

REGISTRO NAZIONALE GEMELLI: ISTITUZIONE E SVILUPPI

Rodolfo Cotichini (a), Valeria Patriarca (a), Cristina D'Ippolito (a), Corrado Fagnani (a), Lorenza Nisticò (a), Rosalia Cirrincione (a), Sonia Brescianini (a), Simonetta Pulciani (b), Emanuela Medda (a), Daniela Rotondi (a), Lorenzo Giannantonio (a), Davide Delfino (a), Silvia Meucci (a), Virgilia Toccaceli (a), Miriam Salemi (a), Maria Antonietta Stazi (a)
(a) *Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità, Roma*
(b) *Dipartimento di Biologia Cellulare e Neuroscienze, Istituto Superiore di Sanità, Roma*

Gli anni '50 hanno visto sorgere in vari paesi europei la raccolta sistematica di informazioni, su base nazionale o locale, riguardanti popolazioni gemellari.

Il registro gemelli più antico al mondo è quello danese, che fu istituito nel 1954, a partire dai registri parrocchiali, con lo scopo principale di studiare le cause del cancro. Nei decenni, si è talmente ampliato da arrivare a comprendere tutte le coppie di gemelli nati nell'arco di 125 anni, dal 1870 al 1996.

I registri gemelli costituiscono una fonte preziosa di informazioni per gli studi di epidemiologia genetica. In essi sono archiviati dati anagrafici, sul fenotipo, sulle abitudini di vita e sullo stato di salute di un elevato numero di coppie di gemelli (1).

L'istituzione dei primi registri gemelli è stata dettata dalla necessità di sviluppare specifici progetti di ricerca su patologie psichiche o fisiche.

L'esperienza positiva di questi registri iniziali ha stimolato l'istituzione di registri gemelli da parte di un gran numero di nazioni, in Europa e nel mondo.

L'Italia vanta una grande tradizione nell'ambito della ricerca gemellare. Una tappa fondamentale nel consolidamento del metodo gemellare è senza dubbio rappresentata dalla pubblicazione, nel 1950, del libro di Luigi Gedda "Studio dei gemelli", un lavoro monumentale che ha sancito la nascita della "gemellologia" e nel quale sono presentati, con una mirabile ricchezza di esempi, gli aspetti cruciali di questo affascinante approccio investigativo (2).

A Luigi Gedda si deve non solo la paternità della gemellologia ma anche l'imponente arruolamento, tramite l'Istituto Mendel di Roma, di una coorte volontaria di circa 40.000 gemelli.

Questa coorte, seguita fin dal 1950 per circa quaranta anni, ancora oggi costituisce una preziosa banca di informazioni per studi su caratteri complessi, benché non siano disponibili dati su indicatori molecolari.

Tuttavia, quello che mancava in Italia era un registro nazionale di gemelli su base di popolazione.

Database dei "probabili gemelli"

La creazione di un registro di popolazione di gemelli italiani è stata possibile grazie a una collaborazione tra l'Istituto Superiore di Sanità (ISS), l'Università di Roma "La Sapienza" e il Ministero delle Finanze (3).

Negli anni '70 è stato introdotto in Italia il Codice Fiscale (CF) come strumento di identificazione del cittadino nei confronti del fisco (DPR 605/1973).

Il CF è un “codice identificativo univoco” composto da sedici caratteri, che viene attribuito dall’Anagrafe Tributaria ad ogni cittadino (DM n. 345/1976).

Al fine di consentire lo scambio di informazioni tra gli organi della pubblica amministrazione, l’utilizzo del CF come codice identificativo di ogni cittadino è stato esteso anche ad altre amministrazioni, quali ad esempio: Comuni; INPS; Servizio Sanitario Nazionale (DM n. 404/1989) (3, 4).

Dal 1994 i comuni, al momento della prima iscrizione nel registro della popolazione residente, attribuiscono alle persone fisiche il CF, la cui generazione è gestita attualmente dal Ministero delle Finanze.

Detta amministrazione, avvalendosi nell’ambito del proprio sistema informativo di una banca dati di riferimento contenente i codici già attribuiti, genera il codice sulla base dei dati trasmessi dai comuni, garantendone l’unicità (DPCM 5-5-1994).

