

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

TESI DI DIPLOMA

IN STATISTICA E INFORMATICA PER LE AMMINISTRAZIONI PUBBLICHE

**VALUTAZIONE TRAMITE RECORD LINKAGE DELLE
PROCEDURE DI VALIDAZIONE DELLE
INFORMAZIONI SANITARIE INFORMATIZZATE**

Relatore: ch.mo prof. Lorenzo Simonato

Diplomanda: Cristina Rodella

ANNO ACCADEMICO 2002-2003

a Vincenzo

Ringrazio la mia famiglia per avermi permesso pazientemente di arrivare fin qui.

Ringrazio Antonella e Cristiano per avermi sempre sostenuto.

Ringrazio Il prof. Lorenzo Simonato per la disponibilità dimostratami.

*Desidero fare un ringraziamento particolare alla dott.ssa Luisa Manci e a tutta l'U.O. Statistica, Flussi Informativi e Comunicazione nelle persone di Michela Apostoli, Francesca Bruni, Rosetta Castellin, Martina Donato, Federica Fornasiero, Maria Previato, Paola Reffo, Tiziana Zannantonio Martin per avermi accolto, sopportato e appoggiato fino alla fine. Grazie!
Un grazie speciale a Tina e a Fedè.*

INDICE

Introduzione	1
Capitolo 1 - L'Unità Locale Socio-Sanitaria n.17 della regione Veneto	3
Capitolo 2 - L'Anagrafe Sanitaria	9
2.1 Dati disponibili	10
Capitolo 3 - Archivio di Anatomia Patologica	13
3.1 Modalità di compilazione e classificazione.	14
3.2 Dati disponibili (descrizione dell'archivio originale)	15
3.3 Procedura per la sistemazione dell'archivio di Anatomia Patologica	20
3.3.1 Premessa	20
3.3.2 Introduzione alla procedura di sistemazione dell'archivio di Anatomia Patologica	22
3.3.3. Descrizione degli incroci effettuati	23
3.3.4 Schema Riassuntivo	37
3.4 Dati da considerare	39
Capitolo 4 - Archivio dei Ricoveri	51
4.1 Modalità di compilazione	53
4.2 Dati disponibili (descrizione dell'archivio iniziale 290.357 record)	63

4.3 Introduzione alla procedura di sistemazione dell'archivio delle Dimissioni (1995-2002 ULSS 17)	67
4.3.1 Descrizione degli incroci effettuati	68
4.3.2 Schema Riassuntivo	77
4.4 Dati da considerare	79
Conclusioni	85
Bibliografia	89

INTRODUZIONE

Questa tesi è nata da un'esperienza di stage svolto presso l'Unità Operativa Statistica, Flussi Informativi e Comunicazione dell'ULSS 17 della Regione Veneto, diretto dalla dott.sa Luisa Mancini.

La tesi si propone di spiegare un possibile metodo atto alla validazione dei soggetti in archivi sanitari tramite l'utilizzo dell'Anagrafe Assistiti.

L'interesse per questo lavoro nasce dall'esigenza di trovare delle procedure standardizzate di record linkage tra dati sanitari e Anagrafe Assistiti, in quanto questo studio si colloca all'interno di un progetto ben preciso che è quello di realizzare un registro tumori dell'ULSS 17. Il registro tumori viene costruito sulla base dell'incrocio di tre fonti: l'Anatomia Patologica, le schede di Dimissioni Ospedaliere e le schede di Morte.

Il lavoro si è articolato principalmente in tre parti:

1. è stato verificato l'effettivo contenuto dello scarico dell'Anagrafe Assistiti utilizzata. Si sono quindi resi i record, in essa contenuti, univoci e ben identificabili per quanto concerne le variabili nome, cognome, data di nascita e tessera sanitaria.
2. è stato verificato l'effettivo contenuto dello scarico dell'Anatomia Patologica utilizzato. Si è quindi proceduto con una serie di linkage tra l'archivio suddetto e l'Anagrafe Assistiti sistemata precedentemente. Si è poi deciso quali dati recuperare e quindi di quali link tener conto. Sui record così recuperati e corretti si sono proposte delle tabelle e dei grafici riassuntivi e descrittivi.
3. è stato verificato l'effettivo contenuto dello scarico delle Dimissioni utilizzato. Si è quindi proceduto con una serie di linkage (gli stessi eseguiti tra l'Anatomia Patologica e l'Anagrafe Sanitaria) tra l'archivio

suddetto e l'Anagrafe Assistiti sistemata precedentemente. Si sono poi considerati i dati recuperati dai link di cui si era precedentemente deciso di tener conto. Sui record così recuperati e corretti si sono proposte delle tabelle e dei grafici riassuntivi e descrittivi.

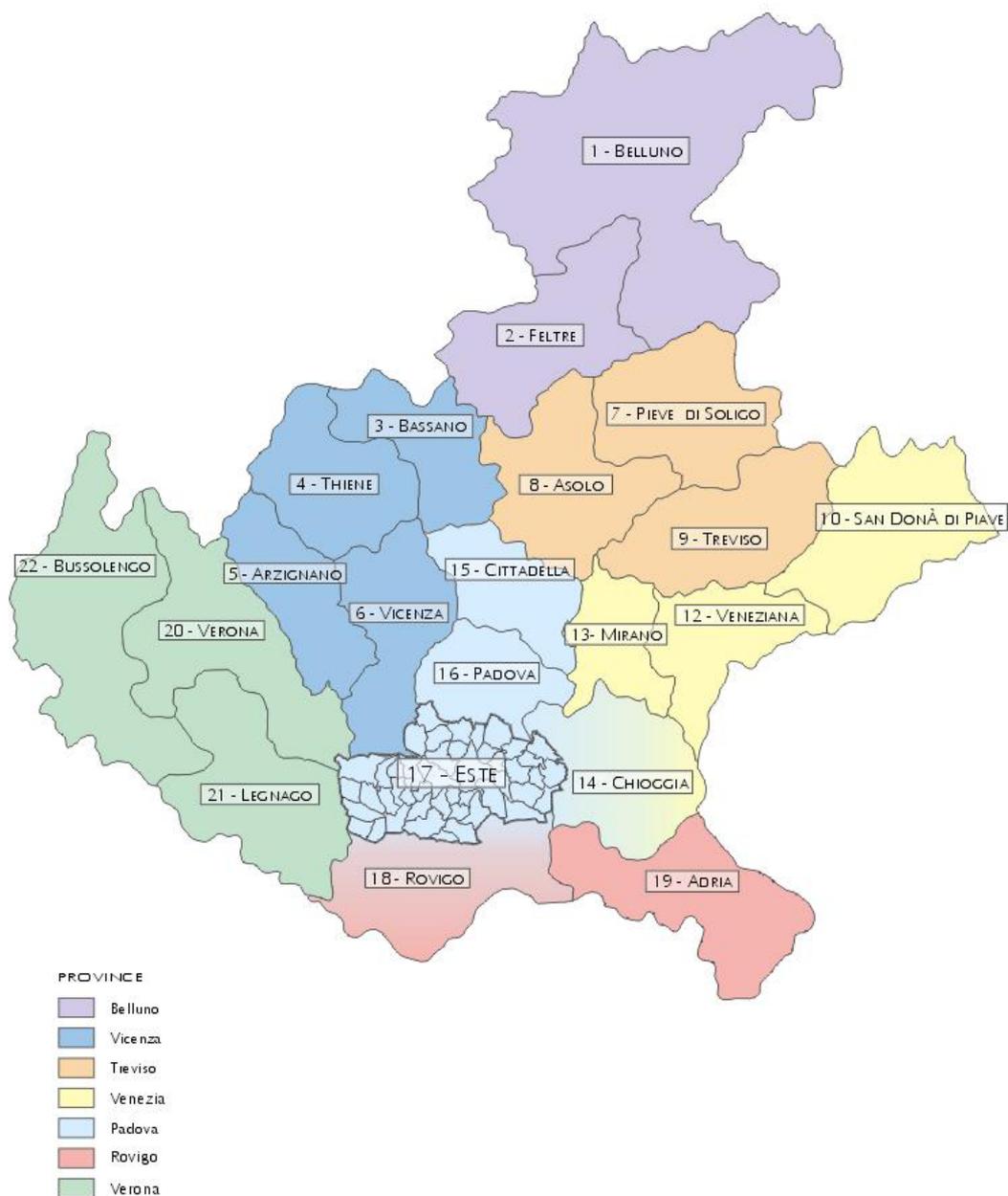
CAPITOLO 1

L'UNITÀ LOCALE SOCIO-SANITARIA N.17 DELLA REGIONE VENETO

Figura n°1.1

REGIONE VENETO

- Suddivisione per ULSS e per Provincia -



Elaborazioni: U.O. Statistica, Flussi Informativi e Comunicazione - Azienda Ulss 17

L'Unità Locale Socio-Sanitaria n. 17 di Conselve, Este, Monselice, Montagnana della Regione Veneto, è situata nella parte meridionale della provincia di Padova e confina con le ULSS di Padova, Rovigo, Legnago, Vicenza e Chioggia-Piove di Sacco (Figura n°1.1). Il territorio, composto da 46 comuni, comprende gran parte dei Colli Euganei, delle zone termali e una parte della Pianura Padana.

Istituita l'uno gennaio 1995 con l'aggregazione dei territori della ex ULSS n.22 e della parte più rilevante della ex ULSS n.23 (Conselvano e Monselicense), l'ULSS n.17 risultava a tale data composta da 48 comuni, divenuti 47 a partire dall'uno marzo 1995, poiché i comuni di Carrara San Giorgio e Carrara Santo Stefano si sono fusi nell'unico comune Due Carrare, e divenuti infine 46 con il passaggio di Maserà di Padova all'ULSS n.16 di Padova a partire dal 18/02/1996.

I comuni dell'ULSS n.17 sono suddivisi in due distretti definiti in base alla Deliberazione n. 782 del 21 agosto 2001:

1. Distretto di Monselice-Conselve formato da 20 comuni: Agna, Anguillara Veneta, Arquà Petrarca, Arre, Bagnoli di Sopra, Battaglia Terme, Bovolenta, Candiana, Cartura, Conselve, Due Carrare, Galzignano Terme, Monselice, Pernumia, Pozzonovo, San Pietro Viminario, Solesino, Stanghella, Terrassa Padovana e Tribano;
2. Distretto di Este-Montagnana composto da 26 comuni: Baone, Barbona, Carceri, Casale di Scodosia, Castelbaldo, Cinto Euganeo, Este, Granze, Lozzo Atesino, Masi, Megliadino San Fidenzio, Megliadino San Vitale, Merlara, Montagnana, Ospedaletto Euganeo, Piacenza d'Adige, Ponso, Saletto, Santa Margherita d'Adige, Sant'Elena, Sant'Urbano, Urbana, Vescovana, Vighizzolo d'Este, Villa Estense e Vò.

Il territorio dell'ULSS n.17 ha una estensione di 887,9 kmq. I comuni sono prevalentemente piccoli con viabilità impervia e disagiata, con aggravio dovuto alla scarsità di mezzi pubblici di trasporto persone.

Quanto alla dimensione delle comunità territoriali, si è rilevato che tra i comuni che compongono l'Azienda ULSS n.17 ben 39 hanno meno di 5.000 abitanti ed in questi vive più della metà della popolazione (58,8%). Inoltre la popolazione che vive in comuni sotto i 3.000 abitanti rappresenta il 24,8% del totale. Mentre sono solo due i comuni che contano più di 15.000 abitanti, Monselice ed Este, in cui vive il 19,2 % della popolazione (Tabella n°1.1).

Tabella n°1.1

Dimensione dei comuni dell'ULSS 17

Numero abitanti	COMUNI			POPOLAZIONE		
	valori assoluti	composizione percentuale	frequenze cumulate	valori assoluti	composizione percentuale	frequenze cumulate
fino a 3.000 ab.	23	50,0	50,0	44.279	24,8	24,8
3.001-5.000 ab.	16	34,8	84,8	60.926	34,1	58,8
5.001-10.000	5	10,9	95,7	39.265	22,0	80,8
10.001-15.000	—	—	95,7	—	—	80,8
15.001-20.000	2	4,3	100,0	34.338	19,2	100,0
totale	46	100,0	—	178.808	100,0	—

Al 31 dicembre 2002 l'ULSS n.17 della Regione Veneto contava 178.808 residenti, con una densità abitativa di 201 abitanti per chilometro quadrato. I comuni più abitati risultano essere quelli di Monselice (17.565 ab.), Este (16.773 ab.), Montagnana (9328 ab.) e Conselve (9145 ab.) che, ad eccezione del comune di Conselve, corrispondono ai comuni più estesi territorialmente. Se si considera la densità abitativa, i comuni che hanno una densità maggiore rispetto agli altri sono Solesino (688 ab/kmq), Battaglia Terme (663 ab/kmq), Este (512 ab/kmq). Mentre quelli con

densità molto bassa sono i comuni di Vescovana (71 ab/kmq), Sant'Urbano (70 ab/kmq) e Vighizzolo d'Este (53 ab/kmq).

Si osservino ora i dati relativi alla popolazione aggregando i comuni in base al distretto di appartenenza.

Come si può vedere il distretto più popolato è quello di Conselve-Monselice dove risiedono 96.437 abitanti pari al 54% della popolazione dell'ULSS, con una densità abitativa di 240 ab/kmq. Il distretto di Este-Montagnana che è anche quello più esteso territorialmente conta 82.371 residenti pari al 46% della popolazione totale, con una densità abitativa di 170 ab/kmq (Tabella n°1.2).

Tabella n°1.2

Distribuzione della popolazione e del territorio per distretto

Distretti	Numero Comuni	Popolazione			% di popolazione sul totale	Superficie territoriale (Km ²) (*)	densità (ab/Km ²)
		maschi	femmine	totale			
Conselve-Monselice	20	47.344	49.093	96.437	54	401,97	240
Este-Montagnana	26	40.396	41.975	82.371	46	485,88	170
Tot. Ulss 17	46	87.740	91.068	178.808	100	887,85	201

(*) Fonte: ISTAT - 13° censimento generale della popolazione - 21.10.1991

Come si è già detto, i residenti dell'ULSS n.17 sono 178.466, di questi il 49,1% è di sesso maschile, mentre il 50,9% è di sesso femminile.

Se si considera la distribuzione per età, si può notare che il 19,5% della popolazione ha più di 65 anni, mentre solo il 13% della popolazione ha meno di 14 anni. In particolare mentre per i giovani (0-14 anni) non c'è una grossa differenza tra maschi e femmine, per i più anziani (65 anni e più) c'è una notevole differenza tra i sessi, sia nella numerosità che nella distribuzione della popolazione tra classi di età. Per quanto riguarda,

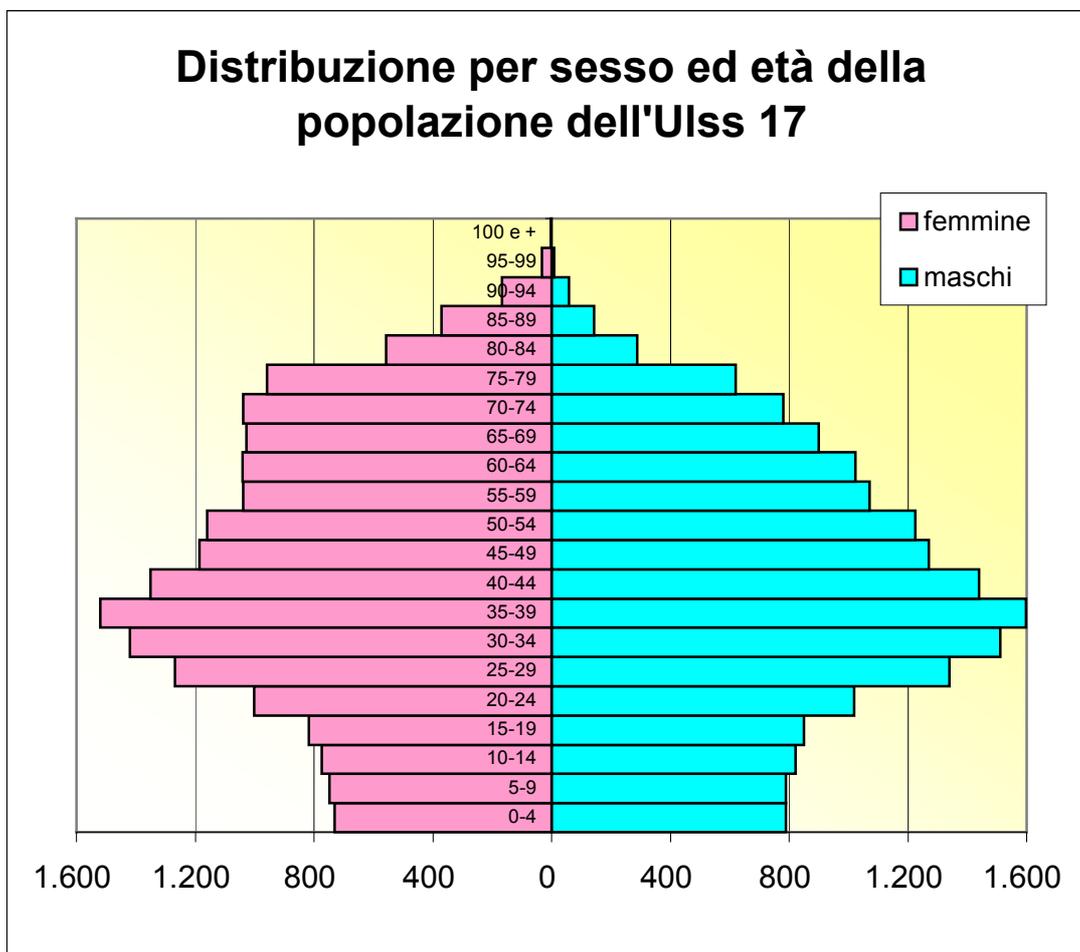
invece, le classi di età intermedie c'è una leggera differenza di distribuzione tra maschi e femmine; per entrambi comunque le classi d'età più numerose sono quelle tra i venticinque e quarant'anni (Tabella n°1.3 e Figura n°1.2).

Tabella n°1.3

Distribuzione per gruppi di età della popolazione dell'Ulss n.17

Età	Femmine	% sul Totale	Maschi	% sul Totale	Totale	% sul Totale
0-14	11272	6,3	11997	6,7	23269	13,0
15-64	59019	33,0	61738	34,5	120756	67,5
65 e +	20777	11,6	14006	7,8	34783	19,5
Totale	91068	50,9	87740	49,1	178808	100,0

Figura n°1.2



CAPITOLO 2

L'ANAGRAFE SANITARIA

L'Anagrafe Sanitaria (o Anagrafe Assistiti) dell'ULSS 17 è un archivio che contiene le schede di tutti coloro che si sono iscritti al Servizio Sanitario nazionale presso l'ULSS suddetta scegliendo un medico esercitante nella stessa.

Tali schede riportano dati anagrafici e d'altro genere. Per iscriversi all'Anagrafe Sanitaria basta dimostrare di essere residente nel territorio dell'ULSS ed eventualmente compilare un modulo per la scelta del medico dal quale si vuole venire assistiti, poiché è consentita la sola iscrizione al servizio sanitario nazionale senza la scelta di alcun medico curante. Se non si è residenti nell'ULSS per iscriversi bisogna compilare un apposito modulo nel quale si motiva la scelta di avvalersi dei servizi sanitari di tale ULSS e si esprime la scelta del medico curante. L'ULSS d'iscrizione accerta se l'individuo sia già iscritto al Servizio Sanitario presso un'altra ULSS e se ne è il caso dà comunicazione dell'avvenuta iscrizione all'ULSS "di provenienza" comprovandola inviando copia della tessera sanitaria personale del nuovo iscritto ed esplicitando le caratteristiche dell'iscrizione siffatta. Un'iscrizione di tal genere viene classificata come "in deroga", come esempio si potrebbe immaginare il caso di uno studente universitario che prenda domicilio nell'ULSS per poter frequentare più agevolmente l'università e decida di fruire dei servizi offertogli dall'ULSS per il periodo dei suoi studi.

