

*Settore
Sanità e Previdenza*

**La codifica
automatica delle
cause di morte in
Italia:
aspetti metodologici e
implementazione della
ICD X**

A cura di: Luisa Frova; Stefano Marchetti; Monica Pace.

Hanno collaborato nella stesura dei capitoli: Roberta Crialesi, Luisa Frova (capitolo 1)
Luisa Frova, Stefano Marchetti, Monica Pace (capitolo 2)
Roberta Crialesi, Luisa Frova, Stefano Marchetti (capitolo 3)
Silvia Bruzzone, Alessandra Battisti (capitolo 4)
Luisa Frova, Stefano Marchetti, Monica Pace, Stefania Macchia (capitolo 5)
Luisa Frova, Monica Pace (capitolo 6)
Silvia Bruzzone, Luisa Frova (Appendice A)
Silvia Bruzzone, Monica Pace (Appendice C)

Ha collaborato: Paola Giorgetti (sviluppo database Oracle e Capo progetto del sistema Monitor)
Simona Cinque (predisposizione della versione on-line del volume, editing del volume)

Per chiarimenti sul contenuto della
pubblicazione rivolgersi a:
Istat, Servizio "Sanità e Assistenza"
06.46737382 – 06.46737396

Sanità e Previdenza

La codifica automatica delle cause di morte in Italia: aspetti metodologici e implementazione della ICD X

Il presente volume nasce dall'esigenza di aggiornare una precedente pubblicazione (Istat, 2001, *La Nuova Indagine sulle Cause di Morte*, Metodi e Norme) tenendo conto dei nuovi sviluppi concettuali e metodologici nell'ambito della produzione delle statistiche ufficiali di mortalità per causa in Italia.

Come noto, la Rilevazione delle Cause di Morte è stata oggetto di una notevole attività di ristrutturazione che ha interessato molteplici fasi dell'indagine. Il processo innovativo più importante è stato l'adozione della codifica automatica delle cause di morte (basata essenzialmente sulla procedura Micar-Acme, sviluppata dall'NCHS, USA). Il 77% dei decessi è oggi codificato automaticamente secondo i criteri della Classificazione Internazionale delle Malattie, Traumatismi e Cause di Morte (ICD), IX Revisione.

Attualmente è in fase di realizzazione il passaggio alla X Revisione, un passaggio tra i più importanti e delicati e già compiuto in numerosi Paesi del mondo. L'aver adottato procedure automatiche per la codifica della causa iniziale di morte agevererà considerevolmente il passaggio da una revisione alla successiva.

L'esperienza effettuata in precedenza con l'adozione della codifica automatica per l'ICD IX, consente adesso all'Italia di essere tra i pochi paesi in Europa all'avanguardia in questo campo, con la possibilità di proporsi come termine di riferimento e confronto per i paesi che si stanno avviando verso la medesima strada. Da rilevare che l'uso del sistema ACME per l'attribuzione della causa iniziale di morte dovrà essere implementato in tutta Europa entro il 2005.

Il passaggio alla nuova revisione dell'ICD rende comunque necessario introdurre alcune novità anche nel sistema di codifica automatica. Tra le più importanti si ricordano la traduzione in lingua italiana del dizionario delle espressioni diagnostiche sulla base della nuova classificazione da adottare e l'implementazione di un nuovo software (ACTR) per l'accoppiamento della terminologia medica con codici numerici.

Tra i temi affrontati meritano attenzione anche le strategie per migliorare l'importante aspetto della qualità della certificazione e la comparabilità internazionale.

Sito internet: <http://www.istat.it>

La codifica automatica delle cause di morte in Italia: aspetti metodologici e implementazione della ICD X

INDICE

Introduzione

1	IL REENGINEERING DELLE STATISTICHE DI MORTALITÀ PER CAUSA IN ITALIA	5
2	LE CAUSE DI MORTE E IL SISTEMA DI CODIFICA AUTOMATICA	7
2.1	LA CAUSA INIZIALE DI MORTE	7
2.2	LE CAUSE MULTIPLE DI MORTE.....	9
2.3	VANTAGGI E PROBLEMATICHE DERIVANTI DALL'USO DI UN SISTEMA AUTOMATICO	10
2.3.1	<i>Impatto sulla qualità</i>	10
2.3.2	<i>Impatto sulla tempistica</i>	11
2.3.3	<i>Informazione sulle cause multiple di morte</i>	11
2.4	LA CODIFICA AUTOMATICA IN ITALIA	12
2.5	LA PROCEDURA ACME	13
2.6	LA PROCEDURA MICAR.....	13
2.7	LA PROCEDURA TRANSAX.....	15
3	LA CODIFICA AUTOMATICA E L'ADEGUAMENTO ALLA TERMINOLOGIA MEDICA ITALIANA	16
3.1	L'ESPERIENZA ISTAT.....	16
3.2	LA REGISTRAZIONE DELLE SCHEDE DI MORTE	18
3.3	LA PROCEDURA CORREMOR.....	19
3.4	LA PROCEDURA GSM	20
3.5	LA CODIFICA MANUALE.....	21
3.5.1	<i>La codifica delle cause di morte di natura violenta</i>	22
3.5.2	<i>La codifica dell'AIDS</i>	22
3.6	EFFICACIA DEL SISTEMA DI CODIFICA AUTOMATICA	23
4	BRIDGE CODING TRA CODIFICA MANUALE E AUTOMATICA	24
4.1	IL CAMPIONE	24
4.2	RISULTATI PRINCIPALI	24
4.3	IMPATTO DEL SISTEMA AUTOMATICO SUI DATI DI MORTALITÀ PER CAUSA	26
5	ADEGUAMENTO DEL SOFTWARE PER LA TRANSIZIONE VERSO LA DECIMA REVISIONE DELLA ICD	30
5.1	LA TRADUZIONE DEL DIZIONARIO DEGLI ERNS.....	31
5.1.1	<i>Problematiche linguistiche</i>	31
5.1.2	<i>Il database per la traduzione della terminologia medica</i>	32
5.1.3	<i>Traduzione automatica provvisoria di alcuni ERN s</i>	34
5.2	LA CODIFICA DEL TESTO CON ACTR.....	36
5.3	ACTR APPLICATO ALLA TERMINOLOGIA MEDICA	38
5.3.1	<i>Obiettivi</i>	38
5.3.2	<i>L'implementazione di ACTR: due diversi approcci per la sperimentazione</i>	38
5.3.3	<i>Il criterio adottato per l'accoppiamento</i>	39
5.3.4	<i>Passi operativi e risultati principali</i>	39

6	FUTURI SVILUPPI.....	42
6.1	TEMPESTIVITÀ	42
6.2	QUALITÀ	43
6.2.1	<i>Completezza della raccolta dei dati.....</i>	<i>43</i>
6.2.2	<i>Validazione del dato.....</i>	<i>43</i>
6.3	ARMONIZZAZIONE INTERNAZIONALE	43
6.3.1	<i>Qualità della certificazione.....</i>	<i>43</i>
6.3.2	<i>Implementazione della ICD X.....</i>	<i>44</i>
6.4	VALORIZZAZIONE DELLE FONTI INFORMATIVE ESISTENTI IN ITALY	45
	RIFERIMENTI BIBLIOGRAFICI.....	47
	APPENDICE A: IL FLUSSO INFORMATIVO	51
	APPENDICE B. I MODELLI ISTAT PER LA CERTIFICAZIONE DEI DECESSI: D4, D5, D4BIS AND D5BIS	53
	APPENDICE C: ISTRUZIONI PER LA COMPILAZIONE DEI CERTIFICATI DI MORTE (DISPONIBILE SUL RETRO DEI MODELLI)	58

La codifica automatica delle cause di morte in Italia: aspetti metodologici e implementazione della ICD X

Introduzione

Il presente lavoro vuole essere una importante testimonianza dell'esperienza dell'Istituto Nazionale di Statistica italiano nell'ambito del rinnovamento dell'Indagine sulle Cause di morte. L'elemento più importante è stato l'inserimento di procedure automatiche per la codifica delle cause di morte. Fino ai decessi del 1994 la fase di codifica era completamente manuale e questo aveva inevitabili ricadute sia in termini di qualità che di tempestività nel rilascio dei dati. L'introduzione di procedure automatiche ha richiesto, di necessità, una revisione delle diverse fasi di lavorazione del dato. La Rilevazione sulle Cause di Morte è stata quindi oggetto di una attività di ristrutturazione talmente intensa e radicale da configurarsi come una vera e propria reingegnerizzazione dell'intero percorso di formazione del dato di mortalità per causa.

Gli interventi di ristrutturazione, effettuati cogliendo le opportunità offerte dagli avanzamenti tecnologici nel panorama degli strumenti informatici, hanno interessato diversi livelli dell'indagine, dall'ampliamento dei contenuti rilevati, al trattamento dei dati, alla progettazione e realizzazione di nuove forme di diffusione.

Nell'ambito del dibattito internazionale sono emerse alcune aree particolarmente problematiche che devono essere tenute sotto controllo dai quei Paesi che utilizzano o che si avviano ad utilizzare il processo di automazione della fase di codifica; tra queste si ricordano:

- la figura professionale del nosologo, la sua preparazione e il suo aggiornamento;
- la conoscenza del processo di codifica automatica da parte del personale preposto ad essa e la disponibilità di supporto tecnico;
- i meccanismi di controllo di qualità e di aggiornamento delle tavole decisionali utilizzate dalla codifica automatica;
- il bridge coding;
- la registrazione dei dati;
- le cause di morte violenta;
- le problematiche linguistiche;
- le problematiche di implementazione del sistema automatico.

Obiettivo di questo volume è descrivere in quale modo l'Istat ha implementato il sistema di codifica automatico e come si è confrontato con le difficoltà emerse in relazione alle aree problematiche succitate ed indicare le diverse soluzioni via via scelte (registrazione controllata, sviluppo di diversi software, analisi del bridge coding, formazione del personale) (capitoli 1-4).

Negli ultimi due capitoli (capitoli 5-6) del volume vengono descritti i traguardi più importanti da raggiungere nel prossimo futuro e i passi più importanti già avviati in questa direzione. Il riferimento è in particolare all'implementazione della X revisione della Classificazione, all'utilizzo e diffusione dati di multi-morbosità e all'uso di tecniche per la validazione ed la correzione di dati demografici e sanitari.

1 IL REENGINEERING DELLE STATISTICHE DI MORTALITÀ PER CAUSA IN ITALIA

Negli anni più recenti l'Indagine sulle Cause di Morte, condotta dall'Istat, è stata oggetto di una grande ristrutturazione che ha visto in primo piano l'introduzione di un sistema automatico per la codifica della causa di morte e la completa revisione delle molteplici fasi di lavorazione del dato di mortalità (registrazione dei modelli, controlli quantitativi e qualitativi e forme di pubblicazione e diffusione del dato finale). Gli elementi innovativi introdotti hanno permesso di ridurre notevolmente i tempi di rilascio dei dati definitivi. Tuttavia, la fase di raccolta dei modelli è ancora in forma cartacea e risente pesantemente del lungo percorso che la modulistica è tenuta a seguire prima di giungere alla sede centrale dell'Istat (cfr. Appendice A). Una volta rientrati all'Istat i modelli cartacei¹, il Servizio Sanità e Assistenza, Unità Operativa Mortalità per Causa, si occupa interamente di tutto il processo di lavorazione, codifica e pubblicazione del dato.

Fino ai decessi avvenuti nel corso del 1994 il processo di codifica della causa di morte era manuale e la fasi di inserimento, controllo e correzione del dato erano successive alla codifica e alla registrazione numerica del codice-causa. Ogni anno in Italia ci sono in media oltre cinquecentocinquantamila decessi e ovviamente la preponderanza di fasi di lavorazione manuali causava negative ricadute in termini di qualità finale, di risorse impegnate e dei tempi nel rilascio del dato.

Il processo di organizzazione dell'indagine, profondamente rinnovato nell'ottica di attuare una crescente automazione, ha coinvolto quasi tutte le fasi della lavorazione.

È stata eliminata la lunga e onerosa fase di revisione manuale delle schede di morte in cui l'operatore effettuava una prima verifica sulla corretta compilazione del questionario nella parte demografica, completava le informazioni mancanti se deducibili da altri elementi della scheda e correggeva le incompatibilità più evidenti (ad esempio età al decesso e data di nascita). Tale fase è stata sostituita da una revisione automatica quantitativa dei modelli e da un piano di registrazione controllata progettato per individuare e correggere *on-line* i più ricorrenti errori *formali*² e *sostanziali*³ presenti nel modello, nonché gli errori ortografici derivanti dalla registrazione

¹ Modelli D4, D5, D4bis, D5bis.

Una riproduzione dei modelli viene riportata in Appendice B.

² Per *errore formale* si intende quello derivante dalle norme di compilazione del modello; si pensi ad esempio alle regole relative ai codici non ammissibili nelle variabili precodificate.

³ Per *errore sostanziale* si intende quello derivante da incompatibilità con informazioni derivanti da variabili diverse, indipendentemente dalla struttura del modello.

alfabetica di tutte le entità morbose riportate sulla scheda di morte così come richiesto dalle procedure di codifica automatica.

Una delle linee guida fondamentali è stata quella di sostituire il modello cartaceo con una 'mappa' informatizzata che riproduce su video le informazioni della scheda di morte nella sua totalità.

Questo ha portato ad anticipare di molto la fase di registrazione rispetto alla vecchia indagine. Da questa fase in poi tutto il processo è stato automatizzato con la parziale eccezione della gestione degli scarti della codifica automatica per i quali è stata sviluppata una procedura interattiva di codifica manuale sulla scheda informatizzata.

2 LE CAUSE DI MORTE E IL SISTEMA DI CODIFICA AUTOMATICA

2.1 LA CAUSA INIZIALE DI MORTE

Storicamente le statistiche di mortalità si basano sul concetto di causa iniziale di morte. La scelta di usare una sola causa per lo studio della mortalità è stata formalizzata nel 1900, quando a Parigi, nel corso della prima conferenza internazionale per la revisione della classificazione delle malattie, si evidenziò l'importanza di individuare, tra le diverse patologie insorte nel soggetto (successivamente deceduto), la 'causa primaria' di morte. La selezione di tale causa doveva avvenire sulla base di alcune specifiche regole delineate da Bertillion secondo la scuola francese. Nel corso di questo secolo sono state prodotte diverse revisioni della classificazione internazionale. Merita un'evidenza particolare l'innovazione apportata con la VI revisione del 1948 con cui si predispose un'unica classificazione da adottare sia nella mortalità che nella morbosità. Sempre in questa conferenza internazionale, condotta con la supervisione dell'OMS, la revisione subisce forti modifiche e si giunge ad una definizione ancora oggi attuale di 'causa iniziale di morte' ovvero 'la malattia o il traumatismo che avvia il concatenamento degli eventi morbosi che conduce direttamente alla morte', o 'l'insieme delle circostanze dell'accidente o della violenza che hanno provocato la lesione traumatica mortale'⁴.

Il concetto di causa iniziale di morte è oggi ben assimilato e abitualmente usato in tutto il mondo per la misurazione della mortalità.

Molto spesso l'individuazione della causa iniziale di morte è difficile perché più stati morbosi contribuiscono alla morte. L'identificazione della causa iniziale, tra le molteplici indicate su una scheda di morte, risulta pertanto un compito particolarmente delicato che richiede una specifica competenza in campo medico, una approfondita conoscenza delle regole di selezione indicate dall'OMS e di dover tenere conto di tutte le informazioni demografiche e sanitarie riportate sulla scheda di morte e dei possibili concatenamenti tra le diverse cause.

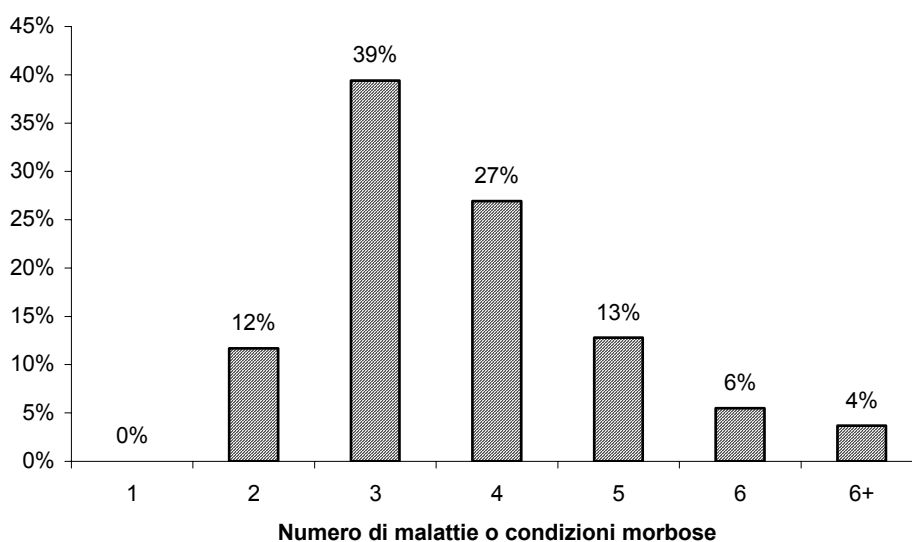
In genere la parte sanitaria del modello di rilevazione è ampiamente documentata e vengono compilati, oltre al primo quesito relativo alla 'causa iniziale', anche altri quesiti (Quesito 2: 'Causa intermedia o complicazione'; Quesito 3: 'Causa terminale'; Quesito 4: 'Altri stati morbosi rilevanti')⁵. In particolare si è osservato che, su un campione di 323 mila schede del 1995, il numero

⁴ OMS *World Health Organisation, International Statistical Classification of Diseases and related Health Problems, 10th revision*, Ginevra 1993.

⁵ Da notare che il formato del certificato di morte italiano non è conforme a quello indicato dall'OMS, in quanto l'ordine dei quesiti è invertito e le morti per cause naturali e quelle per causa di natura violenta devono essere descritte su colonne distinte.

medio di patologie o stati morbosi rilevanti o espressioni diagnostiche riportato sulle schede di morte è pari a 3,7. Quasi il 40% delle schede contiene 3 patologie, il 27% ne ha quattro e nel 3,7% dei certificati di morte sono riportate più di 6 diseases or conditions (cfr. Figura 1).

Figura 1 – Distribuzione percentuale delle schede di morte per malattie o condizioni morbose in Italia. Anno 1995.



Il medico certificatore è tenuto a compilare in modo appropriato la scheda di morte ed in particolare ad indicare, nel Quesito 1 del modello Istat, la causa maggiormente responsabile del decesso. Tuttavia, in molti casi la scheda non viene correttamente compilata e risulta incoerente con le raccomandazioni fornite dall'OMS. Per l'individuazione del codice-causa è quindi necessario far riferimento alle regole di selezione e modifica tra le diverse patologie previste nella IX o nella X Revisione della ICD. Tuttavia l'assegnazione di un solo codice è sottoposta ad una variabilità dovuta ai diversi soggetti che lavorano sul dato di mortalità e alla differente interpretazione che essi possono dare di uno stesso processo di malattia-morte. Talvolta è proprio il codificatore che non si attiene rigidamente alle regole di codifica e tende ad 'interpretare' la trascrizione del medico certificatore secondo un quadro di morbosità per lui più coerente.

Alle difficoltà nella codifica legate all'elevato numero di patologie, di espressioni diagnostiche e di termini medici, si aggiungono quelle non trascurabili dovute alla presenza sul certificato di morte di terminologia indecifrabile a causa della grafia dei medici. Questi ultimi, sebbene siano tenuti a compilare in stampatello il modello, spesso non si attengono alle indicazioni fornite dall'Istat. Ad aumentare infine la variabilità nella codifica della causa iniziale di morte intervengono altri importanti fattori legati, da un lato, al diverso uso di espressioni diagnostiche da parte dei medici (a seconda della regione di appartenenza o della scuola seguita), dall'altro alla

qualità della certificazione, ovvero ad un maggiore o minore dettaglio nella descrizione del quadro morboso del deceduto e alla pertinenza con l'evento fatale delle informazioni sanitarie riportate sulla scheda di morte.

