The Italian Twin Project: From the Personal Identification Number to a National Twin Registry

Maria Antonietta Stazi[1], Rodolfo Cotichini[1], Valeria Patriarca[1], Sonia Brescianini[1], Corrado Fagnani[1], Cristina D'Ippolito[1], Stefania Cannoni[2], Giovanni Ristori[2], and Marco Salvetti[2]

[1] Department of Epidemiology and Biostatistics, Istituto Superiore di Sanità, Rome, Italy
[2] Department of Neurosciences, University of Rome "La Sapienza", Rome, Italy

The unique opportunity given by the "fiscal code", an alphanumeric identification with demographic information on any single person residing in Italy, introduced in 1976 by the Ministry of Finance, allowed a database of all potential Italian twins to be created. This database contains up to now name, surname, date and place of birth and home address of about 1,300,000 "possible twins". Even thought we estimated an excess of 40% of pseudo-twins, this still is the world's largest twin population ever collected. The database of possible twins is currently used in population-based studies on multiple sclerosis, Alzheimer's disease, celiac disease, and type 1 diabetes. A system is currently being developed for linking the database with data from mortality and cancer registries. In 2001, the Italian Government, through the Ministry of Health, financed a broad national research program on twin studies, including the establishment of a national twin registry. Among all the possible twins, a sample of 500,000 individuals are going to be contacted and we expect to enrol around 120,000 real twin pairs in a formal Twin Registry. According to available financial resources, a sub sample of the enrolled population will be asked to donate DNA. A biological bank from twins will be then implemented, guaranteeing information on future etiological questions regarding genetic and modifiable factors for physical impairment and disability, cancers, cardiovascular diseases and other age related chronic illnesses.

Italy has a great tradition in studies on twins. In particular, Luigi Gedda's book, *Studio dei Gemelli*, first published in 1951, represents a milestone in twin research. Moreover, Gedda's pioneering work on twins at the Mendel Institute in Rome, where detailed phenotypic information has been collected on approximately 40,000 volunteer twins since the 1950s, is still a valuable resource for genetic studies on complex traits, despite the lack of data on molecular markers.

In 1996, a database was created to register all possible twins in Italy. Possible twins are identified by means of the "codice fiscale" (fiscal code), a personal identification number (PIN) based on the first and last name and place and date of birth. Specifically, those persons whose PIN indicates that they could have the same last name and date and place of birth are defined as "possible twins". The database is currently used for detecting twins among persons enrolled in population-based studies, with the aim of studying the weight of genetic, environmental, and behavioral factors in the etiopathogenesis of multi-factorial diseases. Although approximately 1,300,000 possible twins have been registered as of 1996, making this the world's largest twin database, Italy still lacks a national population-based twin registry. The data provided by such registries, as observed in northern European countries, can greatly contribute to genetic epidemiological studies, in a way that is not possible with regular family or case-control studies (Lichtenstein et al., 2000).

In consideration of the extreme importance of twin studies in medical research, the Italian Government recently designated a regular fund for a comprehensive national research program on twin studies, which includes the establishment of a national twin registry based on the existing database of possible twins. The present work is a description of the current activities for implementing this registry, with particular emphasis on evaluating the reliability of the data registered in the database of possible twins.

The Database of Possible Twins

As mentioned, the Database of Possible Twins is based on the PIN, which consists of 16 characters, specifically: 3 letters from the surname, 3 from the first name, 2 numbers for the year of birth, 1 letter for the month of birth, 2 numbers for the day of birth and gender (1 to 31 for males, 41 to 71 for females), 1 letter and 3 numbers for the place of birth, and a 16th letter for confirming the accuracy of the entire algorithm. The PIN was first introduced in 1976 and was initially assigned only upon request for fiscal purposes. Since 1989, it has been assigned by default to all new-borns and retrospectively to all residents born since January 1971, and it is now currently used by the National Health Service.

