cohort	Focus	Selection	N (Individuals)
Q 61, 63, 70	1960–1970	1886–1925	Health and exposures	Same-sex, complete pairs	21,890
Q 73	1973	1926–1958	Health and exposures	Same sex	36,536
SALT	1998–2002	1944–1958	Health older adults	No	45,000
TwinGene	2004–2008	<1943	CVD, RA, Diabetes	SALT participants	14,598
SALTY	2009	<1958	Economic behavior, health	SALT participants	11,743
STAGE	2004–2006	1959–1985	Health in adulthood	No	25,484
YATSS	2013	1985–1992	Health in young adulthood	No	6870
CATSS	2004–	>1992	Neurodevelopment, asthma, allergy	No	32,570

Screening Across the Lifespan Twin Study (SALT) — all twins born 1958 or earlier. A large nationwide effort to contact all available twins, including for the first time, opposite-sex pairs, was conducted between 1998 and 2002 (Lichtenstein et al., 2002; Pedersen et al., 2002). The project involved extensive data collection conducted over the telephone by trained interviewers. If self-response was not possible or if the twin performed poorly on a short mental status screen conducted on all twins 65 years old or older, informants such as spouse or relatives were interviewed. The data collection consisted of similarity questions used for the zygosity algorithm and questions about the amount of contact with co-twin, demographics (including occupation and education), checklist of common diseases, use of prescribed and nonprescribed medications, health behaviors (physical activity, alcohol and tobacco consumption) and permission to collect medical records. By using branching with detailed follow-up questions to participants who responded positively to initial key items, the specificity of disease screening was improved for a number of disorders. The selection of items for each disease domain was done in collaboration with experts in each field, with an intended aim to use recognized and established instruments whenever available. When the data collection for the SALT study was ended in 2002, 44,919 twins had participated.

Study of Twin Adults: Genes and Environment (STAGE), all twins born in 1958–1985. During 2005 and 2006, the STR contacted twins born in Sweden 1958–1985 with an invitation letter to respond to a web-based survey (this was the first attempt to collect data by a web-based interphase; Furberg, Lichtenstein, Pedersen, Thornton et al., 2008; Lichtenstein et al., 2006). The sample was restricted to twins whose co-twin survived until at least 1 year of age. The letter contained personal log-in details to the survey, containing questions about a wide variety of common complex health problems as well as exposures (including detailed assessment of tobacco use and quitting attempts). In total, contacts with invitations have been made with 42,582 (both same- and opposite-sex) twin individuals. For twins, so choosing, the questionnaire could also be completed as a telephone interview with a trained interviewer. The purpose was to screen for most common complex diseases and to measure exposures relevant during young adulthood and midlife. The final response rate from combining web and telephone answers was close to 60%.

TwinGene (SALT participants). Between 2004 and 2008, twin-pairs who had participated in SALT were invited to respond to a paper health questionnaire about common diseases and to

provide a blood sample and have their blood pressure, height and weight measured at their local healthcare facility. Out of 22,000 twins invited, close to 14,600 (65%) responded to the questionnaire. Cardiovascular diseases were covered with questions about the occurrence and year of onset for angina pectoris, coronary infarct, hypertension, high cholesterol or high triglycerides, claudication, venous thrombosis, stroke or blood clot and transient ischemic attacks. Cardiac bypass or angioplastic surgery and heart medications were also recorded. Questions about diabetes included type of diabetes and diabetes medication and whether (and when) or not a medical doctor had established the diagnosis. Further, questions about pain, migraine or recurring headache, knee or hip arthrosis and rheumatoid arthritis were asked. For rheumatoid arthritis, follow-up questions included whether or not a medical doctor had established the diagnosis, if the subject went to regular check-ups because of the condition and whether medication for arthritis besides pain remedy was used (Magnusson et al., 2013; Rahman et al., 2009).

SALTY (extended questionnaire to younger subjects in SALT). The SALTY study was a collaborative effort between STR and researchers in epidemiology, medicine, political science and economics. The target population was the younger part of the SALT cohort born between 1943 and 1958. The data collection began in the fall of 2008 by sending an extensive questionnaire to 24,914 Swedish twins, and it was completed in the summer of 2010. The survey generated a total of 11,647 responses (47%). Out of these, 11,482 (98.6%) respondents gave informed consent to have their responses stored and analyzed. The data collection consisted of three parts: (1) an extensive self-report paper questionnaire; (2) saliva collection for DNA extraction (see under biobanking) and (3) a request to participate in a web-based investigation that included questions on musical experience, tendency to experience psychological flow and creative achievement, as well as tests of cognitive and motor performance. A total of 3070 twins aged between 50 and 67 (mean 58.9) years chose to participate in that web-based extension of the SALTY study (Mosing et al., 2012).