Nel caso invece che uno cambi residenza e voglia iscriversi nell'ULSS in cui è immigrato, l'ULSS d'immigrazione provvederà a informare con

le modalità già viste l'ULSS di provenienza che prenderà atto dell'avvenuto trasferimento del suo assistito per aggiornare il suo archivio.

2.1 Dati disponibili

Siccome l'Anagrafe Sanitaria è continuamente aggiornata si è lavorato con uno scarico di tale archivio che è stato fornito in data 19/12/2003 dalla U.O. Statistica, Flussi Informativi e Comunicazione. L'archivio è stato convertito in un file.dbf di access. Di seguito è riportato il tracciato record dei campi presi in considerazione nello scarico dell'archivio.

Campo	Uso campo	tipo	dimensione
ACODICE	numero tessera sanitaria	testo	16
ACOFIS	codice fiscale	testo	16
ACOGN	cognome	testo	25
ANOME	nome	testo	25
Data nascita	data di nascita	gg/mm/aaaa	9
ASEX	sex	testo	1
ACONASCI	comune di nascita	testo	6
ACORESAN	comune di residenza	testo	6
ASOSP	motivo della sospensione	testo	1
Data decorrenza	data di decorrenza dell'iscrizione	gg/mm/aaaa	9

Nel campo ASOSP sono codificate con un numero i seguenti motivi della sospensione del medico:

1. Trasferimento (se uno emigra)
2. Scadenza iscrizione (riguarda coloro che sono stranieri e temporaneamente presenti in ULSS)
3. Decesso assistito
4. Cessazione medico

5. Servizio militare, trasferimento nell'ambito dell'ULSS, temporaneamente all'estero
6. Doppia iscrizione (nel caso in cui fossero stati inseriti dati anagrafici sbagliati perciò per la stessa persona abbiamo due registrazioni pensando che si tratti di due individui omonimi)
7. Cessata età pediatrica
8. Cessazione iscrizione in deroga
9. Cessazione scelta in deroga
10. Doppia iscrizione
11. Scadenza scelta in deroga
22. Perdita diritto
23. Accorpamento comuni nuovi distretti
24. Cessazione T.P. per tessera definitiva
25. Cessazione rapporto di fiducia
26. Uscita residenza protetta
27. Cambio distretto di residenza
95. Cessazione del medico
96. Deroga out
97. Nuova scelta
98. Cessata età pediatrica

Ovviamente se non si verifica nessuna di queste situazioni resta memorizzato il codice 0.

L'archivio iniziale dell'Anagrafe Sanitaria era composto da 249.807 record 904 dei quali duplicati: a stesso cognome, nome, data di nascita e comune di nascita corrispondevano due tessere sanitarie differenti. Verificato che i record con i dati anagrafici uguali e tessere sanitarie differenti non si riferissero ad omonimi si è assegnato agli individui

due volte presenti nel registro una delle due tessere secondo un semplice criterio di scelta: si sono guardati i campi *ASOSP* e *data decorrenza* e si è scelta la tessera ancora in vigore o se entrambe scadute l'ultima ad essere stata utilizzata (*Esempio n°2.1*).

Esempio n°2.1

ACOGNO	ANOME	data nascita	ACONASCI	ACODICE	ASOSP	data decorrenza	tessera assegnata
ROSSI	ANDREA	01/01/1990	28055	200088165	1	01/05/1996	no
ROSSI	ANDREA	01/01/1990	28055	200088156	0	10/05/1996	si

I record relativi alle 456 Tessere Sanitarie scartate sono stati tolti dalla tabella **anagrafe** ed inseriti in una tabella chiamata **AnagrafeTessereScadute**.

Al termine di questo procedimento la tabella **anagrafe** contiene 249.351 record (249.807-456) tutti univocamente identificabili dalla tessera sanitaria.

CAPITOLO 3

ARCHIVIO DI ANATOMIA PATOLOGICA

I dati utilizzati per le elaborazioni sono stati forniti dal U.O. Sistema Informatico dell'ULSS 17 attraverso uno scarico, avvenuto in data 13/01/2003, dell'attività dell'U.O. Anatomia Patologica per tutti quegli anni in cui era possibile recuperare record.

L'archivio messo a disposizione su PC era in formato.txt, per la sua gestione è stato convertito in un file.dbf di Access.

L'Archivio iniziale è composto da 202.014 record relativi a tutti gli esami effettuati dai laboratori dell'Anatomia Patologica dell'ULSS 17 dal 1991 ai primi tredici giorni del 2003 (non è stato possibile considerare anni precedenti al 1991 in quanto l'archiviazione dei dati non era ancora automatizzata).

Di seguito è riportato il tracciato record dell'archivio .dbf.

CAMPO	USO CAMPO		dimensione
cognome	cognome	testo	40
nome	nome	testo	40
data nascita	data di nascita	(ggmmaaaa)	8
nosologico	codice ricovero	testo	10
esmo	E/M (Este o Monselice)	testo	1
tessera	numero di tessera sanitaria	testo	9
*tipologia	tipo di esame	testo	1
*numero progressivo	un numero progressivo	testo	6
*anno competenza	anno di competenza	testo	8
data referto	data referto	(ggmmaaaa)	8
materiale prelievo	materia prelievo	testo	50
diagnosi codificata	diagnosi codificata	testo	61
sexo	sexo(f/m)	testo	1
codfis	codice fiscale	testo	16

CAMPO	USO CAMPO		dimensione
sno1	codice SNOMED	testo	8
sno2	codice SNOMED	testo	8
sno3	codice SNOMED	testo	8
sno4	codice SNOMED	testo	8
sno5	codice SNOMED	testo	8
sno6	codice SNOMED	testo	8
sno7	codice SNOMED	testo	8
sno8	codice SNOMED	testo	8
sno9	codice SNOMED	testo	8
sno10	codice SNOMED	testo	8
sno11	codice SNOMED	testo	8
sno12	codice SNOMED	testo	8
codice memo	codice memo	testo	11
comune residenza	comune di residenza(codice ISTAT)	testo	5
comune nascita	comune di nascita(Codice ISTAT)	testo	5

3.1 Modalità di compilazione e classificazione

Tipologia: codice relativo al tipo di esame effettuato. Di seguito la tabella di decodifica (Tabella n°3.1).

Tabella n°3.1

Decodifica del campo *Tipologia*

Tipologia esame	
1	istologico
2	citologico
5	screening
6	citologico vaginale
9	screening

Esmo: identifica l'ospedale in cui si effettuata l'analisi **E**ste o **M**onselice

Anno di competenza: identifica l'anno in cui è stata eseguita l'analisi

Sno1...Sno12: in questi campi sono inseriti i codici SNOMED. Codici atti ad identificare P-procedura, T-topografia, M-morfologia, G il medico che ha eseguito l'analisi. L'operatore addetto al caricamento dei dati introduce un codice ed in automatico vengono inseriti i codici T e P univoci per ogni record.

Sesso: il sesso è codificato con **F** per le femmine e **M** per i maschi

Numero progressivo: è un numero progressivo per anno e tipologia inserito dalla procedura automaticamente nel momento del caricamento dell'esame nell'archivio di Anatomia Patologica.

I campi contrassegnati con l'asterisco: *tipologia, anno competenza e numero progressivo* considerati insieme identificano univocamente il record.

3.2 Dati disponibili (descrizione dell'archivio originale)

L'archivio iniziale così come consegnato non può essere utilizzato per elaborazioni affidabili a causa di imprecisioni e incompletezze dell'informazione facilmente riscontrabili almeno per quanto concerne la parte anagrafica del database quale: cognome, nome, codice sanitario, data di nascita, comune di residenza, comune di nascita, codice fiscale e sesso.

Probabilmente la procedura informatica utilizzata per l'inserimento dei dati è stata costruita per soddisfare esigenze gestionali e non in previsione di raccolta dati per elaborazioni.

Di seguito sono riportate alcune tabelle (Tabella n°3.2) relative alla situazione iniziale dei record dal punto di vista della completezza dell'informazione anagrafica .

Tabella n°3.2

Composizione iniziale dei campi di tipo anagrafico

Campo sesso		%
F	113.091	56,0
M	49.735	24,6
non compilate	39.188	19,4
totale	202.014	100,0

Campo Nome		%
compilate	202.014	100,0
non compilate	0	0,0
totale	202.014	100,0

Campo Cognome		%
compilate	202.014	100,0
non compilate	0	0,0
totale	202.014	100,0

Campo Data di nascita		%
compilate	201.104	99,5
bianche	910	0,5
totale	202.014	100,0

Campo Comune nascita		%
compilate	161.585	80,0
non compilate	40.429	20,0
totale	202.014	100,0

Campo Comune Residenza		%
compilate	162.824	80,6
bianche	39.190	19,4
totale	202.014	100,0

Campo tessera		%
compilate	170.095	84,2
non compilate	31.919	15,8
totale	202.014	100,0

Campo Codice fiscale		%
compilate	152.558,0	75,5
non compilate	49.456,0	24,5
totale	202.014,0	100,0

Si noti che i campi *nome* e *cognome* sono i soli ad essere sempre compilati anche se ciò non implica la correttezza delle informazioni in essi contenuti.

Nel capitolo successivo verrà approfondito il problema della qualità del dato mostrando gli errori più comuni che si creano in fase di caricamento delle informazioni e proponendo un procedimento per il recupero delle stesse però corrette.

Per quanto riguarda gli altri campi (sempre relativi alle informazioni anagrafiche) si noti come i record non compilati variano da una percentuale dello 0,5% per il campo *data di nascita* ad un 24% per il *codice fiscale*. Anche per questi campi rimane valido il discorso sopradetto che la completezza non implica la correttezza.

È evidente che se è facile individuare errori di mancata compilazione non è altrettanto banale identificare e recuperare inesattezze di digitazione o di errata informazione.

Siccome per quanto riguarda l'aspetto anagrafico l'archivio presenta delle lacune non si crede corretto (in questa fase di descrizione dell'archivio iniziale) proporre tabelle riassuntive relative agli utenti dell'Anatomia ma forse è più giusto dare un'idea della quantità e del tipo di prestazioni effettuate nei laboratori dell'Anatomia Patologica dell'ULSS 17 negli anni considerati.

La tabella n°3.3 mostra il numero di esami effettuati negli anni 1991-2003 nei laboratori di Anatomia Patologica ULSS 17 divisi per anno e tipologia di esame.

Tabella n°3.3

Distribuzione della tipologia di esame per anno di competenza

anno	TIPOLOGIA ESAME					Totale
	1 istologico	2 citologico	5 screening	6 citologico vaginale	9 screening	
1991	1					1
1992	4.006					4.006
1993	4.699					4.699
1994	4.997	1				4.998
1995	7.017	1				7.018
1996	13.870	497				14.367
1997	13.945	3.685				17.630
1998	14.389	3.420	1.951		25	19.785
1999	14.785	4.805	5.368	7.862	162	32.982
2000	14.276	5.042	5.337	9.498	366	34.519
2001	14.118	5.311	7.270	3.819	309	30.827
2002	13.567	5.051	6.113	5.901	372	31.004
2003	67	102	1	7	1	178
Totale	119.737	27.915	26.040	27.087	1.235	202.014

Come si vede dalla tabella n° 3.3 le informazioni riguardanti gli anni 1991 e 2003 sono minime ma questo non a causa di un'inattività dei laboratori di Anatomia Patologica. Del 2003, infatti, si sono scaricati record relativi solo ai primi 13 giorni di gennaio (dal momento che lo scarico è avvenuto il 13/01/2003) mentre la mancanza d'informazioni riguardanti il 1991 sono imputabili all'esistenza e all'utilizzo fino a quell'anno di un altro metodo d'archiviazione delle informazioni.

Si noti anche come tra il 1995 ed il 1996 l'attività degli esami effettuati in Anatomia Patologica sia raddoppiata. Questo è dovuto in parte ad una tendenza di crescita continua (che c'è stata fino al 2000) di esami effettuati dai laboratori di Anatomia e in parte dal fatto che il territorio di riferimento nel corso del 1995 è cambiato. Nel corso del 1995 si è completata l'aggregazione dei territori della ex ULSS 22 (i dati precedenti al 1996 si riferiscono ad assistiti di questo territorio) con la parte più rilevante della ex ULSS 23 (Conselvano e Monselicense).

Dal 1996 al 1999 la crescita di esami effettuati è stata continua e rapida. Nel corso degli anni ad incidere sul numero di prestazioni effettuate sono state la richiesta di nuove tipologie di analisi:

1996 esame citologico (tipologia 2 presente anche precedentemente ma in quantità non rilevante).

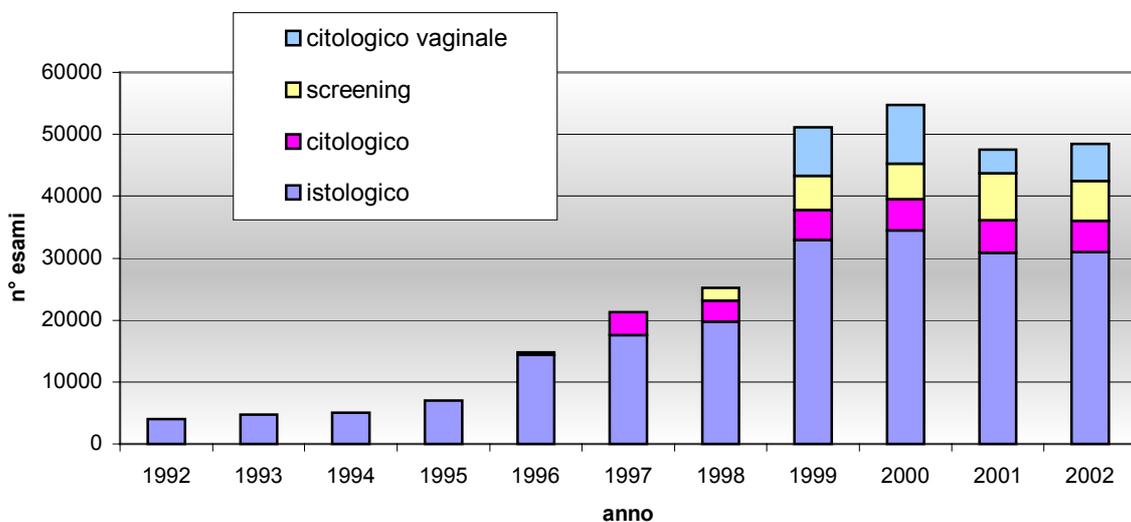
1998 esame screening (tipologia 5 o 9. Applicazione di un test ad una popolazione apparentemente sana al fine di diagnosticare precocemente una malattia, prima della comparsa di sintomi, e per avviare tempestivamente il trattamento terapeutico).

1999 esame citologico vaginale (tipologia 6).

Il grafico n°3.1 mostra l'attività svolta nei laboratori dell'Anatomia Patologica della ULSS 17 per anno e tipologia. Le informazioni relative agli anni 1991 e 2003 non sono stati inseriti nel grafico perché irrilevanti.

Grafico n° 3.1

Distribuzione degli esami effettuati nei laboratori di Anatomia Patologica 1992-2002
ULSS 17



3.3 Procedura per la sistemazione dell'archivio di Anatomia Patologica

3.3.1 Premessa

Dopo una breve fase descrittiva dell'archivio iniziale e non corretto ora si passa ad un'altra in cui viene proposto ed eseguito un percorso atto a ripulire l'archivio da inesattezze e mancanze anagrafiche.

La fase di validazione dei soggetti è avvenuta principalmente attraverso una serie di linkage automatici (e solo in parte manuali con controllo a video) con uno scarico (avvenuto in data 19/12/2002) dell'archivio dell'Anagrafe Assistiti dell'ULSS 17 (contenente anche lo storico anagrafico).

Per questa fase di rifinitura dell'archivio, che si propone di sistemare i record anche per un eventuale successivo utilizzo nel processo automatizzato dell'archivio tumori, è stato necessario proporsi una linea guida atta a standardizzare il più possibile i metodi usati dal personale addetto nelle varie realtà circostanti. A tal proposito è stato fatto un incontro in cui ogni partecipante ha esposto le proprie esperienze pratiche in questo campo ed illustrato gli eventuali limiti causati per lo più dai dati e dagli archivi disponibili nella realtà in cui si opera.

Alla fine si è giunti a stilare una proposta d'incroci consigliabile ma non vincolante che di seguito è riportata.

1. Tessera sanitaria (o l'identificativo locale) +Cognome + Nome +
Data di Nascita
2. Tessera sanitaria (o l'identificativo locale) + Cognome +
Data di Nascita
3. Tessera sanitaria (o l'identificativo locale) + Cognome + Nome

4. Tessera sanitaria (o l'identificativo locale) + Nome + Data di Nascita
5. Cognome + Nome + Data di Nascita

Nel lavoro di rifinitura degli archivi di Anatomia Patologica e dei Ricoveri che verrà proposto nei prossimi capitoli si è seguito un procedimento molto simile a quello qui sopra consigliato. Però non bisogna erroneamente pensare che questo sia l'unico tipo di percorso possibile, infatti, un'altra proposta di linkage a cui si era approdati inizialmente prevedeva solo due tipi di incrocio:

1. Tessera sanitaria
2. Nome + Cognome + Data di Nascita

Si sarebbero dovuti incrociare i due archivi (Anatomia e Anagrafe) tramite la variabile *tessera sanitaria* ed il controllo successivo sarebbe dovuto avvenire solo sui record non recuperati nel primo link. Si è cercato di calcolare la quantità di errore che si sarebbe creato se si fosse seguito questo percorso. Si sono confrontati la quantità di record che si sarebbero agganciati se si fosse considerata solo la tessera sanitaria con quelli agganciatesi per Nome – Cognome - Tessera sanitaria - Data di Nascita (NCTDN). La tabella n° 3.4 mostra il risultato.

Tabella n°3.4

ANATOMIA PATOLOGICA (202014 record iniziali)		
		% sul totale 202014
Solo tessera	162450	80,42
NCTDN	158403	78,41
differenza	4047	2,00
scarto di errore	2,49	$(4047/162450)*100$

3.3.2 Introduzione alla procedura di sistemazione dell'archivio di Anatomia Patologica

Di seguito verranno descritti e mostrati gli incroci utilizzati per la sistemazione dell'archivio di Anatomia Patologica dell'ULSS 17.

L'unico tipo di informazioni sui cui è stato possibile fare un controllo sono state quelle anagrafiche, infatti, utilizzando l'archivio dell'Anagrafe Assistiti dell'ULSS 17 e confrontando la parte anagrafica è stato possibile recuperare alcune informazioni riguardanti: numero di tessera sanitaria, data di nascita, cognome, nome, comune di nascita, comune di residenza, codice fiscale e sesso.

Ovviamente questo tipo di controllo permette di recuperare solo informazioni relative agli assistiti dell'ULSS 17.

I controlli eseguiti possono essere divisi in due categorie: quelli avvenuti in maniera totalmente automatica (incroci che vanno dall'uno al cinque) e quelli invece in cui ad una prima selezione avvenuta meccanicamente ne è seguita una seconda "manuale" (incroci che vanno dal sei al dieci).