In base a quanto sopra e pur con i limiti ampiamente documentati in letteratura gli indicatori basati sulla mortalità per causa continuano a rappresentare il mezzo più utilizzato dagli epidemiologi e dagli operatori sanitari al fine di misurare lo stato di salute di una popolazione e monitorare e/o programmare possibili interventi di salute pubblica (Pollard A. H. 1980; Hansluwska H. 1986; Vigotti et al. 1986; Preston S. H. 1987; CISM 1989; Barchielli A. et al. 1992; Barchielli et al. 1996).

A livello internazionale i dati di mortalità sono uno strumento fondamentale per effettuare confronti tra i diversi Paesi in termini di salute. Tuttavia, differenze nei dati di mortalità per causa possono dipendere non solo dal diverso quadro nosologico delle cause di morte ma anche da sostanziali differenze tra i Paesi nella certificazione, individuazione e codifica della causa iniziale di morte.

2.2 LE CAUSE MULTIPLE DI MORTE⁶

Le statistiche di mortalità basate sulla sola causa di morte permettono di osservare solo una parte di un quadro morboso letale molto più complesso. Questo è vero quanto più il profilo della mortalità è caratterizzato da patologie croniche a lungo decorso.

In questo secolo l'Italia è stata interessata da una rapida evoluzione sociale e sanitaria che ha avuto effetti diretti sui livelli di mortalità (riducendoli drasticamente) e sul suo profilo per causa. All'inizio del secolo le cause di morte prevalenti erano legate a malattie di natura infettiva, in particolare nelle età infantili. La drastica riduzione della mortalità per queste cause ha permesso ad un numero sempre crescente di persone di raggiungere le età più avanzate. Ma in genere è proprio con l'aumentare dell'età che l'organismo comincia a presentare le sue prime disfunzioni. Il passaggio da una condizione di salute ad una di malattia ed eventualmente di morte ben difficilmente si manifesta con un preciso evento. Esso è piuttosto rappresentabile come l'avvio di un processo - non necessariamente tutto palese - che, attraverso varie fasi, si sviluppa nel corso del tempo e dell'età dell'individuo verso diversi esiti (guarigione, cronicizzazione, morte). Durante il processo, l'individuo può attraversare fasi di diversa gravità della malattia, sia in termini di rischio di morte ad essa connesso sia in termini di limitazioni indotte sull'autonomia e sulla qualità della vita. Soprattutto in età avanzata, infine, difficilmente lo stato di malattia è riconducibile ad un unico processo morboso, più spesso diversi processi si cumulano e interagiscono nello stesso individuo,

⁶ La terminologia 'cause multiple di morte' proviene dalla traduzione dall'inglese di 'multiple causes', definizione internazionalmente riconosciuta e correntemente usata anche nei Paesi di lingua latina.

dando luogo a situazioni ancora più complesse. Sovente sono più concause che sinergicamente contribuiscono all'esito fatale.

L'uso delle cause multiple di morte nell'analisi della mortalità non solo permette di far luce su possibili relazioni tra diverse malattie e quindi anche sulle problematiche relative ai rischi in competizione, ma consente di studiare quella morbosità che, anche se non direttamente responsabile del decesso, ha contribuito sostanzialmente al deterioramento dello stato di salute dell'individuo (Chamblee & Evans 1979, 1982, Cornfield 1957, Dorn & Moriyama 1964, Goodman et al. 1982, Israel et al. 1986, Pavillion & Jouglà 1997). Nell'analisi delle cause multiple, in Italia ad esempio, il morbo di Parkinson e l'Alzheimer compaiono come concausa con una frequenza doppia rispetto alle volte che essi sono indicati come 'causa iniziale di morte'. In altre parole, solo un caso su due di malattia può essere considerato direttamente responsabile della morte.

2.3 VANTAGGI E PROBLEMATICHE DERIVANTI DALL'USO DI UN SISTEMA AUTOMATICO

2.3.1 *Impatto sulla qualità*

Il passaggio da una codifica di tipo manuale ad una di tipo automatico riduce la variabilità dovuta alla diversa interpretazione che i differenti soggetti che lavorano sul dato sanitario possono dare ad uno stesso processo di morbosità-mortalità. Viene pertanto eliminata la variabilità intracodificatori. A questo proposito alcuni lavori hanno focalizzato la loro attenzione sullo studio del confronto dei dati di mortalità a livello internazionale e sulle possibili distorsioni che si possono avere a causa delle diverse certificazioni o della variabilità nella codifica delle cause di morte (Jouglà, 1997). In particolare alcuni di questi lavori hanno evidenziato problemi di comparabilità legati esclusivamente ai processi di codifica tra i diversi Paesi (Percy 1978, 1989; Kelson 1983, 1987, Jouglà 1992, Balkau 1993, Mackenbach et al. 1987, Jouglà et al. 1992).

La codifica automatica permette di ridurre l'interpretazione soggettiva legata alla codifica manuale e di aumentare quindi la comparabilità dei dati di mortalità anche a livello internazionale.

Tuttavia esiste una quota di record che viene generalmente scartata dalla procedura automatica che deve essere codificata manualmente e che richiede tempi lunghi e personale altamente specializzato. Gli scarti riguardano infatti decessi che presentano un quadro morboso complesso (con un elevato numero di patologie riportate) oppure riconducibili a cause che richiedono un doppio codice (cfr. Paragrafo 3.5).

Con l'adozione della codifica automatica viene completamente eliminato l'errore dovuto alla incorretta registrazione del codice della causa (apportato manualmente sulla scheda) che si verificava sul 2% circa delle schede. La procedura automatica richiede infatti la registrazione alfanumerica su supporto magnetico dell'intera parte sanitaria della scheda e la codifica della causa

iniziale di morte avviene in una fase successiva alla registrazione. Tuttavia la fase di registrazione deve rispettare degli elevati standard qualitativi per poter assicurare una buona performance del sistema automatico (cfr. Capitolo 3).

2.3.2 Impatto sulla tempistica

L'automazione della procedura di codifica permette sicuramente di ridurre i tempi legati alle fasi manuale di individuazione e applicazione delle regole di selezione della causa iniziale di morte.

A parità di personale si riducono quindi enormemente i tempi di codifica, sebbene come visto, la codifica automatica richieda tempi di registrazione più lunghi. E' infatti necessario prevedere una registrazione alfanumerica di tutte le espressioni mediche riportate sul certificato di morte. Tuttavia i tempi più lunghi richiesti dalla fase di registrazione vengono ampiamente compensati dall'implementazione del sistema automatico di codifica. Un altro aspetto a favore della codifica automatica è che la maggior mole di lavoro che ora grava sulla registrazione ha alleggerito la fase di codifica manuale che necessita di professionalità più elevata.

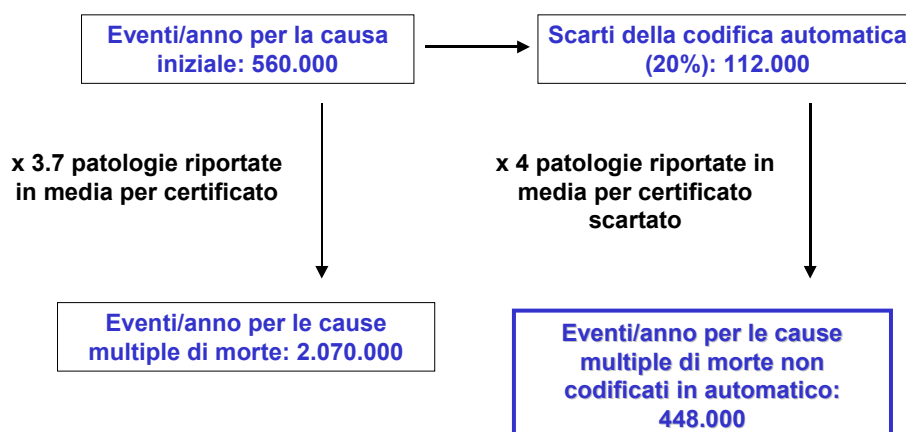
2.3.3 Informazione sulle cause multiple di morte

La codifica automatica delle cause di morte richiede, come visto, che tutta l'informazione della parte sanitaria del modello sia registrata su supporto magnetico. Essa attribuisce automaticamente un codice ad ogni malattia o condizione morbosa riportata sul certificato di morte. L'informazione disponibile comprende quindi ogni stato morboso o malattia che ha avuto rilevanza⁷ nel contribuire al decesso dell'individuo (cfr. Appendice C).

L'adozione della multimorbosità per l'intavolazione dei dati di mortalità se da un lato è, come visto, di estrema importanza da un punto di vista di incremento informativo, dall'altro comporta un grosso impegno di risorse preposte alla codifica. Il numero di malattie, lesioni o traumi che dovrebbe essere codificato annualmente in Italia supera i 2 milioni di unità, tenendo conto di una media di 3-4 malattie per ogni scheda di morte (Figura 2).

⁷ Secondo l'opinione del medico preposto alla certificazione

Figura 2 - Aumento del volume di lavoro con l'introduzione della codifica delle cause multiple.



Tuttavia l'adozione di un sistema automatico per la codifica delle multiple cause di morte permette di ridurre i tempi di lavorazione. In questo caso è necessario ricorrere alla codifica manuale solo per quelle entità morbose che non hanno superato la fase automatica. L'uso di procedure automatizzate fornisce quindi una ottima occasione per elaborare e diffondere dati sulle multiple cause. Tuttavia ancora oggi non esiste uno standard internazionale di riferimento che indichi i criteri di codifica e di intavolazione dei dati. Prezioso in questo campo è il lavoro concettuale e tecnico sviluppato dall'NCHS che per primo ha ideato un impianto di regole e un software (MICAR, ACME, TRANSAX) validi per la codifica della multiple cause. Pochi sono i Paesi che attualmente rilasciano dati ufficiali sulla multiple cause, tra questi si annoverano, oltre agli Stati Uniti, l'Australia.

2.4 LA CODIFICA AUTOMATICA IN ITALIA

In Istat, a partire dai dati relativi al 1995, il 76% circa di tutti i decessi viene codificato automaticamente⁸. Il sistema di codifica utilizzato è una versione integrata, modificata e adattata alla lingua italiana di MICAR (Mortality Medical Coding Indexing, Classification and Retrieval) e ACME (Automated Classification of Medical Entities), di programmi già ampiamente sperimentati e applicati da oltre 30 anni dal National Center of Health Statistics (NCHS) degli Stati Uniti di America⁹. La scelta di usare una versione modificata di MICAR-ACME è legata principalmente ai tempi e ai costi dell'implementazione. E' noto infatti che qualsiasi innovazione in un processo comporta, proprio in quanto tale, un incremento di risorse e/o una perdita di produttività nella fase di transizione in vista di futuri vantaggi. L'adattamento al caso italiano di procedure già esistenti ha

⁸ Per le eccezioni e per una descrizione più dettagliata si rimanda al Capitolo successivo

⁹ <http://www.cdc.gov/nchs/about/major/dvs/medsof.htm>

sicuramente consentito di avvalersi dell'esperienza maturata nello stesso campo da altri Paesi e ha permesso di ridurre significativamente i tempi ed i costi di produzione. Una delle problematiche emerse nell'ambito dell'International Collaborative Effort on Automating Mortality Statistics (ICE group) e maggiormente sentite dai Paesi che si avviano all'automatizzazione, è legata proprio agli alti costi da sostenere per la programmazione e lo sviluppo di software appropriato (CDC, 1999). Un modo per ovviare, seppure parzialmente, a questo problema è la soluzione adottata dall'Istat, ovvero lo sviluppo di programmi da integrare alla versione americana di MICAR -ACME.

2.5 LA PROCEDURA ACME

Il sistema ACME utilizza delle opportune tavole di decisione per tradurre le regole dettate dall'OMS in passato seguite e applicate manualmente dai nosologi e/o dai codificatori. Ciò permette di ridurre i tempi di codifica e, allo stesso tempo, di annullare la variabilità soggettiva dovuta ai diversi operatori che partecipano al processo di codifica manuale per individuare la causa iniziale di morte.

L'attuale procedura ACME utilizzata in Italia seleziona automaticamente la causa iniziale di morte e associa ad essa un codice della IX revisione della ICD (gli U.S.A utilizzano oggi la versione di ACME aggiornata ai codici della X revisione e hanno pubblicato i dati secondo la nuova revisione).

La procedura ACME ha un importantissimo pregio: non dipende dalla lingua parlata dal Paese in cui è stata progettata (Inglese). Le informazioni sanitarie necessarie per la procedura di codifica automatica vengono infatti trascritte sotto forma di codici della IX revisione (oppure della X). Per ciascuna malattia o lesione riportata sulla scheda di morte i codici possono essere attribuiti manualmente o tramite la procedura automatica MICAR.

ACME è in grado di individuare la causa iniziale di morte nel 97,5% dei casi sottoposti alla procedura.

2.6 LA PROCEDURA MICAR

MICAR è stato sviluppato dall'NCHS negli anni 1992-1993 (versione definitiva 1994), molti anni dopo la procedura ACME.

MICAR è un programma complesso composto da due moduli:

1. Il **primo modulo** (MICAR 100) attribuisce ad ogni causa di morte, entità morbosa o lesione, un codice ERN (Entity Reference Number) attraverso un dizionario (The Big Book of Death, BBD) che comprende una lista di cause chiamate 'entities', ovvero entità mediche. Un'entità medica può essere una malattia, un disturbo, un processo morboso, una anomalia, un sintomo, un avvelenamento, una modalità della morte. L'obiettivo di questo sistema, basato sui codici

ERN, è quello di avvicinarsi il più possibile alla terminologia comunemente più usata, tenendo conto, laddove è possibile, del differente modo di parlare e di esprimersi dei medici, indipendentemente dalla terminologia adottata nella ICD. Nel BBD si è quindi cercato di includere tutta la potenziale terminologia e conseguentemente ad ogni singolo ERN possono corrispondere diverse diciture simili (ALIAS).

Se al primo tentativo non si ottiene l'accoppiamento tra entità medica trascritta sulla scheda di morte e dizionario di MICAR (e quindi codici ERN), la procedura MICAR tiene conto di una lista di termini 'ininfluenti' ai fini della codifica e procede alla loro eliminazione. Tuttavia per aumentare la probabilità di associazione tra patologia ed ERN, senza far 'esplodere' le dimensioni del dizionario, la terminologia medica riportata sul certificato viene standardizzata 'a priori' in accordo con il dizionario utilizzato da MICAR. Tale dizionario è composto da migliaia di termini medici ai quali corrispondono i codici ERN. La versione americana del dizionario è costruita in modo indipendente dalla revisione adottata, esso contiene infatti i codici ERN che sono appunto indipendenti sia dalla IX che della X rev. Tuttavia MICAR 100, che si avvale di informazioni alfabetiche, dipende necessariamente dalla lingua utilizzata. Un lavoro particolarmente oneroso e delicato è stato quello di tradurre il dizionario inglese e di costruirne un altro adatto alla terminologia medica italiana (composto da circa 160.000 termini)¹⁰. Alle difficoltà dovute alle differenze sintattiche delle due lingue si devono aggiungere le difficoltà di traduzione dei modi di esprimersi dei medici americani in quelli tipici dei medici italiani. L'attuale versione italiana del dizionario utilizzato da MICAR 100 dipende dalla IX revisione, in quanto non è stato sempre possibile attribuire a ciascuna terminologia utilizzata dai medici italiani il corrispondente ERN del dizionario americano ed in questi casi si è scelto di attribuire all'espressione medica italiana (non presente nel dizionario americano) un ERN coerente a livello di codice della IX revisione.

Attualmente la percentuale di accoppiamento tra entità morbosa ed ERN è del 94% dei casi.

2. Il **secondo modulo** (MICAR 200) trasforma gli ERN in codici della IX revisione della Classificazione Internazionale delle Malattie (ICD) utilizzando specifiche regole di codifica. Un codice della IX è la conseguenza di uno o più ERN a seconda della relazione che intercorre tra questi. Le specifiche regole di codifica utilizzate da MICAR 200 tengono anche conto di alcuni modificatori tra i quali età, sesso o tipologia delle malattia (ovvero se di origine perinatale, o reumatica o se postumo). Tali modificatori fanno sì che ad uno stesso ERN possano talvolta corrispondere differenti codici della IX revisione della ICD.

¹⁰ La traduzione del dizionario è stata effettuata dal Dott. Giuseppe Feola, Istituto Nazionale di Statistica.

Attualmente, alla fine dell'intera procedura MICAR (100 e 200), 80 records su 100 hanno tutte le espressioni diagnostiche 'tradotte' in codici numerici (IX revisione).

2.7 LA PROCEDURA TRANSAX

La procedura è stata sviluppata dall'NCHS (NCHS 1986) per ovviare all'oneroso lavoro necessario per produrre le statistiche di mortalità secondo una codifica multipla, da parte di personale specializzato.

In sostanza TRANSAX (TRANSlation of AXis) è il software complementare ad ACME che permette di passare dall'asse delle entità morbose all'asse dell'individuo, ovvero, a partire dall'insieme delle entità morbose trascritte in una scheda di morte, consente di rappresentare tale scheda secondo quel percorso logico 'essenziale' che conduce l'individuo verso un esito fatale. In particolare, TRANSAX si occupa di verificare e rimuovere contraddizioni, eliminare duplicazioni e, più in generale, permette di produrre una accurata ed esaustiva codifica delle cause di malattia e/o morte in accordo con la struttura generale della classificazione internazionale delle malattie (ICD IX o X). Per fare un esempio, nel caso in cui l'asse delle entità contenesse due informazioni molto simili e una di queste fosse maggiormente specifica, TRANSAX eliminerebbe l'entità morbosa più generica. Se in una scheda di morte, per esempio, fossero indicati 'infarto del miocardio' e 'insufficienza cardiocircolatoria', TRANSAX procederebbe all'eliminazione dal record dell'insufficienza cardiocircolatoria perché ridondante.

Attualmente, ad eccezione degli USA e dell'Australia, la procedura TRANSAX non è adottata come standard di uso corrente nei paesi che adottano la codifica automatica; in Italia è stata effettuata presso l'Unità operativa 'Mortalità per Causa' dell'Istat una prima sperimentazione che riguarda le cause multiple di mortalità in presenza di malattia di Alzheimer (Istat, 2003).