Young Adult Twins in Sweden Study (YATSS) twins born in 1986–1992. The (YATSS) survey was conducted in 2013, when invitations were sent out to twin-pairs born in May 1986–June 1992. Along with the invitation came information about the study, a personal identification code and log-in details for the survey website. Twins who did not respond to the survey received one reminder via telephone. To ensure that twins who did not have access to the internet could partake in the survey, the option of

participating in a telephone version of the survey was also offered. Twins who participated received either a cinema ticket or a gift voucher worth 120 SEK (15 USD). Respondents also had the chance to win a tablet computer. Out of 16,244 invited, 6870 participated (42%). The survey was similar to the Study of Twin Adults: Genes and Environment (STAGE) and contained items evaluating exposures and common complex health problems and diseases such as gastrointestinal, arthritis, asthma, allergies, affective disorders, obsessive-compulsive disorder, hoarding, eating disorders, fibromyalgia and chronic fatigue syndrome (Ivanov et al., 2017).

The Child and Adolescent Twin Study in Sweden (CATSS). Since 2004, the STR has continuously contacted the parents (or other guardians) of Swedish-born twin children by the time the twins turn 9 years of age. The purpose of the contacts has been dual, both to include new twins into the STR and as a part of a research program focusing on the development of common health and behavioral problems during childhood and adolescence. During the first three collection years (2004–2007), 12-year-old twins were also approached, meaning that all twins born since July 1, 1992 have been targeted (Anckarsater et al., 2011). The interview contains questions about social environment and somatic health problems in general, as well as a detailed screening instrument for all major clinical diagnostic criteria in child and adolescent psychiatry called the ‘Autism — Tics, ADHD and other Comorbidities inventory’ (Hallerod et al., 2010; Larson et al., 2013).

By May 2019, 16,476 parental interviews concerning 32,952 twins had been completed, with an overall response rate of 69%. All twins not opting out of further contacts are recontacted at age 15 (twins and parents), 18 (twins and parents) and 24 (twins only). Thus, the twins get the opportunity to join or ask to not be contacted again when they become legal adult citizens. The research program about mental health development during childhood and adolescence is the largest and most extensive external effort based on STR and has its own steering group and management. Data collected in the CATSS study are primarily part of the research program of CATSS and not incorporated into the core STR database until used by the CATSS principal investigators.

Biobanking

Chromosomal DNA constitutes the core biomaterial in the STR biobank. Venous blood is the preferred source for most purposes (due to amount, homogeneity and possibilities to standardize/robotize handling). However, venous blood from venipuncture demands the assistance of healthcare professionals. Among the elderly already retired participants of STR, it was feasible to ask them to book a visit with the local healthcare provider, while this was harder for the younger twins in education or the active workforce.

Two alternative ways of collecting chromosomal DNA through self-sampling were investigated in 2007. Two hundred twins were randomized to be sent either kits for capillary blood fingertip puncture (Sarstedt, Microvette 200 µl) or saliva collection kits (Oragene). The response rates were 48% for capillary blood and 56% for saliva. Heterogeneity in terms of amount was extensive for capillary blood. The results indicated saliva as the best alternative.

Blood-based DNA and serum. In 2004, TwinGene was initiated to undertake clinical baseline measurements and to collect blood for DNA and serum biobanking among older Swedish-born twins who had participated in the SALT study (see above). Additional

inclusion criteria were that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study if they previously had declined participation in future studies or if they had been enrolled in other smaller DNA sampling projects linked to the STR. Our logistic capacity of sample handling allowed a rate of approximately 200 invitation letters to be sent per month. As the project was ongoing for several years, we chose to conduct the sampling in an age-ordered manner, where the oldest twins were approached first (to minimize the risk of death before being invited). The response rate was high for twins in the nonworking age groups; however, as we entered birth years still in the workforce (i.e., 65 or younger), the response rate dropped dramatically. We believe this most likely reflects how few work-active persons are willing to take time off in order to go to a local healthcare provider for blood sampling. Therefore, systematic invitations were only sent to the SALT participants born in 1943 or earlier, and the sample collection closed in 2008, with 22,000 twins contacted and an overall response rate of 56%. Along with the invitations and study information came consent forms and a health questionnaire. Materials for blood sampling were sent to the participants after signed consent forms had been returned. Consenting participants were asked to make an appointment at their local healthcare facility Monday to Thursday mornings and not the day before a national holiday to ensure that the sample would reach the KI biobank the following morning by overnight mail. The subjects were instructed to fast from 8 pm the previous evening. By venipuncture, a total of 50 ml of blood was drawn from each subject. Blood for DNA extraction was collected in a 10-ml EDTA tube, while blood for serum storage was collected in three 10-ml gel tubes, inverted five times immediately, followed by 30 min for coagulation in room temperature and thereafter centrifugation. Tubes with serum and blood for biobanking as well as for clinical chemistry tests were sent to Karolinska Institutet by mail. After arrival to the KI biobank, the serum was stored in liquid nitrogen. One 7-ml EDTA tube of whole blood was stored at -80°C , while the DNA in a second 7-ml EDTA tube of blood was extracted using a Puregene extraction kit (Gentra systems, Minneapolis, MN, USA). The DNA was thereafter stored at -20°C .

Saliva DNA. The largest part of the STR biobank consists of DNA samples extracted from saliva. Since the collection started in 2009, saliva from 43,000 participants has been collected. We have consistently used the same provider of kits (Oragene OG-500, DNA Genotek Inc., Canada) and utilized a custom barcoding of both the tubes and the containers. Kits and requests for saliva donations were sent to the participants together with the paper questionnaire in SALT, while for STAGE, YATSS and CATSS, the kits have been sent after completion of a web questionnaire or phone interview. The amount of saliva is 2 ml, and the kits have automatic dispensing of pre-extraction preservation buffer, keeping the sample stable also at room temperature. Donated samples were transported back to Karolinska Institutet by mail.