Di seguito è riportata la lista degli incroci effettuati:

1. tessera sanitaria + cognome + nome + data di nascita + comune di nascita
2. tessera sanitaria + cognome + nome + data di nascita
3. tessera sanitaria + cognome + data di nascita
4. tessera sanitaria + cognome + nome
5. cognome + nome + data di nascita
6. tessera sanitaria + nome + data di nascita
7. tessera + data di nascita
8. tessera

9. cognome + data di nascita
10. nome + data di nascita

3.3.3 Descrizione degli incroci effettuati

1° incrocio. *tessera sanitaria + cognome + nome + data di nascita + comune di nascita*

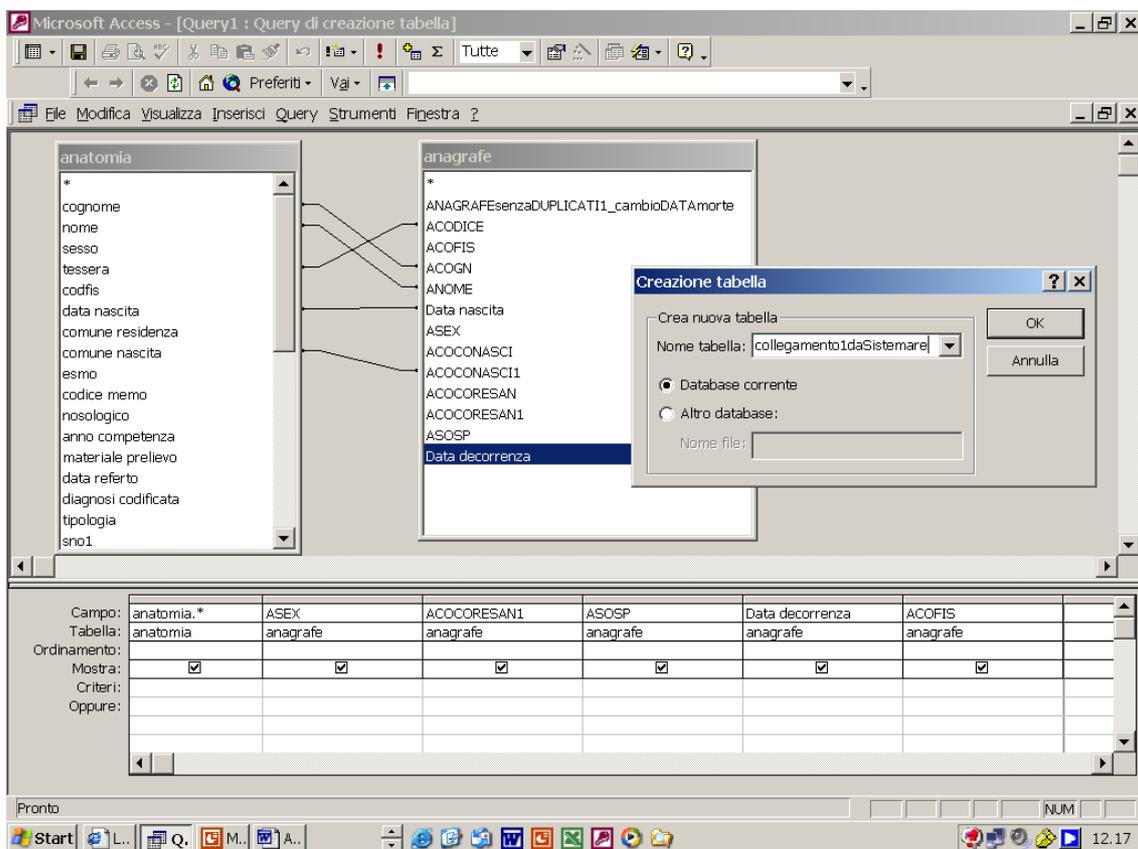
✓ *Primo step*

Sono state collegate le tabelle **Anatomia** e **anagrafe** attraverso un join uno ad uno (con questo join vengono selezionate solo le righe in cui i campi collegati da entrambe le tabelle sono uguali; in questo caso *tessera sanitaria, cognome, nome, data di nascita e comune di nascita*).

I 160.270 record così selezionati sono stati inseriti in una tabella chiamata **collegamento1daSistemare**. La tabella **collegamento1daSistemare** contiene tutti i campi presenti in **Anatomia** ed in più le variabili derivanti dalla tabella **anagrafe** non utilizzate nel join (*ACOFIS, ASEX, ACORESANI, ASOSP, DATA DECORRENZA*).

La schermata n° 3.1 mostra i join eseguiti nel primo incrocio e la creazione della tabella **collegamento1daSistemare**.

Schermata n° 3.1

✓ Secondo step

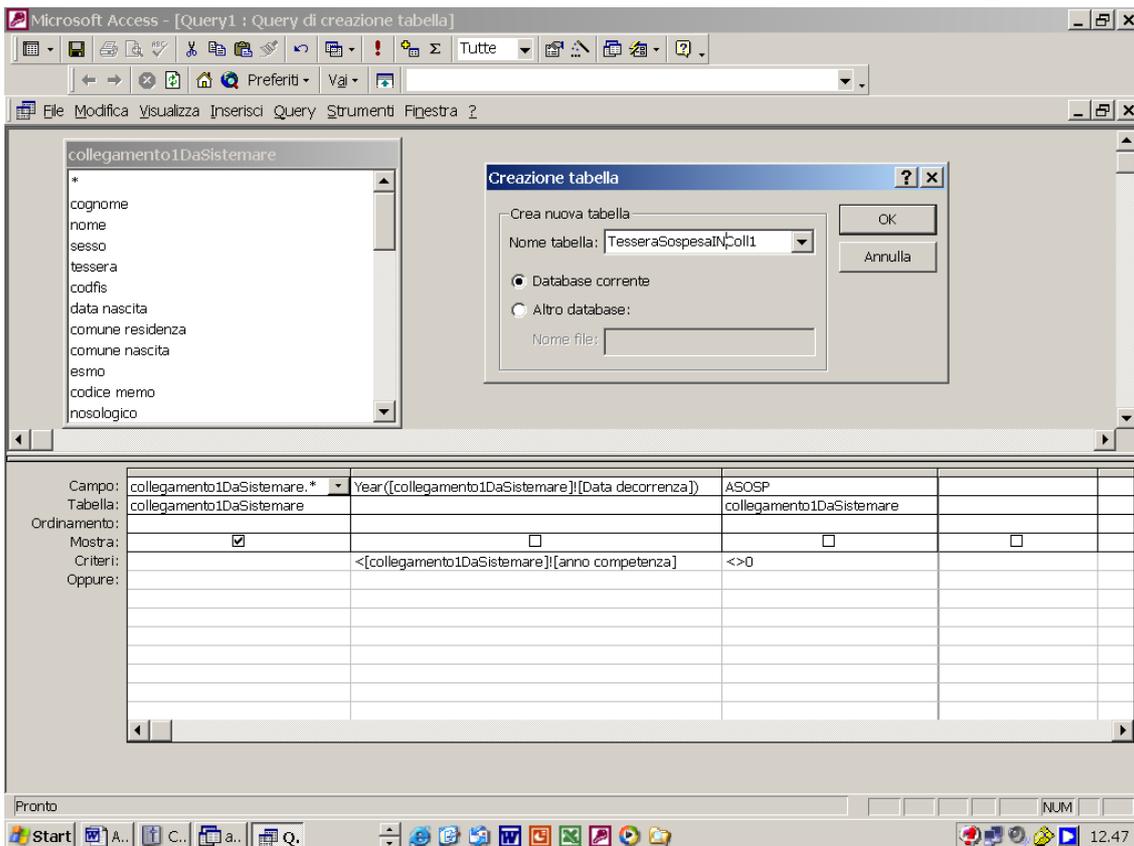
Successivamente si è fatto un controllo sui record della tabella appena creata, atto ad eliminare quelli relativi ad individui con tessera sanitaria scaduta precedentemente l'anno dell'analisi, per evitare di considerare persone non più assistite dall'ULSS 17.

Quest'operazione è avvenuta in maniera automatizzata in due fasi:

- 1) Si è creata una tabella contenente gli individui non assistiti dall'ULSS 17 al momento dell'analisi
- 2) Si è eliminata questa tabella da **collegamento1DaSistemare**

1) Nella tabella **collegamento1DaSistemare** si sono individuati i record in cui l'anno della data di decorrenza della tessera sanitaria, per una causa di sospensione differente da zero ($ASOSP <> 0$), risultava inferiore all'anno di competenza (l'anno in cui è stata eseguita l'analisi in Anatomia Patologica). Si è specificato che la causa di sospensione fosse diversa da zero in quanto quando il campo *ASOSP* è codificato con zero sta a significare che la tessera è ancora in vigore e che l'individuo ad essa associata è ancora un assistito dell'ULSS considerata. I 1892 record così selezionati dalla tabella **collegamento1DaSistemare** sono stati inseriti in una tabella chiamata **TesseraSospesaINColl1**.

Schermata n° 3.2



La schermata 3.2 mostra come sono stati selezionati i record nella tabella **collegamento1DaSistemare** per la creazione di **TesseraSospesaINColl1**.

Esempio n°3.1

Esempio di record presenti nella tabella: TesseraSospesaINColl1

cognome	nome		data nascita	anno competenza	Data decorrenza	ASOSP
C*****	A*****	10/12/36	2000	01/07/1996	1
F*****	S*****	25/01/1952	1998	02/08/1997	10

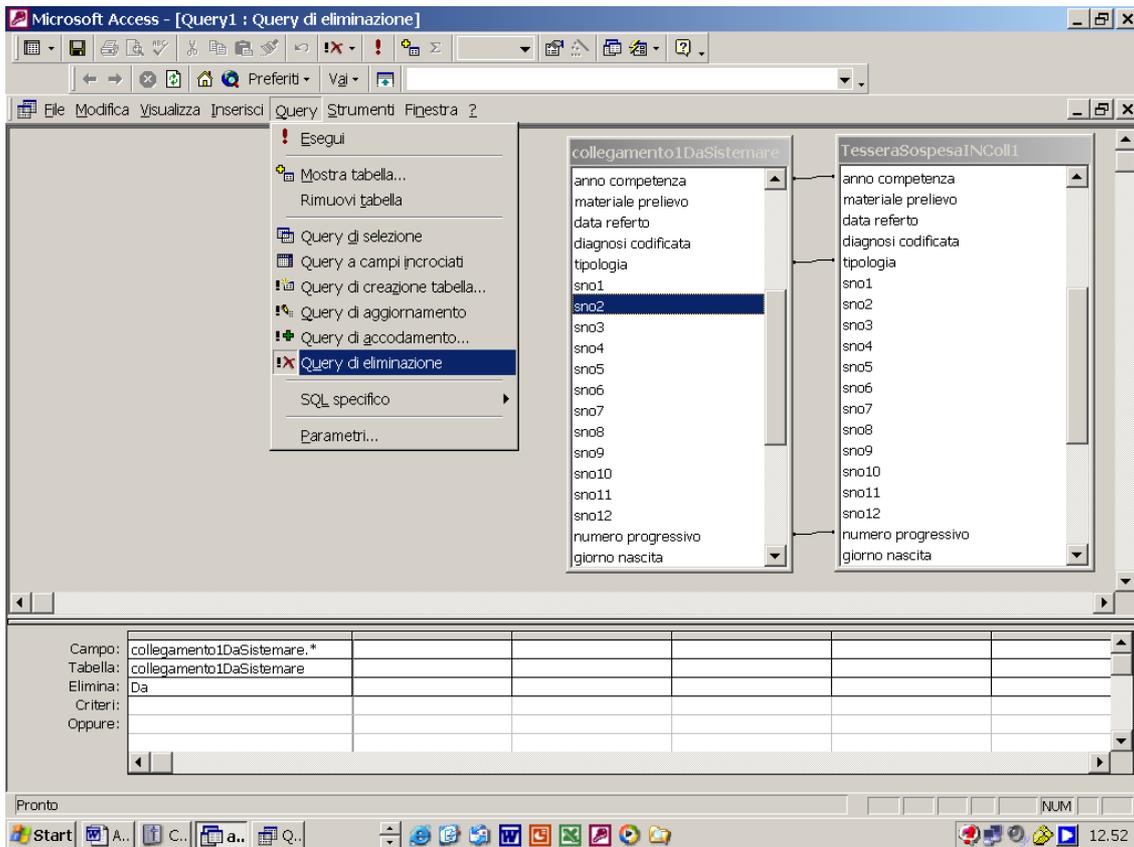
2) Si sono rimossi i record contenuti nella tabella **TesseraSospesaINColl1** dalla tabella **Collegamento1DaSistemare**.

Questo passaggio è avvenuto attraverso una query di eliminazione. Le variabili *anno competenza*, *tipologia* e *numero progressivo* relative alle tabelle **collegamento1DaSistemare** e **TesseraSospesaINColl1** sono state unite attraverso un join uno ad uno (con questo join vengono selezionati solo le righe in cui i campi collegati da entrambe le tabelle sono uguali) e si è specificato quale tabella dovesse essere sottratta all'altra.

Importante è stato l'utilizzo delle tre variabili del link: *anno competenza*, *tipologia* e *numero progressivo* in quanto questi tre campi sono sufficienti ad identificare univocamente i record e quindi nella query di eliminazione si è certi di quali record si vanno a sottrarre. Nelle query di eliminazione è, infatti, necessario che i record della tabella che si va ad eliminare non presentino duplicati.

La schermata n° 3.3 mostra l'impostazione della query di eliminazione sopra descritta.

Schermata n° 3.3



Alla fine di questi interventi nella tabella **collegamento1DaSistemare** sono rimasti 158.378 record (circa il 78,3% dei record contenuti nel database iniziale).

✓ Terzo step

La tabella **collegamento1DaSistemare** è stata rinominata e chiamata **collegamento1**.

✓ Quarto step

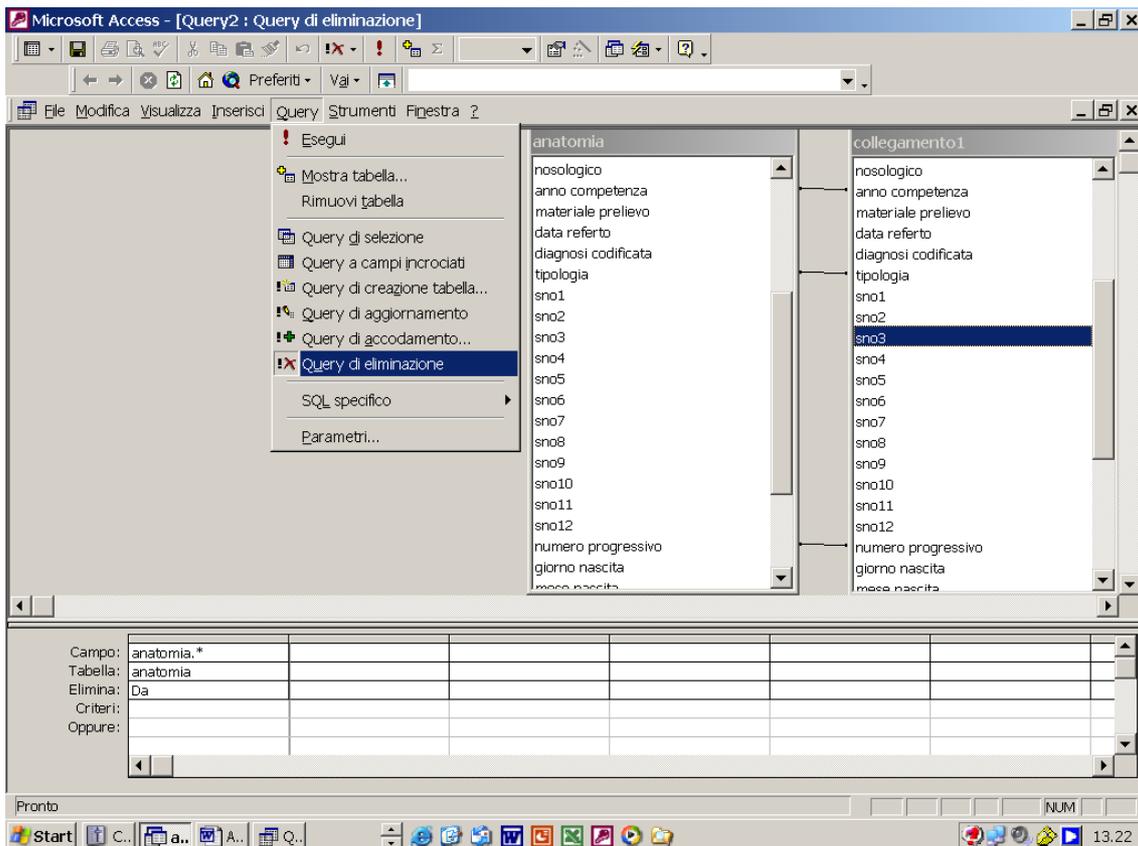
Prima di proseguire con il secondo incrocio si sono tolti i record contenuti in **collegamento1** (158.378 record) dalla tabella **anatomia** (202.014

record) per permettere che il controllo successivo (in cui saranno incrociati nuovamente le tabelle **anagrafe** e **anatomia**) potesse avvenire su record diversi rispetto a quelli contenuti e già “corretti” in **collegamento1**.

Questo passaggio è stato eseguito con una query di eliminazione. Le variabili *anno competenza*, *tipologia* e *numero progressivo* relative alle tabelle **anatomia** e **collegamento1** sono state unite attraverso un join uno ad uno (con questo join vengono selezionate solo le righe in cui i campi collegati da entrambe le tabelle sono uguali) e si è specificato quale tabella dovesse essere sottratta all'altra.

A questo punto in **anatomia** sono rimasti 43.636 record.

Schermata n° 3.4



La schermata n° 3.4 mostra la query di eliminazione effettuata tra la tabella **anatomia e collegamento1**

N.B.

I quattro step descritti nel primo incrocio sono stati eseguiti identicamente nei quattro linkage che seguiranno quindi successivamente si daranno per scontati i passaggi che hanno portato alla creazione di **collegamento2**, **collegamento3**, **collegamento4**, **collegamento5** per soffermarsi invece sui tipi di problematiche (incompletezze e inesattezze dei record) che con questo procedimento si sono cercati di risolvere.

2° incrocio. *tessera sanitaria + cognome + nome + data di nascita*

Con questo incrocio si sono recuperati 25 record che sono stati inseriti in una tabella chiamata collegamento2.

I record così recuperati sono quelli in cui il comune di nascita o era scritto in modo scorretto o mancava.

3° incrocio. *tessera sanitaria + cognome + data di nascita*

Sono risultati 1.484 record che sono stati inseriti in una tabella chiamata **collegamento3**.

Con questo linkage si sono recuperati i record in cui il nome era stato inserito in modo scorretto (basta anche un solo spazio, carattere differente per non far funzionare l'incrocio) non solo per errori comuni di digitazione, come possono essere le doppie o caratteri senza senso come è mostrato nella esempio n°3.2.

Esempio n°3.2

Esempio di errori dovuti a doppie o a caratteri senza senso

nome (anatomia)	ANOME (anagrafe)
BERTILA	BERTILLA
FRANCESCA/	FRANCESCA

Gli errori più frequenti verificatisi sono relativi a nomi composti (esempio n°3.3) o a nomi in cui compaiono alcune precise lettere che vengono scambiate con altre (esempio n°3.4).

Esempio n°3.3

Esempio di errori in un nome composto

ANNA MARIA che può anche essere scritto:

A.MARIA, A. MARIA, ANNA-MARIA, ANNAMARIA, MARIA, ANNA.

Esempio n°3.4

Esempio di errori dovuti a scambi di lettere

J e I, K e C, O con lo 0 (zero)

nome (anatomia)	ANOME (anagrafe)
IOLANDA	JOLANDA
CATIA	KATIA
JOLANDA	IOLANDA
KATIA	CATIA
FRANCO	FRANCO

In alcuni casi si è trovato che certe inesattezze sono imputabili direttamente al registro dell'Anagrafe. Per esempio si è osservato che quattro nomi in Anagrafe hanno al posto della O lo 0 (zero), due nomi invece terminano con una serie di puntini di sospensione e altri due hanno al loro interno una virgola (esempio n°3.5).

Esempio n°3.5

Esempio del tipo di inesattezze riscontrate nell'archivio dell'Anagrafe Assistiti

ANOME(anagrafe)
FRANCO
CINZIA...
MAR,CO

Non avendo la competenza per correggere il registro dell'Anagrafe ci si è limitati a far presente queste imprecisioni.