To identify possible twins, in 1997 all records that could be matched for the first three characters (i.e., those based on the last name) and the seventh to 15th characters (those referring to the date and place of birth) were extracted from PIN database. Among these, only individuals with the same

Address for correspondence: Maria Antonietta Stazi, Department of Epidemiology and Biostatistics, Istituto Superiore di Sanità, Viale Regina Elena, 299; 00161 Rome, Italy. Email: stazi@iss.it

exact surname were recorded in the Database of Possible Twins. A total of 1,600,000 possible twins were detected; however (Salvetti et al., 1997) the attribution of two codes to the same individual accounted for an excess of approximately 15%, and the database was thus restricted to 1,300,000 records. To verify the reliability of this database, we conducted several comparisons:

1. We compared the number of possible twins alive in 1995 to the 1995 Italian resident population in terms of the distribution by: year of birth, sex-ratio and geographical area of residence.

Table 1 shows the number of same-sex and opposite-sex possible twin pairs registered in the database by year of birth and geographic area (i.e., northern, central, and southern Italy), as compared to the 1995 Italian general population. The age distribution of possible twins was comparable among male/male, male/female, and female/female twin-pairs. The opposite-sex twin pairs were over-represented (37.5% vs. 31.2% for both male/male and female/female twin pairs, constituting a difference of approximately 20%); this was especially evident in southern Italy. The twin rates calculated on the 1995 Italian population were higher than expected, especially in the older age groups and, again, in southern Italy (3.1% in southern Italy vs. 1.8% and 1.9% in northern and central Italy, respectively).

Figure 1 shows the distribution by year of birth of the twin population and the 1995 Italian population (ISTAT). The two distributions are quite similar: the decrease of births during the First and the Second World Wars and the "baby boom" of the 1960s can both be easily recognised.

Figure 2 shows the male-to-female ratio by year of birth for the twins registered in the Database of Possible Twins and for the 1995 Italian population. For the older age-groups, the number of males seems to be higher in the twin database, compared to the general population.

2. We compared the possible twins to the actual new-born twin pairs recordered by the National Institute of Statistics (ISTAT) in terms of the number of same-sex and opposite-sex twins born between 1981 and 1995.

According to ISTAT, 80,955 twin-pairs were born in this period, compared to 112,384 twin-pairs registered in the database, which represent an excess of 39%. This excess, however, steadily decreased over time, from 49% in 1981 to 12% in 1995. Moreover, the excess was lower for same-sex pairs (26% overall) than for opposite-sex pairs (68% overall).

3. We determined whether or not the possible twins had the same home address, as registered in the database, and analyzed this concordance by year of birth.
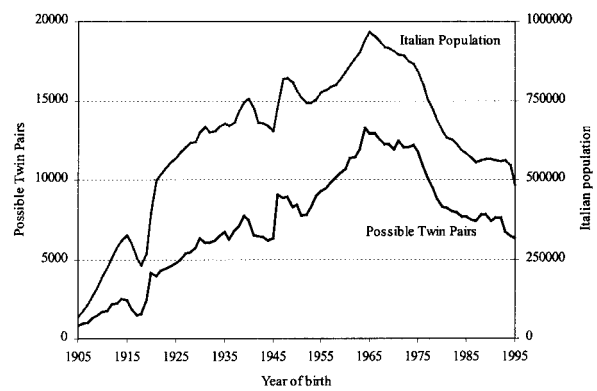

Figure 1

Number of possible twin pairs as compared to the Italian population (ISTAT, 1995) by year of birth.

Table 1

Distribution of Possible Twin Pairs According to Sex-Pair, by Year and Geographical Area of Birth, 1900–1995