3 LA CODIFICA AUTOMATICA E L'ADEGUAMENTO ALLA TERMINOLOGIA MEDICA ITALIANA

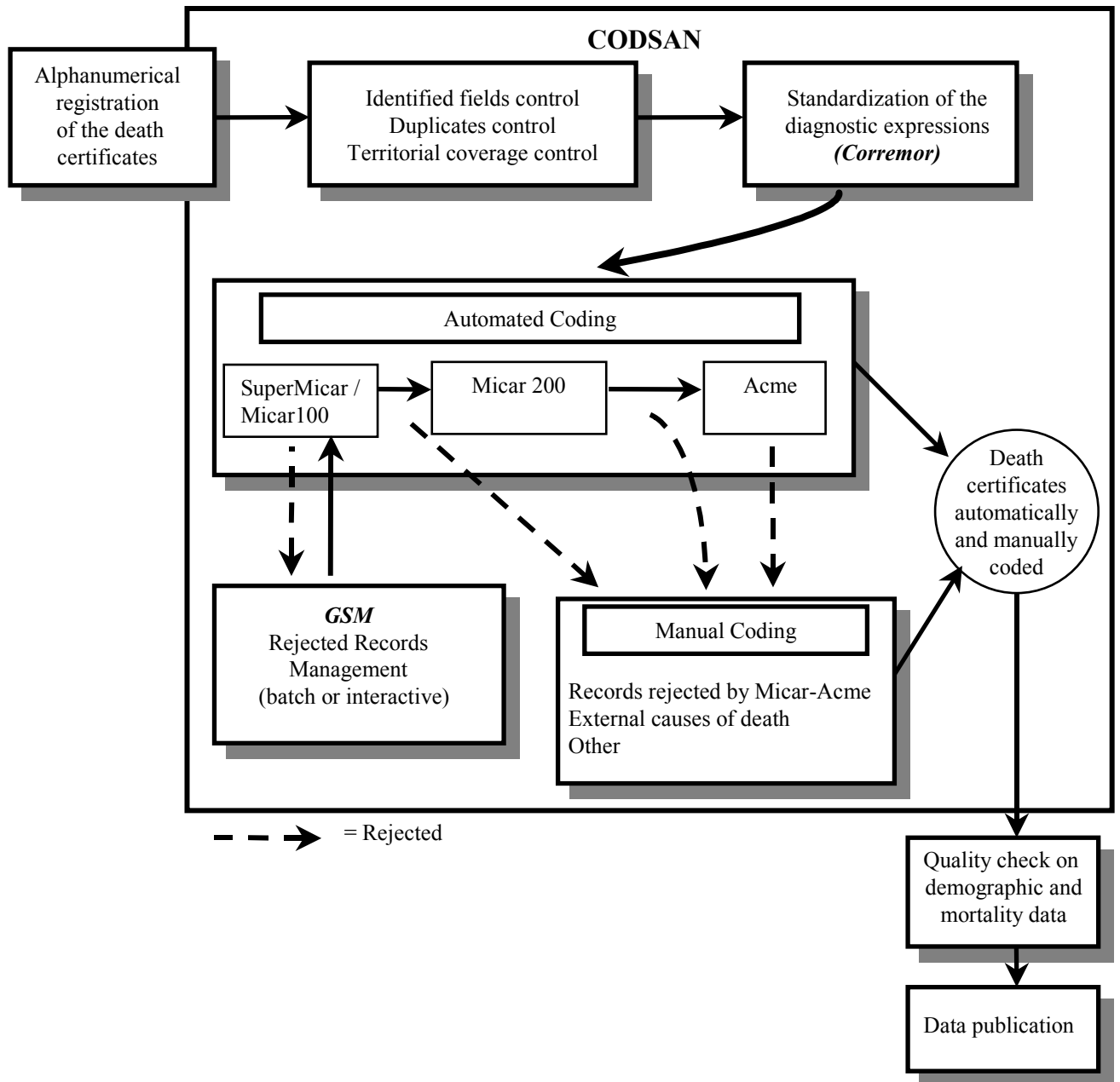
3.1 L'ESPERIENZA ISTAT

Nel precedente capitolo è stato descritto il cuore della codifica automatica per le cause di morte: l'insieme dei programmi che associano un codice a ciascuna delle cause riportate sul certificato di decesso e successivamente, attraverso l'uso di tavole decisionali, individuano il codice della causa iniziale di morte.

L'ottimizzazione del processo di produzione del dato, oltre ad aver richiesto la revisione delle diverse fasi dell'indagine, ha tuttavia reso opportuno implementare nuovi programmi e procedure per sfruttare al meglio le potenzialità della codifica automatica (Figura 3). In particolare è stato sviluppato un software (CODSAN) che permette di effettuare una serie di controlli qualitativi e quantitativi sul materiale pervenuto, di gestire in maniera integrata il flusso dei file e di intervenire sugli scarti della codifica automatica. Tre le funzioni più importanti si evidenziano:

- controllo degli identificativi;
- controllo delle distribuzioni dei quantitativi per provincia di decesso;
- standardizzazione della terminologia medica (procedura CORREMOR);
- gestione degli scarti (GSM);
- codifica on-line delle schede non processate con il sistema automatico.

Figura 3 - Le fasi della codifica in Italia, ICD IX.



Nelle prime applicazioni di codifica automatica, in cui alle potenzialità di MICAR-ACME (versione italiana) si aggiungevano solo i vantaggi di una registrazione controllata¹¹, la percentuale delle schede codificate era del 61%. Attualmente, con la procedura 'MICAR-ACME integrata', il 76% delle schede di morte (circa 450 mila) è codificato automaticamente. Tale percentuale è in linea con le statistiche di lavorazione della maggior parte dei Paesi che utilizzano la procedura automatica.

¹¹La registrazione delle schede di morte deve comunque seguire dei criteri standard ed utilizzare una particolare simbologia necessaria per processare MICAR.

3.2 LA REGISTRAZIONE DELLE SCHEDE DI MORTE

Fino ai decessi del 1994 la registrazione su supporto informatico dei modelli delle cause di morte era completamente numerica. Con l'avvio della procedura di codifica automatica è nata l'esigenza di trascrivere tutta l'informazione sanitaria (ad esempio nel caso di morte naturale: causa iniziale, cause intermedie, causa finale e altri stati morbosi rilevanti) e di conseguenza è stato rivisto interamente il piano di registrazione.

Uno dei maggiori limiti della procedura di codifica automatica è che richiede un lavoro accurato ed oneroso in termini di tempi e costi nella fase di inserimento dei dati:

- *accurato*, perché l'eventuale scarsa qualità della fase di inserimento dei dati ha un notevole impatto sui tempi di lavoro delle fasi successive;
- *oneroso*, perché comporta la digitazione di tutte le informazioni riportate sulla scheda. La procedura automatica richiede inoltre l'uso di una particolare simbologia, necessaria per identificare correttamente la sequenza degli eventi morbosi o le eventuali concause, e, laddove presente, richiede che venga registrata anche la durata della/e malattie e degli stati morbosi trascritti sulla scheda di morte. Il numero medio di caratteri per la registrazione di una intera scheda di morte è ora di circa 200 (con la registrazione numerica il numero dei caratteri era meno della metà); ne consegue che anche i costi di registrazione siano molto più elevati.

Come detto, uno dei presupposti principali per l'uso della procedura automatica MICAR-ACME è la corretta trascrizione dei termini medici indicati nella scheda. La presenza di errori ortografici o di errate interpretazioni della grafia dei medici riduce ovviamente la probabilità di associazione tra espressione diagnostica e codici ERN (del dizionario di MICAR 100) e di conseguenza diminuisce il numero di schede codificate in automatico. Per minimizzare i possibili errori derivanti dalla fase di registrazione si è scelto di verificare la qualità del dato alfanumerico direttamente nella fase di data entry attraverso il confronto dei singoli termini registrati (e conseguente correzione degli eventuali errori) con i termini presenti su un dizionario medico appositamente preparato dall'Istat (circa 50.000 voci). In fase di registrazione il controllo della corrispondenza dell'intera entità morbosa con il dizionario (BBD) non è applicabile perché richiederebbe personale qualificato e rallenterebbe eccessivamente la fase di inserimento dati.

Considerata l'elevata specificità delle notizie riportate nella parte sanitaria della scheda, la fase di inserimento dei dati deve essere preceduta da un corso finalizzato all'apprendimento delle principali regole da seguire e alla comprensione della grafia dei medici e della terminologia trascritta sulla scheda di morte. Il personale proposto alla registrazione dei modelli (non necessariamente codificatore, medico o nosologo) deve quindi conoscere adeguatamente la terminologia riportata sulle schede di morte ed essere abile nell'interpretare la grafia dei medici. In ogni caso, per ridurre

al minimo possibili errori derivanti dalla registrazione è opportuno che in fase di digitazione venga riprodotta fedelmente l'intera parte sanitaria della scheda di morte, evitando il ricorso a modifiche del testo suscettibili di interpretazioni personali.

L'obiettivo di ridurre il numero delle schede difficilmente decifrabili o addirittura completamente illeggibili è stato perseguito anche cambiando la grafica dei modelli. Al fine di facilitare la scrittura e quindi la lettura dei termini medici, è stata ampliata la dimensione della parte sanitaria dei D4, D5, D4bis e D5bis. In particolare, è stato lasciato maggiore spazio ai quesiti alfanumerici relativi alle cause di malattia e/o di morte, e sono state inserite delle caselle per incentivare l'uso di caratteri in stampatello da parte dei medici.

Nell'ambito di una revisione del piano di registrazione si è migliorata anche la fase di data entry delle informazioni demografiche del deceduto: è stato introdotto un controllo in tempo reale della digitazione numerica di dati errati (ad esempio digitazione di codici non ammessi, incompatibilità tra il codice comune ed il codice provincia o tra data di nascita e l'età alla morte).

La predisposizione di un nuovo piano di registrazione controllata delle informazioni, sia per la parte demografica sia per la parte sanitaria contenente le espressioni diagnostiche, ha permesso di ridurre in modo consistente gli errori di registrazione. Questo ha comportato una riduzione del 70% delle incompatibilità sui dati demografici. Da notare che con l'introduzione del sistema di codifica automatica gli errori sanitari si sono ridotti del 75%.

3.3 LA PROCEDURA CORREMOR

Come visto nel paragrafo precedente, la fase di registrazione prevede che vengano trascritte fedelmente tutte le entità morbose riportate sul certificato di morte. Poiché è possibile per una stessa patologia utilizzare termini diversi, gli scarti saranno tanto minori quanto più la terminologia adoperata dai medici nella compilazione delle schede di morte e il dizionario utilizzato dalla procedura MICAR sono simili.

Per comprendere l'infinità varietà dei casi che si possono presentare è utile fornire alcuni esempi.

Per *adenocarcinoma* è possibile trovare una numero elevato di possibili trascrizioni nelle diverse schede di morte (compresi gli errori ortografici¹²):

ADENOCARCINOMA, ADENOCA, ADK, ADENOK, ADENOCAR

così come per *metastasi*:

METASTASI, METASTATIZZATO, METASTATIZZATE, M, MTS, MTX, METASTATIZATO , ETC.

o per *destro* (analogamente per *sinistro*):

¹² La registrazione controllata limita la possibilità di errore ma non la esclude totalmente.

DESTRO, DESTRA, DESTRE, DESTRI, DX, DS, DES.

Considerando le possibili combinazioni tra le varie dizioni di adenocarcinoma, metastasi, destro (o sinistro), l'aggiunta di altre specificazioni (quali ad esempio la sede di insorgenza del tumore), oltre all'eventuale uso di articoli, congiunzioni, preposizioni ed altre parole ininfluenti, si capisce come la dimensione del dizionario di MICAR (che dovrebbe contenere tutte le espressioni diagnostiche utilizzate dai medici italiani che certificano il decesso) sarebbe quasi priva di limite.

L'esecuzione della procedura MICAR viene fatta quindi precedere da una fase preliminare (CORREMOR) che riduce la variabilità linguistica attraverso una correzione automatica delle parole errate più frequenti ed una gestione dei sinonimi, delle abbreviazioni, dei singolari e dei plurali e delle parole ininfluenti. Tale normalizzazione, applicata sia al testo riportato nel certificato di morte che al dizionario usato da MICAR (Big Book of Death, BBD), permette di aumentare la percentuale dei record codificati automaticamente.

Per realizzare la normalizzazione si è delineato un profilo dei casi più frequenti di mancato accoppiamento analizzando i record scartati dalla procedura automatica, particolare attenzione è stata dedicata alle espressioni diagnostiche scartate ma presenti almeno quattro volte in un mese di lavorazione, e da questo si è creato un file con le standardizzazione da realizzare.

Per fare qualche esempio, la procedura sostituisce sempre i termini:

METASTASI, METASTATIZZATO, M, MTX, METASTATIZATO,	con	MTS
SINISTRO, SIN, SINISTRE, SINISTRI	con	SX
INSUFFICIENZA, INSUFF, INSUFFICENZA, INS, INSUF	con	INSUFF

L'aumento di schede codificate in automatico conseguente all'introduzione della fase di standardizzazione è di 8 punti percentuali.

3.4 LA PROCEDURA GSM

Come abbiamo visto, CORREMOR effettua una standardizzazione a priori delle singole parole e permette di aumentare la probabilità che un determinato record venga accettato dalla codifica automatica. Tuttavia, la quota degli scarti dovuti al mancato accoppiamento con il dizionario di una, o più, espressioni diagnostiche presenti sulla scheda era ancora elevato (il 95% del 25% circa di scarto complessivo). Si è quindi implementata una procedura di gestione di tali scarti che permette di trasformare, ovviamente senza alterarne il significato, le diciture non accoppiate.

Nel progettare tale procedura (GSM) si è ragionato su due livelli:

- i) correzioni batch: è stata analizzata la distribuzione degli scarti di un intero anno di lavorazione e a partire dagli scarti più frequenti sono state individuate, quando possibile, le

corrispondenti diciture presenti sul dizionario. A questo livello, si è potuto intervenire solo laddove i singoli scarti erano da ritenersi ‘autoreferenziali’, non era cioè necessario conoscere cosa altro era riportato sul certificato. Il concetto di ‘autoreferenzialità’ di uno scarto è probabilmente più comprensibile riportando un esempio di scarto ‘non-autoreferenziale’: una sede anatomica senza l’indicazione della patologia o del traumatismo. E’ stato quindi costruito un file con due colonne: dicitura originale e corrispondente sostituzione. Questo file viene elaborato in maniera batch permettendo una correzione rapida degli errori ricorrenti su tutti i record di ciascun mese.

- ii) correzioni individuali: per gli scarti non gestiti dalla modalità batch viene visualizzato l’intero record e vengono evidenziate le patologie non presenti nel dizionario; viene data la possibilità di correggere le espressioni diagnostiche consultando on-line il dizionario utilizzata da MICAR.

Le due modalità di intervento sono compatibili e possono essere effettuate più volte su ciascun mese in lavorazione. Il programma permette inoltre di ampliare continuamente la casistica delle correzioni più frequenti (da effettuare in modalità batch), attraverso una semplice gestione interattiva dell’elenco degli errori ricorrenti. Attualmente è raro che in un mese di lavorazione ci siano errori ‘autoreferenziali’ con frequenza superiore a tre e la sola opzione batch di questa procedura permette di recuperare circa il 7% di schede mensili. Tali schede vengono sottoposte nuovamente alla fase di codifica automatica. Di conseguenza, le performance della procedura di codifica automatica sono valutabili intorno ad una percentuale di accoppiamento del 76%.

3.5 LA CODIFICA MANUALE

In questa fase vengono codificati on-line i record definitivamente scartati da MICAR-ACME ed i record che seppure codificati automaticamente hanno almeno una delle patologie riconducibili ad una ‘causa esterna’ o a cause mal gestite dal software. Quindi, il 23-25% dei record mensili vengono codificati manualmente in modo interattivo visualizzando su schermo le informazioni demografiche e sanitarie riportate sulla scheda necessarie per l’individuazione della causa iniziale di morte. Per questo insieme di schede, al fine di contenere i tempi del rilascio del dato, non è prevista la codifica delle cause multiple e verrà codificata solo la causa iniziale di morte.

L’introduzione della codifica automatica richiede comunque personale altamente preparato per codificare manualmente il 24% degli scarti. Questo perché i decessi che vengono codificati manualmente presentano generalmente un quadro morboso più complesso dei decessi in cui la codifica automatica ha avuto esito positivo (hanno infatti una numero medio di patologie più elevato). Trattandosi, inoltre, di decessi frequentemente riconducibili a:

- cause di natura violenta
- conseguenze di atti chirurgici
- conseguenze letali dovute all'assunzione di farmaci e altre sostanze chimiche

il tempo di lavorazione risulta più elevato poiché in questi casi è richiesto il doppio codice, uno relativo alla classificazione N (natura della lesione) e l'altro relativo alla classificazione supplementare delle cause esterne, dei traumatismi e degli avvelenamenti (codici E).

3.5.1 La codifica delle cause di morte di natura violenta

Nella fase di studio della procedura di codifica automatica si è visto che la codifica effettuata da MICAR-ACME per le cause di natura violenta non era soddisfacente: gestire l'ampia casistica e la relativa terminologia delle modalità di un evento esterno e delle sue conseguenze in modo da attribuire i codici automaticamente era già di per se una operazione lenta e faticosa.

Per questi motivi si è evitato di destinare risorse umane alla fase di pre-codifica degli eventi traumatici e di conseguenza molte delle schede relative alle cause di morte di natura violenta vengono direttamente scartate dalla procedura automatica in quanto contenenti termini (in particolare quelli usati per la descrizione della lesione) non presenti nel BBD. Inoltre, a causa della scarsa qualità anche la quota di schede codificate automaticamente ma che contengono almeno una patologia con codice ICD-IX>800.0, vengono estratte e codificate da personale specializzato.

3.5.2 La codifica dell'AIDS

Nella fase di studio della procedura di codifica automatica si è eseguita un'analisi approfondita delle differenze esistenti tra il sistema di codifica manuale e quello automatico. In tale analisi si è riscontrata una limitata capacità del sistema di codifica automatica ad individuare i decessi la cui causa iniziale è riconducibile all'AIDS. In particolare, la codifica automatica non riesce ad individuare alcuni casi di morte per AIDS correttamente individuati invece dalla codifica manuale. La versione MICAR-ACME per mainframe (adottata in Italia per codificare in decessi del 1995, 1996 e 1997) sottostimava l'Aids del 10% circa. Tale percentuale si è ridotta con la versione per PC. Per capire come ovviare a questa limitazione è stato condotto uno studio qualitativo sui casi in cui il codice AIDS assegnato manualmente non trovava riscontro dalla codifica automatica. Si è visto che la codifica automatica riusciva ad individuare una o più patologie Aids correlate¹³ ma le tavole di decisione l'assegnazione del codice per la causa iniziale non ne tenevano conto in maniera soddisfacente.

¹³ ICD IX: 0072; 0319; 0429; 0439; 0449; 0539; 0785; 1128; 130; 1363; 1398; 1739; 2793; 3632; 7573; 7958.

Per questo motivo, nella fase corrente di produzione del dato, vengono sottoposti a codifica manuale anche i record con codice automatico di causa iniziale diverso dall'AIDS ma contenenti nella parte relativa alle cause multiple di morte almeno un codice di patologie 'AIDS-correlate'.

3.6 EFFICACIA DEL SISTEMA DI CODIFICA AUTOMATICA

Nei paragrafi precedenti sono stati descritti gli accorgimenti adottati per aumentare la percentuale di schede codificate automaticamente (registrazione controllata, CORREMOR e GSM) e per migliorare qualitativamente le performance dell'intero processo di codifica (re-codifica manuale di determinate tipologie di decessi). Per chiarire meglio quali sono state le conseguenze delle varie scelte in termini di percentuale di codifica automatica si può far riferimento alla seguente Tabella 3.1:

Tabella 3.1 - Efficacia del sistema di codifica automatica. Percentuali di accoppiamento.

	SUPERMICAR /MICAR100	MICAR200	ACME	Percentuale di records accettati
Registrazione controllata	65%	64%	62%	60%
Registrazione controllata e CORREMOR	74%	73%	71%	69%
Registrazione controllata, CORREMOR e GSM	81.5%	80%	78%	76%

4 BRIDGE CODING TRA CODIFICA MANUALE E AUTOMATICA

4.1 IL CAMPIONE

Il sistema di codifica automatico ha permesso senza dubbio la ristrutturazione ed ottimizzazione dell'intero processo di produzione dei dati, ma ha condotto anche alla creazione di una discontinuità nei dati sulle cause di morte tra il 1994 ed il 1995. Per questo motivo è stato progettato e realizzato un esteso **bridge coding** sottoponendo a doppia codifica, manuale ed automatica, i decessi relativi a sette dei dodici mesi del 1995 (gennaio, febbraio, marzo, maggio, luglio, settembre, novembre) opportunamente scelti, tenendo conto delle eventuali variazioni dovute all'effetto della stagionalità, per un totale di circa 300.000 schede di morte. L'applicazione della procedura automatica ha avuto successo, comunque, solo nel 77% dei casi, circa 250.000 schede di morte. I decessi per cause violente e per Aids sono stati esclusi dall'analisi in quanto sottoposti solo a codifica manuale.

4.2 RISULTATI PRINCIPALI

I gruppi di cause sono quelli adottati dalla ICD IX. Dei 245.999 records individuali analizzati si osserva una discordanza nell'11,1% dei casi, ovvero circa 27.000 decessi presentano una differenza 'sostanziale' nel codice attribuito dai due metodi di codifica, differenze dovute, come vedremo in seguito, alle molteplici condizioni morbose più o meno letali riportate su una scheda di morte ed alle connessioni tra loro (Tabella 4.1).