All DNA extractions from saliva have been performed at the KI biobank using the automatic systems of Puregene (Gentra system) or Chemagic (STAR instrument, Hamilton Robotics). Due to large variation in non-host DNA (e.g., bacteria, virus and fungi) content between the samples, no normalization of the concentrations has been performed.

In-person health check-up. When TwinGene participants visited their local healthcare facility for fasting blood sampling, a healthcare professional also performed a simple health check-up: the

subjects were asked to rest for 5 min in a sitting position before the systolic and diastolic blood pressures were measured twice. Without shoes and in light clothes, the subjects' weight, height, hip and waist circumference were recorded. Data are available for a similar number of participants as for serum ($N = 12,500$).

Blood biochemistry. Clinical blood chemistry assessments were performed directly after the sample collection in the TwinGene project. The Karolinska University Laboratory performed the analyses using standard protocols for the following biomarkers: total cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, C-reactive protein, glucose, apolipoprotein A1, apolipoprotein B, hemoglobin and hemoglobin A1C (Rahman et al., 2009). All laboratory results were evaluated by a specialist in cardiology and internal medicine, together with the information from the health questionnaire. Within 3 weeks, the participating twins were informed about the results and clinical recommendations for a follow-up at their local healthcare center were given when judged motivated.

Serum in TwinGene as an open resource for new measurements. The frozen serum samples gathered within TwinGene project create, together with the longitudinal phenotypic information, a platform well equipped to investigate the genetic influences on a broad spectrum of health-related traits and common diseases. An average of five 900- μ l master tubes is available for each participating twin. Aliquots from single master tubes from the whole collection of participants have been derived in two rounds thus far. The first time, 100- μ l aliquots in 96-well foil-sealed PCR plate formats were created. This is a well-functioning layout for applications when all samples are to be analyzed, which has been done for IgA level, cystatin C, kreatinin and anticitrullinated protein antibodies. However, it prohibits use in cherry-picking projects as the whole plate has to be thawed even if just one or a few samples in the plate are targeted. The second master-tube aliquotation was done in 2018, relying on individual 2D bar-coded microtubes (100 μ l). This system allows removal and thawing of individual aliquots of cherry-picked samples.

Genotyping

Before the amazing development of genomewide genotyping methods over the last decade, more rudimentary and smaller scale methods have been applied in projects focusing on candidate gene-association studies. As history tells, these were rarely successful. Furthermore, when candidate genotyping was used for assignment of zygosity, it could result in considerable problems. At the time, an easily done mistake was to interpret any twin-pair genotype mismatch as valid evidence for dizygosity. By now, it is quite obvious that the use of genotypic data for the purpose of determining zygosity relies on a sound appreciation of the genotyping error rate. If only a few markers are available and a zero genotyping error is assumed, it will typically result in increased false dizygotic (DZ) assignments. Genotyping dedicated for zygosity determination accompanied by an algorithm taking laboratory error into account was introduced into the STR in 2007 (Hannelius et al., 2007). The properties of this test are excellent, and thus far the method has been run on close to 9600 pairs. Even though it demands fairly large batches to be cost efficient (resulting in rather long waiting times for twins interested in the results), it has been shown to be better than most genomewide association study platforms in this respect.

Genomewide single nucleotide polymorphisms array genotyping. The vast majority of all STR participants who have donated DNA samples of adequate amount and quality have now been genotyped with a genomewide single nucleotide polymorphisms (SNPs) array. All genotyping has been performed by SNP& SEQ Technologies Uppsala, Sweden, using Illumina arrays. For monozygotic (MZ) twins, imputation based on one randomly selected member of each pair has been implemented. Sample size, versions of arrays used and other details are provided below for the various substudies. As the STR is a dynamic register from which the participants may withdraw their participation, the numbers may change, and among the imputed MZ twins, only those who provided consent will be possible to link to phenotypic information.

TwinGene. First, in 2007, and as an integrated part of the GenomEUtwin collaboration (Surakka et al., 2012), 301 female MZ TwinGene participants were genotyped using the Infinium II assay on the 317K HumanHap300-Duo Genotyping BeadChips (Illumina Inc., San Diego, CA, USA). The signal intensity data were converted into genotypes using Illumina Beadstudio 2.0 software. Second, after exclusion of these already genotyped MZ females and redundant MZ co-twins, DNA from 9896 individual subjects was genotyped in 2009 and 2010 with the 700K Illumina OmniExpress BeadChip. Genotyping results for 9835 subjects passed the initial laboratory-based quality control. Genotypes for paired MZ co-twins to genotyped individuals were imputed for 1111 pairs, making the full dataset to consist of genotypes for 10,946 individual twins.