Dall’anno della sua introduzione sono stati attribuiti in Italia 75 milioni di CF. L’evidente soprannumero di CF rispetto alla popolazione residente (56 milioni) è dovuto principalmente a due motivi: in primo luogo, è possibile che dei CF facciano riferimento a persone decedute (in quanto non è prevista la cancellazione del CF in caso di decesso); in secondo luogo, due o più CF possono corrispondere alla stessa persona (per esempio a causa di errori nella trascrizione del CF su documenti o dichiarazioni).

La banca dati dei CF contiene informazioni quali il cognome, nome, data e luogo di nascita, ed il più recente indirizzo di residenza.

L’idea di base che ha indotto ad utilizzare la banca dati dei CF per l’identificazione dei gemelli è che tutti i gemelli:

- sono nati nello stesso luogo;
- sono nati nella stessa data;
- hanno lo stesso cognome.

Queste tre condizioni sono state trasformate in un criterio di ricerca all’interno della banca dati dei CF.

Dai 75 milioni di CF, che rappresentano le persone nate in Italia dal 1 gennaio 1897 al 31 dicembre 1995 e ivi residenti, sono stati estratti 1.600.000 record che rispondevano in pieno ai tre criteri di selezione.

La ricerca ha prodotto non solo coppie ma anche triplete, quadruplette, ecc.

L’esclusione, da questo set di dati, di tutti i record con CF diverso ma che rappresentavano la stessa persona (errore nella trascrizione del nome, per esempio Iessica al posto di Jessica) e delle triplete, quadruplette, ecc., ci ha portato a definire una coorte di 650.000 probabili coppie di gemelli, nati e residenti in Italia al 31 dicembre 1995.

Nella Figura 1 sono mostrate le distribuzioni per anno di nascita delle probabili coppie di gemelli e della popolazione italiana al 1995.

Le due distribuzioni hanno andamenti abbastanza simili, ed in esse si può facilmente riconoscere sia il calo delle nascite durante le due guerre mondiali, sia il “baby boom” degli anni Sessanta.

Nella Tabella 1 è riportata la distribuzione per anno di nascita e area geografica (nord, centro, sud e isole) delle probabili coppie di gemelli in base al sesso della coppia (maschio-maschio, maschio-femmina e femmina-femmina).

Le coppie di sesso diverso sono leggermente sovra-rappresentate rispetto alla percentuale attesa del 33,3% (37,5% per le coppie maschio-femmina contro 31,2% sia per le coppie maschio-maschio che per le coppie femmina-femmina), e ciò è evidente in modo particolare nel Sud Italia. Nell’ultima colonna viene anche riportato il tasso di gemellarità sulla popolazione italiana al 1995.

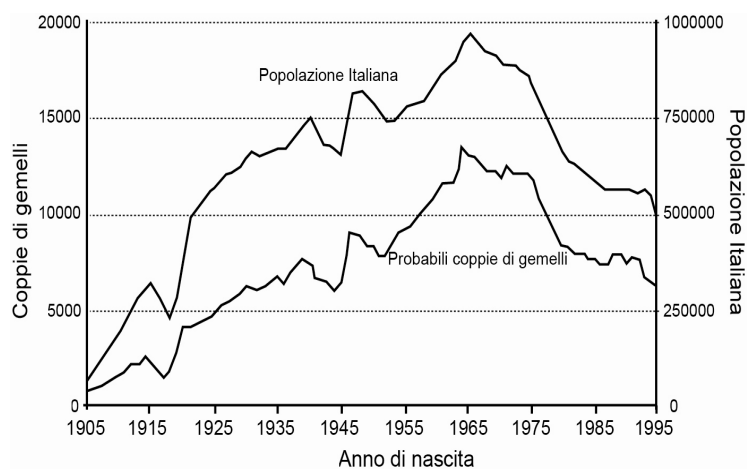


Figura 1. Numero delle probabili coppie di gemelli e popolazione italiana al 1995 per anno di nascita

Tabella 1. Distribuzione delle probabili coppie di gemelli per anno e area geografica di nascita

