



# SUGItalia '98

## Atti del Convegno

XIV Convegno annuale  
di SAS Institute

Centro Congressi Lingotto  
Torino  
28-30 ottobre 1998

**SAS**  
SAS Institute

## Sanità

### **Il “Web-powered” warehouse: Manuale di sopravvivenza**

*Roberto Gregori - Azienda Ospedaliera, Arcispedale S. Anna, Ferrara* ..... pag. 373

### **Un’applicazione dell’utilizzo della Scheda di Dimissione Ospedaliera**

*Ettore Bidoli, Diego Serraino, Silvia Franceschi - Centro Riferimento Oncologico, Aviano (PN)* .... pag. 378

### **Costruzione di un Registro Tumori Regionale automatizzato in Friuli - Venezia Giulia**

*Ettore Bidoli, Luigino Dal Maso - Centro Riferimento Oncologico, Aviano (PN)*

*Giorgio Simon, Loris Zanier - Regione Friuli Venezia Giulia, Udine*

*Beatrice Delfrate, Pierantonio Romor - Insiel, Udine* ..... pag. 383

### **Dal progetto DRG Analyst alla realizzazione di strumenti ad hoc per la analisi statistica nel contesto di una rete informativa del sistema sanitario regionale su base distribuita**

*Fabrizio Carinci*

*Consorzio Mario Negri Sud, Istituto Ricerche Farmacologiche, S: Maria Imbaro (CH)* ..... pag. 397

### **Rete Informativa Distributiva (RID) e Sistemi di Analisi Statistica avanzata come piattaforma di sviluppo dei Sistemi Informativi Istituzionali**

*Fabrizio Carinci, Donatella Corrado, Antonio D’Ettore, Angelo Menna, Fabio Pellegrini*

*Consorzio Mario Negri Sud, Istituto di Ricerche Farmacologiche - S. Maria Imbaro (CH)* ..... pag. 405

### **SCHARP (Survival Curve and Hazard Ratio Program)**

#### **Una applicazione SAS per la conduzione di metanalisi basate su dati individuali**

*Angelo Tinazzi, Valter Torri - Istituto di Ricerche Mario Negri, Milano*

*Jayne Tierney, Lesley Stewart, Max Parmar - MRC Cancer Trial Office, Cambridge (UK)* ..... pag. 412

### **Il monitoraggio delle prestazioni sanitarie e la gestione delle compensazioni per la mobilità**

*Aldo Bellotto, Paolo Romairone—Datasiel, Genova* ..... pag. 422

### **Verso il sistema informativo integrato in ambito sanitario**

#### **Massimo Arcà, Claudio Grego - Osservatorio Epidemiologico del Lazio**

*Vittoria Liguori, Maria Carla Terzi - Sistemi Territoriali*

*Edoardo Bracci - Istituto CNUCE-CNR* ..... pag. 435

## Dal progetto DRG Analyst alla realizzazione di strumenti ad hoc per la analisi statistica nel contesto di una rete informativa del sistema sanitario regionale su base distribuita

Fabrizio Carinci

Unità di Statistica e Sistemi Informativi

Dipartimento di Farmacologia Clinica ed Epidemiologia

Istituto di Ricerche Farmacologiche Mario Negri, Consorzio Mario Negri Sud, S.Maria Imbaro (CH)

(contatti: <http://www.cmns.mnegri.it/it/special/RISS.html>, e-mail: [carinci@cmns.mnegri.it](mailto:carinci@cmns.mnegri.it))

### Abstract

Per dare un saggio delle potenzialità insite dai data-base ospedalieri, nel 1996 abbiamo realizzato un primo prototipo di sistema informativo, scritto in linguaggio macro SAS<sup>®</sup>, moduli BASE, STAT, GRAPH e IML, basato su un ristretto insieme di caratteristiche estratte dalla SDO. Nel presente lavoro si descrive il passaggio di tale prototipo di sistema informativo sanitario dalla originaria logica di "DRG Analyst" ad una impostazione nuova, basata su di una logica condivisa e su un metodo di lavoro "distribuito" che prevede la permanenza dei files a livello periferico e lo svolgimento di una analisi dei dati "frammentata".

Tale sistema, oggi in fase di realizzazione presso l'Unità di Statistica e Sistemi Informativi (USSI) del Dipartimento di Farmacologia Clinica ed Epidemiologia del Consorzio Mario Negri Sud (DFCE), nel contesto del progetto RISS-H, sta supportando le attività degli osservatori epidemiologici della Regione Basilicata e della Regione Puglia.