Year of birth	Possible Twin Pairs								1995 Italian Population		Twin rate
	Male Male		Male Female		Female Female		Total				
	n	%	n	%	n	%	n	%	n	%	%
1900–1914	6584	3.2	5205	2.1	8395	4.1	20,184	3.0	201,765,2	3.5	2.0
1915–1924	10,048	4.8	10,481	4.2	11,156	5.4	31,685	4.8	388,551,1	6.8	1.6
1925–1934	18,757	9.0	21,053	8.5	18,035	8.7	57,845	8.7	627,300,4	10.9	1.8
1935–1944	21,511	10.4	25,969	10.4	20,418	9.9	67,898	10.2	699,294,5	12.2	1.9
1945–1954	25,058	12.1	32,762	13.2	25,012	12.1	82,832	12.5	759,748,1	13.3	2.2
1955–1964	32,597	15.7	42,566	17.1	32,631	15.7	107,794	16.2	839,514,6	14.7	2.6
1965–1974	37,604	18.1	48,433	19.5	37,479	18.1	12,3516	18.6	912,012,6	15.9	2.7
1975–1984	29,038	14.0	33,597	13.5	28,566	13.8	91,201	13.7	691,971,3	12.1	2.6
1985–1994	23,920	11.5	26,593	10.7	23,588	11.4	74,101	11.2	561,624,9	9.8	2.6
1995	2145	1.0	2234	0.9	1979	1.0	6358	1.0	482,953	0.8	2.6
Geographical area of birth											
Northern	75,713	36.9	78,817	31.7	76,821	37.2	231,351	35.0	254,430,09	44.4	1.8
Central	33,418	16.3	39,527	15.9	34,003	16.5	106,948	16.2	109,886,00	19.2	1.9
Southern	96,246	46.9	130,277	52.4	95,787	46.4	322,310	48.8	208,691,71	36.4	3.1
Total	207,262	100.0	248,893	100.0	207,259	100.0	663,414	100.0	573,007,80	100.0	2.3
%	31.2		37.5		31.2		100.0				

M.A. Stazi, R. Cotichini, V. Patriarca, S. Brescianini, C. Fagnani, C. D'Ippolito, S. Cannoni, G. Ristori, and M. Salvetti
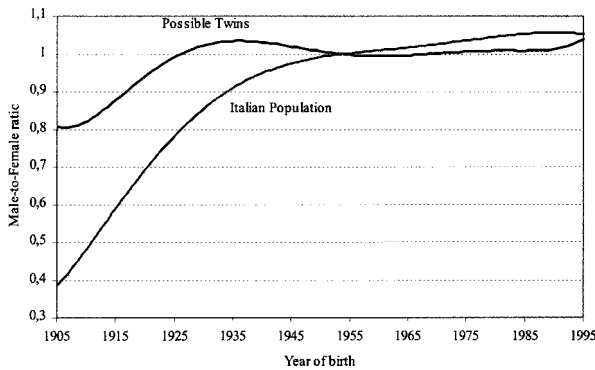
Figure 2

Male-to-female ratio in the possible twins and in the Italian population (ISTAT, 1995) by year of birth.

The distribution of possible twins according to "concordance" of home addresses within the pair as exactly digitized in the original Ministry of Finance database is shown on figure 3. Assuming that twin children are more likely to live in the same home, we could estimate the proportion of "pseudo" twins in the youngest. In figure 3 is also reported the number of new-born twin pairs from ISTAT for 1981–1995. The shape of that curve is highly similar to that of our possible twins with the concordant address.

4. We compared the age and gender distribution of the possible twins residing in the city of Rome to those registered in the Twin Epidemiological Registry of Rome (TERRY), maintained by the University of Sport and Movement Sciences (IUSM) of Rome.

The comparison between a sample of the data set (i.e., possible twins both resident in the city of Rome) and the Twin Epidemiological Registry of Rome, which tracks, starting from the Rome Civil Registration System, all pairs with both members resident in the city, and was considered the gold standard, evidenced an overall excess of twin pairs in our data, localized in the elderly.
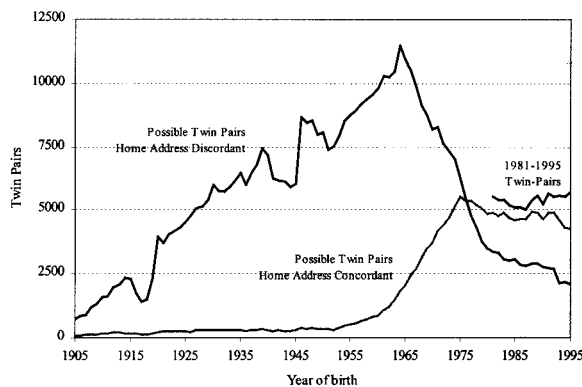

Figure 3

Number of possible twin pairs according to home-address concordance and number of new-born twin pairs (ISTAT, 1981–1995) by year of birth.