Riferimenti di nascita	Probabili coppie di gemelli				Popolazione italiana al 1995 (ISTAT)	Prevalenza probabili gemelli
	MM	MF	FF	Totale		
	<i>n.</i>	<i>n.</i>	<i>n.</i>	<i>n. (a)</i>	<i>n. (b)</i>	% [(2*a)/b]
Anno						
1900-1914	6.584	5.205	8.395	20.184	2.017.652	2,0
1915-1924	10.048	10.481	11.156	31.685	3.885.511	1,6
1925-1934	18.757	21.053	18.035	57.845	6.273.004	1,8
1935-1944	21.511	25.969	20.418	67.898	6.992.945	1,9
1945-1954	25.058	32.762	25.012	82.832	7.597.481	2,2
1955-1964	32.597	42.566	32.631	107.794	8.395.146	2,6
1965-1974	37.604	48.433	37.479	123.516	9.120.126	2,7
1975-1984	29.038	33.597	28.566	91.201	6.919.713	2,6
1985-1994	23.920	26.593	23.588	74.101	5.616.249	2,6
1995	2.145	2.234	1.979	6.358	482.953	2,6
Area geografica						
Nord	75.713	78.817	76.821	231.351	25.443.009	1,8
Centro	33.418	39.527	34.003	106.948	10.988.600	1,9
Sud e Isole	96.246	130.277	95.787	322.310	20.869.171	3,1
Totale	207.262	248.893	207.259	663.414	57.300.780	2,3

Dal confronto del rapporto tra i sessi (maschio/femmina) per anno di nascita, nel database dei probabili gemelli e nella popolazione italiana al 1995 (Figura 2), si osserva che nelle fasce di età più anziane vi è un eccesso di maschi rispetto alla popolazione italiana.

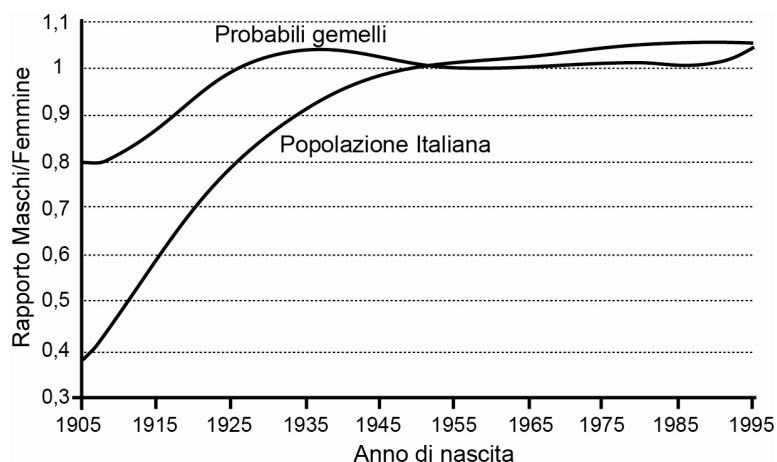


Figura 2. Rapporto maschio/femmina nei probabili gemelli e nella popolazione italiana (ISTAT 1995) per anno di nascita

Ciò potrebbe essere dovuto sia al meccanismo di assegnazione del codice fiscale nei primi anni della sua istituzione, per cui tra coloro che ne facevano richiesta (lavoratori, pensionati e capifamiglia) la percentuale degli uomini era più alta.

Nella Figura 3 sono descritte le distribuzioni per anno di nascita delle probabili coppie di gemelli classificate secondo il criterio di “stesso indirizzo” o “indirizzo diverso” e delle coppie di gemelli secondo i dati ISTAT riferiti al periodo 1981-1995.