Tabella 4.1 - Concordanza e discordanza tra il metodo di codifica automatico e quello manuale.¹⁴

	Numero Casi	Percentuale ^(a)
Numero totale casi analizzati	245.999	100,0
Discordanza	26.259	11,1
Concordanza di gruppo	219.740	89,3
<i>Di cui:</i>		
<i>Concordanza puntuale alla III cifra</i>	<i>189.796</i>	<i>77,2</i>
<i>Concordanza puntuale alla IV cifra o perfetta</i>	<i>176.320</i>	<i>71,7</i>

(a) La percentuale è calcolata sul totale dei casi

¹⁴ *concordanza puntuale alla IV cifra o concordanza perfetta*: perfetta coincidenza fra i codici attribuiti ai due sistemi di codifica fino alla IV cifra;
concordanza puntuale alla III cifra: perfetta coincidenza fra i codici attribuiti dai due sistemi di codifica almeno alla III cifra;
concordanza di gruppo: coincidenza nel gruppo ma non necessariamente nel singolo codice a tre cifre;
discordanza: le due codifiche non coincidono né a livello di gruppo, né a livello di categoria.

Per i casi più complessi può essere infatti difficile individuare univocamente la causa del decesso per la presenza nella scheda di più malattie che possono aver condotto alla morte, quali tumore, malattie dell'apparato circolatorio o respiratorio ecc.

I due metodi di codifica hanno individuato esattamente lo stesso codice (fino alla IV cifra) nel 71,8% dei casi e tale percentuale sale al 77,3% quando si considera la concordanza puntuale a livello di categoria (III cifra).

Sebbene la concordanza sia complessivamente elevata si riscontrano differenze anche notevoli nell'ambito dei diversi gruppi considerati (Tabella 4.2).

Tabella 4.2 - Distribuzione dei decessi per settore della ICD IX secondo i due metodi di codifica.

Gruppi di cause (ICD IX)	Metodi di codifica		Concordanza			Concordanza(%) (a)		
	Manuale	Automatica	Gruppo	III cifra	IV cifra	Gruppo	III cifra	IV cifra
Malattie Infettive (1-139)	734	1.148	461	405	369	62,8	55,2	50,3
Tumori (140-239)	70.793	69.089	67.923	63.912	61.344	95,9	90,3	86,7
Malattie ischemiche del cuore (410-414)	37.413	37.013	33.418	31.829	30.725	89,3	85,1	82,1
Altre malattie del sistema circolatorio (390-409; 415-459)	81.781	82.340	73.053	55.793	52.285	89,3	68,2	63,9
Malattie dell'apparato respiratorio (460-519)	16.556	17.484	13.779	10.626	10.384	83,2	64,2	62,7
Malattie dell'apparato digerente (520-579)	12.517	11.539	10.287	9.613	8.516	82,2	76,8	68,0
Altre malattie (b) (codice < 800)	26.205	27.386	20.819	17.618	12.697	79,4	67,2	48,5
Totale cause	245.999	245.999	219.740	189.796	176.320	89,3	77,2	71,7

(a) La percentuale è uguale al rapporto dei casi concordanti (a livello di gruppo, di III, di IV cifra) ed i casi codificati manualmente.

(b) I decessi per Aids non sono inclusi.

Per i tumori e le malattie del sistema circolatorio, che insieme costituiscono oltre il 70% della mortalità complessiva, la percentuale di concordanza di gruppo (ottenuta rapportando i casi concordanti a livello di record con i casi codificati manualmente) è infatti molto elevata e rispettivamente del 95,9% e del 89,3%.

Nel caso dei tumori, inoltre, anche la concordanza puntuale, a livello sia di categoria che di sottocategoria, è molto elevata e rispettivamente del 90,3% e del 86,7%.

Per quanto riguarda le altre cause di morte si osserva una maggiore discordanza nel gruppo delle malattie infettive, delle ghiandole endocrine, nutrizione e metabolismo, delle malattie del sangue, dei disturbi psichici, delle malattie del sistema nervoso, delle malattie della pelle e del sistema osteomuscolare e connettivo: queste cause costituiscono nel loro insieme poco più del 6% dei casi analizzati.

Dai dati presi in considerazione è possibile affermare che la concordanza tra i due sistemi di codifica è sicuramente maggiore nel caso delle malattie più diffuse e quindi più note sia al medico certificatore (agevolato in una più facile certificazione della morte e compilazione della scheda) sia al codificatore Istat.

Una spiegazione della non corrispondenza tra codifica manuale ed automatica, ed in particolare della discordanza completa tra le codifiche, potrebbe essere individuata nella diversa esperienza professionale dei medici certificatori e nella soggettività diagnostica del processo di morbosità.

Altre discordanze infine possono dipendere non dal diverso sistema di codifica, bensì da errori casuali intervenuti nella fase di registrazione delle schede nel momento in cui il codice viene apposto manualmente sul modello cartaceo.

Per una migliore comprensione delle differenze riscontrate tra i due metodi di codifica sono stati applicati dei modelli di analisi multivariata che tenessero conto di variabili quali il sesso, l'età, la ripartizione geografica in cui è avvenuto il decesso, il numero delle patologie presenti sulla scheda di morte e la causa iniziale secondo la codifica automatica (Cralesì et al., 1998).

I risultati più interessanti possono essere così brevemente riassunti:

- I due criteri di codifica risultano indipendenti dalle caratteristiche demografiche del deceduto.
- Le maggiori differenze riscontrate tra i due metodi di codifica dipendono dalla causa di decesso e dalla complessità del quadro patologico riportato sulla scheda, indirettamente misurato dal numero di patologie. Come evidenziato dall'analisi descrittiva, i gruppi di cause che presentano probabilità di discordanza più elevate risultano essere le malattie infettive. Al contrario i tumori e le malattie dell'apparato digerente presentano una discordanza estremamente contenuta.
- In presenza di un elevato numero di patologie sulla scheda di morte la discordanza è sensibilmente più elevata quando la codifica automatica individua come causa iniziale di morte una malattia infettiva. Nel caso dei tumori, viceversa, anche in un quadro patologico complesso, i due diversi metodi di codifica portano comunque a risultati concordanti nella maggior parte dei casi: basti pensare che anche in corrispondenza della classe 'più di 5 patologie' vi è una consistente riduzione della discordanza rispetto alla media generale.
- La discordanza cresce al crescere dell'età per tutte le cause, fatta eccezione per le malattie infettive che è invece valida la relazione inversa.

4.3 IMPATTO DEL SISTEMA AUTOMATICO SUI DATI DI MORTALITÀ PER CAUSA

Per dare un'indicazione dell'entità dei cambiamenti avvenuti nella serie storica di mortalità per causa si è ritenuto opportuno calcolare, sulla base del campione sottoposto a doppia codifica, riferito all'anno di decesso 1995, un coefficiente di raccordo K calcolato nel seguente modo:

$$K_i = \frac{i d_N}{i d_M}$$

i = causa di morte;

$i d_N$ = numero di decessi codificati con il nuovo sistema di codifica nel gruppo i-esimo;

d_M = numero di decessi codificati manualmente nel gruppo i-esimo.

Il coefficiente è stato calcolato per i 15 gruppi di cause di morte, presenti nella IX revisione (esclusi i due gruppi: Complicazioni della gravidanza, del parto e del puerperio ed Alcune condizioni morbose di origine perinatale) e per i 51 sottogruppi proposti dall'Eurostat¹⁵ ed utilizzati anche dall'Istat nella pubblicazione 'La mortalità in Italia nel periodo 1970-1992: evoluzione e geografia'¹⁶. Per dare un'indicazione sulla stabilità dei coefficienti di raccordo viene fornita nelle tabelle 4.3 e 4.4 anche la numerosità dei decessi ottenuta con i due sistemi di codifica.

Sebbene la concordanza sia complessivamente elevata si riscontrano differenze, talvolta consistenti, nell'ambito dei diversi gruppi considerati. Per il complesso dei tumori e delle malattie del sistema circolatorio, che insieme costituiscono circa il 70% della mortalità totale, il coefficiente di raccordo è prossimo all'unità, indicando una sostanziale omogeneità tra i due sistemi di codifica a livello di grandi gruppi.

Per i tumori, tuttavia, esistono alcune differenze non trascurabili nell'analisi della sede di insorgenza. La maggiore discordanza si riscontra per i tumori maligni delle gengive e per i tumori di comportamento incerto e di natura non specificata.

Per quanto riguarda le altre cause di morte si osserva una maggiore discordanza nei seguenti gruppi: malattie infettive, malattie del sangue, disturbi psichici, sintomi segni e stati morbosi mal definiti. Queste cause costituiscono nel loro insieme poco più del 3% dei casi analizzati. Sembra opportuno sottolineare inoltre la significativa riduzione, con l'introduzione della codifica automatica, del numero dei decessi dovuti a sintomi segni e stati morbosi mal definiti (cod. ICD IX: 780-799) e il corrispondente aumento dei casi classificati nel gruppo tumori di comportamento incerto e di natura non specificata (cod. ICD IX: 235-239).

Nel caso delle malattie infettive è infine importante evidenziare come il passaggio da codifica manuale ad automatica abbia determinato uno spostamento di una quota della mortalità per cirrosi epatica (cod. ICD IX: 571) verso l'epatite virale (cod. ICD IX: 070): calcolando il coefficiente di raccordo per questo gruppo di cause, dopo avere isolato il codice in questione (cod. ICD IX: 070 Epatite virale) si può notare un avvicinamento dei due metodi di codifica comprovato dall'abbassamento del coefficiente K.

¹⁵ Eurostat, November 1998, Causes of death, "European shortlist"- final list_August 1998.

¹⁶ Istat-Iss, s.d.

Tabella 4.3 – Coefficienti di raccordo - Decessi per grandi gruppi di cause secondo il diverso sistema di codifica e coefficienti di raccordo K.

Cause di morte	Codici ICD IX	Decessi secondo codifica		K ^(b)
		Manuale	Automatica ^(a)	
Malattie infettive e parassitarie ^(c)	001-139	1.230	1.644	1,337
Tumori	140-239	90.554	88.850	0,981
Malattie delle ghiandole endocrine, del metabolismo e disturbi immunitari ^(d)	240-279	11.233	11.581	1,031
AIDS	279.1	2.724	2.724	1,000
Malattie del sangue e degli organi ematopoietici	280-289	1.126	1.296	1,151
Disturbi psichici	290-319	3.862	4.368	1,131
Malattie del sistema nervoso e degli organi dei sensi	320-389	6.280	6.770	1,078
Malattie del sistema circolatorio	390-459	143.481	143.640	1,001
<i>di cui Malattie ischemiche del cuore</i>	<i>410-414</i>	<i>44.874</i>	<i>44.474</i>	0,991
<i>Malattie cerebrovascolari</i>	<i>430-438</i>	<i>42.015</i>	<i>41.250</i>	0,982
Malattie dell'apparato respiratorio	460-519	19.794	20.722	1,047
Malattie dell'apparato digerente	520-579	16.676	15.698	0,941
Malattie dell'apparato genitourinario	580-629	4.045	4.190	1,036
Malattie della pelle e del tessuto sottocutaneo	680-709	369	394	1,068
Malattie del sistema osteomuscolare e del tessuto connettivo	710-739	1.030	1.059	1,028
Malformazioni congenite	740-759	483	506	1,048
Sintomi segni e stati morbosi mal definiti	780-799	4.924	4.369	0,887
Traumatismi ed avvelenamenti	800-999	16.216	16.216	1,000
Totale cause	001-999	324.027	324.027	1,000

(a) In questo caso si utilizza il termine ‘codifica automatica’ sebbene il 23% dei casi sia codificato manualmente (Aids, Cause violente e records scartati dalla procedure MICAR-ACME).

(b) Il coefficiente K è stato calcolato nel seguente modo: $K_i = \frac{i d_N}{i d_M}$ dove i = causa di decesso, $i d_N$ = numero di decessi codificati con il nuovo sistema di codifica nel gruppo i-esimo, $i d_M$ = numero di decessi codificati manualmente nel gruppo i-esimo.

(c) Escludendo nella codifica automatica il codice ICD IX 070 (Epatite virale), il coefficiente di raccordo K risulta pari a 1,138.

(d) Non contiene il codice ICD IX 279.1.

Tabella 4.4 - Decessi per sottogruppi di cause secondo il diverso sistema di codifica e coefficienti di raccordo K.

Cause di morte	Codici ICD IX	Decessi secondo codifica		K ^(b)
		Manuale	Automatica ^(a)	
Tubercolosi	010-018,137	432	459	1,063
Infezione meningococcica	36	8	8	1,000
Epatite virale	70	62	375	6,048
Tumori maligni	140-208	88.250	85.411	0,968
Tumori maligni delle labbra, della cavità orale e della faringe	140-149	1.768	1.724	0,975
Tumori maligni delle gengive	143	14	24	1,714
Tumori maligni del pavimento della bocca	144	70	70	1,000
Tumori maligni di altre e non specificate parti della bocca	145	361	358	0,992
Tumori maligni dell'esofago	150	1.257	1.222	0,972
Tumori maligni dello stomaco	151	7.531	7.248	0,962
Tumori maligni del colon	153	6.400	6.236	0,974
Tumori maligni del retto, della giunzione rettosigmoidea e dell'ano	154	2.765	2.732	0,988
Tumori maligni del fegato e dei dotti biliari intraepatici	155	5.911	5.651	0,956
Tumori maligni del pancreas	157	4.020	3.935	0,979
Tumori maligni della laringe, della trachea, dei bronchi e dei polmoni	161-162	19.348	19.037	0,984
Melanoma maligno della pelle	172	802	780	0,973
Tumori maligni della mammella	174-175	6.605	6.534	0,989
Tumori maligni del collo dell'utero	180	274	260	0,949
Tumori maligni in altre parti dell'utero	179,182	1.533	1.423	0,928
Tumori maligni dell'ovaio e degli altri annessi uterini	183	1.672	1.539	0,920
Tumori maligni della prostata	185	4.138	3.810	0,921
Tumori maligni del testicolo	186	52	50	0,962
Tumori maligni della vescica	188	3.272	2.940	0,899
Tumori maligni del rene e di altri e non specificati organi urinari	189	1.849	1.734	0,938
Tumori maligni dell'encefalo	191	1.770	1.565	0,884
Tumori maligni dei tessuti linfatico ed ematopoietico	200-208	7.128	6.848	0,961
Malattia di Hodgkin	201	286	283	0,990
Leucemie	204-208	3.065	2.945	0,961
Tumori di comportamento incerto e di natura non specificata	235-239	1.909	3.050	1,598
Diabete mellito	250	10.251	10.118	0,987
Abuso di alcool (incluse le psicosi alcoliche)	291,303	196	178	0,908
Farmacodipendenza, abuso di droghe senza dipendenza	304-305	515	543	1,054
Meningiti	320-322	88	70	0,795
Malattie ischemiche del cuore	410-414	44.874	44.474	0,991
Altre malattie del cuore	420-423, 425-429	32.248	33.796	1,048
Disturbi circolatori dell'encefalo	430-438	42.015	41.250	0,982
Polmonite	480-486	3.712	4.219	1,137
Influenza	487	392	541	1,380
Malattie polmonari croniche ostruttive	490-494,496	2.102	2.869	1,365
Asma	493	715	993	1,389
Ulcera gastrica, duodenale e gastroduodenale	531-534	1.504	1.263	0,840
Ernie addominali	550-553	334	407	1,219
Cirrosi ed altre malattie croniche del fegato	571	8.521	7.821	0,918
Litiasi biliare	574	378	455	1,204
Malattie del pancreas	577	566	577	1,019
Malattie del rene e dell'uretere	580-594	3.638	3.762	1,034
Insufficienza renale cronica ed insufficienza renale non specificata	585-586	2.865	2.150	0,750
Iperplasia alla prostata	600	190	193	1,016
Artrite reumatoide ed osteoartrosi	714-715	380	437	1,150
Malformazioni congenite del sistema nervoso	740-742	123	85	0,691
<u>Malformazioni congenite del sistema circolatorio</u>	<u>745-747</u>	<u>194</u>	<u>212</u>	<u>1,093</u>

(a) Cfr. nota (a) Tabella 4.3.

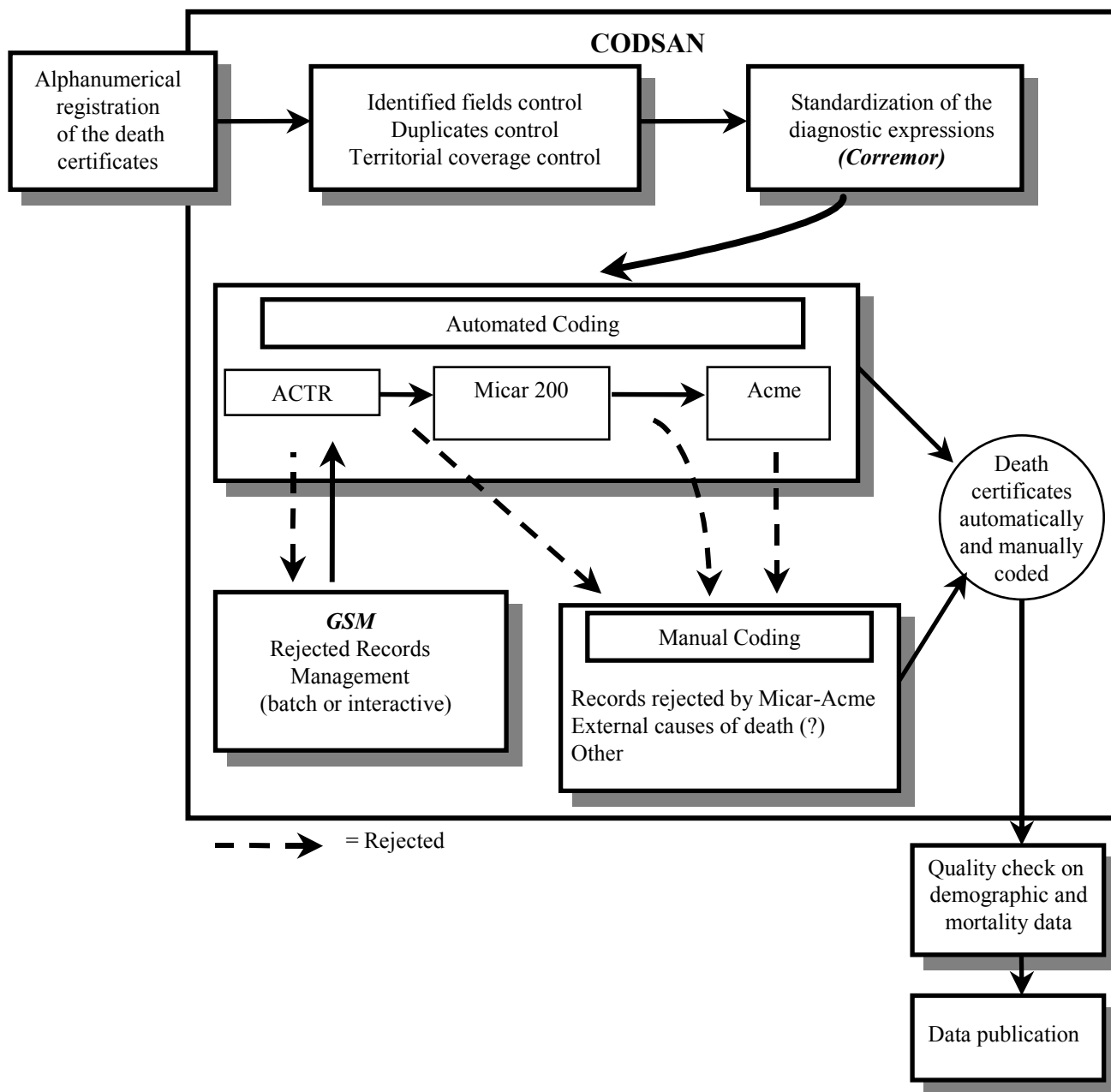
(b) Cfr. nota (b) Tabella 4.3.