SALTY, CATSS (first round). Genotyping was performed in 2014–2015 using the Illumina 550K PsychArray BeadChip, in 18 batches. Content for the PsychArray includes 265,000 proven tag SNPs found on the Infinium Core-24 BeadChip, 245,000 markers from the Infinium Exome-24 BeadChip and 50,000 additional markers associated with common psychiatric disorders. Calling of genotypes was initially done at the genotyping center using Illumina GenCall algorithm. As it is well known that rare variants are poorly called using GenCall, a rare variant calling algorithm, zCall, was applied in order to increase sensitivity for heterozygotes of rare variants. Genotypes of rare variants from zCall were subsequently integrated with genotypes of common variants from GenCall. Samples from 17,898 participants passed quality controls. Imputed genotypes for 3854 MZ co-twins inferred from their paired genotyped twin made the full dataset to consist of 21,752 individuals.

STAGE, YATSS, CATSS (second round). In 2017–2018, the two remaining larger study collections (STAGE and YATSS) were genotyped with the 650K Illumina Global Screening Array (GSA) BeadChip. Also, new samples from the CATSS study collected as the latest genotyping with PsychArray were included in the order. Delivery of the last batch of results occurred in May 2019. After an initial quality control of excluding all samples with call rates below 98%, there were 8468 from STAGE, 2742 from YATSS and 4064 from CATSS, giving a total of 15,274 genotyped genomes. Among these, there are 4399 MZ co-twins whose genotypes can be imputed based on the genotypes of their MZ co-twin, making the total number of twins genotyped with the GSA 19,673 individuals.

SweGen whole genome sequencing. In 2014, there was a national call for a suitable sample collection that could constitute a reference of Swedish genomes for whole genome sequencing aiming to make available a population-based genetic variant dataset. Among the potential sample collections, TwinGene was considered to reflect the best Swedish population geographically. The project denoted SweGen was based on Illumina HiSeq X sequencing and was finished in 2016. The resulting variant frequencies are available through a search engine open to researchers and clinicians (<https://swefreq.nbis.se/dataset/SweGen>). Average autosomal read-depth coverage was 37×. The number of single-nucleotide variants and indels identified were 29.2 million and 3.8 million, respectively. Of all the variants, 9.9 million were not present in other current databases. Each sample contributed with an average of 7199 individual-specific variants. Comparative analyses showed that the genetic diversity within Sweden appears substantial in relation to continental European populations, underscoring the relevance of establishing a local reference data set (Ameur et al., 2017). Almost all (94%) of the samples included in the SweGen project come from a random selection of TwinGene twin-pairs (one twin per selected pair). Thus, no disease information has been used for the selection. Individual positions in the genome can be viewed using the Graphical Browser. A variant frequency file may be downloaded upon registration.

National Register Linkages

As in other countries in northern Europe, Sweden has a long history of using personal identity numbers for all citizens. The number is linked to the home address, used for keeping monthly updated information about the residence of all twins. Furthermore, it allows linkages to the national health registers. Together with the government-funded healthcare system, these registers enable collection and use of data on health issues such as diagnoses, drug prescriptions and surgical procedures, with good nationwide coverage (Ludvigsson et al., 2011). The STR holds such medical information through regular updates with the Patient Register, containing in-patient and out-patient (from nonprivate specialized care) diagnosis codes, as well as the Medical Birth, Prescribed Drug, Cancer and the Causes of Death Registers.

STR holds health register data in two versions, one where the key to the personal number is available and one where data are anonymized and the key has been destroyed so the data are no longer considered personal data. For all twins who have actively participated in at least one of the STR studies and not asked to be withdrawn, updated health register data from the National Board of Health and Welfare are linked every 2–3 years. These data can then be analyzed against other data in the STR, including existing and new measurements from biological samples in the biobank. For twin studies solely aimed at variance decomposition from the pairwise distribution of particular diagnoses (ICD-codes), prescriptions or surgical procedures for MZ and DZ twins, the anonymized version may suffice. Worth noting is that as the health register-based information is independent of whether the twins have chosen to participate in an STR study or not, they are also free of participation bias. However, it is important to be aware that the zygosity is given (no matter whether they have participated in a study or not) for all opposite-sex DZ twins, while for same-sex twins, the zygosity information requires active participation. This difference in selection among twins with information about zygosity needs to be handled when opposite-sex pairs are included in the model.

Today, the Medical Birth Registry, held by the Swedish Board of Health and Welfare, constitutes the source of information of twin births in Sweden. Information about new twins is requested in batches, approximately every 7 years. The STR contains personal data solely on the twins themselves, meaning that name, personal identity numbers, addresses, register linkages, questionnaire and interview data are available only for twins and not for parents, siblings or other relatives. Thus, despite the fact that initial consent and baseline information about health and behavior are provided by the guardians, the data itself constitute personal data on the twins. For research questions that rely on data on other relatives to the twins (e.g., parents, offspring and siblings), it may be possible to retrieve such information from *de novo* linkage to the national multigeneration register (MGR, Statistics Sweden). In such instances, the data will become anonymized, meaning that the key to personal data is destroyed so the dataset will not be possible to update with additional personal data.

Developments

The STR mission to facilitate research has to be balanced against the potential risk of harm to participants. Each study conducted that utilizes the content of STR has to be subjected to ethical vetting. When twins are contacted, they are informed about that participation is completely voluntary.