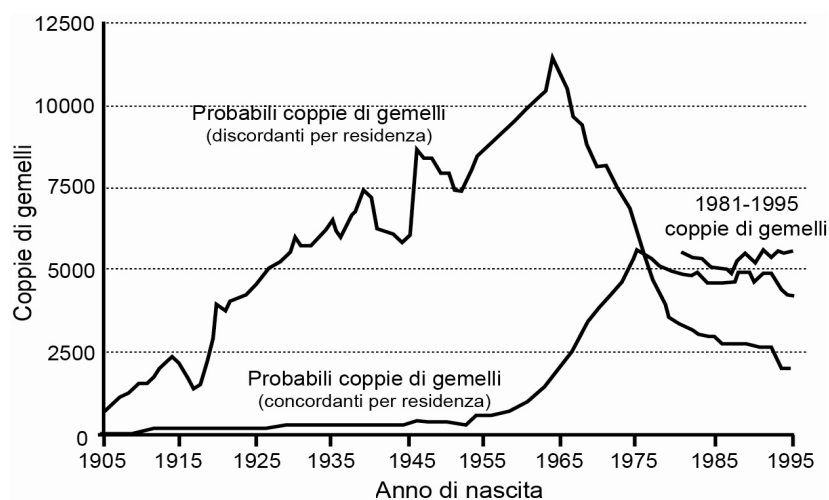


Figura 3. Numero delle probabili coppie di gemelli, concordanti e discordanti per residenza, e numero di coppie di gemelli (ISTAT 1981-1995) per anno di nascita

Nel periodo 1981-1995, le probabili coppie di gemelli concordanti per residenza seguono una distribuzione simile a quella delle coppie secondo l'ISTAT; inoltre, si osserva che la

distribuzione delle coppie di gemelli discordanti per residenza è paragonabile a quella dei nati da parti singoli (vedi Figura 3).

In conclusione, i vari confronti effettuati evidenziano la presenza di falsi gemelli nel nostro database, che si traduce in un eccesso medio stimabile intorno al 40%, localizzato soprattutto nelle fasce di età più anziane, al Sud Italia e tra le coppie di sesso diverso. Fermo restando che la probabilità di creare falsi gemelli è, in ultima analisi, una funzione della frequenza delle nascite e dei cognomi nelle diverse aree geografiche, essa naturalmente dipende anche dalle procedure di assegnazione e dai meccanismi di verifica della qualità dei dati che non sono sempre stati omogenei sul territorio nazionale.

Registro Nazionale Gemelli: banca dati e banca biologica

Nel 2000 il Ministero della Salute, attraverso i fondi per la Ricerca Finalizzata, ha finanziato un progetto di ricerca intitolato “Fattori genetici e ambientali nelle malattie multifattoriali: istituzione di un registro nazionale dei gemelli” (4, 5).

L’obiettivo principale di questo progetto era quello di creare, a partire dal database dei “probabili” gemelli, un registro di popolazione di gemelli italiani, da utilizzare per studi di epidemiologia genetica su caratteri complessi. Sebbene il target iniziale del registro fosse costituito da tutte le coppie di gemelli appartenenti a tre specifiche coorti di nascita, e cioè quelle nate tra il 1925 e il 1934 (anziani), tra il 1955 e il 1964 (adulti), e tra il 1985 e il 1994 (adolescenti), successivamente l’arruolamento ha interessato numerose altre coorti di gemelli, per far fronte alle diverse esigenze ed opportunità di ricerca che sono emerse nel corso degli anni.

Inoltre, allo scopo di aggiornare ed integrare il database dei “probabili” gemelli, si è avviata con le Anagrafi comunali una procedura di verifica ed aggiornamento dei dati anagrafici dei gemelli.

Il Registro Nazionale Gemelli (RNG) è accessibile a gruppi di ricerca istituzionali che intendano valutare il peso relativo di fattori ambientali, comportamentali e genetici nell’eziopatogenesi di malattie multifattoriali nel rispetto delle direttive sancite dal Comitato Etico dell’ISS e dal Garante della Privacy.

L’accertamento della “gemellarità” è effettuato attraverso l’invio tramite posta, ai probabili gemelli, di un primo questionario.

Ogni probabile gemello, insieme al questionario, riceve una lettera contenente informazioni generali sul progetto, ed un modulo per esprimere il consenso a partecipare all’iniziativa. Il questionario include domande per confermare la condizione di gemello e valutarne la zigosità, nonché domande sul livello di istruzione e il tipo di occupazione.

Il materiale inviato ai gemelli in questa fase è presentato nell’Appendice A e B.

Di recente, è stata anche implementata una procedura per l’arruolamento online, basata sulla compilazione di moduli e questionari in forma elettronica disponibili sul sito web del registro: <http://www.gemelli.iss.it>.