4° incrocio. *tessera sanitaria + cognome + nome*

Sono risultati 184 record che sono stati inseriti in una tabella chiamata **collegamento4**.

Con questo link si sono recuperati i record che presentavano errori nella data di nascita. Le inesattezze rilevate in questo campo sono dovute per lo più a errori di digitazione su tastiera. Molto spesso caricando nel database la data di nascita sono state inserite al posto di quelle esatte numeri che sulla tastiera risultavano occupare posizioni "confinanti" (esempio n°3.6).

Esempio n°3.6

Esempio di errori in valori numerici

12/08/1999 al posto di 12/08/1996

7	8	9
4	5	6
1	2	3
0	+/-	.

Altre inesattezze che si sono recuperate con questo link riguardano la data di nascita di alcune persone anziane partorite in casa e registrate in anagrafe solo qualche giorno dopo con data di nascita sbagliata.

5° incrocio. *cognome + nome + data di nascita*

Sono risultati 6.907 record che sono stati inseriti in una tabella chiamata **collegamento5**.

Con questo link si sono recuperati i record che presentavano inesattezze o mancanze nel codice sanitario.

Gli errori in questo campo sono per lo più dovuti ad una sbagliata digitazione su tastiera ma a volte anche ad errata associazione di tessera all'individuo.

Per esempio ad uno stesso individuo possono essere associate contemporaneamente due tessere sanitarie, una valida ed una scaduta o sospesa ed al momento del caricamento dei dati è stata scelta una di queste ultime due. Inoltre può verificarsi l'evenienza che al momento della prestazione la tessera inserita era valida ma al momento del controllo dell'archivio questa risultava decaduta.

N.B.

Con il quinto incrocio si è conclusa la fase di sistemazione e di recupero di informazioni anagrafiche avvenuta utilizzando solamente procedure automatizzate.

Fino a qui sono stati selezionati e corretti 166.978 record circa 82,7% dell'archivio iniziale.

Ora saranno invece proposti dei linkage, che definiremo “semi-automatici”, dove ad una iniziale selezione meccanica segue un intervento di cernita

manuale affidata alle capacità decisionali dell'operatore addetto al compito di ripulitura dell'archivio.

Quindi è evidente, ma lo sarà maggiormente man mano che verranno descritti gli incroci successivi, come operatori differenti possano alla fine arrivare a risultati discordanti.

Si veda precisamente di cosa si tratta.

6° incrocio. *tessera sanitaria + nome + data di nascita*

Dopo avere eseguito i primi due *step*, descritti precedentemente e comuni a tutti gli incroci (anche a quelli che seguiranno), la tabella **collegamento6daSistemare** (creatasi durante i due *step*) conteneva 271 record. In realtà dopo un controllo a video in cui sono stati affiancati il campo *cognome* (proveniente dall'archivio di anatomia) ed il campo *ACOGNO* (proveniente dall'archivio dell'anagrafe) e confrontato la similarità si sono scartati 28 record perché si è ritenuto plausibile che non si trattasse della stessa persona in quanto i cognomi risultavano decisamente diversi. Si sono recuperati invece i record in cui i cognomi si scostavano solo per alcuni caratteri (esempio n°3.7).

Esempio n°3.7

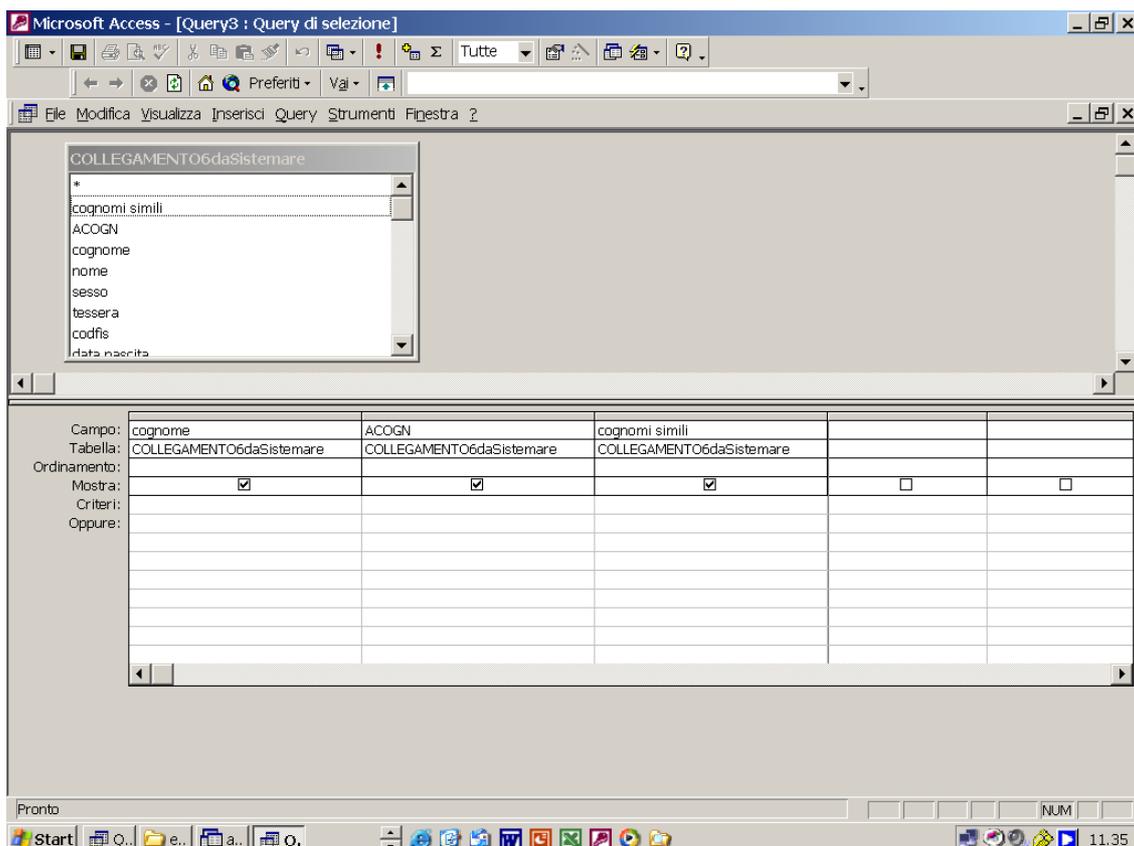
Esempio di cognomi considerati simili o diversi

ACOGN	cognome	simili
PRINI	BEHANOVA	no
CUIBATRE	TIENE	no
NOGUEIRA DA	NOGUEIRA DA	si
CAPUZZO	CAPPUZZA	si
PAJOLA	PAIOLA	si
BOMBOVA'	BOMBAVA'	si

Non sempre però è vero che cognomi diversi si riferiscano obbligatoriamente a persone differenti poichè può accadere che un individuo venga registrato in un archivio con il cognome del padre ed in un altro con quello della madre. Non avendo la possibilità di verificare queste ipotesi si è preferito perdere le suddette informazioni piuttosto di rischiare di lavorare con dati imprecisi.

La schermata n° 3.5 mostra come sono stati affiancati i campi *cognome* e *ACOGNO*. Il campo *cognomi simili* è stato compilato con sì o no a seconda che i record nei due campi confrontati fossero simili o meno.

Schermata n° 3.5



Nella tabella chiamata **collegamento6** sono stati inseriti i 243 record così controllati.

7° incrocio. *tessera + data di nascita*

Dopo aver eseguito i primi due *step* i record recuperati risultano 23. Si sono poi affiancati i campi *nome* e *cognome* dell'**anatomia** con *ACOGNO* e *ANOME* dell'**anagrafe** (esempio n°3.8) e scelti i **15** record in cui i campi confrontati a due a due risultavano molto simili e scartati 8 record in cui invece risultavano, a due a due (ed è il caso che si è verificato) o anche singolarmente, differenti (1 record in cui questi campi sono uguali sono già stati recuperati negli incroci precedenti n°3 o n°6).

Esempio n°3.8

Esempio di confronto tra cognomi e nomi di anatomia e dell'anagrafe

<i>ACOGNO</i>	<i>cognome</i>	<i>ANOME</i>	<i>nome</i>	<i>simili</i>
MICHELON SANTINELLO	MICHELON SANTIN	ANNA	ELLO ANNA	si
DAVIES	GILLIAN	GILLIAN	DAVIES	si
CILLO	FRATUCELLO	ANTONIO	NATALIA	no
FORNASIERO	FRANCESCO	FORNASIERO	MICHELA	no

La tabella **collegamento7** contiene i 15 record recuperati.

8° incrocio. *tessera*

I record incrociatesi sono 58 quelli recuperati, dopo aver seguito un procedimento uguale a quello sopraccitato sono 8. Si sono affiancati i campi *nome*, *cognome* e *data di nascita* provenienti dall'**anatomia** con quelli provenienti dall'**anagrafe** e se ne è osservata la similarità (esempio n°3.9).

Esempio n°3.9

Esempio di confronto tra cognomi, nomi e data di nascita di anatomia e dell'anagrafe

ACOGN	cognome	ANOME	nome	ADATA NASCITA	data nascita	simili
VASCON	PADOVAN	LINO	BAT	17-mar-38	17-ago-11	no
GIACOMIN	GIACOMIN	IOLE	JOLE	25-ago-23	24-ago-23	si
PADOVAN	CAPUZZO	BARBARA	ADELE	19-ago-69	01-lug-18	no
MANTOVAN	MANTOVAN	MARIA	ANNA	10-nov-29	11-nov-29	si
PERAZZOLO	PERAZZOLO	DEBHORA	DEBORA	11-nov-72	21-nov-72	si
MALIQATI	MALIGATI	DRITAN	DRITAN	21-mar-77	02-mar-77	si
AGOSTINI	CAVESTRO	TA	PIERINA	31-mag-29	16-set-40	no

La tabella **collegamento 8** contiene gli 8 record recuperati

9° incrocio. *cognome + data di nascita*

I record incrociatesi sono 5.043. Si è confrontata la somiglianza del campo *nome* proveniente dall'archivio dell'**anagrafe** con quello proveniente dall'archivio di **anatomia** e sono stati quindi recuperati 4.037 poi inseriti in una tabella chiamata **collegamento9**.

10° incrocio. *nome + data di nascita*

I record incrociatesi sono 1.787. Si è confrontata la somiglianza del campo *cognome* proveniente dall'archivio dell'**anagrafe** con quello proveniente dall'archivio di **anatomia** e sono stati quindi recuperati 1.654 poi inseriti in una tabella chiamata **collegamento10**.

3.3.4 Schema Riassuntivo

Di seguito è riportato uno schema riassuntivo (tabella n°3.5 degli incroci eseguiti e fin qui descritti.

Le celle contrassegnate in giallo indicano le variabili che sono state utilizzate per i link. Le celle contrassegnate in fucsia indicano le variabili su cui è stato eseguito un controllo manuale dopo aver avuto il risultato dell'incrocio.

Passaggi per la pulitura dell'archivio di Anatomia Patologica (1991-gennaio 2003)
 numero record iniziali: 202.014

Tabella n°3.5

collegamento	cognome	nome	tessera	data di nascita	comune di nascita	N. record	N. record rimanenti
collegamento 1						158.378 (78,3%)	
collegamento 2						25	
collegamento 3						1.484	
collegamento 4						184	
collegamento 5						6.907	
totale parziale						166.978 (82,7%)	35.036 (17,3%)
collegamento 6	simili					243	
collegamento 7	simili	simili				15	
collegamento 8	simili	simili				8	
collegamento 9		simili				4.037	
collegamento 10	simili					1.654	
totale finale						172.935 (85,6%)	29.079 (14,4%)

legenda	uguali	simili ¹⁾
---------	--------	----------------------

1)quando i dati differiscono per un numero o per poche lettere o hanno al loro interno caratteri impropri

3.4 Dati da considerare

Come già accennato precedentemente, con questo lavoro di ripulitura dell'archivio di Anatomia Patologica, si è cercato di stabilire e capire quale percorso e quali link potessero rendere il più standard possibile il recupero di informazioni imprecise o mancanti, in modo da poter applicare lo stesso procedimento ad altri archivi rendendoli più facilmente confrontabili ed utilizzabili per eventuali elaborazioni di più archivi aggregati.

È evidente che, se per le informazioni recuperate con i primi cinque linkage (precedentemente descritti), non c'è spazio per un intervento decisionale e discrezionale dell'operatore addetto alla sistemazione dell'archivio, questo non avviene per i record recuperati con gli incroci che vanno dal sesto al decimo. È quindi su queste ultime proposte di link definite "semiautomatiche" che si è più discusso: quanto si è disposti a perdere in oggettività pur di recuperare una quantità maggiore di record?

Si è deciso di considerare i dati recuperati e corretti fino all'incrocio numero sei.

L'archivio di Anatomia Patologica dopo il procedimento di ripulitura appena descritto contiene 167.221 record (82,8% dei record iniziali) provenienti dall'accodamento delle sei tabelle create durante gli *step* dei primi sei linkage:

collegamento1: 158.378record

collegamento2: 25 record

collegamento3: 1.484 record

collegamento4: 184 record

collegamento5: 6.907 record

collegamento6: 243 record

Le sei tabelle sono state unite tenendo conto per ogni collegamento di accodare i campi in modo da sostituire le informazioni anagrafiche imprecise o incomplete provenienti dall'Anatomia con quelle presumibilmente corrette provenienti dall'Anagrafe Sanitaria.

Seguono delle tabelle riassuntive dell'archivio di Anatomia Patologica contenenti i 167.221 record selezionati e corretti.

La tabella n°3.6 mostra gli esami effettuati negli anni 1991-2003 nei laboratori di Anatomia Patologica dell'ULSS 17 divisi per anno di competenza e tipologia di esame.

Tabella n° 3.6

Distribuzione per tipologia esame e per anno di competenza degli esami effettuati nei laboratori di Anatomia Patologica 1991-2003 (considero l'archivio corretto 167.221 record)

anno	TIPOLOGIA ESAME					Totale
	1 istologico	2 citologico	5 screening	6 citologico vaginale	9 screening	
1991	1					1
1992	3.266					3.266
1993	3.528					3.528
1994	3.776	1				3.777
1995	5.123	1				5.124
1996	9.993	373				10.366
1997	9.929	2.801				12.730
1998	10.397	2.558	1.949		25	14.929
1999	11.582	4.181	5.356	6.757	162	28.038
2000	11.772	4.410	5.324	8.604	362	30.472
2001	11.608	4.746	7.252	3.454	303	27.363
2002	11.179	4.552	6.096	5.275	367	27.469
2003	54	96	1	6	1	158
Totale	92.208	23.719	25.978	24.096	1.220	167.221

Anche la tabella n° 3.7 mostra il numero di esami effettuati negli anni 1991-2003 nei laboratori di Anatomia Patologica dell'ULSS 17 ma distinguendoli per anno di competenza e sesso.

Tabella n°3.7

Distribuzione per anno di competenza e per sesso degli esami effettuati nei laboratori di Anatomia Patologica 1991-2003 (considero l'archivio corretto 167.221 record)

anno competenza	F	M	Totale
1991	1		1
1992	2.106	1.160	3.266
1993	2.074	1.454	3.528
1994	2.122	1.655	3.777
1995	2.971	2.153	5.124
1996	5.996	4.370	10.366
1997	7.403	5.327	12.730
1998	9.957	4.972	14.929
1999	20.802	7.236	28.038
2000	23.026	7.446	30.472
2001	19.370	7.993	27.363
2002	19.795	7.674	27.469
2003	75	83	158
Totale	115.698	51.523	167.221

Si noti che il numero di analisi riguardanti il sesso femminile nel totale è più del doppio rispetto a quello delle analisi riguardanti gli uomini e questo distacco si nota maggiormente se si considerano singolarmente gli anni successivi al 1997 cioè quelli che hanno visto l'introduzione dell'analisi screening e successivamente dell'analisi citologico vaginale (esami esclusivamente femminili).

Il numero di analisi effettuate sono 167.221 relative a 73.415 utenti differenti il 68% dei quali sono femmine (tabella n°3.8).

Tabella n°3.8

Distribuzione per sesso degli utenti presenti nell'archivio di Anatomia Patologica 1991-2003 (considero l'archivio corretto 167.221 record)

sesso	conteggio	%
F	49774	67,8
M	23641	32,2
Totale	73415	100,0

La proporzione tra il numero di record riguardanti maschi e femmine e la proporzione tra il numero di soggetti uomini e donne è molto simile.

Si veda tabella n° 3.9.

Tabella n°3.9

	maschi	femmine	n°M / n°F
numero record	51.523	115.698	0,45
numero utenti	23.641	49.774	0,47

Le descrizioni successive riguardano i 73.415 individui presenti nell'archivio. Di essi si è considerato il primo anno di competenza con cui sono presenti nell'archivio, la fascia di età in cui ricadevano al momento dell'analisi ed il sesso.

La tabella n° 3.10 mostra il numero di utenti suddiviso per sesso, per anno di competenza e per fascia di età. Per ogni anno è stata evidenziata, in rosa per le femmine ed in azzurro per i maschi, la classe di età considerata modale nella distribuzione.

Tabella n°3.10

Distribuzione per anno di competenza, per fascia d'età e per sesso degli individui presenti nell'archivio di Anatomia Patologica anni 1991-2003 (73.415 record)

anno competenza	Fascia d'età																		Totale				
	0-4		5-9		10-14		15-24		25-34		35-44		45-54		55-64		65-74		75 e più		f	m	
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	
1991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1992	2	5	6	8	24	23	75	217	103	329	106	505	110	194	199	137	127	127	127	137	127	1.812	926
1993	9	23	20	24	22	17	74	224	90	271	105	325	124	179	213	127	95	127	95	127	95	1.525	914
1994	16	18	23	29	20	20	93	222	0	223	104	236	124	175	213	130	92	130	92	130	92	1.337	992
1995	29	38	24	22	26	31	92	256	147	317	158	408	165	293	262	195	151	262	195	151	151	2.004	1.320
1996	58	101	60	85	66	58	238	596	288	692	336	816	324	507	601	360	345	601	360	345	345	4.049	2.837
1997	54	79	62	70	47	59	199	690	331	686	325	701	370	545	555	425	361	555	425	361	361	4.151	2.849
1998	47	72	46	52	55	49	205	925	318	1.145	314	957	352	558	471	444	316	471	444	316	316	5.371	2.603
1999	52	87	65	70	55	57	247	2.236	324	2.416	342	1.799	379	892	590	539	356	590	539	356	356	10.143	2.934
2000	52	87	51	66	60	60	209	1.939	314	1.837	365	1.395	387	753	524	373	373	524	373	373	373	8.496	2.867
2001	57	78	48	72	46	38	197	1.464	329	1.617	360	1.165	412	455	531	439	334	455	439	334	334	6.847	2.808
2002	38	69	43	66	46	45	170	971	332	780	347	455	365	361	434	295	295	361	460	295	295	4.025	2.572
2003	0	1	0	0	0	0	0	0	1	4	1	1	2	4	10	2	2	4	10	2	2	13	19
totale	414	658	448	564	467	457	3.615	9.740	2.577	10.317	2.863	8.764	3.114	7.312	4.033	4.916	4.603	3.781	2.847	3.781	2.847	49.774	23.641

Si noti che se per gli uomini la fascia di età più popolosa risulta essere quasi sempre quella dei 65-74 anni per le donne c'è una maggiore variazione. Fino al 1997 la moda ricadeva sempre nella fascia di età 45-54 anni per poi spostarsi negli anni successivi in quelle dei 35- 44 o 25-34 anni.

Questo può essere avvenuto per diversi fattori anche se sembrerebbe logico ipotizzare che lo screening abbia influito notevolmente in questo cambiamento .

Per dare un'idea di come possa avere influito lo screening si sono selezionati solo gli individui che risultavano avere come primo accesso nell'archivio di Anatomia Patologica un'analisi di questo tipo e sono stati inseriti poi in una tabella tenendo conto dell'anno di competenza, della fascia di età e del sesso.