Requests to Quit Participation

According to both Swedish law as well as the recent EU General Data Protection Regulation, the participants in research have the right to quit their participation at any time and say no to future contacts. The overall proportion of STR participants who have requested this is around 1%. The rate, however, varies substantially over birth year (Figure 1). Abrupt shifts over birth year correspond well to the various studies, twins of different birth-year intervals have been asked to participate in. The associations probably reflect a combination of trends in attitude over secular time (i.e., when the studies were conducted), birth cohort differences, as well as variation between studies in how demanding and/or provoking the content has been to the invited twins.

Participation in CATSS

For the past 15 years, invitation to CATSS and the STR has been performed through a parental telephone interview around the time the twins turn 9 years of age (Anckarsater et al., 2011). The CATSS study has been highly productive during these years and is now becoming mature and saturated with some of the phenotypic data it has focused on collecting. Furthermore, a trend of reduced participation has been noticed (Figure 2). Despite the inherent lag in response due to the attempts to contact twins' parents or guardians, which are continued for up to 2 years after the twins' ninth birthday, the reduction during the latest years are worrying.

However, we have noticed that it has become increasingly difficult to get hold of the parents of twins by telephone for the interview, as the habit of not answering calls from unknown numbers has become a common phenomenon (to avoid calls from telemarketing salespersons). In addition, the length of the interview (1 h) may be part of the problem. STR has, therefore, decided to change the mode of data collection to a web-based questionnaire with a somewhat altered and reduced content. The plan is to have implemented these changes by 2020.

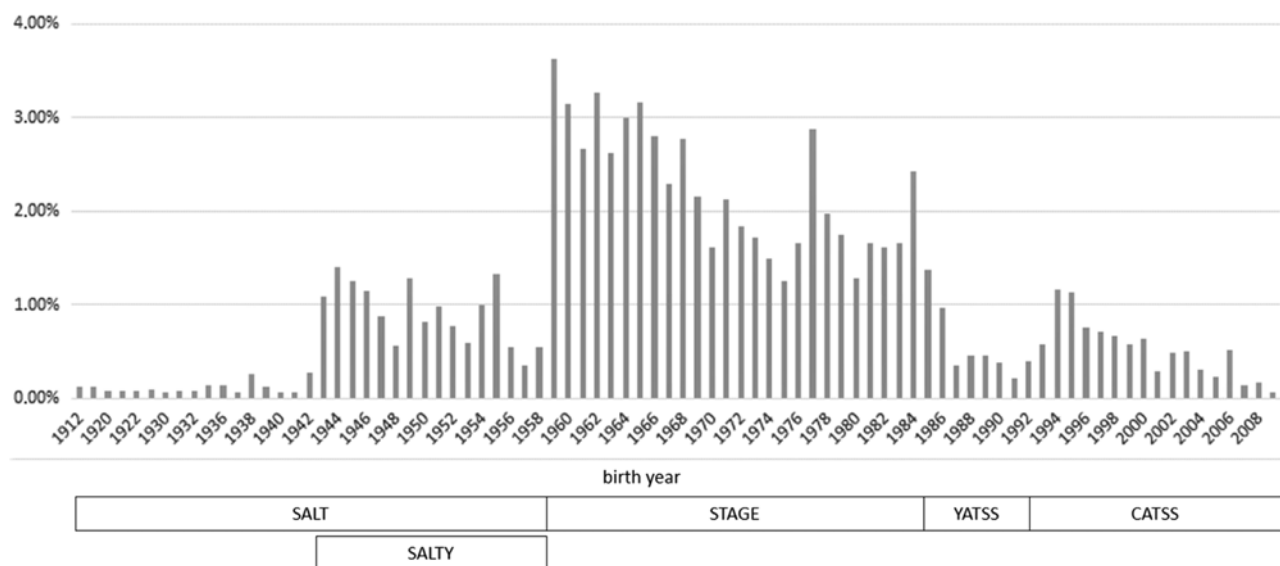


Fig. 1. The proportion of twins demanding to quit their participation in STR. Target birth cohort intervals of the main studies are indicated by the horizontal bars.

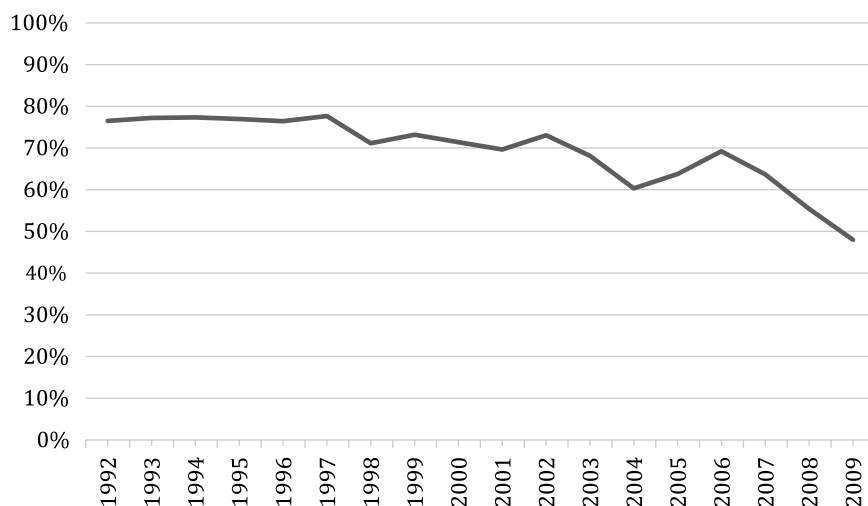


Fig. 2. Participation rate over birth year in CATSS.