5 ADEGUAMENTO DEL SOFTWARE PER LA TRANSIZIONE VERSO LA DECIMA REVISIONE DELLA ICD

Attualmente è in fase di realizzazione il passaggio alla ICD X Revisione (OMS 1992), un passaggio tra i più importanti e delicati e già compiuto in numerosi Paesi del mondo. L'aver adottato procedure automatiche per la codifica della causa iniziale di morte agevolerà considerevolmente il passaggio da una revisione alla successiva.

Per il passaggio alla nuova versione sarà adottato ancora il software dell'NCHS (ovviamente nella sua versione per la nuova revisione, la decima) sostituendo alcune delle sue componenti, in particolare quelle maggiormente condizionate dall'uso di differenti linguaggi (MICAR 100), con un software di codifica che meglio permetta di sfruttare le proprietà della lingua italiana. Per quest'ultimo obiettivo si è individuato il prodotto ACTR v3 (*AUTOMATED CODING BY TEXT RECOGNITION*), un software di codifica generalizzato (cfr. figura 3).

Figura 4 - Le fasi previste per la codifica, ICD X.



5.1 LA TRADUZIONE DEL DIZIONARIO DEGLI ERNS

5.1.1 Problematiche linguistiche

L'NCHS suggerisce ai Paesi di lingua non inglese che vogliono adottare il software MICAR-ACME di creare un collegamento tra le singole parole della lingua del Paese stesso e quelle della lingua inglese. In pratica si tratterebbe di gestire la traduzione attraverso la definizione di sinonimi per le singole parole utilizzabili dai medici per la certificazione delle cause di morte (Hart J. *Technical Aspects of Language Conversion: Details of the NCHS MICAR System*, in CDC, 1999). Le

caratteristiche della lingua italiana e della terminologia medica usata hanno fatto ritenere non percorribile questa strada e si è preferito effettuare la più onerosa traduzione delle intere entità morbose.

La necessità di eseguire la traduzione del dizionario in lingua inglese utilizzato per la codifica automatica dagli USA trova fondamento in alcuni aspetti non trascurabili: innanzitutto la struttura sintattica delle lingue è diversa, in secondo luogo l'uso di terminologia specifica, espressioni gergali ed abbreviazioni variano da lingua a lingua (es. l'espressione 'di' in inglese è l'abbreviazione di 'diabetes', mentre in italiano esiste una doppia corrispondenza, sia come abbreviazione del termine 'diabete' che come preposizione); ed infine, l'accezione di alcuni termini varia da lingua a lingua (es. l'espressione 'tumore' ha generalmente significato implicito ai fini della codifica di 'tumore benigno' in inglese, mentre lo stesso non vale per l'italiano). Queste considerazioni di carattere generale hanno spinto l'Istat, già con l'adozione del software per l'ICD IX, a realizzare un proprio dizionario che rispondesse all'uso effettivo delle espressioni diagnostiche riportate sui certificati di morte.

Tuttavia, come visto nel paragrafo 2.6, il dizionario MICAR nella sua versione italiana per la codifica delle cause di morte in IX revisione dipende dalla revisione adottata. L'attuale versione italiana di tale dizionario non è quindi direttamente utilizzabile per la codifica automatica in X revisione. Inoltre, molte categorie e blocchi in ICD X sono stati ampliati od assegnati a differenti capitoli rispetto all'ICD IX. Di conseguenza si è reso necessario un grande lavoro di revisione che ha portato alla riassegnazione di espressioni diagnostiche ad ERNS diversi.

E' stato quindi progettato e sviluppato un software per la gestione e la traduzione della terminologia medica e la corretta assegnazione del corrispondente ERN. L'obiettivo era costruire un dizionario italiano indipendente dalla classificazione che facilitasse la manutenzione e la gestione dei futuri sviluppi della classificazione in uso.

Per diminuire il numero di ALIAS da tradurre, si è molto lavorato sulla standardizzazione del dizionario, realizzando una migliore gestione dei sinonimi (rispetto al lavoro svolto in precedenza per l'uso in ICD9) che ha consentito una riduzione delle dimensioni del dizionario di circa 15.000 records. Gli alias da tradurre sono quindi circa 170.000.

5.1.2 Il database per la traduzione della terminologia medica¹⁷

Per la traduzione delle espressioni diagnostiche utilizzate dai medici per la certificazione delle cause di morte è stato realizzato un database che permette di confrontare in linea 6 diversi dizionari:

¹⁷ Questo paragrafo è indirizzato ad un lettore interessato a conoscere in dettaglio le problematiche metodologiche affrontate e le soluzioni tecniche adottate nella fase di traduzione del dizionario.

1. ERN_{10,ING}: è il dizionario degli ERN in lingua inglese utilizzato in X revisione; per ogni ERN contiene un solo ALIAS¹⁸, generalmente quello con la dicitura più corretta e completa dal punto di vista medico;
2. BBD_{9,ITA}: è il dizionario usato in Italia per la codifica automatica in IX revisione. È stato ottenuto traducendo il corrispondente dizionario statunitense. Nella fase di traduzione si è rispettata la corrispondenza ‘patologia-codice ICD IX’ piuttosto che la più dettagliata corrispondenza ‘patologia-codice ERN’ e questo rende il dizionario dipendente dalla revisione;
3. BBD_{9,ING}: è il dizionario inglese per la codifica automatica con la IX revisione;
4. BBD_{10,ING}: è il dizionario inglese per la codifica automatica con la X revisione;
5. ICD₉: è la IX revisione della Classificazione Internazionale delle Malattie, in italiano;
6. ICD₁₀: è la X revisione della Classificazione Internazionale delle Malattie, in italiano.

I vari dizionari sono collegati tra loro mediante le chiavi ‘ERN’, ‘collegamento ERN -ICD IX’ e ‘collegamento ERN -ICD X’¹⁹.

Il database è stato realizzato con l’obiettivo di costruire il BBD_{10,ITA} (il dizionario italiano da usare per la codifica automatica con la X Revisione) attraverso una traduzione assistita dei singoli ERNS contenuti in ERN_{10,ING} sfruttando le informazioni contenute in BBD_{9,ITA} e mantenendo traccia delle decisioni prese.

Ognuno degli altri dizionari in linea permette di affrontare le problematiche che via via si presentano in fase di traduzione. Ad esempio, la variabilità linguistica presente nel dizionario BBD_{10,ING} può essere di aiuto nei casi di interpretazione dubbia della patologia presente in ERN_{10,ING}; la terminologia presente in BBD_{9,ING} può aiutare a comprendere perché nel BBD_{9,ITA} alcune patologie hanno determinati ERN s; i collegamenti ai codici delle due Classificazioni forniscono indicazioni sulle conseguenze che una data traduzione può avere sulla codifica finale.

La traduzione assistita avviene utilizzando una serie di maschere per la consultazione *on-line* dei dizionari presenti nel database. Il prodotto realizzato consente inoltre di individuare, tramite appositi *flags*, il traduttore che ha eseguito l’immissione, l’avvenuta revisione da parte del supervisore, nonché i criteri con i quali la stringa è stata tradotta. I criteri adottati per eseguire la traduzione/revisione del nuovo dizionario italiano per l’ICD X sono rigorosi, concordati e condivisi tra tutti gli operatori coinvolti nel processo. In particolare è stato sviluppato uno schema decisionale a cui i traduttori sono tenuti ad attenersi con scrupolo per giungere all’assegnazione degli ERN s. Sono possibili le seguenti modalità per ciascuna stringa tradotta:

- traduzione ‘certa’;

¹⁸ Ricordiamo che per ALIAS si intendono le diverse accezioni linguistiche di una stessa espressione diagnostica.

¹⁹ I collegamenti sono stati estratti dai files CVTERN usati da Micar 200 nelle due revisioni.

- traduzione con ‘dubbio’;
- traduzione effettuata mantenendo la ‘compatibilità’ solo a livello di codice ICD X;
- traduzione non effettuata.

Le traduzioni eseguite vengono revisionate in modo sistematico, con possibilità di confermare o modificare sia il livello di accettazione della traduzione che la traduzione stessa (assegnazione di un nuovo ERN). Nel caso di mancata traduzione, il traduttore è tenuto a tenere traccia del processo logico con il quale è giunto alla decisione di non tradurre e, quando possibile, a fornire dei suggerimenti di possibili ERNS da attribuire alla stringa in esame.

5.1.3 Traduzione automatica provvisoria di alcuni ERN s

La migrazione completa del dizionario usato per l’ICD IX nel software per la codifica in ICD X non è stata possibile poiché la precedente traduzione era dipendente dalla Revisione (cfr. paragrafo 2.6). Tuttavia, laddove nel passaggio da una Revisione alla successiva un codice ICD IX è confluito in un solo codice ICD X, è stato possibile tradurre automaticamente i corrispondenti ERNS con una compatibilità a livello di classificazione analitica ICD X. Successivamente, in vista dei futuri periodici aggiornamenti della classificazione, si dovrà andare a migliorare il dettaglio della compatibilità fino ad arrivare all’indipendenza del dizionario dalla revisione.

Per individuare gli ERNS traducibili automaticamente sono state utilizzate le informazioni provenienti da:

- il file rilasciato dall’OMS (OMS, 1997) nel quale sono riportate le Tavole di Corrispondenza tra le due Revisioni. Da tali tavole è possibile ricavare le associazioni elementari e/o complesse (‘1 a 1’, ‘1 a molti’, ‘molti a 1’, ‘molti a molti’) tra i codici delle due classificazioni analitiche;
- i files dei collegamenti dell’NCHS²⁰. Da questi files sono state ricavate analoghe tavole di Corrispondenza tra le due revisioni ma a livello di singolo ERN.

Si ricorda che, indipendentemente dalla revisione adottata, ad un ERN possono corrispondere uno o più codici della ICD (dipende da quanti codici sono necessari per descrivere la patologia collegata all’ ERN). Per individuare gli ERNS traducibili in automatico sono stati considerati solo gli quelli che danno luogo a singoli codici sia in ICD IX che in ICD X. Pertanto, su un totale complessivo di 106.000 ERNS, 72.581 ERNS sono oggetto di questa analisi (a cui corrispondono 149.472 ALIAS). Quindi, dal file dei collegamenti dell’NCHS è stato possibile ricavare le distribuzioni degli ERNS e degli ALIAS per ‘tipo di relazione’ e per ‘tipo di traduzione’ (Tabelle 5.1 e 5.2).

²⁰ <http://www.cdc.gov/nchs/about/major/dvs/medsof.htm>

Tabella 5.1 – Distribuzione degli ERN e degli ALIAS per Tipo di Relazione ICD IX→ICD X

Tipo di Relazione	Numero di ERNs	Numero di ALIAS
1 a 1	4967	9015
Molti a 1	3319	6243
1 a molti	31588	66561
Molti a molti	32707	67653
Totale	72581	149472

Tabella 5.2 – Distribuzione degli ERN e degli ALIAS per Tipo di Traduzione

Tipo di Traduzione	Numero di ERNs	Numero di ALIAS
Automatica	8286	15258
Manuale	64295	134214
Totale	72581	149472

Dai collegamenti indicati dall'NCHS emergono quindi 15.258 ALIAS (ovvero 8.286 ERNS) potenzialmente traducibile in automatico. Tuttavia, prima di accettare queste traduzioni, è stata analizzata la loro coerenza con le indicazioni fornite dall'OMS (OMS, 1997). Nelle Tabelle 5.3 e 5.4 sono pertanto riportate le distribuzioni doppie (NCHS/OMS) rispettivamente per il 'tipo relazione' e per 'tipo di traduzione'.

Tabella 5.3 – Distribuzioni doppie degli ALIAS per Tipo di Relazione ICD IX→ICD X

NCHS	OMS				Relazione non indicata	Totale
	1 a 1	Molti a 1	1 a molti	Molti a molti		
1 a 1	6646	496	1486	202	185	9015
Molti a 1	2604	2087	435	758	359	6243
1 a molti	44181	1839	17387	2808	346	66561
Molti a molti	23446	8293	7186	22026	6702	67653
Totale	76877	12715	26494	25794	7592	149472

Tabella 5.4 – Distribuzioni doppie degli ALIAS per Tipo di Traduzione

NCHS	OMS			Totale
	Automatica	Manuale	Relazione non indicata	
Automatica	11833	2881	544	15258
Manuale	77759	49407	7048	134214
Totale	89592	52288	7592	149472

Dei 15.258 ALIAS che secondo i collegamenti indicati dall'NCHS possono essere tradotti in automatico, 11.833 sono coerenti con le indicazioni date dall'OMS.

L'analisi qualitativa degli ALIAS relativi ai 3.425 (2.881+544) casi in cui le indicazioni dell'NCHS sono diverse da quelle dell'OMS, ha permesso di evidenziare che le incoerenze sono in parte giustificate dal fatto che i collegamenti NCHS sono relativi a codici provvisori (modificabili da

MICAR200 sulla base delle altre informazioni presenti sul certificato) ed in parte dovute ai mancati aggiornamenti sulle relazioni indicate sia dall’OMS che dall’NCHS.

Tuttavia la scelta operativa è stata quella di accettare tutte le 15.528 traduzioni automatiche individuate tramite i collegamenti forniti dall’NCHS, con l’accortezza, di mantenere traccia sul record tradotto delle eventuali incorenze rispetto alle indicazioni OMS. Si puntualizza che questa prima traduzione automatica ha avuto lo scopo di accelerare le simulazioni con ACTR e di facilitare il lavoro di traduzione.

5.2 LA CODIFICA DEL TESTO CON ACTR

Il sistema di codifica ACTR presenta la caratteristica di essere assolutamente generalizzato, indipendente cioè dalla classificazione considerata e dalla lingua. La logica di base del sistema è ispirata alla metodologia sviluppata originariamente presso US Census Bureau (Hellerman, 1982) ed utilizza degli algoritmi di ricerca e matching testuale messi a punto successivamente da ricercatori di Statistics Canada (Wenzowski, 1988).

ACTR è disponibile in versioni adatte per varie piattaforme (Windows 95 & Windows NT; Unix) ed è stato utilizzato con successo per applicazioni nell’ambito della statistica ufficiale (Tourigny e Moloney, 1995).

In Istat è stato utilizzato con risultati soddisfacenti su diverse variabili e in numerose indagini:

Variabili	Indagini
Attività Economica	Cens. Pop. 1991 – Ind. Qualità Cens. Pop. 1991 Indagine pilota Forze di Lavoro 1999 I e II indagine pilota Cens. Pop. 2001 Cens. Intermedio dell’Industria 1998
Professione	Cens. Pop. 1991 – Ind. Qualità Cens. Pop. 1991 Indagine sulla Salute 1994 4 trimestri indagine Forze di Lavoro 1998 (test) Indagine pilota Forze di Lavoro 1999 I e II indagine pilota Cens. Pop. 2001
Titolo di studio	Cens. Pop. 1991 – Ind. Qualità Cens. Pop. 1991 I e II indagine pilota Cens. Pop. 2001
Cittadinanza	II indagine pilota Cens. Pop. 2001
Comune	II indagine pilota Cens. Pop. 2001
Natura giuridica imprese	Cens. Intermedio dell’Industria 1998
Cause di morte	Cause di morte 1999 (test)

ACTR è inoltre in corso di utilizzo nell'ambito di entrambi i censimenti 2001 ('Popolazione e Abitazioni' e 'Industria e Servizi').

Il dizionario per la codifica è generalmente una versione ampliata del manuale ufficiale della classificazione e deve essere costruito in modo da avvicinare il linguaggio di questi manuali a quello utilizzato dai rispondenti e poter essere elaborato da un software. Ovviamente, dovrebbe contenere esclusivamente descrizioni non ambigue e la sua qualità e la sua ricchezza influiranno direttamente sui risultati della codifica.

Una delle peculiarità di ACTR rispetto ad altri sistemi è la notevole flessibilità e possibilità di personalizzazione del processo di *parsing* che precede il confronto tra le stringhe. Il *parsing* consiste in una standardizzazione dei dati testuali avente lo scopo di rimuovere la variabilità grammaticale o sintattica che differenzia frasi con lo stesso contenuto semantico e che è irrilevante ai fini della codifica. Tale attività di standardizzazione è completamente controllata dall'utente che ha il compito di adattarla al particolare contesto applicativo (lingua, classificazione, tipologia di rispondente) avendo a disposizione 14 diverse funzioni (mappatura dei caratteri, rimozione di spazi multipli, eliminazione di parole ininfluenti, definizione di sinonimi per parole singole o per gruppi di parole, rimozione di suffissi, rimozione di prefissi, etc...).

L'accoppiamento tra la stringa in input ed il dizionario può avvenire sulla base di una perfetta corrispondenza tra le stringhe (match 'esatto') o sulla base di funzioni di similarità tra i testi (match 'parziale'), quantificate tramite appositi algoritmi per l'attribuzione di punteggi ai testi del dizionario che contengono almeno una parola in comune con il testo da codificare. In questo secondo caso, l'utente ha la possibilità di definire diversi parametri, relativi ai citati punteggi, quali la soglia minima e la soglia massima di accettazione e la distanza necessaria per discriminare tra le eventuali diverse risposte ricadenti nella regione di accettazione definita dalle soglie. Nella definizione dei parametri si deve raggiungere un compromesso tra l'accuratezza e l'efficacia del processo di codifica.

A seconda delle finalità il processo di codifica può essere:

- Automatico: viene analizzato in batch un file contenente l'insieme di risposte testuali e la finalità è di individuare ed estrarre dal dizionario una singola descrizione che realizzi il match con quella da codificare;
- Assistito: il software costituisce un supporto interattivo per il codificatore, facilitandone la navigazione nel dizionario della classificazione di riferimento. Viene individuato nel dizionario un set di descrizioni, simili alla descrizione da codificare, lasciando al codificatore l'individuazione di quella corretta (ad oggi non è disponibile la funzione di selezione ed attribuzione del codice

selezionato).

5.3 ACTR APPLICATO ALLA TERMINOLOGIA MEDICA

5.3.1 Obiettivi

Inizialmente si è riflettuto se utilizzare ACTR per sostituire completamente la procedura MICAR, per ottenere cioè, a partire dalle descrizioni alfanumeriche delle patologie, il file con i codici ICD da dare in input ad ACME. Sembrava infatti naturale utilizzare l'indice analitico della nuova classificazione per assegnare a ciascuna patologia il corrispondente codice ICD. Tuttavia le regole di codifica non prevedono sempre una corrispondenza univoca tra patologia e codice: l'attribuzione del codice ad una patologia può dipendere infatti da una serie di modificatori quali l'età, il sesso, la durata della malattia, la presenza o meno di altre patologie, la posizione sul certificato. Si è visto che la gestione di questo insieme di regole era particolarmente oneroso e che si doveva in pratica implementare una sorta di MICAR 200.

Pertanto si è scelto di abbandonare tale approccio e di utilizzare ACTR per sostituire solo il primo modulo della procedura MICAR (MICAR 100) che effettua la traduzione dall'alfanumerico al sistema degli ERNS. I vantaggi dell'adozione di ACTR rispetto a MICAR 100 sono la sua maggiore flessibilità e la facilità di apportare aggiornamenti.

Ovviamente, la decisione di abbandonare MICAR 100 e passare ad ACTR è stata presa non soltanto in funzione della più facile gestione del sistema, ma anche del fatto che, in seguito ad una sperimentazione effettuata (cfr. paragrafo 5.3.4), si ritiene di poter raggiungere risultati, in termini di efficacia ed accuratezza, superiori a quelli ottenuti con il sistema preesistente.