Si veda tabella n° 3.11.

Tabella n°3.11

Distribuzione per anno di competenza, per fascia d'età e per sesso degli individui presenti nell'archivio di Anatomia Patologica anni 1991-2003 avendo considerato come prima competenza solo le analisi di tipo screening (15.042 record)

anno competenza	Fascia d'età														Totale							
	0-4		5-9		10-14		15-24		25-34		35-44		45-54		55-64		65-74		75 e più			
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m		
1991																						
1992																						
1993																						
1994																						
1995																						
1996																						
1997																						
1998					325						507					356					366	
1999					958						1.229					956					849	
2000					810						889					717					839	
2001					1.021						1.253					918					1.024	
2002					574						479					288					333	
2003																						
totale					3.688						4.357	1				3.235					3.411	
																						15.041
																						1

Anche in questa tabella è stata selezionata in rosa la classe con maggior numero di individui per anno di competenza che come si vede risulta sempre essere inferiore a 45-54 anni.

Come si vede gli utenti che hanno avuto come primo accesso un'analisi di tipo screening sono 15.042. Si nota subito la presenza di un maschio avvenimento alquanto singolare visto che il tipo di analisi selezionata riguarda solamente le donne . Dopo un controllo eseguito sull'individuo "discriminato" si è trovato che l'errore era dovuto ad una scorrettezza presente nello scarico dell'Anagrafe con cui sono stati eseguiti i controlli di "rifinitura" dell'archivio iniziale. Infatti erroneamente quell'individuo in Anagrafe era stato codificato come uomo pur essendo una donna.

Successivamente sono state create altre due tabelle con struttura identica alla tabella n°3.10 ma con numero di utenti differenti.

Nella tabella n°3.12 il primo anno di comparsa nell'archivio di Anatomia Patologica per ogni individuo è stato calcolato togliendo a priori dall'archivio tutte le analisi di tipologia screening. Quindi nella tabella non compaiono i soggetti presenti nell'archivio di Anatomia solamente per il tipo di esame sopraccitato.

Nella tabella n° 3.13 invece sono stati tolti tutti gli individui il cui primo anno di accesso era legato ad un esame di screening anche se poi successivamente comparivano nell'archivio per un esame differente.

Ovviamente, siccome lo screening è un esame rivolto solamente alle donne e che è presente nell'archivio dal 1998, la differenza del numero di soggetti è riscontrabile solamente a partire dal 1998 per il sesso femminile.

Tabella n°3.12

Distribuzione per anno di competenza, per fascia d'età e per sesso degli individui presenti nell'archivio di Anatomia Patologica anni 1991-2003 avendo escluso a priori tutte le analisi di tipo screening (62.008 record)

anno competenza	Fascia d'età																		Totale			
	0-4		5-9		10-14		15-24		25-34		35-44		45-54		55-64		65-74		75 e più		f	m
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m
1991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1992	2	5	6	8	24	23	118	75	217	103	329	106	505	110	280	170	194	199	137	127	1.812	926
1993	9	23	20	24	22	17	132	74	224	90	271	105	325	124	216	149	179	213	127	95	1.525	914
1994	16	18	23	29	20	20	118	93	222	126	223	104	236	124	174	173	175	213	130	92	1.337	992
1995	29	38	24	22	26	31	204	92	256	147	317	158	408	165	252	254	293	262	195	151	2.004	1.320
1996	58	101	60	85	66	58	350	238	596	288	692	336	816	324	544	461	507	601	360	345	4.049	2.837
1997	54	79	62	70	47	59	400	199	690	331	686	325	701	370	541	500	545	555	425	361	4.151	2.849
1998	47	72	46	52	55	49	358	205	621	318	689	314	633	352	495	454	537	471	444	316	3.925	2.603
1999	52	87	65	70	55	57	629	247	1.426	324	1.399	341	1.003	379	694	482	767	590	539	356	6.629	2.933
2000	52	87	51	66	60	60	593	209	1.315	314	1.230	365	949	387	633	482	654	524	523	373	6.060	2.867
2001	57	78	48	72	46	38	331	197	612	329	636	360	488	412	421	457	465	531	439	334	3.543	2.808
2002	38	69	43	66	46	45	381	170	578	332	588	347	395	365	366	449	422	434	460	295	3.317	2.572
2003	0	1	0	0	0	0	1	0	0	1	4	1	2	2	2	2	4	10	2	2	15	19
totale	414	658	448	564	467	457	3.615	1.799	6.757	2.703	7.064	2.862	6.462	3.114	4.618	4.033	4.742	4.603	3.781	2.847	38.368	23.640

Tabella n°3.13

Distribuzione per anno di competenza, per fascia d'età e per sesso degli individui presenti nell'archivio di Anatomia Patologica anni 1991-2003 avendo tolto le prime competenze dello screening (record 59.730)

anno competenza	Fascia d'età														Totale							
	0-4		5-9		10-14		15-24		25-34		35-44		45-54		55-64		65-74		75 e più			
	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m
1991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1992	2	5	6	8	24	23	118	75	217	103	329	106	505	110	280	170	194	199	137	127	1.812	926
1993	9	23	20	24	22	17	132	74	224	90	271	105	325	124	216	149	179	213	127	95	1.525	914
1994	16	18	23	29	20	20	118	93	222	126	223	104	236	124	174	173	175	213	130	92	1.337	992
1995	29	38	24	22	26	31	204	92	256	147	317	158	408	165	252	254	293	262	195	151	2.004	1.320
1996	58	101	60	85	66	58	350	238	596	288	692	336	816	324	544	461	507	601	360	345	4.049	2.837
1997	54	79	62	70	47	59	400	199	690	331	686	325	701	370	541	500	545	555	425	361	4.151	2.849
1998	47	72	46	52	55	49	358	205	621	318	689	314	633	352	495	454	537	471	444	316	3.925	2.603
1999	52	87	65	70	55	57	629	247	1.399	324	1.328	341	949	379	661	482	761	590	539	356	6.438	2.933
2000	52	87	51	66	60	60	593	209	1.203	314	1.046	365	769	387	514	482	622	524	523	373	5.433	2.867
2001	57	78	48	72	46	38	331	197	513	329	476	360	326	412	285	457	425	531	439	334	2.946	2.808
2002	38	69	43	66	46	45	381	170	433	332	332	347	189	365	183	449	351	434	460	295	2.456	2.572
2003	0	1	0	0	0	0	1	0	0	1	4	1	1	2	1	2	4	10	2	2	13	19
totale	414	658	448	564	467	457	3.615	1.799	6.374	2.703	6.393	2.862	5.859	3.114	4.146	4.033	4.593	4.603	3.781	2.847	36.090	23.640

Il grafico n° 3.2 sintetizza graficamente i risultati delle tabelle n° 3.10-3.12-3.13 considerando solo le donne e gli anni 1998-2002 (si veda tabella n°3.14).

Grafico n°3.2

Confronto fra il numero di individui (osservando solo il genere femminile) al primo accesso in Anatomia Patologica considerando o meno lo screening.

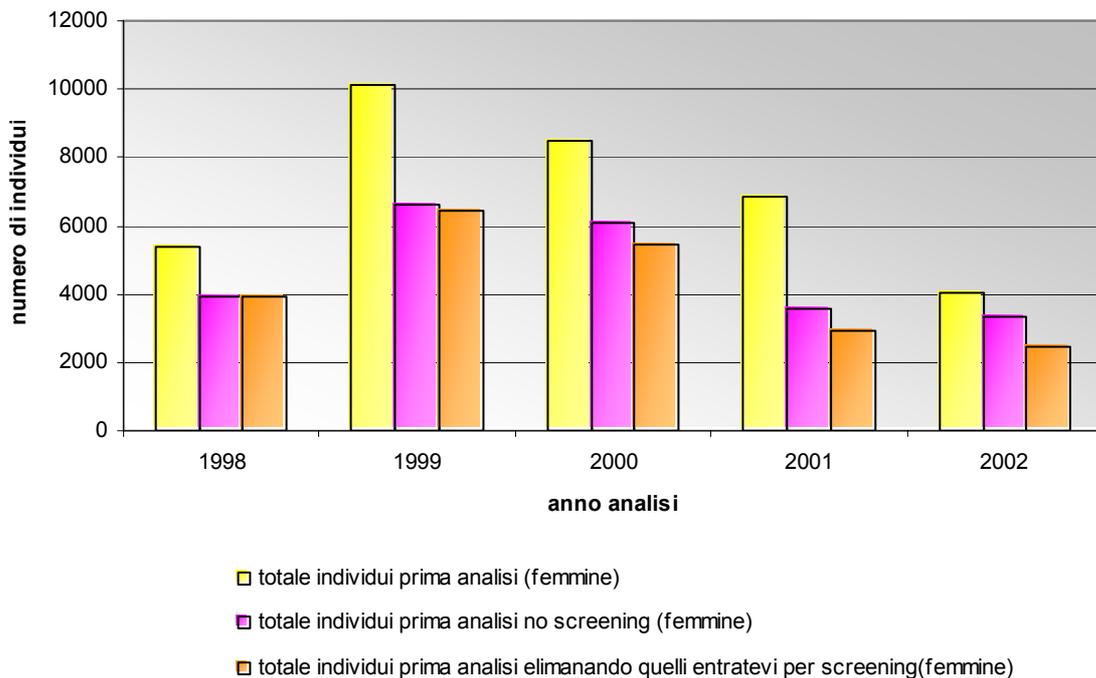


Tabella n° 3.14

Distribuzione per anno di competenza degli individui (osservando solo il genere femminile) al primo accesso in Anatomia Patologica considerando o meno lo screening

Anno analisi	totale individui prima analisi (femmine)	totale individui prima analisi no screening (femmine)	totale individui prima analisi eliminando quelli entratevi per screening (femmine)
1998	5.371	3.925	3.925
1999	10.143	6.629	6.438
2000	8.496	6.060	5.433
2001	6.847	3.543	2.946
2002	4.025	3.317	2.456
Totale	34.882	23.474	21.198

CAPITOLO 4

ARCHIVIO DEI RICOVERI

L'archivio dei ricoveri è stato fornito dall'U.O. Statistica, Flussi Informativi e Comunicazione in formato .dbf di Access.

L'archivio iniziale è composto da 290.357 record relativi a tutti i ricoveri dei residenti dell'ULSS 17 avvenuti in Veneto dal 1995 ai primi nove mesi del 2002 più i ricoveri avvenuti nell'ULSS 17 dei suoi residenti per gli ultimi tre mesi del 2002.

Di seguito è riportato il tracciato record dell'archivio .dbf.

Nome Campo	Tipo	Dimensione
Anno	testo	50
RegioneInviante	testo	50
UlssInviante	testo	50
Istituto	testo	50
Hsp11Bis	testo	50
SchedaSDO	testo	50
Cognome*Nome	testo	50
Cognome	testo	50
Nome	testo	50
CodiceFiscale	testo	50
CodiceAssistito	testo	50
ComuneNascita	testo	50
DataNascita	data/ora	8
Sesso	testo	50
ComuneResidenza	testo	50
Cittadinanza	testo	50
StatoCivile	testo	50
RegioneResidenza	testo	50
UlssResidenza	testo	50
RegimeRicovero	testo	50
CodiceMedico	testo	50
DataRicovero	data/ora	8
RepartoRicovero	testo	50
OnereDegenza	testo	50
Provenienza	testo	50
TipoRicovero	testo	50
MotivoRicovero	testo	50
MotivoDH	testo	50

Traumatismo	testo	50
DataTRasferimento1	data/ora	8
RepartoTrasferimento1	testo	50
DataTRasferimento2	data/ora	8
RepartoTrasferimento2	testo	50
DataTRasferimento3	data/ora	8
RepartoTrasferimento3	testo	50
DataTRasferimento4	data/ora	8
RepartoTrasferimento4	testo	50
GiorniPermesso	testo	50
GiorniPresenzaDH	testo	50
DataDimissione	data/ora	8
RepartoDimissione	testo	50
ReDimi	testo	50
ModalitàDimissione	testo	50
RiscontroAutoptico	testo	50
Codifica	testo	50
Diagnosi1	testo	50
Diagnosi2	testo	50
Diagnosi3	testo	50
Diagnosi4	testo	50
Diagnosi5	testo	50
Diagnosi6	testo	50
DataIntervento1	data/ora	8
Intervento1	testo	50
RepartoIntervento1	testo	50
DataIntervento2	data/ora	8
Intervento2	testo	50
RepartoIntervento2	testo	50
DataIntervento3	data/ora	8
Intervento3	testo	50
RepartoIntervento3	testo	50
DataIntervento4	data/ora	8
Intervento4	testo	50
RepartoIntervento4	testo	50
DataIntervento5	data/ora	8
Intervento5	testo	50
RepartoIntervento5	testo	50
DataIntervento6	data/ora	8
Intervento6	testo	50
RepartoIntervento6	testo	50
AccessiDH	testo	50
DRG	testo	50
Importo	precisione doppia	8
PosizioneContabile	testo	50
DistrettoResidenza	testo	50
DistrettoResidenzaDescrizione	testo	50

DistrettoResidenzaDescrizione	testo	50
GiornateDegenza	testo	50
EtaDimissione	testo	50
AnnoNascita	testo	50
TrimestreDimissione	testo	50
TipoDH	testo	50
TesseraElena	testo	50
MMGoPLS	testo	50

4.1 Modalità di compilazione.

Anno: l'anno della dimissione (aaaa)

RegioneInviante: codice a 3 caratteri che identifica la regione che invia i dati

UlssInviante: codice a tre caratteri che identifica l'ULSS

Istituto: Codice del Ministero della sanità. E' composto da 3 caratteri che identificano la regione, 3 caratteri che costituiscono una numerazione progressiva all'interno della regione e 2 caratteri che costituiscono una numerazione progressiva attribuita agli istituti facenti parte della stessa struttura di ricovero ("subcodice"). I primi 6 caratteri andranno rilevati nel campo "Istituto"; i rimanenti 2 andranno rilevati nel campo "Hsp11Bis".

Per la rilevazione fino al 1999 si continuerà ad utilizzare il codice a 6 caratteri. A partire dall'anno 2000, le aziende che abbiano compilato modelli HSP11bis, utilizzeranno il codice a 8 caratteri.

Hsp11Bis: Da valorizzare a partire dall'anno 2000.

SchedaSDO: Coincide con il numero identificativo della cartella clinica. Le prime 2 cifre identificano l'anno, le rimanenti 6 un numero progressivo

all'interno dell'anno.

Cognome*Nome: Cognome e nome divisi da un asterisco (questo tipo di codifica è utilizzata per gli anni 1995 e 1996)

Cognome: Il cognome del paziente

Nome: Il nome del paziente

CodiceFiscale: Un codice a 16 cifre che identifica univocamente un individuo

CodiceAssistito: Un codice regionale individuale di 9 cifre. Campo da valorizzare solo per gli assistiti della Regione Veneto.

ComuneNascita: Se si tratta di comune italiano, si adoperano i codici ISTAT dei comuni (3 campi per la provincia e 3 campi per i comuni). Se si tratta di stato estero, si indica 999 nei primi 3 campi e il codice ISTAT degli stati esteri nei secondi 3.

DataNascita: La data in formato gg/mm/aaaa

Sesso: Il maschio è codificato con 1, la femmina è codificata con 2

ComuneResidenza: si vedano istruzioni per il *ComuneNascita*

ComuneResidenzaDescrizione: Il nome del comune di residenza

Cittadinanza: Codificato con 100 per i pazienti con cittadinanza italiana. Codice a 3 caratteri definito dal ministero dell'interno per i pazienti con cittadinanza estera.

StatoCivile: Codificato fino al 31/12/1999 con:

1 non coniugato/a

2 coniugato/a

3 separato/a

4 divorziato/a

5 vedovo/a

Dal 1/1/2000 codificato con:

1 celibe/nubile

2 coniugato/a

3 separato/a

4 divorziato/a

5 vedovo/a

6 non dichiarato

RegioneResidenza: è la regione a cui appartiene il comune in cui il paziente risiede. E' impiegato un codice a 3 caratteri definito dal Ministero della Sanità. Nel caso in cui il paziente risieda all'estero si indica il codice dello stato estero secondo la classificazione del Ministero dell'Interno.

UlssResidenza: è l'Azienda ULSS che comprende il comune in cui il paziente risiede. E' impiegato un codice a 3 caratteri definito dal Ministero della Sanità. La variabile non è compilata nel caso in cui il paziente sia residente all'estero.

RegimeRicovero: distingue il ricovero ordinario dal ricovero diurno. È codificato con:

1 ricovero ordinario

2 ricovero diurno

CodiceMedico: è utilizzato il codice regionale dei medici o il codice fiscale

del medico specialista

DataRicovero: E' utilizzato il formato gg/mm/aaaa. Identifica la data del ricovero. Nei casi di ricovero diurno, si intende per data di ricovero la data di inizio del ciclo di trattamento.

RepartoRicovero: Identifica il reparto in cui il paziente viene ricoverato. Nel caso di ricovero diurno indica il reparto di trattamento. E' identificato da un codice a 4 cifre: le prime 2 identificano la disciplina, le rimanenti il progressivo di reparto.

OnereDegenza: Identifica chi deve sostenere le spese per il ricovero.

1 SSN

2 SSN + eventuali spese alberghiere (non valido per i ricoveri diurni)

3 Rimborso a carico del SSN per ricovero in casa di cura non convenzionata

4 ricovero a carico del paziente

9 Altro

Dal 1/1/2000 si utilizza la seguente codifica:

1 SSN

2 prevalentemente a carico SSN con parte delle spese a carico del paziente (differenza alberghiera)

3 Rimborso successivo del SSN

4 Ricovero senza oneri per SSN

5 prevalentemente a carico SSN con parte delle spese a carico del paziente (libera professione)

6 prevalentemente a carico SSN con parte delle spese a carico del paziente (libera professione e differenza alberghiera)

7 SSN (pazienti stranieri provenienti da paesi convenzionati con il SSN)

8 SSN (pazienti stranieri con dichiarazione di indigenza)

Provenienza: identifica la provenienza del paziente. È codificato per la degenza ordinaria con:

1 paziente che acceda all'istituto senza proposta di ricovero formulata da un sanitario

2 paziente inviato all'istituto dal medico/pediatra di base

3 paziente inviato all'istituto di cura da uno specialista

4 paziente trasferito da istituto di ricovero e cura pubblico

5 paziente trasferito da istituto di ricovero e cura privato

6 paziente trasferito dal ricovero diurno

9 altro

È codificato per i dati di ricovero diurno con:

1 trasferito da regime di degenza ordinaria

2 ricovero diurno permanente

9 altro

Dal 1/1/2000 è stata utilizzata la seguente codifica:

1 paziente che acceda all'istituto senza proposta di ricovero formulata da un medico

2 paziente inviato all'istituto con proposta del medico di base

3 ricovero precedentemente programmato dallo stesso istituto di cura

4 paziente trasferito da istituto di ricovero e cura pubblico

5 paziente trasferito da istituto di cura privato accreditato

6 paziente trasferito da istituto di cura privato non accreditato

7 paziente trasferito da altro regime di ricovero nello stesso istituto

9 altro

Da non compilare nella scheda di dimissione del neonato relativa al ricovero in cui è avvenuta la nascita

TipoRicovero: Identifica il tipo di ricovero. È codificato con:

- 1 ricovero programmato non urgente
- 2 Ricovero urgente
- 3 TSO
- 4 TSV

Dal 1/1/2000 la codifica è la seguente:

- 1 ricovero programmato non urgente
- 2 Ricovero urgente
- 3 TSO ricovero per trattamento sanitario obbligatorio
- 4 ricovero programmato con preospedalizzazione.

Da non compilare nella scheda di dimissione del neonato relativa al ricovero in cui è avvenuta la nascita.