Should STR Move to Invitation at Birth Instead of at Age 9?

One idea currently investigated is to invite mothers detected pregnant with twins during routine ultrasound. Such a system could combine the invitation and consent with a free DNA-based zygosity test for which the biological sample is taken directly at the time of delivery. It would open up new research questions regarding healthy development in utero and the first years of life. This time period is of particular interest for twins as birth complications, premature births and small size for gestation are increased.

Through our contacts with twins and parents of twins, it has become obvious that uncertainty is common and they often live with a mistaken belief about the true zygosity. In addition, it is also clear that information about zygosity is of important medical and psychological value to the twins and their families. Further, rates and types of complications after birth differ substantially between MZ and DZ twins (West et al., 1999). The rates of MZ and DZ twins change with in-vitro fertilization (IVF), and they also differ

substantially between countries and ethnic groups (Smits & Monden, 2011).

Recent trends in utilization of IVF and in immigration from various parts of the world to Sweden we argue speak for a monitoring of zygosity rates and, together with the value to the individual twin-pair, we argue an addition of DNA-based zygosity to the standard battery of medical work-up of multiple births is warranted.

STR is planning a pilot study in which the mode and timing of the invitation, data and biosample collection (questionnaire, registers, blood from umbilical cord) of newborn twins will be evaluated. Although routine DNA-based zygosity would be of great value for both the twins themselves and their families, it would be of large importance to the research community. In particular, if zygosity would be incorporated into the Swedish multi-generation register (Statistics Sweden), it would open up for new opportunities for twin model-based research on virtually any register-based data accessible through the vast Swedish health and social registers.

Table 2. Examples of influential findings based on data from the Swedish Twin Registry over the decades.

Decade	Driving researchers/ Directors	Focus	Finding	Article
1960s, 1970s*	Lars Friberg, Rune Cederlöf	Cardiovascular disease — environmental factors	Within twin pair results consistent with causal effects of alcohol	Hrubec et al. (1976)
1980s	Rune Cederlöf, Birgitta Floderus, Anders Ahlbom	Personality	Heritability of personality	Floderus-Myrhed et al. (1980)
		Blood biomarkers	Heritability of blood cholesterol levels	Dahlén et al. (1983)
		Smoking — mortality	Smoking — mortality association holds within twin-pairs	Floderus et al. (1988)
1990s	Anders Ahlbom, Nancy Pedersen	Radiation — cancer	Results consistent with in utero X-rays as a riskfactor for childhood cancer	Rodvall et al. (1990)
		Bereavement — depression	Anticipation effect — increased depression not only after but also before death of spouse	Lichtenstein et al. (1996)
		IQ	High heritability for IQ late in life	McClearn et al. (1997)
		MZ twinning	No paternal effect on MZ twinning	Lichtenstein et al. (1998)
2000s	Nancy Pedersen, Paul Lichtenstein	In utero growth	Heritability of birth weight and gestational length	Clausson et al. (2000)
		Cancer	Low heritability of most forms of cancers	Lichtenstein et al. (2000)
		Food — prostate cancer	Association between fatty fish consumption and reduced risk of prostate cancer	Terry et al. (2001)
		Alzheimer disease	High heritability, same genes in males and females	Gatz et al. (2006)
		Anorexia nervosa (AN) — neuroticism	High heritability of AN, neuroticism predicts development of AN	Bulik et al. (2006)
		Major depression	Consistent effects in birth-cohorts spanning over six decades, higher heritability in women.	Kendler et al. (2006)
		Snus — smoking cessation	Snus use — the strongest independent correlate of smoking cessation	Furberg, Lichtenstein, Pedersen, Bulik et al. (2008)
2010s	Paul Lichtenstein, Patrik Magnusson	Blood lipids — GWAS	Identification of 95 loci for blood lipids levels	Teslovich et al. (2010)
		Childhood neuropsychiatric disorders	Different neuropsychiatric disorders have a common genetic etiology	Lichtenstein et al. (2010)
		Autism spectrum disorders (ASD)	Etiological similarity between ASD and autism like traits in the general population	Lundström et al. (2012).
		Educational attainment — GWAS	First successful GWAS for a behavioral trait with associations that hold within DZ pairs. Strong overlap with IQ demonstrated in the STR sample.	Rietveld et al. (2013)
		Autism	Increased autism prevalence more due to administrative rather than etiological factors	Lundström et al. (2015)

Note: *Studies were mainly published as monographs during the 1960s and 1970s, and are not included here. GWAS = genome-wide association study.

Management

The STR has, during the six decades it has been operational, received financial support for running the core functions channeled through the budget of Karolinska Institutet, which is one of the universities supported by the Swedish government. Additional external funds from both industry and research project grants have been critical to boost data collections in specific cohorts. A part of the operating costs for STR is levied through fees for access to the resource (see below).

In 2018, the STR received, for the first time, funding from the Swedish Research Council for the specific purpose of constituting a national research infrastructure (VR-2017-00641). This grant runs

over 5 years until 2022 and covers up to 50% of the core budget during these years. Besides Karolinska Institutet (the host), the universities of Lund, Gothenburg, Jönköping, Linköping, Örebro, Uppsala, Umeå and the Stockholm School of Economics signed a final agreement of collaboration in 2018 and share interest in this national infrastructure.