A ciascun componente della coppia che aderisce al progetto, viene attribuito un Codice Identificativo Univoco utilizzato nei successivi contatti postali che prevedono il trasferimento di informazioni sensibili (vedi ad esempio i questionari studio-specifici) e al contempo per attuare la separazione dei dati identificativi dai dati sensibili in accordo con il DL.vo 196/2003.

Nel rispetto dello stesso decreto (già Legge n. 675 del 1996 sulla “Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali”), non sono previste nella fase di

arruolamento domande di carattere sanitario o cosiddette “sensibili”. Successivamente, si procede all’approfondimento e all’integrazione delle informazioni relative ad aspetti sanitari generali.

In una fase iniziale del progetto è stata avviata un’indagine pilota su un campione di probabili gemelli, per effettuare una prima valutazione del tasso di rispondenza e dell’efficienza delle procedure di raccolta delle informazioni, nonché per verificare la validità di alcune domande inserite nel questionario.

Ad oggi sono iscritti all’RNG circa 11.200 gemelli, appartenenti a diverse fasce di età ed aree geografiche. In Tabella 2 è riportata la distribuzione, per anno di nascita, zigosità e sesso, delle coppie di gemelli iscritte all’RNG ad ottobre 2007. Si tenga presente che, nel totale delle coppie riportato in Tabella 2, sono incluse anche coppie in cui uno solo dei gemelli si è iscritto.

Tabella 2. Distribuzione per anno di nascita, zigosità e sesso delle coppie di gemelli iscritte all’RNG ad ottobre 2007

Zigosità (*)	Fino al 1940	1980-1941	1988-1981	1996-1989	dopo il 1996	Totale
Monozigote						
MM	91	249	579	122	7	1048
%	15.58	16.54	17.26	15.23	3.91	16.32
FF	99	399	875	105	10	1488
%	16.95	26.51	26.09	13.11	5.59	23.17
Dizigote						
MM	101	189	411	155	71	927
%	17.29	12.56	12.25	19.35	39.66	14.43
FF	100	228	547	190	56	1121
%	17.12	15.15	16.31	23.72	31.28	17.45
MF	178	371	883	228	35	1695
%	30.48	24.65	26.33	28.46	19.55	26.39
Non determinata						
Coppie	15	69	59	1	0	144
%	2.57	4.58	1.76	0.12	0.00	2.24
Totale						
Coppie	584	1505	3354	801	179	6423
%	100.00	100.00	100.00	100.00	100.00	100.00

(*) MM = maschio-maschio; FF = femmina-femmina; MF = maschio-femmina

Anche se il numero totale di gemelli arruolati fino ad ora è nettamente al di sotto delle previsioni, le procedure adottate stanno garantendo l’arruolamento di più di 2000 nuovi gemelli ogni anno (6).

Sul versante normativo per la tutela della privacy, nel luglio 2007 l’Istituto Superiore di Sanità ha prodotto il proprio “Regolamento per il trattamento dati sensibili” (<http://www.iss.it/privacy>). Questo regolamento fornisce una cornice normativa ben delineata alle attività dell’RNG, ricomprendendo, in maniera sintetica, le finalità ed i flussi informativi di acquisizione dati personali dalle Anagrafi italiane (scheda 4 del Regolamento: “Attività di ricerca scientifica”). Nello specifico, la richiesta di dati personali alle Anagrafi Comunali (art. 6 c. 4 Codice di Deontologia e buona condotta per il trattamento di dati personali per scopi statistici e scientifici 2004), per gli aggiornamenti dell’RNG, è finalizzata alla conduzione di specifici progetti di ricerca epidemiologica che vengono sottoposti al vaglio del Comitato Etico dell’ISS. Tali studi, sia per le finalità sia per la metodologia, infatti, non possono essere svolti senza l’uso di dati personali. Nel rispetto del principio di pertinenza e non eccedenza dei dati,

l’RNG adotta le misure tecniche per la definizione del campione di studio, per la messa in sicurezza degli archivi (art. 31 e seguenti del DL.vo 196/03 e art. 15 Codice Deontologico) e pubblicizza lo svolgimento degli studi sul proprio sito istituzionale: www.gemelli.iss.it (art. 6 commi 4 e 5 Codice Deontologico).