5.3.2 L'implementazione di ACTR: due diversi approcci per la sperimentazione

Per valutare i vantaggi e le problematiche derivanti dall'uso di ACTR è stata effettuata una sperimentazione. Due sono stati gli approcci alternativi individuati:

- i) circoscrivere la sperimentazione ad un sottoinsieme del dizionario (ad esempio caricare un dizionario corrispondente soltanto alle patologie più frequenti, oppure ad una sola lettera dell'alfabeto) sviluppando un ambiente applicativo capace di sfruttare al massimo le funzioni di normalizzazione dei testi fornite da ACTR;
- ii) considerare direttamente tutto il dizionario utilizzato da MICAR 100 (circa 187.000 termini) nella versione italiana della ICD IX e applicare una strategia di parsing necessariamente ridotta.

La prima soluzione, pur consentendo di mettere in luce tutte le soluzioni linguistiche realizzabili e di formare il personale addetto all'utilizzo del sistema, non avrebbe consentito l'uso di un campione

rappresentativo per la validazione della sperimentazione: sui certificati di morte si presenta la totalità della patologie e non soltanto quelle previste nel sottoinsieme del dizionario. Si è deciso quindi di seguire la seconda strada.

Nel dizionario utilizzato da MICAR 100 nella versione italiana della ICD IX (il BBD_{9,ITA}) ciascun termine medico è riportato in tutte le sue possibili varianti linguistiche: i sinonimi, le abbreviazioni ed i vari modi di esprimere lo stesso concetto (con e senza articoli, preposizioni e congiunzioni).

Affinchè il BBD_{9,ITA}, costruito originariamente come una serie di stringhe continue nelle quali le parole non erano separate da spazi, fosse utilizzabile da ACTR è stato necessario effettuare un passo preliminare per la suddivisione delle parole. Tale operazione è stata effettuata con l'ausilio di procedure batch che, seppur riducendo i tempi rispetto ad un lavoro record per record, hanno comunque richiesto un forte impegno nella preparazione dei files con le indicazioni delle modifiche. La sperimentazione è stata effettuata secondo una strategia di *parsing* inizialmente estremamente ridotta, limitata all'eliminazione di articoli, preposizioni e congiunzioni e ad un limitato numero di suffissi.

5.3.3 Il criterio adottato per l'accoppiamento

Un aspetto importante ha riguardato la definizione dei parametri per l'accoppiamento. In ACTR le misure di similarità tra stringhe vengono costruite a partire dal grado di 'informatività' delle singole parole il quale è a sua volta inversamente proporzionale alla frequenza con cui queste si presentano nel dizionario. Questo assunto, in base al quale una parola ha meno potere 'discriminatorio' tanto più è di uso comune, non è in realtà adatto nel nostro contesto: basti pensare, ad esempio, ai termini che descrivono le patologie e le sedi anatomiche che hanno una frequenza nettamente superiore a quella di termini modificatori e sono allo stesso tempo molto più informative. Per questo motivo si è scelto di accettare solamente gli abbinamenti 'esatti' (direct match) nell'assegnazione dei codici alle patologie.

5.3.4 Passi operativi e risultati principali

Il campione

Come detto precedentemente, per garantire una confrontabilità tra le performance di MICAR 100 ed ACTR, la sperimentazione è stata condotta su un insieme di schede reali. Per garantire la variabilità della casistica sono stati considerati tutti i decessi (circa 40.000) di uno specifico mese²¹ per un totale di circa 150.000 espressioni diagnostiche.

²¹ Settembre, 1998.

Il file di input

ACTR, a differenza di MICAR 100 che è un software appositamente costruito per la codifica delle schede di morte, è progettato per codificare una sola stringa per record in input. Per questo motivo è stato necessario trasformare ciascun record del file di input di MICAR 100 in più records (uno per ogni patologia riportata) e compattare successivamente l'output di ACTR per ottenere l'input di MICAR 200. In questi passaggi, al fine di ricostruire correttamente il record, è stato necessario tenere memoria per ciascuna patologia di tutta una serie di informazioni: identificativo del record di partenza, casella del certificato, posizione nella casella, durata, eventuale legame di causa-effetto con la patologia successiva.

Risultati

Le diverse applicazioni sono state condotte parallelamente a progressivi miglioramenti sia della strategia di parsing che nella organizzazione del dizionario. I principali risultati sono riportati nella tabella 5.5:

Tabella 5.5 – Risultati delle applicazioni di ACTR.

Elaborazione	Occorrenze del dizionario	Record rifiutati in fase di caricamento	Efficacia del sistema ^(a)	Percentuale di schede codificate ^(b)
I	187.000	Non disponibile	35%	4%
II	187.000	14.000	48%	11%
III	187.000	14.000	52%	14%
IV	187.000	15.000	69%	33%
V	187.000	17.000	81%	53%
VI	187.000	18.000	89%	71%
VII ²²	187.000	34.000	92%	78%

(a) Per efficacia del sistema si intende la percentuale di testi in cui il sistema è riuscito ad associare univocamente un codice.

(b) Una scheda è considerata codificata quando tutte le patologie in essa presenti sono state codificate

In fase di caricamento del dizionario il sistema rifiuta automaticamente i records incoerenti o ridondanti:

- per incoerenti si intendono due testi uguali (o per errore nel dizionario base o divenuti tali a seguito delle procedure di parsing) associati a codici diversi;
- per ridondanti si intendono due testi uguali (divenuti tali a seguito delle procedure di parsing) associati allo stesso codice.

In generale la maggior parte dei records rifiutati in fase di caricamento sono ridondanti ed anche una parte delle incoerenze può essere equiparate alle ridondanze in quanto il livello di dettaglio dei codici ERN è eccessivo rispetto al codice finale che si vuole ottenere. In particolare, dei 34.000

²² Questa applicazione è stata effettuata dopo la standardizzazione del dizionario menzionata nel paragrafo 5.1. Gran parte dell'aumento dei record rifiutati in fase di caricamento è quindi dovuto all'aumento delle ridondanze sul dizionario in input.

records rifiutati nell'ultima applicazione, 25.000 sono risultati ridondanti e 9.000 incoerenti a livello di ERN. Analizzando a livello di codici ICD i 9.000 record incoerenti si è visto che 8.300 sono in realtà ancora ridondanze e solo 900 sono incoerenze. L'analisi qualitativa di queste incoerenze ha messo in luce che esse potranno essere risolte con un lavoro più accurato sul parsing e sul dizionario.

Per quanto riguarda le percentuali di codifica, si vede che i risultati si avvicinano progressivamente a quelli raggiunti con MICAR 100, grazie ai miglioramenti sul dizionario e sulla strategia di *parsing* effettuati ad ogni successiva elaborazione. Continuando a lavorare sull'organizzazione del dizionario e sulla strategia di *parsing* si potrà migliorare ulteriormente l'efficacia del sistema. In particolare, si ritiene che siano raggiungibili almeno le attuali *performance* di MICAR (efficacia 95% e percentuale di schede codificate 80%).

Oltre agli aspetti di flessibilità e di efficacia altre caratteristiche di ACTR ci inducono a giudicare positivamente la sua adozione:

- la possibilità di analizzare interattivamente gli accoppiamenti 'parziali' potrà essere utilizzata per aiutare la codifica manuale delle patologie non accoppiate in modo 'esatto';
- la forte flessibilità nelle procedure di manipolazione delle stringhe potrà essere utilizzata per la codifica delle modalità degli accidenti. Si potrà provare a costruire un secondo dizionario e la relativa strategia di *parsing* per riuscire a ridurre la forte variabilità linguistica nella descrizione degli accidenti.

6 FUTURI SVILUPPI

Le trasformazioni avvenute nel processo di lavorazione del dato di mortalità grazie all'uso di procedure automatiche per la codifica della causa di morte sono state indubbiamente di notevole entità. Come visto, molteplici fasi sono state modificate alla luce delle novità introdotte.

Nel corso degli ultimi anni questo Settore è stato inoltre fortemente coinvolto anche in diverse attività offerte dal panorama internazionale. A questo proposito si fa riferimento alle attività svolte nell'ambito dell'ICE group (International Collaborative Effort on Automating Mortality Statistics), alla partecipazione ai lavori del European Technical Group on Causes of Death e al coordinamento del progetto internazionale incluso nel Community Statistical Programme 2003-2007 'EU training package on certification of causes of death (OJ of EC n° 2002/S 170-136139)'.

Tuttavia la strada da percorrere è ancora lunga e diversi sono gli obiettivi che devono essere raggiunti nei prossimi anni. L'intensa attività di ricerca nell'ambito delle statistiche di mortalità per causa è volta al miglioramento della qualità del dato, alla valorizzazione delle fonti informative esistenti e all'adozione di nuovi strumenti di classificazione internazionale.

La natura Pubblica dell'Istat, ovvero operante al servizio della collettività, deve tenere conto delle esigenze informative che possono avere i diversi utenti finali. E' necessario dunque che il dato 'ufficiale' sia di qualità, sia in termini di tempestività del rilascio del dato che di trasparenza delle procedure di raccolta, controllo ed eventuale correzione sia del dato sanitario che demografico.

6.1 TEMPESTIVITÀ

Oltre all'adozione di procedure automatizzate per la codifica delle cause di morte che, come visto, ha ampiamente ridotto i tempi di rilascio del dato finale, si è sviluppato un sistema di anticipazione dei dati di mortalità per causa a livello regionale sulla base di dati parziali integrati da altre fonti informative. Tali stime, effettuata con cadenza annuale, si riferiscono all'anno precedente e sono disponibili sul sito www.istat.it.²³

E' stata inoltre realizzata una banca dati in Oracle per le interrogazioni al nostro data base di mortalità a partire dall'anno 1980. Tale banca dati è attualmente disponibile per uso interno e sono in fase di valutazione le modalità per l'interrogazione tramite internet.

²³ La metodologia utilizzata è disponibile sul sito sopra menzionato.

6.2 QUALITÀ

6.2.1 Completezza della raccolta dei dati

L'insieme delle strutture preposte al controllo²⁴ del flusso informativo (cfr. Appendice A) garantiscono in genere una buona copertura sul territorio del dato di mortalità per causa. Tuttavia, si può verificare a livello locale una caduta in termini di qualità della copertura. Nel corso di quest'anno è stato sviluppato un sistema di monitoraggio quantitativo (MONITOR²⁵) per la valutazione della copertura dell'indagine. Il sistema permette di controllare a livello comunale (in Italia i comuni sono circa 8100) il numero di modelli pervenuti all'Istat e confrontare il dato con altre fonti informative (ad esempio dati riepilogativi mensili degli eventi riferiti alla popolazione residente distinti per sesso e entro/oltre il 1° anno di vita.). Il sistema, sviluppato in Oracle, permette di aggregare per gruppi di mesi o per aree geografiche più ampie, di confrontare i dati con serie storiche precedenti, di individuare i duplicati e di calcolare indicatori di qualità.

6.2.2 Validazione del dato

L'attuale procedura di controllo e correzione è effettuata con programmi deterministici Cobol su UNIX (predisposti agli inizi degli anni '80). I controlli sono tipo trasversale (tra modalità di variabili appartenenti allo stesso individuo).

Uno degli obiettivi è quello di sostituire l'attuale programma di controllo e correzione degli errori stocastici dei dati demografici con il sistema SCIA (Barcaroli, 1992). SCIA è un sistema per l'editing e l'imputazione automatica di variabili qualitative realizzato autonomamente in Istat a partire dal 1990. SCIA è attualmente operativo in Istat, ed è stato già ampiamente applicato su diverse indagini (es. dati del Censimento degli alloggi e della popolazione, Indagini Multiscopo). I vantaggi offerti dall'uso di SCIA rispetto ai tradizionali piani di incompatibilità di tipo deterministico sono nella tempestività dei risultati e nella qualità dei risultati. La qualità è garantita dalla tecnica di verifica e di correzione utilizzata, che, per la parte probabilistica, è interamente basata sulla metodologia di Fellegi-Holt (Fellegi I.P. e Holt D., 1976).

6.3 ARMONIZZAZIONE INTERNAZIONALE

6.3.1 Qualità della certificazione

La razionalizzazione ed il miglioramento del dato di mortalità è un obiettivo da perseguire a tutti i livelli; tra questi un aspetto determinante è rappresentato dagli interventi da realizzare per il

²⁴ Prefetture e Uffici Regionali Istat

²⁵ Capo Progetto Paola Giorgetti, Istat. Il prodotto è stato sviluppato in collaborazione con RTI-Akros.

miglioramento della certificazione relativa alla parte medica (parte A) della scheda di morte. Ad oggi si è provveduto a realizzare delle istruzioni per la compilazione dei certificati in forma sintetica che, dopo essere state sottoposte al vaglio di alcuni medici legali e del ministero della sanità, sono state stampate sul retro dei certificati medesimi a partire dall'edizione 2002. Un grosso impegno è attualmente sostenuto dall'Istat nel gestire un progetto europeo voluto e finanziato dalla Commissione Europea tramite EUROSTAT per lo sviluppo di materiale didattico-formativo per i medici certificatori che possa migliorare la qualità dei dati di mortalità a livello europeo. Nell'ambito di tale progetto è prevista la realizzazione di un manuale, di un prodotto per il web ed di una brochure sintetica che funga da guida di rapida consultazione; il termine del progetto è previsto per l'estate 2004 e l'adozione di questi strumenti per il training dei medici è da realizzarsi a cura dei singoli paesi coinvolti (40 paesi europei).

6.3.2 Implementazione della ICD X

Il valore aggiunto della codifica automatica

L'esperienza effettuata in precedenza con l'adozione della codifica automatica per l'ICD IX, consente adesso all'Italia di essere tra i pochi paesi in Europa all'avanguardia in questo campo, con la possibilità di proporsi come termine di riferimento e confronto per i paesi che si stanno avviando verso la medesima strada; l'uso del sistema NCHS per l'attribuzione della causa iniziale di morte (modulo ACME) dovrà essere inoltre implementato in Europa per il 2005. Il valore dell'esperienza maturata in questo innovativo settore va quindi a compensare il ritardo con il quale il nostro paese si affaccia all'introduzione della decima revisione dell'ICD. L'esperienza maturata nell'uso e nella gestione di un sistema complesso quale quello della codifica automatica ha fornito le basi per ottenere una implementazione del dizionario per l'ICD X.

Bridge coding

Il bridge coding ('comparability study' negli USA) è un valido strumento che sta affermandosi a livello internazionale quale momento di studio al passaggio di revisione (NCHS, 2001a; NCHS, 2001b; Office for National Statistics, 2002) o di sistema di codifica (Istat, 2000). Per la transizione all'ICD X verrà effettuato uno studio di questo tipo per consentire agli utilizzatori dei dati di mortalità di raccordare agevolmente i salti nelle serie storiche delle cause di morte. In alcuni casi le cause di morte potranno essere analizzate individualmente (alla quarta cifra), a livello di categorie (alla terza cifra) e di blocchi (raggruppamenti alla terza cifra). E' attualmente in corso uno studio di fattibilità in cui viene valutata l'opportunità di codificare con entrambe le classificazioni (ICD IX e

ICD X) tutto l'anno di riferimento o di effettuare un campionamento dei circa 550.000 certificati disponibili. Viene inoltre valutato il livello di dettaglio da fornire per l'analisi delle diverse cause.

Sensibilizzazione all'uso della ICD X

L'introduzione della decima revisione presuppone una notevole mole di lavoro in termini di formazione ed aggiornamento degli operatori preposti alla codifica e degli utilizzatori dei diversi settori coinvolti. La decima revisione infatti introduce sostanziali novità per la complessità del dettaglio fornito e per i cambiamenti introdotti nelle regole di selezione e modifica nell'attribuzione della causa iniziale. Sarà quindi indispensabile approntare un piano formativo da realizzarsi a livello centrale e periferico per il corretto uso della classificazione.

Reimpostazione delle pubblicazioni ufficiali e nuovo piano di diffusione

In ultimo, un ulteriore aspetto da curare e riorganizzare con il passaggio di revisione consiste nella nuova impostazione da utilizzare nelle pubblicazioni ufficiali, che deve tener conto della riorganizzazione dei capitoli e del maggiore dettaglio dell'ICD X rispetto all'ICD IX e del cambiamento dei codici (da numerici ad alfanumerici). Il passaggio di revisione sarà anche un importante stimolo per un ammodernamento del piano di diffusione che sfrutti maggiormente le potenzialità delle metodologie informatiche con particolare attenzione al web.

6.4 VALORIZZAZIONE DELLE FONTI INFORMATIVE ESISTENTI IN ITALY

Le cause multiple di morte: l'esperienza Istat

Come visto ai paragrafi 2.3.3 e 2.7, l'uso di procedure automatizzate fornisce una ottima occasione per studiare, per pianificare la produzione e per diffondere dati sulle cause multiple di mortalità. Tuttavia, al momento attuale non esiste una prassi di lavoro consolidata alla codifica delle cause multiple nemmeno in Istat. L'impianto di regole stabilito dall'NCHS è molto complesso e richiede un periodo di training molto lungo per poter essere adottato con risultati ottimali. L'approccio seguito ad oggi in seno all'Unità Operativa 'Mortalità per Causa' è stato quello di formare direttamente in USA il personale qualificato per la codifica con l'ICD X (formazione per formatori) e successivamente allargare la formazione alla codifica delle cause multiple anche ad altri operatori in Italia.

Si possono individuare due strade per usare i dati sulle cause multiple senza dover ricorrere alla loro codifica manuale:

- utilizzare le schede non scartate dal processo di codifica automatica. Questo approccio introduce però delle distorsioni nei dati in quanto sono generalmente i quadri morbosi più complessi ad

avere le maggiori probabilità di essere scartati, senza contare le limitazioni esistenti nella codifica di alcune cause specifiche con il sistema di codifica automatica;

- ricorrere alla correzione individuale dei record nelle schede scartate e sottoporle al processo di codifica automatica in modo ripetuto, fino a non avere più scarti. Questa seconda possibilità è stata studiata per una patologia specifica, l'Alzheimer con un'analisi relativa ai dati di un anno comprendente circa 4400 schede. Tale approccio, valido ai fini sperimentali, comporta la correzione interattiva dopo il primo passaggio su MICAR, per quanto attiene le modifiche terminologiche, tenendo conto degli aspetti riguardanti la codifica delle cause multiple.

RIFERIMENTI BIBLIOGRAFICI

Balkau B. ed altri (1993), 'European study of the certification and coding of causes of death of six clinical case histories of diabetic patients', *Int J Epidemiol*, n.22, pp. 116-126.

Barcaroli G. (1992), 'An integrated system for edit and imputation of data in the Italian Statistical Institute', *Survey and Statistical Computing*, pp. 167-177.

Barchielli A., Geddes M. (1986), 'The accuracy of local death certificates in cancer of the lung and the stomach', *Tumori*, vol. 72, pp.475-479.

Capocaccia R., Terracini B., Verdecchia A. (a cura di) (1996), *Epidemiologia dei tumori dell'apparato digerente in Italia*, Annali dell'Istituto Superiore di Sanità, vol. 32, n.4.

Capocaccia R., Terracini B., Verdecchia A. (a cura di), (1992) *Epidemiologia dei Tumori dell'apparato respiratorio in Italia*, , Annali dell'Istituto Superiore di Sanità, vol. 28, n.1.

CDC (1999), *Proceedings of the International Collaborative Effort on Automating Mortality Statistics*, vol n.1, Kimberly Peters editor, CDC, NCHS.

CDC (2001), *Proceedings of the International Collaborative Effort on Automating Mortality Statistics*, vol n.2, Arialdi M. Minino and Harry Rosenberg editors, CDC, NCHS.

Chamblee F., Evans C. (1979), *A National Multiple cause of death Data System*, NCHS.

Chamblee F., Evans C. (1982), 'New dimension in cause of death statistics', *Am J Publ Health*, n.72, pp. 1265-1270.

CISM (1989), *Manuale per la raccolta, codifica ed elaborazione dei dati di mortalità*, a cura di Balzi D. ed altri, Firenze.

Cornfield J. (1957), 'The estimation of the probability of developing a disease in presence of competing risks', *Am J Publ Health*, n. 47, pp. 601-607.

Crialesi R., Frova L., Marchetti S. (1998), *The impact on mortality related procedures and data of introducing automated cause coding in Italy*, Contributed Papers, NTTS 98, International Seminar on new Techniques & Technologies for Statistics, 4-6 November, 1998, Sorrento, Italy.