Access to the STR

STR is open to applications from Swedish and international researchers. A steering committee meets four times per year to review these applications and make decisions about which projects will be implemented. Approval requires a project description that

indicates sound scientific methodology and that twins are not contacted unnecessarily. Access to data and resources further requires an ethical approval from the local ethics review board and that the charges to STR are paid. For non-Swedish applicants, collaboration with a Swedish university is required. The current tariffs can be found at the STR homepage <http://ki.se/en/research/the-swedish-twin-registry>.

External Projects Based on New Contacts with Twins

After successful application to the STR steering committee, it is possible to obtain contact information (postal addresses) to twins for invitations to new studies. This opportunity has in recent years been used for external research projects about rheumatoid arthritis, immunoglobulin deficiencies, anxiety, inflammatory bowel disease, autism, obsessive-compulsive disorder, Crohn's, Parkinson's and schizophrenia. In all studies involving disease-based ascertainment, the STR demands that healthy pairs are also included in the recruitment, as otherwise the pair aspect inherent in the twin design would reveal medical information about co-twins. The inclusion of healthy pairs should, therefore, also be mentioned in the information to the participants.

The Basic Operations of the STR Infrastructure and the Research Conducted by Using It — Distinctions and Mutual Dependence

The long-term investment in the STR from Karolinska Institutet has made it possible to establish a sustainable and steady inclusion of new twins over time as well as a transparent system for providing access to data for researchers. Still, significant parts of the available data content of the STR have been collected in research projects (financed through external grants) that have deposited the data collected by these researcher-initiated projects into the register. The STR continues to promote and encourage this approach, as it is an efficient way of building research value by creating new possibilities to investigate measurements in combinations not otherwise available.

Remarks

For six decades, the STR has been an important resource for the study of genetic and environmental aspects of a broad spectrum of traits, phenotypes and disorders. A selection of examples of influential findings over the years is presented in [Table 2](#).

STR is built around the core functions of (1) inclusion and questionnaire data collection of new twins at a steady pace, (2) DNA collection enabling genomewide genotyping and zygosity determination, (3) linkages to national registers and (4) a transparent system for providing researchers with access to data as well as possibilities to contact twins for new data collection. The core functions of STR constitute in that sense a basic machinery which enables research projects to be initiated by a multitude of researchers aiming to answer a multitude of research questions. STR continues to be one of the richest twin resources worldwide in several aspects (van Dongen et al., 2012).

Acknowledgments. SNP&SEQ Technology Platform in Uppsala (www.genotyping.se) has performed the genotyping. The facility is part of the National Genomics Infrastructure supported by the Swedish Research Council for Infrastructures and Science for Life Laboratory, Sweden. The SNP&SEQ Technology Platform is also supported by the Knut and Alice Wallenberg Foundation. The Swedish Twin Registry is managed by

Karolinska Institutet as a core facility and receives funding through the Swedish Research Council under the grant no. 2017-00641.