MotivoRicovero: Identifica il motivo di ricovero secondo un codice ministeriale. È codificato con:

- 1 ricovero elettivo per intervento chirurgico
- 2 Primo ricovero in ambito ospedaliero effettuato per la specifica diagnosi
- 3 Altro

MotivoDH: Identifica il motivo del ricovero in regime diurno. Questo campo viene compilato a partire dal 1/1/2000 e solo in caso in cui la variabile “regime ricovero” sia codificata =2. È codificato con:

- 1 ricovero diurno diagnostico
- 2 ricovero diurno chirurgico
- 3 ricovero diurno terapeutico
- 4 ricovero diurno riabilitativo

Traumatismo: Identifica traumi ed intossicazioni. È codificato con:

- 1 infortunio in ambiente di lavoro

2 infortunio in ambiente domestico

3 incidente stradale

4 altro luogo o tipo di incidente / intossicazione

5 violenza altrui

6 autolesione o tentativo di suicidio

9 altro

Dal 1/1/2000 si utilizza la seguente codifica:

1 infortunio sul lavoro

2 infortunio in ambiente domestico

3 incidente stradale

4 violenza altrui, indipendentemente dal luogo dove è avvenuta

5 autolesione o tentativo di suicidio, indipendentemente dal luogo dove è avvenuta

6 altro tipo di incidente o di intossicazione

DataTrasferimento1...DataTrasferimento4: la data in cui il paziente viene trasferito in reparti interni.

RepartoTrasferimento1...RepartoTrasferimento1: il reparto in cui un paziente viene trasferito (cfr reparto di ammissione)

GiorniPermesso: numero di giorni di permesso

GiorniPresenzaDH: numero di giorni di presenza in ricovero diurno. Dà il numero di accessi dell'utente alla struttura di ricovero nell'ambito dello stesso ciclo di assistenza.

DataDimissione: data di dimissione o di morte

RepartoDimissione: il reparto da cui si è dimessi (cfr reparto di

ammissione)

ReDimi: è codificato con i primi 2 caratteri del codice di “*RepartoDimissione*”, identifica la disciplina.

ModalitàDimissione: identifica la modalità di dimissione. Per la degenza ordinaria:

- 1 deceduto
- 2 dimissione ordinaria al domicilio del paziente
- 3 dimissione ordinaria presso istituto di lungodegenza o RSA
- 4 dimissione protetta al domicilio del paziente
- 5 dimissione volontaria
- 6 trasferimento ad altro istituto di ricovero pubblico o privato per acuti
- 7 trasferimento al day hospital interno dell'ospedale

Per i dati di ricovero diurno impiegare la seguente codifica:

- 1 termine programmato
- 2 il paziente non si presenta agli appuntamenti prefissati
- 3 paziente trasferito in regime di degenza ordinaria
- 4 Day hospital permanente
- 5 deceduto

Dal 1/1/2000 si utilizza la seguente codifica:

Valido sia per la degenza ordinaria sia per il ricovero diurno:

- 1 deceduto
- 2 dimissione ordinaria al domicilio del paziente
- 3 dimissione ordinaria presso Residenza Sanitaria Assistenziale (RSA)
- 4 dimissione al domicilio del paziente con attivazione di ospedalizzazione domiciliare
- 5 dimissione volontaria (da utilizzare anche nei casi in cui il paziente in

ciclo di trattamento diurno non si sia presentato durante il ciclo programmato)

6 trasferimento ad altro istituto di ricovero e cura, pubblico o privato per acuti

7 trasferimento ad altro regime di ricovero nell'ambito dello stesso istituto

8 trasferimento ad altro istituto pubblico o privato di riabilitazione

9 dimissione ordinaria con attivazione di assistenza domiciliare integrata.

RiscontroAutoptico: nel caso di decesso del paziente segnala se è stata eseguita autopsia.

1 è stata eseguita autopsia

2 non è stata eseguita autopsia

Codifica: carattere che identifica la codifica

Diagnosi1: diagnosi principale alla dimissione. Codice ICD9. Dal 1/1/2000 codice ICD9CM

Diagnosi2...Diagnosi6: diagnosi secondaria alla dimissione. Codice ICD9. Dal 1/1/2000 codice ICD9CM

DataIntervento1: data intervento principale o parto

Intervento1: intervento chirurgico principale o parto

RepartoIntervento1: reparto intervento principale o parto

DataIntervento2...DataIntervento6: data altro intervento

Intervento2...intervento6: altro intervento

RepartoIntervento2...RepartoIntervento6: reparto altro di intervento

AccessiDH: è un codice che identifica il numero di accessi al day hospital

DRG: è un codice che identifica il raggruppamento omogeneo di diagnosi.
Si impiega il codice HCFA 10.0

Importo: importo degenza. Si usa il tariffario regionale.

PosizioneContabile: codice ad una cifra che identifica la posizione contabile

DistrettoResidenza: codice ad una cifra che identifica il distretto di appartenenza

DistrettoResidenzaDescrizione: descrizione del distretto di appartenenza

EtaDimissione: l'età al momento della dimissione

AnnoNascita: anno di nascita (formato aaaa)

TrimestreDimissione: unico carattere che identifica il trimestre di dimissione

TipoDH: identifica il tipo di day-hospital.

N.B. Prima di occuparsi della correttezza e della completezza dei dati contenuti in questo database è stato necessario, dove possibile, uniformare le informazioni nei campi tenendo conto delle piccole differenze di codifica esistenti poiché negli anni si sono modificati i metodi d'inserimento dei dati. Per esempio per gli anni 1995 e 1996 non esistono due campi separati per cognome e nome ma queste informazioni si trovano in uno stesso campo divisi da un asterisco (es. Rossi*Mario). Inoltre per questi anni è presente solo il regime di ricovero ordinario e non day-hospital. Per gli anni

che vanno dal 1995 al 1998 i campi attribuiti alla codifica della diagnosi sono 4 mentre successivamente saranno 6.

4.2 Dati disponibili (descrizione dell'archivio iniziale 290.357 record)

Sebbene l'archivio dei ricoveri sia già stato sottoposto ad alcuni controlli e modifiche operati dall'U.O. Statistica, Flussi Informativi e Comunicazione sui *nomi e cognomi* degli stranieri e sui *medici* di base, prima di poter essere utilizzato per elaborazioni affidabili necessita di controlli e correzioni più approfondite visto mancanze o imprecisioni facilmente riscontrabili per quanto concerne le informazioni anagrafiche: nome, cognome, data di nascita, sesso, codice sanitario, codice fiscale, comune di nascita e comune di residenza.

Di seguito sono riportate alcune tabelle relative alla completezza dei record dal punto di vista della completezza anagrafica (Tabella n°4.1).

Tabella n°4.1

Composizione iniziale dei campi di tipo anagrafico

Sesso		%
1	137.042	47,2
2	153.315	52,8
totali	290.357	100

campo nome		%
non compilate	237	0,1
compilate	290.120	99,9
totali	290.357	100,0

campo comune di residenza		%
non compilate	2	0,0
compilate	290.355	100,0
totali	290.357	100,0

campo comune di nascita		%
non compilate/0	86.463	29,8
compilate	203.894	70,2
totali	290.357	100,0

campo cognome		%
compilate	289.951	99,9
non compilate	295	0,1
anonimo	72	0,0
privacy	39	0,0
totali	290.357	100,0

campo data nascita		%
non compilate	40	0,0
compilate	290.317	100,0
totali	290.357	100,0

campo tessera		%
non compilate/0	17.222	5,9
compilate	273.135	94,1
totali	290.357	100,0

campo codice fiscale		%
non compilate/0	110.546	38,1
compilate	179.811	61,9
totali	290.357	100,0

Come si può notare dalle tabelle l'unico campo ad avere un consistente numero di dati mancanti è il *comune di nascita*. Ma questa incompletezza del campo è dovuta anche in parte ad una svista dell'operatore che ha fornito l'archivio in quanto ha accodato in modo impreciso i record relativi al 1998 e per quello anno il campo comune di nascita è risultato vuoto.

La tabella ed il grafico seguenti mostrano la distribuzione dei ricoveri negli anni considerati (Tabella n°4.2 ; Grafico n°4.1).

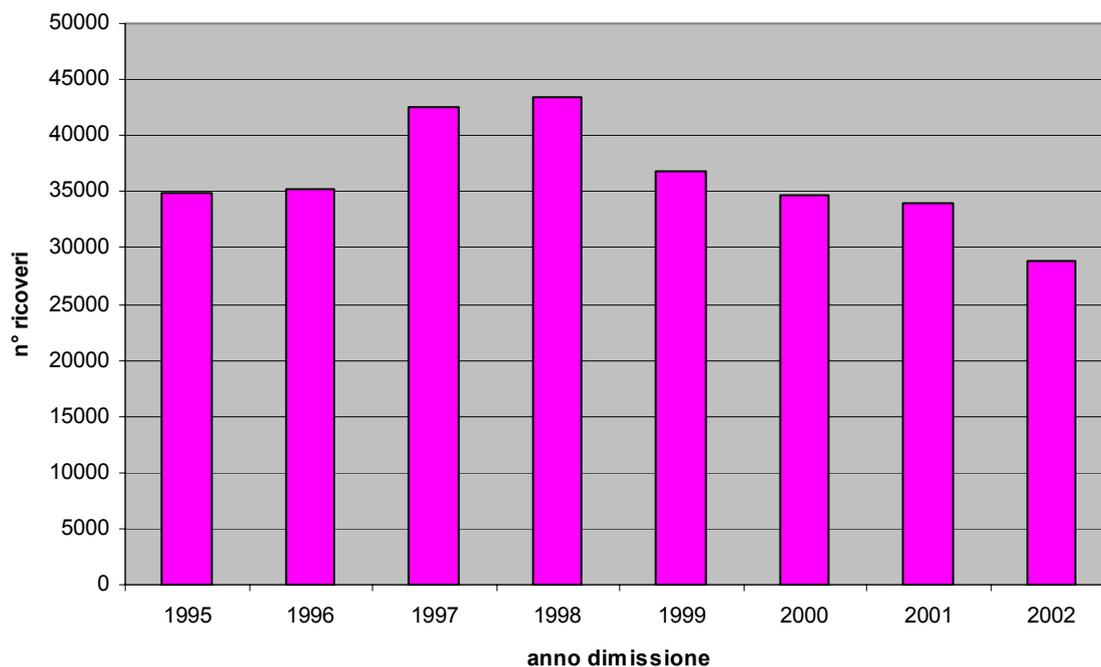
Tabella n°4.2

Distribuzione per anno di dimissione dei ricoveri in ULSS 17

Anno dimissione	totale ricoveri	%
1995	34902	12,0
1996	35161	12,1
1997	42541	14,7
1998	43337	14,9
1999	36869	12,7
2000	34611	11,9
2001	34061	11,7
2002	28875	9,9
totale	290357	100,0

Grafico n° 4.1

Distribuzione dei ricoveri per anno di dimissione 1995-2002



Dal grafico si nota la tendenza di crescita del numero di ricoveri che c'è stata dal 1995 al 1998 ed ha avuto il suo picco nel 1998. Negli anni successivi vi è stata una tendenza inversa, il numero di ricoveri è andato via

via diminuendo. Non può essere considerato l'anno 2002 per il confronto in quanto è incompleto: mancano i record relativi agli ultimi tre mesi del 2002 dei residenti dell'ULSS 17 ricoverati in strutture fuori dall'ULSS di appartenenza.

Il grafico n°4.2 e la tabella n°4.3 successive mostrano sempre la distribuzione dei ricoveri per anno ma considerando anche il tipo di ricovero: ordinario o in day-hospital.

Grafico n° 4.2

Distribuzione dei ricoveri per anno di dimissione 1995-2002 (tenendo conto del regime di ricovero)

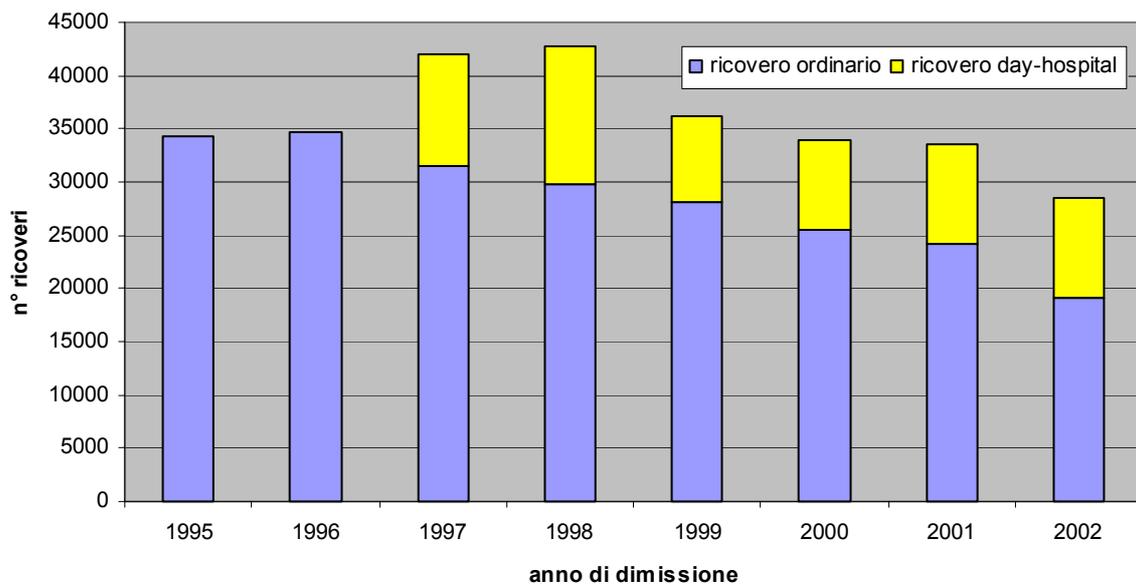


Tabella n° 4.3

Distribuzione delle dimissioni (1995-2002) per anno e regime ricovero

Anno Dimissione	regime ricovero		Totale ricoveri
	1	2	
1995	34.902	0	34.902
1996	35.161	0	35.161
1997	31.752	10.789	42.541
1998	30.045	13.292	43.337
1999	28.573	8.296	36.869
2000	25.867	8.744	34.611
2001	24.540	9.521	34.061
2002	19.355	9.520	28.875

Come si nota dal grafico n° 4.2 con il procedere degli anni sono andati diminuendo i ricoveri ordinari.

4.3 Introduzione alla procedura di sistemazione dell'archivio delle Dimissioni (1995-2002 ULSS 17)

Di seguito vengono descritti gli incroci utilizzati per la validazione dei soggetti nell'archivio delle Dimissioni dell'ULSS 17.

Come per l'archivio di Anatomia Patologica l'unico tipo di informazioni su cui è stato fatto un controllo sono state quelle anagrafiche attraverso un confronto incrociato con uno scarico dell'Anagrafe Assistiti (lo stesso scarico utilizzato per sistemare l'Anatomia Patologica).

I controlli eseguiti sono quasi gli stessi di quelli utilizzati per l'archivio di Anatomia Patologica dell'ULSS 17 con una sola eccezione dovuta a delle incongruenze riscontrate nei dati che verrà spiegata nel prossimo paragrafo.

Di seguito è riportata la lista degli incroci effettuati:

1. tessera sanitaria + cognome + nome + data di nascita + comune di nascita
2. tessera sanitaria + cognome + nome + data di nascita
3. tessera sanitaria + cognome + data di nascita
4. tessera sanitaria + cognome + nome
5. cognome + nome + data di nascita
6. tessera sanitaria + nome + data di nascita
7. tessera sanitaria + data di nascita
8. tessera sanitaria
9. cognome + data di nascita
10. nome + data di nascita

4.3.1 Descrizione degli incroci effettuati

Tessera sanitaria + cognome + nome + data di nascita + comune di nascita

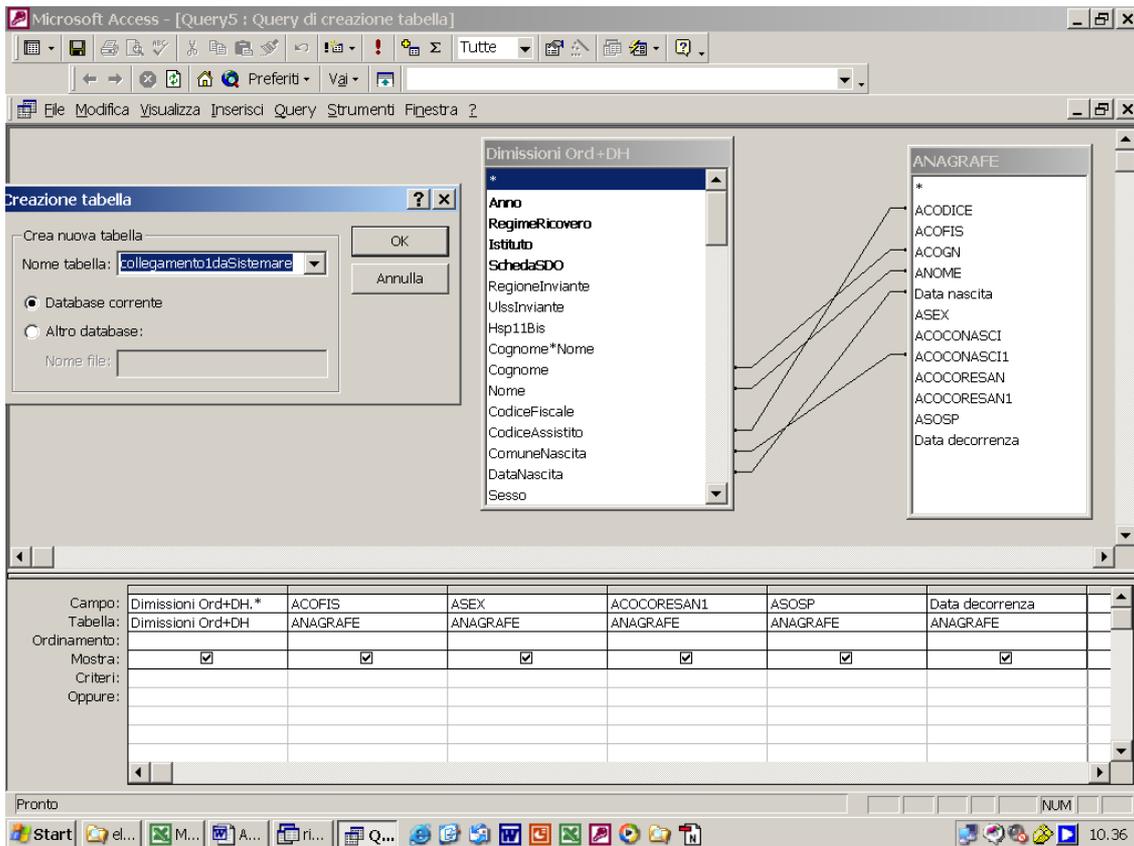
✓ Primo step

Sono state collegate le tabelle **Dimissioni Ord + DH** (tabella contenente tutti i ricoveri ordinari e in day-hospital dal 1995 al 2002) e **Anagrafe** attraverso un join uno ad uno (con questo join vengono selezionate solo le righe in cui i campi collegati da entrambe le tabelle sono uguali; in questo caso *tessera sanitaria, cognome, nome, data di nascita e comune di nascita*).

I 172.570 record così selezionati sono stati inseriti in una tabella chiamata **collegamento1daSistemare**. La tabella **collegamento1daSistemare** contiene tutti i campi presenti in **Dimissioni Ord + DH** ed in più le variabili derivanti dalla tabella **anagrafe** non utilizzate nel join (*ACOFIS, ASEX, ACORESANI, ASOSP, DATA DECORRENZA*).

La schermata n°4.1 mostra i join eseguiti nel primo incrocio e la creazione della tabella **collegamento1daSistemare**.