All the comparisons we have done clearly indicated that our database overestimates the total number of twin pairs. We can roughly estimate an average 40% surplus of pseudo-twins. This excess is mainly localized in opposite sex pairs spread all over the age groups. The likelihood of creating pseudo twins is a function of the frequency of births and of the surnames in the various geographical areas. The high sex ratio in the elderly indicate an excess of pseudo twins among males. This is because in earlier years the code was assigned to the heads of the family (males) who applied for it for fiscal reasons.

The Italian National Twin Registry

As mentioned, in 2001 the Italian Government, through the Ministry of Health, financed a broad national research program on twin studies, including the establishment of a national twin registry. Among all the "possible twin database", three age-groups have been selected, 1925–1934, 1955–1964, 1985–1994, for approximately a total of 500,000 possible twins.

As a first step, a sample of twin pairs belonging to the three age-groups will be contacted by mail for a pilot study and a one page questionnaire will be sent to them. The questionnaire, accompanied by an introductory letter explaining the aims of the Registry and by an informed consent form to be signed by the participant, includes questions about whether the person is a twin, questions about similarity in order to estimate zygosity (for the like-sex pairs), and questions concerning education and occupation, smoking and drinking habits, and leisure-time activities.

To date, a sample of 500 pairs have been contacted by mail. After 6 weeks, before sending the second letter, the responders were 30%. Among them 114 were real twin pairs. The results of the pilot study will orient the future enrolment of twins in the Register.

By the year 2003, we expect to enrol 120,000 twin pairs: 36,000 children, 54,000 adults and 30,000 elderly. General background data will soon be available on zygosity, education, occupation, weight, height, family size and family history of selected diseases. A special attempt will be devoted to enrol twin family members (parents, siblings, children, and spouses).

Based on the availability of financial resources, a sub-sample of the enrolled population will be asked to donate DNA. A biological bank from twins (and twin family members) will then be implemented, guaranteeing information on future etiological questions regarding genetic and modifiable factors for physical impairment and disability, cancer, cardiovascular disease and other age-related chronic illnesses. All of the procedures for collecting data and biological samples are in agreement with Italy's current laws on data confidentiality (ACT no. 675 of 31.12.1996).

Current Projects

The Database of Possible Twins is currently used in population-based studies on multiple sclerosis, Alzheimer's disease, celiac disease, and type 1 diabetes, as described in detail below. A system is currently being developed for linking the database with data from mortality and cancer registries.

Multiple sclerosis (MS). To date, 30,000 persons with MS enrolled in the Italian Multicentric Study of Genetic of Multiple Sclerosis have been matched to the Database of Possible Twins, and more than 200 twin-pairs in which at least one of the twins has MS have been identified, representing the largest number of twin-pairs with MS ever reported. An *ad interim* analysis showed a probandwise concordance rate of 9% for monozygotic twins and of 3% for dizygotic twins (Cannoni et al., 2001), which is lower than the concordance reported for dizygotic twins in most published studies. Although this result must still be confirmed, it may suggest that the relative contribution of environmental and genetic factors to the pathogenesis of MS may be diverse in different populations.

Alzheimer's disease (AD). In the regions of Lazio (Latium) and Toscana (Tuscany), where 5000 persons are estimated to be affected by mild to moderate dementia, a project is currently being implemented to identify all cases of dementia in twins by matching the Database of Possible Twins to the Alzheimer Registry, established for monitoring the pharmacological treatment of the disease.