References

- Ameur, A., Dahlberg, J., Olason, P., Vezzi, F., Karlsson, R., Martin, M., ... Gyllenstein, U. (2017). SweGen: A whole-genome data resource of genetic variability in a cross-section of the Swedish population. *European Journal of Human Genetics*, 25, 1253–1260.
- Anckarsater, H., Lundstrom, S., Kollberg, L., Kerekes, N., Palm, C., Carlstrom, E., ... Lichtenstein, P. (2011). The Child and Adolescent Twin Study in Sweden (CATSS). *Twin Research and Human Genetics*, 14, 495–508.
- Bulik, C. M., Sullivan, P. F., Tozzi, F., Furberg, H., Lichtenstein, P., & Pedersen, N. L. (2006). Prevalence, heritability, and prospective risk factors for anorexia nervosa. *Archives of General Psychiatry*, 63, 305–312.
- Cederlöf, R. (1966). *The twin method in epidemiological studies on chronic disease*. Dissertation, Karolinska Institutet.
- Clauson, B., Lichtenstein, P., & Cnattingius, S. (2000). Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG*, 107, 375–381.
- Dahlén, G., Ericson, C., de Faire, U., Iselius, L., & Lundman, T. (1983). Genetic and environmental determinants of cholesterol and HDL-cholesterol concentrations in blood. *International Journal of Epidemiology*, 12, 32–35.
- Floderus, B., Cederlöf, R., & Friberg, L. (1988). Smoking and mortality: A 21-year follow-up based on the Swedish Twin Registry. *International Journal of Epidemiology*, 17, 332–340.
- Floderus-Myrhed, B., Pedersen, N., & Rasmuson, I. (1980). Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs. *Behavior Genetics*, 10, 153–162.
- Furberg, H., Lichtenstein, P., Pedersen, N. L., Bulik, C. M., Lerman, C., & Sullivan, P. F. (2008). Snus use and other correlates of smoking cessation in the Swedish Twin Registry. *Psychological Medicine*, 38, 1299–1308.
- 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 & Tobacco Research*, 10, 1727–1735.
- Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., & Pedersen, N. L. (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, 63, 168–174.
- Hallerod, S. L., Larson, T., Stahlberg, O., Carlstrom, E., Gillberg, C., Anckarsater, H., ... Gillberg, C. (2010). The Autism—Tics, AD/HD and other Comorbidities (A-TAC) telephone interview: Convergence with the Child Behavior Checklist (CBCL). *Nordic Journal of Psychiatry*, 64, 218–224.
- Hannelius, U., Gherman, L., Makela, V. V., Lindstedt, A., Zucchelli, M., Lagerberg, C., ... Lindgren, C. M. (2007). Large-scale zygosity testing using single nucleotide polymorphisms. *Twin Research and Human Genetics*, 10, 604–625.
- Hrubec, Z., Cederlöf, R., & Friberg, L. (1976). Background of angina pectoris: Social and environmental factors in relation to smoking. *American Journal of Epidemiology*, 103, 16–29.
- Ivanov, V. Z., Nordsletten, A., Mataix-Cols, D., Serlachius, E., Lichtenstein, P., Lundstrom, S., ... Ruck, C. (2017). Heritability of hoarding symptoms across adolescence and young adulthood: A longitudinal twin study. *PLoS One*, 12, e0179541.
- Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *The American Journal of Psychiatry*, 163, 109–114.
- Larson, T., Lundstrom, S., Nilsson, T., Selinus, E. N., Rastam, M., Lichtenstein, P., ... Kerekes, N. (2013). Predictive properties of the A-TAC inventory when screening for childhood-onset neurodevelopmental problems in a population-based sample. *BMC Psychiatry*, 13, 233.
- Lichtenstein, P., Carlström, E., Råstam, M., Gillberg, C., & Anckarsäter, H. (2010). The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *American Journal of Psychiatry*, 167, 1357–1363.
- 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. *Journal of Internal Medicine*, 252, 184–205.
- Lichtenstein, P., Gatz, M., Pedersen, N. L., Berg, S., & McClearn, G. E.** (1996). A co-twin-control study of response to widowhood. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 51, P279–289.
- Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., & Hemminki, K.** (2000). Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343, 78–85.
- Lichtenstein, P., Kallen, B., & Koster, M.** (1998). No paternal effect on monozygotic twinning in the Swedish Twin Registry. *Twin Research*, 1, 212–215.
- Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlstrom, E., . . . Pedersen, N. L.** (2006). The Swedish Twin Registry in the third millennium: An update. *Twin Research and Human Genetics*, 9, 875–882.
- Ludvigsson, J. F., Andersson, E., Ekblom, A., Feychting, M., Kim, J. L., Reuterwall, C., . . . Olausson, P. O.** (2011). External review and validation of the Swedish national inpatient register. *BMC Public Health*, 11, 450.
- Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., & Lichtenstein, P.** (2012). Autism spectrum disorders and autistic like traits: Similar etiology in the extreme end and the normal variation. *Archives of General Psychiatry*, 69, 46–52.
- Lundström, S., Reichenberg, A., Anckarsäter, H., Lichtenstein, P., & Gillberg, C.** (2015). Autism phenotype versus registered diagnosis in Swedish children: Prevalence trends over 10 years in general population samples. *BMJ*, 350, h1961.
- Magnusson, P. K., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., . . . Lichtenstein, P.** (2013). The Swedish Twin Registry: Establishment of a biobank and other recent developments. *Twin Research and Human Genetics*, 16, 317–329.
- McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R.** (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, 276, 1560–1563.
- Mosing, M. A., Magnusson, P. K. E., Pedersen, N. L., Nakamura, J., Madison, G., & Ullen, F.** (2012). Heritability of proneness for psychological flow experiences. *Personality and Individual Differences*, 53, 699–704.
- Pedersen, N. L., Lichtenstein, P., & Svedberg, P.** (2002). The Swedish Twin Registry in the third millennium. *Twin Research*, 5, 427–432.
- 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 Research and Human Genetics*, 12, 286–294.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., & Koellinger, P. D.** (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340, 1467–1471.
- Rodvall, Y., Pershagen, G., Hrubec, Z., Ahlbom, A., Pedersen, N. L., & Boice, J. D.** (1990). Prenatal X-ray exposure and childhood cancer in Swedish twins. *International Journal of Cancer*, 46, 362–365.
- Smits, J., & Monden, C.** (2011). Twinning across the developing world. *PLoS One*, 6, e25239.
- Surakka, I., Whitfield, J. B., Perola, M., Visscher, P. M., Montgomery, G. W., Falchi, M., . . . Genom, E. P.** (2012). A genome-wide association study of monozygotic twin-pairs suggests a locus related to variability of serum high-density lipoprotein cholesterol. *Twin Research and Human Genetics*, 15, 691–699.
- Terry, P., Lichtenstein, P., Feychting, M., Ahlbom, A., & Wolk, A.** (2001). Fatty fish consumption and risk of prostate cancer. *Lancet*, 357, 1764–1766.
- Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., & Kathiresan, S.** (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466, 707–713.
- 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.
- West, C. R., Adi, Y., & Pharoah, P. O.** (1999). Fetal and infant death in mono- and dizygotic twins in England and Wales 1982–91. *Archives of Disease in Childhood: Fetal & Neonatal*, 80, F217–220.