Dorn H.F., Moriyama I.M. (1964), 'Uses and significance of multiple cause tabulations for mortality statistics', *Am J Publ Health*, n. 54, pp. 400-406.

Eurostat (1998), *Causes of death 1994-1995 statistics*, Working Paper 3/1998/E/n° 22, Eurostat E-3, November.

Eurostat (1998), *Final Report on Automation Coding in Member States*, Doc OS/E3/98/COD/3 (EN), Task Force on Causes of Death Statistics, Le Vesinet, INSERM, 8 June.

Eurostat (1999), *Final Report on Automation Coding in Member States*, Eurostat, Eurostat Working Papers – Population and Social Conditions, 3/1999/E/N°1.

Eurostat (2001), *Comparability and Quality Improvement of European Causes of Death Statistics*, European Commission, DG SANCO, N°EDC DGV/F3 SOC 98 20108.

Eurostat (2001), *Final Report on Production of Methods and Tools for Improving Causes of Death Statistics ad Codification level*, Eurostat, Eurostat Working Papers – Population and Social Conditions, 3/2001/E/N°14.

Fellegi I.P. e Holt D. (1976), 'A systematic approach to edit and imputation', *Journal of the American Statistical Association*, vol.71, pp.17-35.

- Feola G., Abbati E. (1995), 'Codifica automatica delle malattie e delle cause di morte', *Quaderni di oncologia*, vol.5 n.1.
- Feola G., Abbati E. (1996), 'Codifica automatica delle malattie e delle cause di morte', in *Atti del VII meeting Medicina e informatica*, Università Cattolica Sacro Cuore di Roma, 15-17 Maggio 1996, pp.189-202.
- Goodman R.A. ed altri (1982), 'Mortality data analysis using multiple-cause approach', *J Am A*, n. 247, pp. 793-796.
- Hansluwska H. (1986)., 'I dati di mortalità in Europa: disponibilità, validità, comparabilità', in *Atti 3° Convegno Nazionale sugli Studi di Mortalità*. CISM, Firenze, pp.9-31.
- Hellermann E. (1982), Overview of the Hellerman I&O Coding System. *Internal document, US Bureau of the Census*.
- Istat (1997), 'Classificazione delle malattie, traumatismi e cause di morte, IX revisione 1975', *Metodi e Norme*, Serie C, n.10, Vol. 1° e 2°, V^a ristampa.
- Istat (2001), 'La Nuova Indagine sulle Cause di morte. La codifica automatica, il bridge coding e altri elementi innovativi', *Metodi e Norme*, n.8.
- Istat (s,d), 'L'analisi della mortalità per la malattia di Alzheimer in Italia. Uno studio di multimorbosità', Argomenti, in preparazione.
- Istat-Iss (s.d), *La mortalità in Italia nel periodo 1970-1992: evoluzione e geografia*. Fuori Collana.
- Jougla E. ed altri (1992), 'Death certificate coding practices related to diabetes in European Countries - the Eurodiab study', *Int J Epidemiol* n. 21, pp. 343-351.
- Jougla E. et al.(1998), 'International comparability of causes of death data Inside the European Community', World Health Organization, Meeting of Heads of OMS Collaborating Centres for the Classification of Diseases, OMS/HST/ICD/C/99.44.
- Jougla E., Pavillon G. (1997), 'International comparability of causes of death data – Methods and results' in *Morbidity and Mortality data: problems of comparability*, Hacettepe University, Institute of Population Studies, Wunsch G. & Hiancioglu A., pp. 75-95.
- Kalpic D. (1994), Automated coding of Census Data, *Journal of Official Statistics*, 10, 449-463.
- Kelson M.C., Farebrother M. (1987), 'The effect of inaccuracies in death certification and coding practices in EEC on international cancer mortality statistics', *Int J Epidemiol* n. 16, pp. 411-414.
- Kelson M.C., Heller R.F. (1983), 'The effect of death certification coding practices on observed differences in respiratory diseases mortality in 8 EEC countries', *Epidemiol Santé Publ* n. 31, pp. 423-432.
- Lyberg L. and Dean P. (1992), Automated Coding of Survey Responses: an international review, *Conference of European Statisticians, Work session on Statistical Data Editing*, Washington D.C.
- Mackenbach J., Van Duyne W., Kelson M. (1987), 'Certification and coding of two underlying causes of death in The Netherlands and other countries of the European Community', *J Epidemiol Comm Health*, n. 41, pp. 156-160.
- Massingham R. (1997), Data capture and Coding for the 2001 Great Britain Census, *XIV Annual International Symposium on Methodology Issues*, Hull, Canada, 5-7 November.
- Nam C.B., Eberstein I.W., Deeb L.C., et al. (1989), 'Infant mortality by cause: a comparison of undelying and multiple cause designations', *Eur J Popul*, vol. 5, pp.45-70.
- NCHS (1969a), *First Progress Report on Final Testing of System from Automated Classification of Medical Entities (ACME)*.

- NCHS (1969b), *Second Progress Report on Final Testing of System from Automated Classification of Medical Entities (ACME)*.
- NCHS (1969c), *Third Progress Report on Final Testing of System from Automated Classification of Medical Entities (ACME)*.
- NCHS (1986), *Transax: the NCHS System for Producing Multiple Cause of Death Statistics*, NCHS Programs and Collections Procedures, Series 1, vol. n.20.
- NCHS (1996a.) *Data entry Instructions for the Mortality Medical Indexing, Classification and retrieval System (MICAR)*, Instruction Manual Part 2g.
- NCHS (2001a), *Comparability of cause of Death between ICD-9 and ICD-10: Preliminary estimates*, Robert N. Anderson, Arialdi M. Minino, Donna L. Hoyert, Harry M. Rosenberg, National Vital Statistics Reports, Vol 49 n°2, Hyattsville, Maryland.
- NCHS (2001b), *Deaths: final data for 1999*, Donna L. Hoyert, Elizabeth Arias, Betty L. Smith, Sherry L. Murphy, Kenneth D. Kochanek, National Vital Statistics Reports, Vol 49 n°8, Hyattsville, Maryland.
- Office for National Statistics (2002), *Results of the ICD-10 bridge coding study, England and Wales*, Health Statistics Quaterly 14, 75-83.
- OMS (1967), *World Health Organisation, International Classification of Diseases, 8th revision*, Ginevra.
- OMS (1977), *World Health Organisation, International Classification of Diseases, 9th revision*, Ginevra.
- OMS (1992), *World Health Organisation, International Statistical Classification of Diseases and related Health Problems, 10th revision*, vol.1, Ginevra.
- OMS (1993), *World Health Organisation, International Statistical Classification of Diseases and related Health Problems, 10th revision*, vol.2, Ginevra.
- OMS (1994), *World Health Organisation, International Statistical Classification of Diseases and related Health Problems, 10th revision*, vol.3, Ginevra.
- OMS (1997), *Translator, Ninth and Tenth Revision, User's Guide to electronic tables (5 diskettes)*, OMS/HST/96.9, Distr.:General.
- OMS (2000), *Classificazione statistica internazionale delle malattie e dei problemi sanitari correlati. Decima revisione*. Vol.1-3, edited by Ministero della Sanità, Italy.
- Pavillion G., Jouglu E. (1997), *Report on the International Collaborative Study on Multiple Cause Analysis*, OMS, ESS/ICD/C/97.18.
- Percy C., Dolman A. (1978), 'Comparison of the coding of death certificates related to cancer in seven countries', *Publ Health Rep* n. 93, pp. 335-350.
- Percy C., Muir C. (1989), 'The international comparability of cancer mortality data', *Am J Epidemiol* n.129, pp. 934-946.
- Pollard A. H. (1980), 'The interaction between morbidity and mortality', *Journal of the Institute of Actuaries*, 107(3):233-302 + summary of discussion, pp. 303-313.
- Preston S. H. (1987), 'Relations among standard epidemiologic measures in a population', *Am. J. of Epidemiology*, vol. 126, n.2, pp.336-345.
- Tourigny J.Y. and Moloney J. (1995), *The 1991 Canadian Census of Population experience with automated coding. Statistical Data Editing, 2. United Nations Statistical Commission*, New York.

Wenzowski M.J. (1988), ACTR - A Generalized Automated Coding System, *Survey Methodology*, 14,2,299-308.

Zanetti R., Viganò C., De Molli S., Colombo S., Cislighi C., 'Comparative completeness and correspondence of cancer mortality data as collected by Istat and cancer registries', *Tumori*, vol.68, pp.457-463.

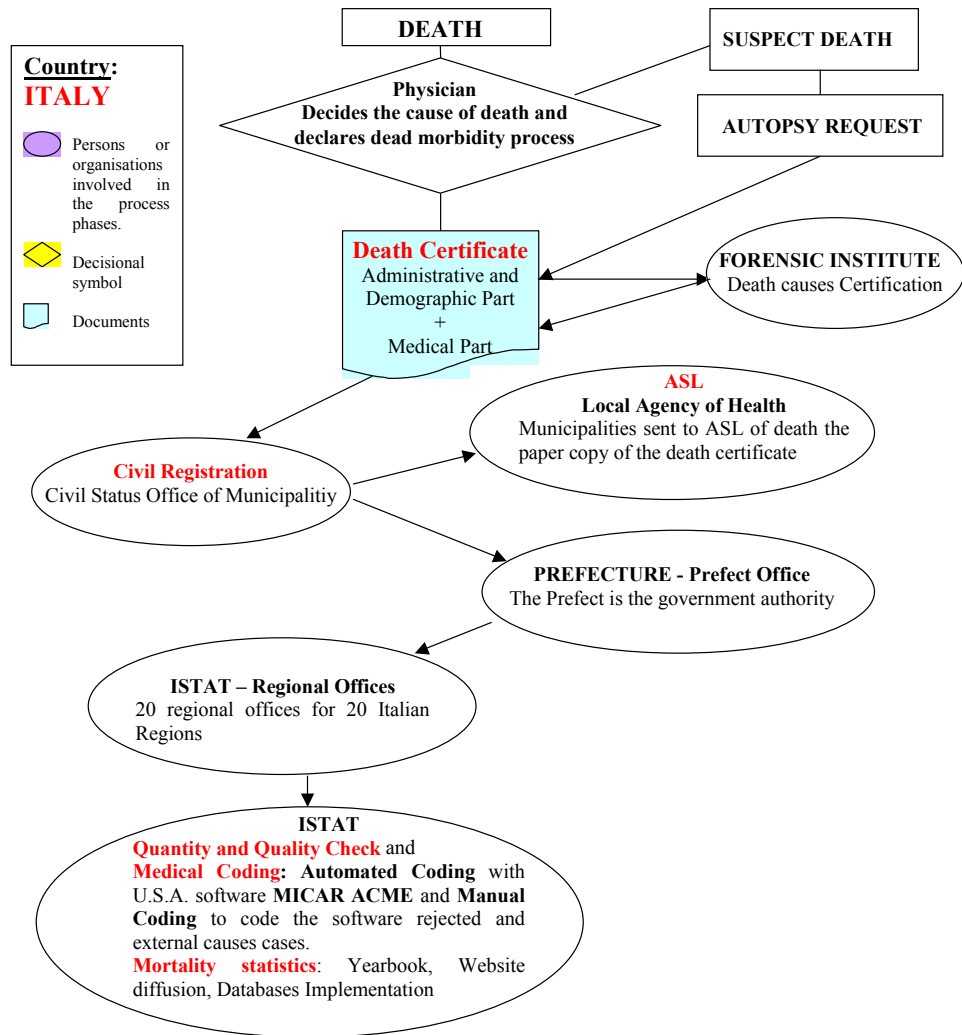
APPENDICE A: IL FLUSSO INFORMATIVO

Prima di seguire attentamente i vari ‘processi innovativi’ introdotti nell’Indagine è utile ricordare in breve il percorso seguito dai modelli prima di giungere compilati all’Istat (Figura 1.1). Ogni anno, l’Istat invia i nuovi modelli (D4, D5, D4bis, D5bis) (cfr. Appendice B) agli Uffici Regionali ed ai Comuni. Le aziende sanitarie locali (ASL), gli ospedali e i medici di famiglia sono tenuti a contattare i Comuni o gli Uffici Regionali i quali mettono a loro disposizione i nuovi modelli. Per ciascun decesso il medico (medico ospedaliero o di famiglia o medico necroscopo) deve indicare sul modello di rilevazione Istat ‘tutte le malattie, gli stati morbosi o i traumatismi che abbiano condotto o contribuito al decesso, e le circostanze dell’accidente o violenza che hanno provocato questi traumatismi’, riportando anche altre notizie di rilievo inerenti al decesso. Nella scheda viene distinto il caso di morte da causa naturale da quello di morte da causa violenta.

Una volta compilata la parte sanitaria da parte del medico, il modello viene inviato ai comuni. L’ufficiale di stato civile deve completare il certificato indicando le informazioni demografiche del deceduto. In particolare le informazioni riguardano, oltre alla data di nascita e di morte, la località di nascita e di residenza, lo stato civile, il grado di istruzione, la condizione professionale, il ramo di attività economica, la cittadinanza e il codice individuale del deceduto. Nel caso in cui il decesso sia avvenuto nel primo anno di vita il modello è leggermente diverso e, per quel che riguarda la parte demografica, le informazioni richieste fanno soprattutto riferimento alla condizione dei genitori.

Il modello, riprodotto in doppia copia, una volta compilato segue due percorsi differenti: una copia viene inviata alla ASL (dove è avvenuto il decesso) mentre l’altra copia, prima di pervenire all’Istat, passa per le Prefetture e gli Uffici regionali, i quali sono tenuti ad effettuare un primo controllo quantitativo del materiale pervenuto.

Figura A.1 – L’Indagine sulle cause di morte: il flusso informativo.



**APPENDICE B. I MODELLI ISTAT PER LA CERTIFICAZIONE DEI DECESSI:
D4, D5, D4BIS AND D5BIS**



Per una corretta compilazione leggere le ISTRUZIONI SUL RETRO della scheda

Cognome e nome del defunto

Età



SPECIMEN

CODICE FISCALE DEL DEFUNTO

ISTAT D.4 - Ediz. 2003 M

COPIA PER L'ISTAT

N. d'ordine della scheda

SCHEDA DI MORTE OLTRE IL 1° ANNO DI VITA PER MASCHIO

PARTE A - A CURA DEL MEDICO

LUOGO DEL DECESSO

- Abitazione, Istituto di cura pubblico, Istituto di cura privato, Struttura socio-assistenziale, Altro

Province and Comune fields, CODICE DELLA ASL DI DECESSO, RISCONTRO DIAGNOSTICO, and table for Riservato all'ISTAT

MORTE DA CAUSA NATURALE

1. Causa iniziale

Scegliere la SOLA patologia che ha dato inizio alla catena dei processi morbosi indicati ai punti 2. e 3.

Form for cause 1 with intervallo (a) field

che ha provocato causa 2

2. Causa intermedia

Eventuali complicazioni o evoluzione della causa precedente, indicata al punto 1.

Form for cause 2 with intervallo (a) field

che ha/hanno provocato causa 3

3. Causa terminale

Patologia che ha immediatamente preceduto il decesso collegata alle cause indicate ai punti 1. e 2.

Form for cause 3 with intervallo (a) field

4. Altri stati morbosi rilevanti che hanno contribuito al decesso

Indicare altre malattie o stati morbosi rilevanti, esclusi dal concatenamento indicato ai punti 1 2 3, ma che hanno contribuito al decesso.

Form for cause 4 with intervallo (a) field

(a) Indicare anche approssimativamente l'intervallo tra l'insorgenza della malattia e la morte

MORTE DA CAUSA VIOLENTA

5. Causa violenta

- Accidentale, Suicidio, Infortunio sul lavoro, Omicidio

6. Descrizione della lesione

Form for cause 6

7. Malattie o complicazioni eventualmente sopravvenute conseguenti alla lesione

Form for cause 7

8. Stati morbosi preesistenti che hanno eventualmente contribuito al decesso

Form for cause 8

9.1 Modalità che ha provocato la lesione indicata al punto 6

Form for cause 9.1

Indicare l'intervallo di tempo tra l'azione violenta e la morte

9.2 Data dell'accidente, infortunio, suicidio, omicidio

Form for cause 9.2

9.3 Luogo dell'accidente, infortunio, suicidio, omicidio

- Abitazione, Istituzione collettiva, Scuola, Luogo dedicato alle attività sportive, Area industriale e di costruzione, Altri luoghi

PARTE B - NOTIZIE A CURA DELL'UFFICIALE DI STATO CIVILE

ATTO DI MORTE

Form for Act of Death: Numero, Parte, Serie

UFFICIO DI STATO CIVILE CIRCOSCRIZIONE

Table for Comune and Provincia with N. distintivo da apporre con il timbro

1. Data di morte

Form for date of death: Ora, Giorno, Mese, Anno

2. Data di nascita

Form for date of birth: Giorno, Mese, Anno

3. Località di nascita

Form for place of birth: Stesso comune di morte, Altro comune, Stato estero

4. Età compiuta

Form for age completed: Anni

5. Stato civile

Form for marital status: Celibe, Coniugato, Vedovo, Divorziato o già coniugato, Separato legalmente

6. Residenza

Form for residence: Stesso comune di morte, Altro comune, Stato estero

7. Grado di istruzione

Form for level of education: Laurea, Diploma universitario, Diploma di scuola media superiore, Licenza di scuola media inferiore, Licenza elementare

8. Professione

9. Condizione professionale o non professionale

Form for professional condition: Occupato, Disoccupato, In cerca di prima occupazione, Ritirato dal lavoro, Casalingo, Studente, Inabile al lavoro, Altro

10. Posizione nella professione

Form for position in profession: Per lavoratori autonomi, Per lavoratori dipendenti

11. Ramo di attività economica

Form for economic activity: Agricoltura, caccia e pesca, Industria, Commercio, pubblica amministrazione, Altri servizi privati

12. Cittadinanza

Form for citizenship: Italiana, Straniera

Form for signature: Firma, nominativo e recapito dell'ufficiale dello Stato Civile

DICHIARO CHE LE CAUSE DELLA MORTE SECONDO SCIENZA E COSCIENZA, SONO QUELLE DA ME SOPRAINDICATE

Data

Timbro e telefono del medico o della struttura

curante

necroscopo

SPECIMEN

(a) Per «già coniugato» si intende la persona che ha ottenuto lo scioglimento o la cessazione degli effetti civili del matrimonio ai sensi della legge 1° dicembre 1970 n. 898 (b) Solo per i Comuni che presentano una suddivisione del territorio in circoscrizioni



Per una corretta compilazione leggere le ISTRUZIONI SUL RETRO della scheda

Cognome e nome della defunta

Età



SPECIMEN

CODICE FISCALE DELLA DEFUNTA

ISTAT D.5 - Ediz. 2003 F

COPIA PER L'ISTAT

N. d'ordine della scheda

SCHEDA DI MORTE OLTRE IL 1° ANNO DI VITA PER FEMMINA

PARTE A - A CURA DEL MEDICO

LUOGO DEL DECESSO

- Abitazione 1
- Istituto di cura pubblico 2
- Istituto di cura privato 3
- Struttura socio-assistenziale 4
- Altro (specificare) 5

Provincia di _____
 Comune di _____
 CODICE DELLA ASL DI DECESSO _____ RISCONTRO DIAGNOSTICO
 Richiesto _____ 1
 Non richiesto _____ 2

Riservato all'ISTAT		
I		
E		

MORTE DA CAUSA NATURALE

1. Causa iniziale (scrivere in stampatello)

Scegliere la SOLA patologia che ha dato inizio alla catena dei processi morbosi indicati ai punti 2. e 3.

_____ intervallo (a) _____
 anni mesi giorni

che ha provocato causa 2

2. Causa intermedia (scrivere in stampatello)

Eventuali complicazioni o evoluzione della causa precedente, indicata al punto 1.