Celiac disease (CD). The first population-based study on twins and CD was conducted in three regions of southern Italy (Cotichini et al., 2001): 6000 persons with CD were screened for twin status: 58 twin-pairs were identified, of whom 47 were screened for anti-endomysial and anti-human-tissue-transglutaminase antibodies; zygosity was verified by DNA fingerprinting, and twins were typed for the HLA class II DRB1 and DQB1 molecules. Probandwise concordance rates for celiac disease differed significantly between monozygotic and dizygotic twins (0.86 and 0.20, respectively). To date, this is the highest concordance reported for a multifactorial disease (Greco et al., 2002).

Type 1 diabetes. A project is underway to link the Database of Possible Twins to the Database of Diabetes Outpatient Clinics of the Lazio Region. This will allow for the identification of all twins among the persons with diabetes living in the region. The extension of this study to other Italian regions, including Sardinia, is under development.

The Italian Twin Registry has also been recently involved in an international collaboration with the Finnish, Swedish, Danish, Norwegian and Dutch Twin Registries, as well as with the MORGAM (Monica, Risk, Genetics, Archiving and Monograph) cohorts for a project on genome-wide analyses of European twins to identify genes predisposing to common diseases. The project, financed by the European Community within the EC Quality of Life Program, aims to use the synergy between the twin cohorts and the representative population cohorts from the same countries in studies of genetic and environmental predictors of traits.

Future Studies

The Italian Twin Project mainly focuses on aging and autoimmune diseases.

Aging. A collaboration has been established with the Danish Twin Registry to join their study on oldest old twins on the genetic and environmental determinants of extreme survival in humans. The main target will be twin-pairs of at least 90 years of age. A genome-wide scans for genetic loci likely to harbour genes of importance for longevity will be performed in both population. The role of gene-regulation for extreme survival will be then studied by performing gene-expression in the oldest Danish and Italian twins and the degree to which gene regulation is genetically determined and preserved at the oldest ages will be estimated.

Autoimmune Diseases. We are currently involved in studies on MS, CD and type 1 diabetes, but twin studies on rheumatoid arthritis, systemic lupus erythematosus and Sjogren syndrome are in our agenda. Several studies have suggested that polymorphisms of multiple genes may control either a predisposition or protection from autoimmune diseases. Shared genetic influences among immune-mediated diseases might define common etiologic pathways. Dealing with large twin studies, from a rather homogenous population such the Italian one, could allow us to investigate and compare the type and magnitude of association in different autoimmune conditions, and thus to contribute to elucidate the pathogenesis of immune-mediated diseases (Salvetti et al., 2000).

Acknowledgments

References

Buzzetti, R., Pozzilli P., Lenetti, F., Di Mario, U., Cotichini, R., & Stazi, M.A. (2001). Disease concordance and genetic susceptibility to type 1 diabetes in italian twins (abstract). *Twin Research, 4*, 174.

Cannoni, S., Cotichini, R. and the Italian Twin Study Group. (2001). A population based study of multiple sclerosis in twins from continental Italy and Sardinia (abstract). *Twin Research, 4,* 174.

Cotichini, R., Greco, L., Romino, R., Coto, I., Tosi, R., & Stazi, M. A. (2001). Progression to coeliac disease in Italian twins (abstract). *Twin Research, 4,* 176.

Giubilei, F., Sepe Monti, M., Giampaoli, S., Vescio, M. F., Ferrucci, L., Iacoviello, L. et al. (2001). Genetic epidemiology of Alzheimer disease in Italian twins (abstract). *Twin Research, 4,* 183.

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M. et al. (2002). The first large population-based twin study of celiac disease. *Gut, 50*(5), 624–628.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M. et al. (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.

Salvetti, M., Ristori, G., Tosi, R., Fieschi, C., & Stazi, M. A. (1997). Italian population yields world's largest twin registry (letter). *Nature Medicine, 3,* 1176.

Salvetti, M., Ristori, G., Bomprezzi, R., Pozzilli, P., & Leslie, R. D. (2000). Twins: Mirrors of the immune system. *Immunology Today, 21,* 342–347.