Ad un sotto-campione della popolazione arruolata, oltre alla compilazione di questionari per la raccolta di informazioni su specifici fenotipi di interesse, è anche richiesta la donazione di un campione biologico (sangue o saliva) per la costituzione della banca biologica dell’RNG. Il progetto per la costruzione della banca biologica dell’RNG è stato sottoposto al vaglio del Comitato Etico dell’ISS nel 2006, in occasione dell’avvio di alcuni studi internazionali (v. GenomEUtwin, Euroclot, ecc.).

La banca biologica dell’RNG ha sede nel Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute (CNESPS) che, già nel 2004, aveva realizzato, nell’ambito del progetto MORGAM (*MONica Risk, Genetics, Archiving and Monograph*) sulle malattie cardiovascolari, una banca biologica in locali adeguati e conformi a caratteristiche standard, e messo a punto procedure per l’archiviazione dei campioni.

La banca biologica dell’RNG attualmente conserva il sangue o la saliva di circa 500 gemelli. Il sangue viene prelevato attenendosi alle normali procedure mediche e utilizzando provette con o senza anticoagulanti (EDTA o sodio citrato) a seconda della frazione ematica che si vuole conservare; il numero delle provette è in funzione del progetto di studio da realizzare. Ogni provetta possiede un’etichetta sulla quale è riportato il codice identificativo del gemello.

Il sangue raccolto è in parte conservato tal quale, ed in parte è processato per separare il siero o il plasma e i leucociti. Il sangue intero e le sue frazioni sono suddivise in provette o in *paillette* di polietilene, ossia in tubicini finissimi simili a cannucce, saldati a caldo alle estremità. Il contenuto delle provette e *paillette* è riconoscibile per il colore del tappo o del manicotto e per la numerazione. Questo tipo di materiale conservato potrà presumibilmente essere utilizzato per studi di genomica, epigenomica, proteomica e metabolomica.

Per la donazione del DNA su larga scala viene utilizzato un sistema commerciale di auto-raccolta della saliva attraverso lo sputo: ai gemelli donatori viene inviato per posta ordinaria un contenitore di plastica – conforme alle linee guida IATA (*International Air Transport Association*) – nel quale possono essere raccolti circa 2 millilitri di saliva. Alla chiusura del contenitore la saliva si mescola con una soluzione tampone stabilizzante che rende possibile la conservazione del DNA a temperatura ambiente per molti mesi. La nota informativa e il consenso per la donazione del DNA sono presentati in Appendice C.

Per quanto riguarda gli aspetti normativi relativi al trattamento dei dati riguardanti la banca biologica, si veda la sezione dedicata in questo rapporto.

Bibliografia

1. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet* 2002;3(11):872-82.
2. Gedda L. *Studio dei gemelli*. Roma: Orizzonte Medico; 1951.
3. Salvetti M, Ristori G, Tosi R, Fieschi C, Stazi A. Italian population yields world’s largest twin registry. *Nat Med* 1997;3(11):1176.
4. Stazi MA, Cotichini R, Patriarca V, Brescianini S, Fagnani C, D’Ippolito C, Cannoni S, Ristori G, Salvetti M. The Italian Twin Project: from the personal identification number to a national twin registry. *Twin Res* 2002;5(5):382-6.

5. Cotichini R, Fagnani C, Patriarca V, Nistico L, Brescianini S, Cirrincione R, D'Ippolito C, Pulciani S, Figà-Talamanca L, Cannoni S, Ristori G, Salvetti M, Tosi R, Stazi MA. Twins in biomedical research and the creation of the "National Twin Registry". *Epidemiol Prev* 2003;27(5):297-302.
6. Fagnani C, Brescianini S, Cotichini R, D'Ippolito C, Dukic T, Giannantonio L, Medda E, Nisticò L, Patriarca V, Pulciani S, Rotondi D, Toccaceli V, Stazi MA. The Italian Twin Register: new cohorts and tools, current projects and future perspectives of a developing resource. *Twin Res Hum Genet* 2006;9(6):799-805.