Schermata n° 4.1

✓ Secondo step

Successivamente si è fatto un controllo sui record della tabella appena creata, atto ad eliminare quelli relativi ad individui con tessera sanitaria scaduta precedentemente alla data del ricovero (nello scarico dell'archivio di Anatomia Patologica non si aveva a disposizione la data dell'analisi ma solo l'anno corrispondente quindi il controllo è stato eseguito sugli anni, naturalmente il risultato era meno preciso), per evitare di considerare individui non più assistiti dall'ULSS 17.

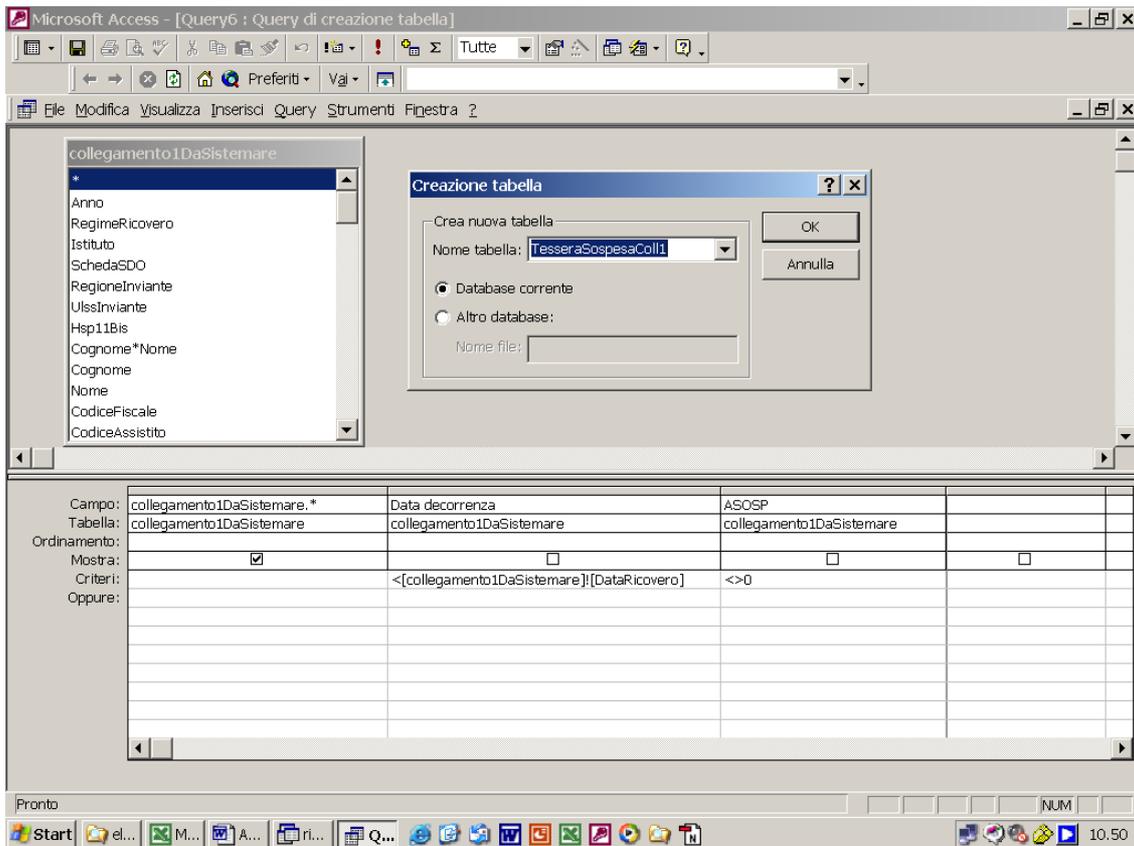
Questa operazione è avvenuta in maniera automatizzata in due fasi:

- 1) Si è creata una tabella contenente gli individui risultati non più assistiti al momento del ricovero
- 2) Si è eliminata questa tabella da **collegamento1daSistemare**

1) Nella tabella **collegamento1daSistemare** si sono individuati i record in cui la data di decorrenza della tessera sanitari, per una causa di sospensione differente da zero ($ASOSP \neq 0$), risultava inferiore alla data del ricovero. Si è specificato che la causa di sospensione fosse diversa da zero in quanto quando il campo *ASOSP* è codificato con lo zero sta a significare che la tessera è ancora in vigore e che l'individuo ad essa associato è ancora un assistito dell'ULSS considerata. I 599 record così selezionati dalla tabella **collegamento1daSistemare** sono stati inseriti in una tabella chiamata **TesseraSospesaColl1**.

La schermata n° 4.2 mostra come sono stati selezionati i record nella tabella **collegamento1daSistemare** per la creazione di **TesseraSospesaColl1**.

Schermata n° 4.2

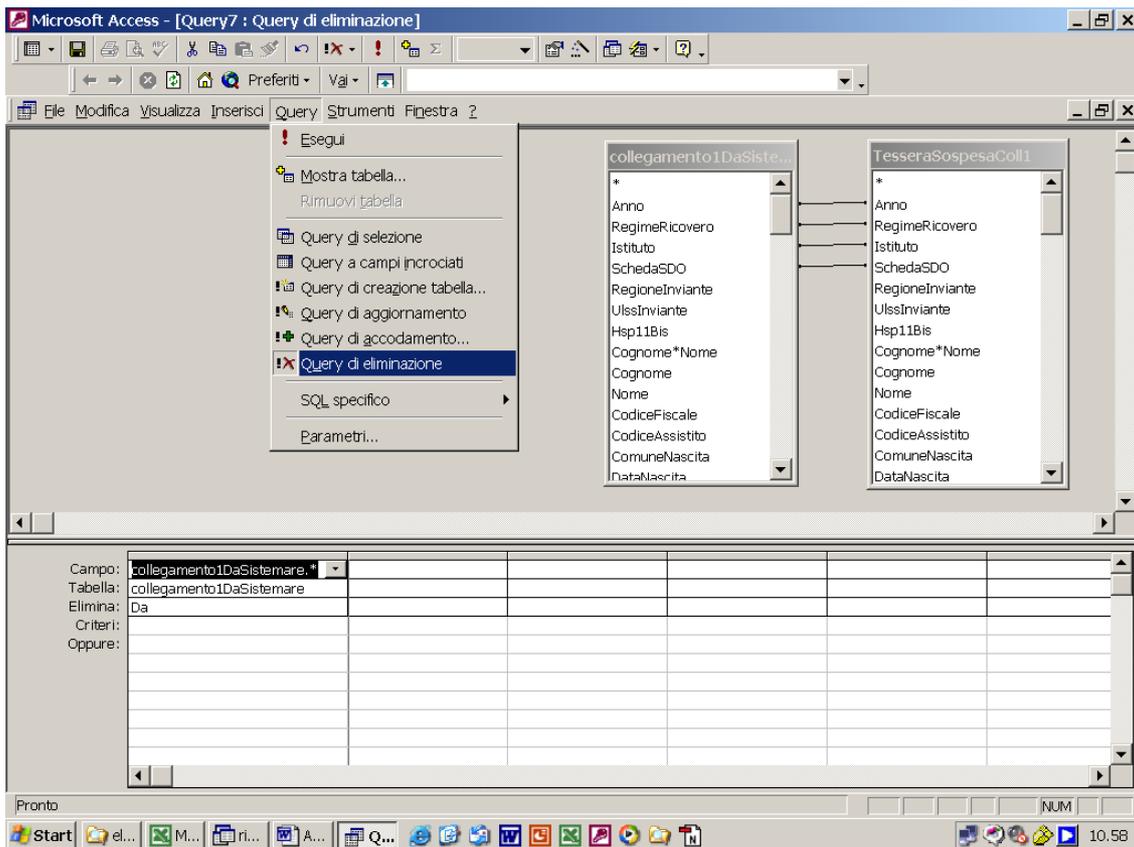


2) Si sono rimossi i record contenuti nella tabella **TesseraSospesaColl1** dalla tabella **collegamento1daSistemare**. Questo passaggio è avvenuto attraverso una query di eliminazione. Le variabili *Anno*, *RegimeRicovero*, *Istituto* e *SchedaSDO* relative alle tabelle **collegamento1daSistemare** e **TesseraSospesaColl1** sono state unite attraverso un join uno ad uno (con questo join vengono selezionate solo le righe in cui i campi collegati da entrambe le tabelle sono uguali) e si è specificato quale tabella dovesse essere sottratta all'altra. Importante è stato l'utilizzo delle quattro variabili del link: *Anno*, *RegimeRicovero*, *Istituto* e *SchedaSDO* in quanto questi quattro campi sono sufficienti ad identificare univocamente i record e quindi nelle query di eliminazione si è certi di quali record si vanno a sottrarre. Nelle query di eliminazione è infatti necessario che i record della

tabella che si va ad eliminare non presentino duplicati.

La schermata n° 4.3 mostra l'impostazione della query di eliminazione sopra descritta.

Schermata n° 4.3



Alla fine di questi interventi nella tabella **collegamento1daSistemare** sono rimasti 171.971 record (circa il 59% dei record contenuti nel database iniziale).

✓ Terzo step

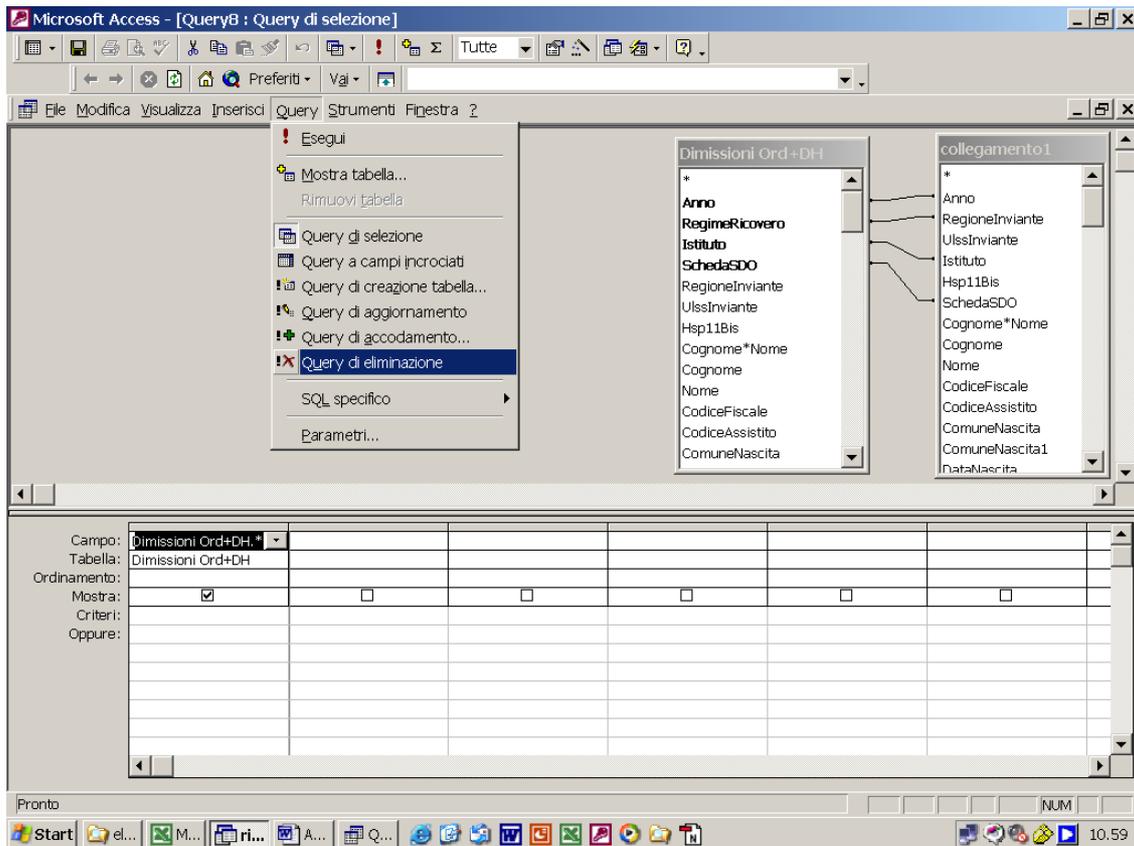
La tabella **collegamento1daSistemare** è stata rinominata e chiamata **collegamento1**.

✓ Quarto step

Prima di proseguire con il secondo incrocio si sono tolti i record contenuti in **collegamento1** (171.971) dalla tabella **Dimissioni Ord + DH (290.357 record)** per permettere che l'incrocio successivo potesse avvenire su record diversi rispetto a quelli contenuti e già corretti in **collegamento1**. Questo passaggio è stato eseguito con una query di eliminazione. Le variabili *Anno*, *RegimeRicovero*, *Istituto* e *SchedaSDO* sono state unite attraverso un join uno ad uno e si è specificato quale tabella dovesse essere sottratta all'altra.

La schermata n°4.4 mostra la query di eliminazione effettuata tra le tabelle **Dimissioni Ord + DH** e **collegamento1**.

Schermata n° 4.4



N.B. Si noti che le uniche differenze accertate tra i quattro step appena eseguiti con quelli descritti nel procedimento di risistemazione dell'archivio di Anatomia Patologica sono riscontrabili nella creazione di **TesseraSospesaColl1** in relazione alle date utilizzate e nelle query di eliminazioni eseguite per quanto riguarda le variabili che rendono identificabile il record e quindi utilizzate nei join delle query. Per il resto la procedura utilizzata per la validazione dei soggetti nell'archivio dei Ricoveri è identica a quella già vista per l'Anatomia Patologica ed anche gli errori e le problematiche riscontrate sono uguali. Quindi si rimanda per la spiegazione al capitolo “**Procedura per la sistemazione dell'archivio di Anatomia patologica**” limitandosi ora ad elencare le tabelle create durante gli incroci eseguiti:

1° incrocio.	collegamento 1 (171.971 record)
2° incrocio.	collegamento2 (80.315 record)
3° incrocio.	collegamento3 (6.436 record)
4° incrocio.	collegamento4 (1.951 record)
5° incrocio.	collegamento5 (23.492 record)
6° incrocio.	collegamento6 (2.052 record)
7° incrocio.	collegamento7 (216 record)
8° incrocio.	collegamento8 (161 record)
9° incrocio.	collegamento9 (809 record)
10° incrocio.	collegamento10 (338 record)

Successivamente si sono guardati i record scartati dai primi sei incroci perché aventi la data di decorrenza della tessera sanitaria precedente alla data del ricovero per un motivo di sospensione diverso da zero. Nello specifico si sono considerati i record in cui il motivo di sospensione della tessera era codificato con 3 (decesso) cioè i record associati ad individui che al momento del ricovero, in Anagrafe, risultavano deceduti. In totale si sono selezionati 41 record. Si è quindi pensato di confrontare la data del decesso pervenuta dall'anagrafe con quella pervenuta dalle schede di morte, che dovrebbe essere quella corretta ed ufficiale, e si è visto che 27 individui avevano effettivamente in Anagrafe la data di morte errata. I 27 record relativi a questi 27 soggetti sono stati recuperati ed inseriti in una tabella chiamata **collegamento11**. I 14 record rimanenti che comunque presentavano la data del ricovero successiva a quella di morte sono stati controllati e se per 2 record relativi al 2002 non è stato proprio possibile verificare che la data del decesso in Anagrafe fosse effettivamente quella corretta poiché non si avevano a disposizione le schede di morte di quell'anno (cioè del 2002) per altri 8 invece le schede di morte non si sono

trovate. Inoltre si è visto che 3 di questi 8 individui avevano in Anagrafe degli omonimi cioè persone con nome e cognome uguale. Degli ultimi 4 record invece si sono trovate le schede di morte ed effettivamente la data del decesso corrispondeva a quella presente in anagrafe. Tutti questi 4 soggetti in anagrafe risultavano avere degli omonimi. Quindi verrebbe quasi da pensare che nell'accettazione dei ricoveri siano stati inseriti i dati anagrafici di persone omonime agli effettivi pazienti ricoverati.

La tabella successiva riassume la situazione appena descritta (Tabella n°4.4).

Tabella n°4.4

Record che risultavano con data del ricovero posteriore alla data di morte

n° Soggetti	Data di morte in anagrafe	Data di morte nelle schede di morte	data ricovero	presenza di omonimi in anagrafe
27	è stata corretta grazie alla scheda di morte	verificata	precedente alla data di morte	irrilevante
2	non si sa se sia corretta o meno	impossibilità di verificare	successivo alla presunta data di morte	si
3	non si sa se sia corretta o meno	non trovata	successivo alla presunta data di morte	si
5	non si sa se sia corretta o meno	non trovata	successivo alla presunta data di morte	no
4	risultava esatta	verificata	successivo alla data di morte	si

4.3.2 Schema Riassuntivo

Di seguito è riportato uno schema riassuntivo degli incroci eseguiti e fin qui descritti (grafico n°4.3).

Le celle contrassegnate in giallo indicano le variabili che sono state utilizzate per i link. Le celle contrassegnate in fucsia indicano le variabili su cui è stato eseguito un controllo manuale dopo aver avuto il risultato dell'incrocio. In **collegamento 11** sono stati inseriti i record recuperati nei primi sei incroci e che inizialmente erano stati scartati perché presentavano data di morte precedente alla data del ricovero.

Grafico n°4.3

Passaggi per la pulitura dell'archivio dei Ricoveri (1995-primi nove mesi 2002)

numero records iniziali: **290.357**

collegamento	cognome	nome	tessera	data di nascita	comune di nascita	N. records	N. records rimanenti
collegamento 1						171.971 (59%)	
collegamento 2						80.315	
collegamento 3						6.436	
collegamento 4						1.951	
collegamento 5						23.492	
totale parziale						284.165 (97,9%)	6.192(2,1%)
collegamento 6	simili					2.052	
collegamento 7	simili	simili				216	
collegamento 8	simili	simili				161	
collegamento 9		simili				809	
collegamento 10	simili					338	
collegamento 11	records reinseriti perché corretta data di morte ²⁾					27	
totale finale						287.768 (99,1%)	2.599 (0,9%)

legenda	uguali	simili ¹⁾
---------	--------	----------------------

1)quando i dati differiscono per un numero o per poche lettere o hanno al loro interno caratteri impropri

2)ho modificato nell'anagrafe la data di morte perché era stata registrata sbagliata

4.4 Dati da considerare

Lo scopo di questo lavoro è quello di creare una procedura di rifinitura degli archivi sanitari che sia il più possibile standard, quindi naturalmente per l'archivio dei Ricoveri è stato deciso di considerare i dati recuperati e corretti fino all'incrocio numero sei (tabelle **collegamento1-collegamento6** + **collegamento11**) come era precedentemente stato fatto per l'archivio di Anatomia Patologica.

L'archivio dei Ricoveri dopo il procedimento di ripulitura appena descritto contiene 286.244 record (98,6% dei record iniziali) provenienti dall'accodamento delle sei tabelle createsi durante gli *step* dei primi sei linkage:

collegamento 1: 171.971 record

collegamento2: 80.315 record

collegamento3: 6.436 record

collegamento4: 1.951 record

collegamento5: 23.492 record

collegamento6: 2.052 record

collegamento11: 27 record

Le sette tabelle sono state unite tenendo conto per ogni collegamento di accodare i campi in modo da sostituire le informazioni anagrafiche imprecise o incomplete provenienti dai Ricoveri con quelle presumibilmente corrette provenienti dall'Anagrafe Sanitaria.

Seguono delle tabelle riassuntive dell'archivio dei Ricoveri contenenti i 286.244 record selezionati e corretti.

La tabella n°4.5 ed il grafico n°4.4 mostrano la distribuzione dei ricoveri per anno di dimissione e sesso.