_____ intervallo (a) _____
 anni mesi giorni

che ha/hanno provocato causa 3

3. Causa terminale (scrivere in stampatello)

Patologia che ha immediatamente preceduto il decesso collegata alle cause indicate ai punti 1. e 2. Non inserire il meccanismo del decesso (arresto cardiovascolare, collasso respiratorio, etc.)

_____ intervallo (a) _____
 anni mesi giorni

4. Altri stati morbosi rilevanti che hanno contribuito al decesso (scrivere in stampatello)

Indicare altre malattie o stati morbosi rilevanti, esclusi dal concatenamento indicato ai punti 1 2 3, ma che hanno contribuito al decesso.

_____ intervallo (a) _____
 anni mesi giorni

(a) indicare anche approssimativamente l'intervallo tra l'insorgenza della malattia e la morte

10. Gravidanza

- Morte avvenuta durante la gravidanza 1 Morte avvenuta entro 42 giorni dopo il parto 2 Morte avvenuta tra 43 giorni ed un anno dopo il parto 3

DICHIARO CHE LE CAUSE DELLA MORTE SECONDO SCIENZA E COSCIENZA, SONO QUELLE DA ME SOPRAINDICATE

Data _____

curante

necroscopo

Firma leggibile del medico

Timbro e telefono del medico o della struttura

SPECIMEN

PARTE B - NOTIZIE A CURA DELL'UFFICIALE DI STATO CIVILE (CONTROLLARE ED EVENTUALMENTE CORREGGERE IL COGNOME E IL NOME DELLA DEFUNTA)

ATTO DI MORTE

Numero _____

Parte 1 I 2 II

Serie 1 B 2 C

UFFICIO DI STATO CIVILE

CIRCOSCRIZIONE _____ (b)

N. distintivo da apporre con il timbro	
COMUNE	
PROVINCIA	

1. Data di morte

_____ Ora _____ Giorno _____ Mese _____ Anno _____

2. Data di nascita

_____ Giorno _____ Mese _____ Anno _____

3. Località di nascita

Stesso comune di morte 1

Altro comune 2

_____ (specificare comune e provincia) _____ (comune) _____ (prov.)

Stato estero 3

_____ (specificare) _____ 0 0 0 _____

4. Età compiuta

Anni _____ 4. _____

5. Stato civile

Nubile 1

Coniugata 2

Vedova 3

Divorziata o già coniugata (a) 4

Separata legalmente 5

Se coniugata o separata legalmente indicare:

l'anno di nascita del coniuge superstite _____

l'anno di matrimonio _____

6. Residenza

Stesso comune di morte 1

Altro comune 2

_____ (specificare comune e provincia) _____ (comune) _____ (prov.)

Stato estero 3

_____ (specificare) _____ 0 0 0 _____

7. Grado di istruzione

Laurea 1

Diploma universitario o laurea breve 2

Diploma di scuola media superiore 3

Licenza di scuola media inferiore 4

Licenza elementare o nessun titolo 5

8. Professione

9. Condizione professionale o non professionale

Occupata 1

Disoccupata 2

In cerca di prima occupazione 3

Ritirata dal lavoro 4

Casalanga 5

Studentessa 6

Inabile al lavoro 7

Altro 8

10. Posizione nella professione

Per lavoratrici autonome

Imprenditrice o libera professionista 1

Lavoratrice in proprio o coadiuvante 2

Altro 3

Per lavoratrici dipendenti

Dirigente o direttivo 4

Impiegata o intermedia 5

Operaia o assimilata 6

Altro (apprendista, lav. a domicilio, ecc.) 7

11. Ramo di attività economica

Agricoltura, caccia e pesca 1

Industria 2

Commercio, pubblici esercizi, alberghi 3

Pubblica amministrazione e servizi pubblici 4

Altri servizi privati 5

12. Cittadinanza

Italiana

Per nascita 1

Acquisita 2

Straniera 3

_____ (specificare) _____

Firma, nominativo e recapito dell'ufficiale dello Stato Civile

Bollo

(a) Per «già coniugato» si intende la persona che ha ottenuto lo scioglimento o la cessazione degli effetti civili del matrimonio ai sensi della legge 1° dicembre 1970 n. 898

(b) Solo per i Comuni che presentano una suddivisione del territorio in circoscrizioni



Per una corretta compilazione leggere le ISTRUZIONI SUL RETRO della scheda



SPECIMEN

ISTAT D.4 bis - Ediz. 2003 M bis

COPIA PER L'ISTAT

Cognome e nome del defunto

Età

CODICE FISCALE DEL DEFUNTO

CODICE FISCALE DEL PADRE

CODICE FISCALE DELLA MADRE

N. d'ordine della scheda

SCHEDA DI MORTE NEL 1° ANNO DI VITA PER MASCHIO

PARTE A - A CURA DEL MEDICO

LUOGO DEL DECESSO

- Abitazione, Istituto di cura pubblico, Istituto di cura privato, Altro

Province, Comune, CODICE DELLA ASL DI DECESSO, RISCANTRO DIAGNOSTICO

Table with 3 rows (I, E, M) and 3 columns (Riservato all'ISTAT)

MORTE DA CAUSA NATURALE

1. Causa iniziale (scrivere in stampatello) Scegliere la SOLA patologia...

2. Causa intermedia (scrivere in stampatello) Eventuali complicazioni o evoluzione...

3. Causa terminale (scrivere in stampatello) Patologia che ha immediatamente preceduto...

4. Altri stati morbosi rilevanti che hanno contribuito al decesso

(a) indicare anche approssimativamente l'intervallo tra l'insorgenza della malattia e la morte

MORTE DA CAUSA VIOLENTA

5. Causa violenta Accidentale, Infanticidio

6. Descrizione della lesione (scrivere in stampatello)

7. Malattie o complicazioni eventualmente sopravvenute conseguenti alla lesione

8. Stati morbosi preesistenti comprese anomalie o malformazioni congenite

9.1 Modalità che ha provocato la lesione indicata al punto 6

Indicare l'intervallo di tempo tra l'azione violenta e la morte

9.2 Data dell'accidente, infanticidio

9.3 Luogo dell'accidente, infanticidio

PARTE B - NOTIZIE A CURA DELL'UFFICIALE DI STATO CIVILE

ATTO DI MORTE

Numero, Parte, Serie, UFFICIO DI STATO CIVILE CIRCOSCRIZIONE

Table with 2 rows (COMUNE, PROVINCIA) and 2 columns (N. distintivo da apporre con il timbro)

1. Data di morte

2. Data di nascita

3. Località di nascita

4. Età compiuta

5. Genere del parto

6. Filiazione

7. Data di nascita della madre

8. Residenza della madre

9. Grado di istruzione dei genitori

10. Professione

11. Condizione professionale o non professionale dei genitori

Occupato, Disoccupato, In cerca di prima occupazione, Ritirato dal lavoro, Casalingo/a, Studente, Inabile al lavoro, Altro

12. Posizione nella professione dei genitori

Per lavoratori autonomi, Per lavoratori dipendenti

13. Ramo di attività economica dei genitori

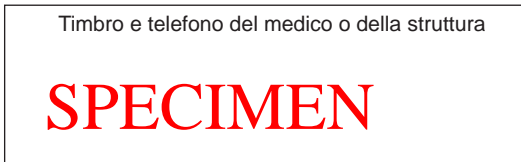
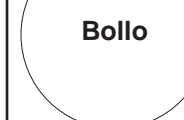
Agricoltura, caccia e pesca, Industria, Commercio, pubblica amministrazione, Altri servizi privati

14. Cittadinanza dei genitori

Italiana, Straniera, Non conosciuta (*)

(*) nel caso dei nati naturali riconosciuti da uno solo dei genitori

Firma, nominativo e recapito dell'ufficiale dello Stato Civile



SPECIMEN

DICHIARO CHE LE CAUSE DELLA MORTE SECONDO SCIENZA E COSCIENZA, SONO QUELLE DA ME SOPRAINDICATE

Data, curante, necroscopo

(b) Solo per i Comuni che presentano una suddivisione del territorio in circoscrizioni



Per una corretta compilazione leggere le ISTRUZIONI SUL RETRO della scheda



SPECIMEN

ISTAT D.5 bis - Ediz. 2003 F bis

COPIA PER L'ISTAT

Cognome e nome della defunta

Età

CODICE FISCALE DELLA DEFUNTA

CODICE FISCALE DEL PADRE

CODICE FISCALE DELLA MADRE

N. d'ordine della scheda

SCHEDA DI MORTE NEL 1° ANNO DI VITA PER FEMMINA

PARTE A - A CURA DEL MEDICO

LUOGO DEL DECESSO

- Abitazione, Istituto di cura pubblico, Istituto di cura privato, Altro

Province, Comune, CODICE DELLA ASL DI DECESSO, RISCANTRO DIAGNOSTICO

Table with 3 columns and 3 rows (I, E, M) for ISTAT reservation

MORTE DA CAUSA NATURALE

1. Causa iniziale (scrivere in stampatello) Scegliere la SOLA patologia che ha dato inizio alla catena dei processi morbosi...

Intervallo (a) in mesi, giorni, ore

che ha provocato causa 2

2. Causa intermedia (scrivere in stampatello) Eventuali complicazioni o evoluzione della causa precedente...

Intervallo (a) in mesi, giorni, ore

che ha/hanno provocato causa 3

3. Causa terminale (scrivere in stampatello) Patologia che ha immediatamente preceduto il decesso collegata alle cause indicate ai punti 1. e 2.

Intervallo (a) in mesi, giorni, ore

4. Altri stati morbosi rilevanti che hanno contribuito al decesso (scrivere in stampatello) Indicare altre malattie o stati morbosi (anomalie o malformazioni congenite) esclusi dal concatenamento indicato ai punti 1 2 3, ma che hanno contribuito al decesso.

Intervallo (a) in mesi, giorni, ore

(a) indicare anche approssimativamente l'intervallo tra l'insorgenza della malattia e la morte

MORTE DA CAUSA VIOLENTA

5. Causa violenta Accidentale, Infanticidio

6. Descrizione della lesione (scrivere in stampatello)

Area for describing the lesion

7. Malattie o complicazioni eventualmente sopravvenute conseguenti alla lesione (scrivere in stampatello)

Area for diseases or complications

8. Stati morbosi preesistenti comprese anomalie o malformazioni congenite che hanno eventualmente contribuito al decesso (scrivere in stampatello)

Area for pre-existing morbid states

9.1 Modalità che ha provocato la lesione indicata al punto 6 (soffocamento da corpo estraneo, soffocamento nella culla, caduta da gradino, contatto con filo percorso da corrente elettrica, ecc.) (scrivere in stampatello)

Area for mode of injury

Indicare l'intervallo di tempo tra l'azione violenta e la morte

9.2 Data dell'accidente, infanticidio

9.3 Luogo dell'accidente, infanticidio

DICHIARO CHE LE CAUSE DELLA MORTE SECONDO SCIENZA E COSCIENZA, SONO QUELLE DA ME SOPRAINDICATE

Data

curante, necroscopo

Timbro e telefono del medico o della struttura

SPECIMEN

PARTE B - NOTIZIE A CURA DELL'UFFICIALE DI STATO CIVILE (CONTROLLARE ED EVENTUALMENTE CORREGGERE IL COGNOME E IL NOME DELLA DEFUNTA)

ATTO DI MORTE

Numero, Parte, Serie, UFFICIO DI STATO CIVILE CIRCOSCRIZIONE

Table for Comune and Provincia with N. distintivo da apporre con il timbro

1. Data di morte

2. Data di nascita

3. Località di nascita

4. Età compiuta

5. Genere del parto

6. Filiazione

7. Data di nascita della madre

8. Residenza della madre

9. Grado di istruzione dei genitori

10. Professione

11. Condizione professionale o non professionale dei genitori

Occupato, Disoccupato, In cerca di prima occupazione, Ritirato dal lavoro, Casalingo/a, Studente, Inabile al lavoro, Altro

12. Posizione nella professione dei genitori

Per lavoratori autonomi, Per lavoratori dipendenti

13. Ramo di attività economica dei genitori

Agricoltura, caccia e pesca, Industria, Commercio, pubblica amministrazione e servizi pubblici, Altri servizi privati

14. Cittadinanza dei genitori

Italiana, Straniera, Non conosciuta (*)

(*) nel caso dei nati naturali riconosciuti da uno solo dei genitori (è sconosciuta la cittadinanza dell'altro) o di filiazione ignota (sconosciuta la cittadinanza di entrambi i genitori)

Firma, nominativo e recapito dell'ufficiale dello Stato Civile

Bollo

(b) Solo per i Comuni che presentano una suddivisione del territorio in circoscrizioni

APPENDICE C: ISTRUZIONI PER LA COMPILAZIONE DEI CERTIFICATI DI MORTE (DISPONIBILE SUL RETRO DEI MODELLI)²⁶

ISTRUZIONI PER LA COMPILAZIONE DELLA PARTE 'A', RELATIVA ALLE CAUSE DI MORTE, NELLE SCHEDE DI MORTE OLTRE IL 1° ANNO DI VITA PER FEMMINA (MOD. ISTAT D.5 - EDIZIONE 2002).

L'accuratezza nel fornire informazioni sulle cause di morte è importante per la corretta impostazione di programmi sanitari a medio-lungo termine finalizzati al miglioramento della salute pubblica.

Il diritto alla riservatezza non viene violato perché tali dati sono forniti ad esclusivo uso statistico e possono essere diffusi esclusivamente in forma aggregata in modo da risultare anonimi. E' fatto altresì obbligo di fornire tutti i dati richiesti sul presente modello, come previsto dalla normativa attuale.

Le informazioni sulle cause di morte DEVONO rappresentare la migliore opinione MEDICA possibile.

Compilare a stampatello LEGGIBILE con inchiostro nero.

Oltre al cognome, nome, età della defunta il medico è tenuto alla compilazione della parte 'A' della scheda di morte, relativa alle cause di morte e comprendente informazioni circa la Provincia, il Comune, il luogo e la ASL dove è avvenuto il decesso e l'eventuale richiesta di riscontro diagnostico. Essa è suddivisa in due riquadri che devono essere compilati in alternativa l'uno all'altro e che descrivono rispettivamente le morti per cause naturali e le morti per cause violente.

Morte da causa naturale: Riportare negli appositi spazi ogni malattia o stato morboso che abbia avuto rilevanza nel contribuire al decesso dell'individuo. Una condizione può essere riportata come 'probabile' se non diagnosticata in modo assolutamente certo. Evitare di riportare sintomi. Se si ritiene che l'abuso di sostanze stupefacenti, di alcool o di tabacco, di cui è nota la relazione causa-effetto abbiano causato la morte o contribuito ad essa, ciò deve essere riportato ove più appropriato. Per ciascuno dei quattro quesiti va riportato nell'apposito spazio l'intervallo di tempo trascorso tra l'insorgere della malattia e la morte, in anni, mesi, o giorni; indicare se l'informazione è approssimativa o sconosciuta, ma NON LASCIARE IN BIANCO.

IL QUESITO 1 dovrebbe riportare la cosiddetta CAUSA INIZIALE, la patologia o stato morboso individuata come iniziale responsabile del concatenamento di patologie che hanno portato alla morte.

IL QUESITO 2 dovrebbe riportare, se presente/i, la/le cosiddetta/e CAUSA/E INTERMEDIA/E o complicazione/i della malattia indicata al quesito 1. Esso dovrebbe indicare la/le sequenza/e che dalla causa iniziale può/possono aver dato luogo allo sviluppo di una ulteriore causa per mezzo di danno ai tessuti o perdita di funzione.

IL QUESITO 3 va sempre riempito e dovrebbe riportare la cosiddetta CAUSA TERMINALE. In esso NON va riportata la modalità o il meccanismo di morte (per es.: arresto cardiaco; insufficienza respiratoria, etc..), ma la condizione morbosa che ha direttamente provocato la morte (per es.: emorragia intracerebrale, encefalopatia epatica, insufficienza della valvola mitralica, etc...). Se non è stata identificata una causa iniziale di morte, è buona norma riportare la causa finale al quesito 3 avendo l'accortezza di riportare al quesito 1 la dicitura: 'CAUSA NATURALE INDETERMINATA'; ciò consente di comprendere che non vi è stata svista od omissione nella compilazione della scheda di morte.

IL QUESITO 4 dovrebbe riportare gli ALTRI STATI MORBOSI RILEVANTI CHE HANNO CONTRIBUITO AL DECESSO, ma che non rientrano nella catena principale di eventi che ha causato la morte indicata ai quesiti 1, 2, 3. Esso dovrebbe fornire ulteriori informazioni sul quadro morboso della defunta, in grado di dare una panoramica più globale sul dato relativo alle cause di morte.

OPPURE

Morte da causa violenta: riportare ogni lesione che abbia provocato un trauma in grado di causare o contribuire alla morte. Una condizione può essere riportata come 'probabile' se non diagnosticata in modo assolutamente certo. Evitare di riportare sintomi. Se si ritiene che l'abuso di sostanze stupefacenti, di alcool o di tabacco, di cui è nota la relazione causa-effetto abbiano causato la morte o contribuito ad essa, questa/e condizione/i devono essere riportate ove più appropriato.

²⁶ Vengono riportate solo le istruzioni relative alla compilazione del modello D5. Le istruzioni degli altri modelli (D4; D4bis; D5bis), a parte opportune modifiche, sono essenzialmente analoghe.

AL QUESITO 5 sono indicate quattro diverse CAUSE all'origine della morte violenta. Barrare SEMPRE una, ed una sola, delle quattro possibilità previste.

IL QUESITO 6 dovrebbe riportare la DESCRIZIONE DELLA LESIONE che ha iniziato la sequenza di eventi che ha portato alla morte. Questa lesione viene generalmente considerata la causa di morte.

IL QUESITO 7 dovrebbe riportare le EVENTUALI MALATTIE O COMPLICAZIONI SOPRAVVENUTE IN SEGUITO ALLA LESIONE iniziale.

IL QUESITO 8 dovrebbe riportare gli STATI MORBOSI PREESISTENTI CHE HANNO CONTRIBUITO AL DECESSO, ma che non rientrano nella catena principale di eventi che ha causato la morte riportata nei quesiti 6 e 7. Esso dovrebbe fornire ulteriori informazioni sul quadro morboso della defunta, in grado di dare una panoramica più completa sul dato relativo alle cause di morte.

IL QUESITO 9.1 dovrebbe riportare le informazioni sul MEZZO O MODO COL QUALE LA LESIONE E' STATA DETERMINATA. Ad esempio, essa può essere stata provocata da armi, oggetti, mezzi meccanici, cadute, sostanze chimiche, farmaci, etc.... Indicare con precisione il mezzo o circostanza che ha causato la lesione. Il quesito 9.1 richiede inoltre che venga riportato nell'apposito spazio l'intervallo di tempo trascorso tra l'avvenuta lesione indicata al quesito 6 e la morte, in anni, mesi, o giorni; segnalare se l'informazione è approssimativa o sconosciuta, ma NON LASCIARE IN BIANCO.

I QUESITI 9.2 e 9.3 forniscono informazioni sulla data e sul luogo in cui il decesso è avvenuto. E' IMPORTANTE COMPILARE QUESTI QUESITI IN TUTTE LE LORO PARTI perché ciò consente una migliore elaborazione dei dati.

Quesito 10: Gravidanza in atto o occorsa nell'anno che precede la morte.

In base alle recenti raccomandazioni dell'OMS è opportuno indicare, ai fini di una migliore codifica delle cause di morte, l'eventuale stato di gravidanza della deceduta secondo le modalità riportate. Barrare la casella corrispondente.

La dichiarazione deve essere resa dal medico curante o necroscopo; riportare la data di compilazione, il timbro con il numero telefonico del medico o della struttura ospitante la defunta e la FIRMA LEGGIBILE del medico che ha compilato la scheda di morte. Se dopo la compilazione della scheda di morte si ottengono ulteriori informazioni, ad esempio tramite riscontri necroscopici, che modificano o individuano con certezza la causa iniziale di morte o il tipo di lesione responsabile della morte, queste vanno tempestivamente comunicate alla ASL di competenza.