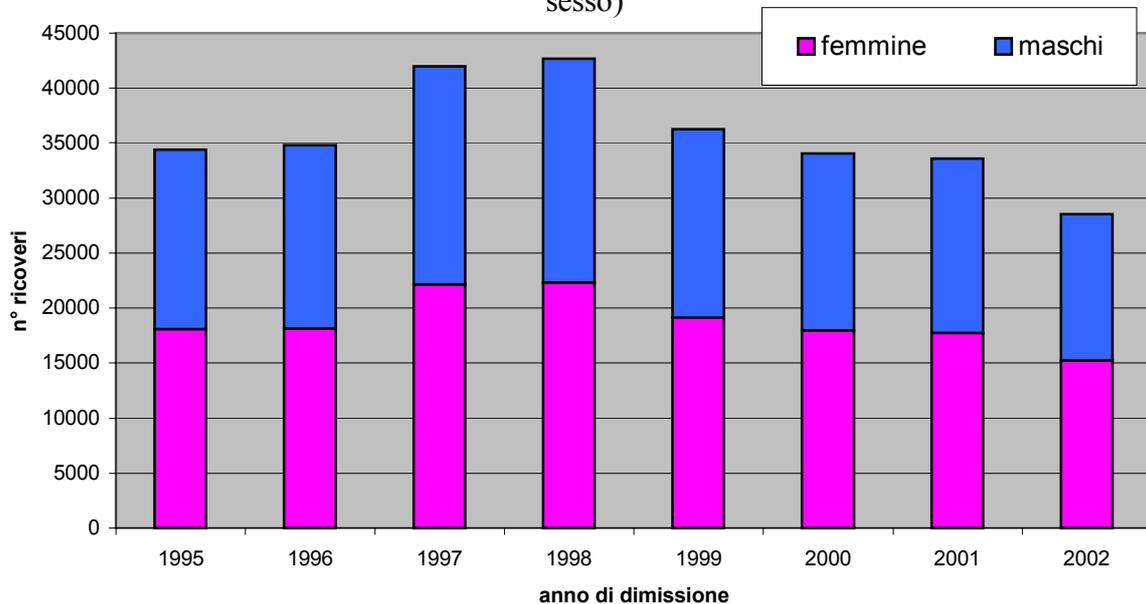
Tabella n°4.5

Numero di dimissioni per anno di dimissione e sesso

Anno	F	F/Totale*100 (%)	M	M/Totale*100 (%)	Totale
1995	18.079	52,6	16.318	47,4	34.397
1996	18.174	52,3	16.602	47,7	34.776
1997	22.148	52,7	19.839	47,3	41.987
1998	22.299	52,2	20.397	47,8	42.696
1999	19.117	52,8	17.119	47,2	36.236
2000	17.973	52,8	16.053	47,2	34.026
2001	17.767	52,9	15.808	47,1	33.575
2002	15.228	53,3	13.323	46,7	28.551
Totale	150.785	52,7	135.459	47,3	286.244

Grafico n°4.4

Distribuzione dei ricoveri per anno di dimissione (tenendo conto del sesso)



Si noti che per ogni anno di dimissione le percentuali dei ricoveri riguardanti donne e uomini è rimasta sempre pressoché costante.

La tabella n°4.6 ed il grafico n°4.5 invece mostrano la distribuzione delle dimissioni per anno di dimissione e per regime ricovero (ordinario o in day-hospital).

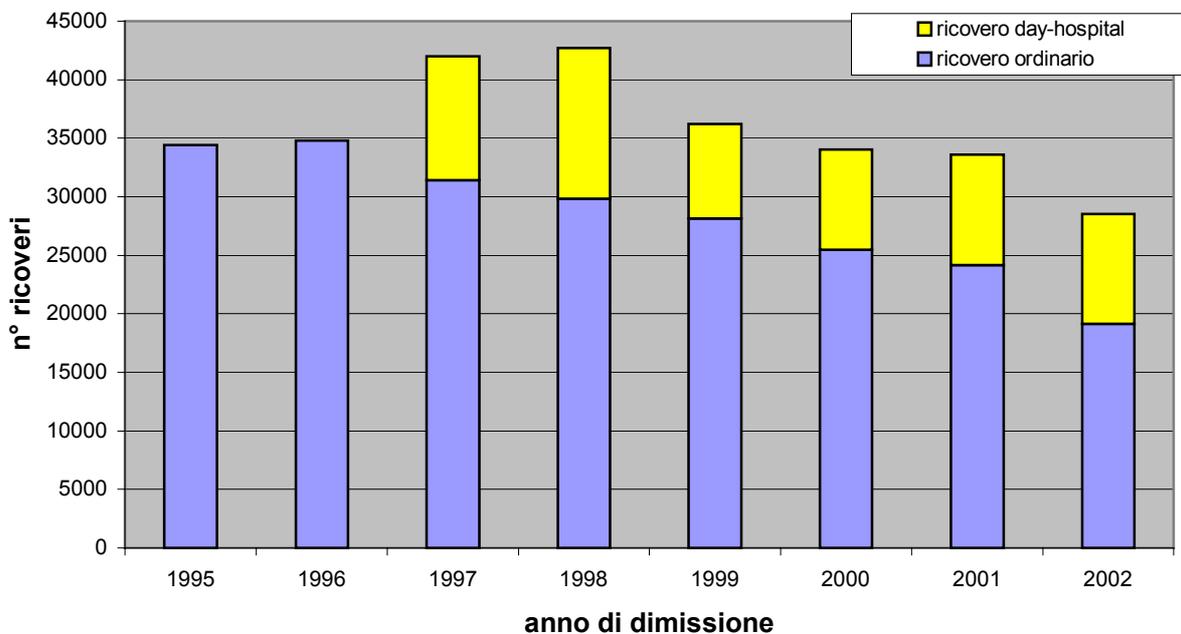
Tabella n°4.6

Distribuzione dei ricoveri per anno di dimissione e regime ricovero

Anno	1=ricovero ordinario	1/Totale*100 (%)	2=ricovero in day-hospital	2/Totale*100 (%)	Totale
1995	34.397	100,0	0	0,0	34.397
1996	34.776	100,0	0	0,0	34.776
1997	31.433	74,9	10.554	25,1	41.987
1998	29.821	69,8	12.875	30,2	42.696
1999	28.125	77,6	8.111	22,4	36.236
2000	25.460	74,8	8.566	25,2	34.026
2001	24.188	72,0	9.387	28,0	33.575
2002	19.162	67,1	9.389	32,9	28.551
Totale	227.362	79,4	58.882	20,6	286.244

Grafico n° 4.5

Distribuzione dei ricoveri per anno di dimissione (tenendo conto del regime di ricovero)



Come si nota dal grafico e dalla tabella dal 1997 è stato previsto il ricovero day-hospital quindi si possono distinguere i ricoveri ordinari da quelli giornalieri (tabella n°4.6, grafico n°4.5).

Nella tabella n°4.7 e nel grafico n°4.6 sono stati considerati i ricoveri per anno di dimissione, per regime di ricovero e per sesso e si è cercato di vedere se tra i due sessi ci fosse o meno una certa differenziazione sul regime di ricovero.

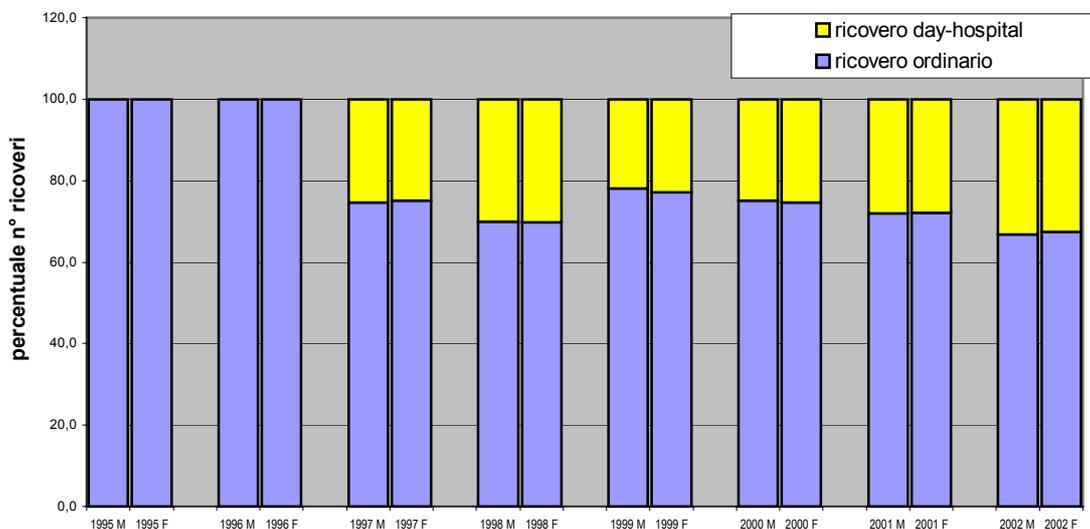
Tabella n°4.7

Numero di ricoveri per anno di dimissione, regime di ricovero e sesso

Anno di dimissione	Femmine					Maschi				
	1: ricoveri ordinari	1/Totale* 100 (%)	2: ricoveri day-hospital	2/Totale* 100 (%)	Totale	1: ricoveri ordinari	1/Totale* 100 (%)	2: ricoveri day-hospital	2/Totale* 100 (%)	Totale
1995	18.079	100,0	0	0,0	18.079	16.318	100,0	0	0,0	16.318
1996	18.174	100,0	0	0,0	18.174	16.602	100,0	0	0,0	16.602
1997	16.633	75,1	5.515	24,9	22.148	14.800	74,6	5.039	25,4	19.839
1998	15.553	69,7	6.746	30,3	22.299	14.268	70,0	6.129	30,0	20.397
1999	14.760	77,2	4.357	22,8	19.117	13.365	78,1	3.754	21,9	17.119
2000	13.399	74,6	4.574	25,4	17.973	12.061	75,1	3.992	24,9	16.053
2001	12.806	72,1	4.961	27,9	17.767	11.382	72,0	4.426	28,0	15.808
2002	10.270	67,4	4.958	32,6	15.228	8.892	66,7	4.431	33,3	13.323
Totale	119.674	79,4	31.111	20,6	150.785	107.688	79,5	27.771	20,5	135.459

Grafico n°4.6

Confronto tra le percentuali di numero di ricoveri per anno di dimissione e per sesso tenendo conto del regime di ricovero



Come si vede dal grafico (grafico n°4.6) per tutti gli anni considerati la percentuale di ricoveri ordinari o in day-hospital per maschi e femmine è pressoché identica.

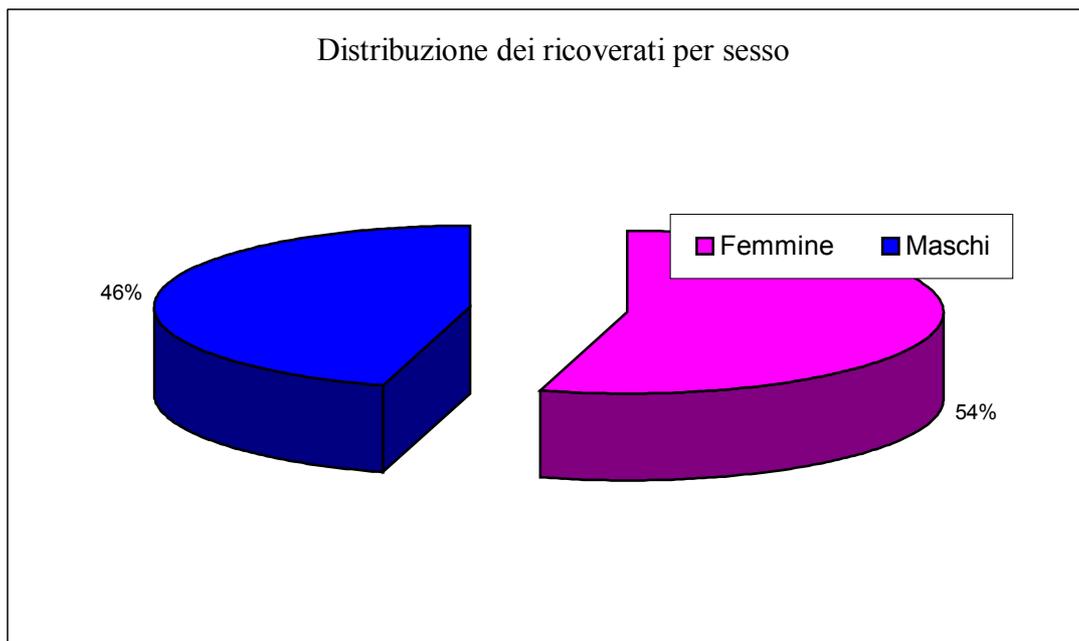
La tabella n°4.8 ed il grafico n° 4.7 mostrano il numero di individui ricoverati tenuto conto del sesso.

Tabella n°4.8

Numero di ricoverati tenuto conto del sesso

Femmine	F/totale*100	Maschi	M/totale*100	Totale
57244	54,4	48073	46,6	105317

Grafico n° 4.7



CONCLUSIONI

La tesi nasce dalla volontà di creare un registro tumori dell'ULSS 17 da qui l'esigenza di disporre di una banca dati che possa prestarsi alla mappatura del cancro nel territorio includendo tutti i nuovi casi di neoplasia maligna incidenti ed escludendo i casi prevalenti. Questa banca dati viene costruita sulla base dell'incrocio di tre fonti: l'Anatomia Patologica, le schede di Dimissioni Ospedaliere e le schede di Morte.

Queste tre fonti sono gestite con procedure differenti e l'inserimento dati viene effettuato da operatori diversi e quindi anche le incongruenze al loro interno possono essere dissimili. Il compito di questo lavoro è stato quello di riuscire a trovare un metodo che potesse rendere i soggetti di tutte e tre le fonti facilmente individuabili e confrontabili utilizzando una stessa procedura di data capture informatizzata e standardizzata non più dipendente dalle capacità di discernere degli operatori addetti alla riqualificazione degli archivi.

La procedura di data capture atta alla validazione delle informazioni sanitarie proposta in questo studio si basa su di una serie di linkage automatici (e solo in parte manuali con controllo a video) tra gli archivi sanitari considerati (archivio di Dimissioni Ospedaliere 1995-2002 e l'archivio di Anatomia Patologica 1991-inizio 2003) e l'Anagrafe Assistiti. È stato utilizzata l'Anagrafe Assistiti (uno scarico, contenente anche lo storico anagrafico, avvenuto in data 19/12/2003) per l'identificazione dei soggetti poichè quest'archivio, affidabile in quanto costantemente aggiornato e corretto, contiene tutte le informazioni anagrafiche relative agli assistiti dell'ULSS e permette attraverso un codice sanitario di identificare univocamente i soggetti.

La procedura di linkage applicata tra l'Anagrafe Assistiti e l'Anatomia Patologica prima e tra l'Anagrafe Assistiti e le schede di Dimissioni Ospedaliere poi è stata la stessa giacchè si è notato che i database da sistemare, sebbene gestiti da strutture diverse, presentavano al loro interno imprecisioni analoghe per lo più imputabili alla stessa modalità di immissione dei dati utilizzata: inserimento manuale.

I link eseguiti, atti alla riqualificazione degli archivi, hanno visto interessate combinazioni di sole variabili anagrafiche quali: nome, cognome, data di nascita, comune di nascita e codice sanitario.

Alla fine del procedimento di linkage è stato recuperato e corretto rispettivamente il 99% ed il 83% dei record relativi agli archivi delle schede di Dimissioni Ospedaliere e dell'Anatomia Patologica.

Di seguito è riportata la lista degli incroci eseguiti:

1. tessera sanitaria + cognome + nome + data di nascita + comune di nascita
2. tessera sanitaria + cognome + nome + data di nascita
3. tessera sanitaria + cognome + data di nascita
4. tessera sanitaria + cognome + nome
5. cognome + nome + data di nascita
6. tessera sanitaria + nome + data di nascita.

Lo studio effettuato si è basato soprattutto sulla ricerca di un procedimento che potesse rendere il recupero e la validazione dei soggetti in archivi sanitari, il più standard e corretto possibile.

Si è cercato di trovare un giusto "compromesso" tra qualità e quantità dell'informazione recuperabile, cioè quanto si è disposti a perdere in quantità pur di avere dei record più precisi?.

Se è vero che il modello qui descritto è nato con il preciso scopo di creare le basi per la costruzione di un registro tumori dell'ULSS 17 è anche vero che sarebbe riduttivo pensarlo attuativo solo in questo ambito. Si è infatti evidenziata più volte l'importanza della standardizzazione del metodo di data capture e come esso sia facilmente applicabile a differenti archivi consentendo di integrare e confrontare dati di diversa tipologia e provenienza. In letteratura purtroppo però sono ancora pochi i casi di aggregazioni di archivi sanitari informatizzati, con obiettivi epidemiologici, dove questo studio potrebbe trovare collocazione e di conseguenza mancano anche studi su modelli per la validazione delle informazioni anagrafiche simili a questo. Da qui l'impossibilità di compararne le qualità ed i limiti. Un esempio comunque di una possibile applicazione del modello ci viene da un recente studio svolto presso l'unità di Epidemiologia dell'ULSS 12 Veneziana che vede la realizzazione di uno "strato" informativo sanitario-epidemiologico di base che consenta di integrare e comparare informazioni relative ad archivi differenti quali quello dell'Anagrafe Sanitaria, dell'Anatomia Patologica, delle Schede di Dimissioni Ospedaliere, dei Decessi, delle Prestazioni Ambulatoriali e del Consumo dei Farmaci ai fini di giungere ad ottenere una stima della frequenza della diffusione delle patologie principali all'interno dell'ULSS stessa e quindi di riuscire ad avere una conoscenza più precisa dell'azienda per in seguito poter migliorare la qualità degli interventi relativi alla salute dei cittadini.

Se lo studio come quello proposto dall'ULSS 12 Veneziana è ancora una realtà a se stante per quanto concerne invece l'archivio informatizzato dei tumori l'interesse è più vivo e molte regioni ed ULSS si sono già rese operative al fine di riuscire ad attivarlo.

Da qui appunto nasce l'esigenza di elaborare delle linee guida che portino ad una standardizzazione del metodo atto all'identificazione corretta dei soggetti presenti nelle diverse fonti. Con questo elaborato si è quindi cercato di dare una prima valida risposta a questa necessità almeno per quanto riguarda la realtà veneta poiché, purtroppo, il reale utilizzo del modello è fortemente condizionato dai limiti causati dai dati e dagli archivi disponibili nel luogo in cui si opera.

La metodologia di archiviazione delle informazioni infatti non è mai stata uniformata e di conseguenza data base dello stesso tipo possono alla fine contenere campi differenti.

È lecito quindi ritenere che bisognerà pensare di uguagliare le informazioni anagrafiche presenti nelle varie fonti affinché si possa delineare una procedura standard per la validazione dei soggetti che sia effettivamente applicabile a tutti gli archivi sanitari.

BIBLIOGRAFIA

Gubitosa L. (2002-2003), “L'utilizzo di un archivio informatizzato delle prescrizioni di farmaci per stimare la frequenza delle malattie nella popolazione dell'ULSS 12 Veneziana”, Tesi di Laurea, Università degli Studi di Padova, Facoltà di Scienze Statistiche

Simionato D. (1997-1998), “Caratteristiche della mortalità nella Bassa Padovana (ULSS 17)”, Tesi di Diploma, Università degli Studi di Padova, Facoltà di Scienze Statistiche

Fornasiero F. (1995-1996), “Analisi dei tassi di ospedalizzazione nell'Ulss 17 di Este – Monselice – Montagnana – Conselve”, Tesi di Diploma, Università degli Studi di Padova, Facoltà di Scienze Statistiche

Donato M., “Analisi dei dati sull'attività del pronto soccorso dell'Ulss 17 della regione Veneto”, Tesi di Laurea, Università degli Studi di Padova, Facoltà di Scienze Statistiche

Ufficio Statistica Azienda ULSS 17 Regione Veneto, *La popolazione residente nel territorio dell'azienda Ulss 17 dal 1981 al 1995*

Bolzanini M., Brancato B., Cristofori A., *Registro tumori mammari della provincia di Trento - Rapporto sul primo anno di attività 1996*, ed. libreria cortina verona