La tecnica impiegata per la realizzazione dello schema del nuovo software è stata definita approccio "Touch+Pile-Up+meta-Analyze (TPA)", facendo riferimento ai passi da compiere per completare il processo che corrispondono oggi a tre elementi ben distinti e ideati per una implementazione in "rete" più moderna e destinata ad avvantaggiarsi dell'uso delle nuove tecnologie. A tal proposito si stanno impiegando i moduli SAS<sup>®</sup>/CONNECT e SAS<sup>®</sup>/AF per un approccio visuale e telematico del quale beneficerà soprattutto l'utente finale per poter effettuare una analisi statistica in tempo reale e soprattutto da una postazione remota.

Approcci del tipo data mining, che in biostatistica ed epidemiologia trovano corrispondenza in tecniche analoghe quali gli alberi di regressione (RECPAM), utili nella classificazione dei pazienti e nella esplicitazione di modelli multivariati per il supporto alle decisioni in salute pubblica, sono oggetto di ulteriore approfondimento presso l'USSI, nell'ottica di un futuro abbinamento a RISS-H come modulo di analisi statistica avanzata.

### Introduzione

Il nuovo assetto organizzativo della rete ospedaliera ha reso necessaria la predisposizione su tutto il territorio di strumenti di raccolta dati dei pazienti quali la scheda di dimissione ospedaliera (SDO), che, partendo da un ottica di contabilità aziendale per sua natura limitata ad un insieme di caratteristiche minimali, rappresenta sicuramente il primo valido supporto di routine atto a predisporre il monitoraggio del servizio sanitario in tutti i sistemi regionali.

La compilazione della SDO ha assunto carattere obbligatorio in Italia a partire dalla introduzione di meccanismi di finanziamento degli ospedali su base prospettica, ovvero secondo uno schema di imputazione dei costi sostenuti basata sui volumi delle prestazioni realmente effettuate, e non già sulla base dei posti letto esistenti. Tale imputazione si basa su una classificazione dei pazienti (guidata da un meccanismo statistico di ripartizione ad albero) originariamente ottenuta come risultato di più ricerche condotte negli Stati Uniti a partire da un campione di circa 5 milioni di ricoveri<sup>1,2</sup>. Questo sistema di classificazione ha assunto il nome di Raggruppamenti Omogenei per Diagnosi (*Diagnosis Related Groups*, o semplicemente *DRG*).

Il DFCE ha da tempo iniziato un percorso di approfondimento di tali tematiche, sia a livello sperimentale che applicativo, proponendo modelli interpretativi nella area della ricerca sui servizi sanitari (*Health Services Research*)<sup>3</sup>.

In questo ambito, ci siamo quindi occupati della esplorazione delle caratteristiche contenute nella SDO a fini di valutazione comparativa della attività ospedaliera e di analisi delle prestazioni, ben consci della relativa valenza scientifica del supporto SDO, il cui utilizzo ai fini della valutazione epidemiologica deve essere comunque preso con estrema cautela, sia per motivi connessi alla discutibile qualità dei dati di routine, sia per la distorsione indotta da un eventuale uso alternativo di dati connessi al sistema di finanziamento.

Ciononostante, la presenza di una registrazione sistematica delle prestazioni rappresenta un motivo di grande novità nel sistema italiano, tanto che indubbiamente i risultati di tale innovazione stanno andando al di là delle motivazioni del legislatore

stesso nel proporre metodologie ed indirizzi nuovi per il servizio sanitario.

Il porre obiettivi quali "...procedere alla elaborazione di un modello di programmazione regionale, capace di soddisfare la domanda espressa, di prevenire ragionevolmente i bisogni futuri, garantire un corretto utilizzo delle risorse e dei finanziamenti disponibili in proposito ..." (Piano Sanitario Regionale, Regione Abruzzo 1994-1996), evidenzia peraltro, anche a livello di piccole aree, la rilevanza con la quale era da tempo avvertita l'esigenza di coniugare lo studio della popolazione ad una gestione attenta e ottimale delle risorse disponibili.

La riforma ha reso possibile una analisi organica dei problemi in chiave di salute pubblica, che troverà la sua migliore applicazione in ambiti territoriali circoscritti quali quelli regionali. In tali aree, infatti, popolazione ed operatori della sanità (presidi ospedalieri, distretti sanitari di base, università, centri di ricerca) possono più proficuamente e facilmente interagire e integrarsi in una collaborazione di studio.

Seppure in alcune regioni si fatichi non poco ad organizzare le basi di dati continuative atte a calibrare gli strumenti di analisi a supporto alle decisioni nel campo sanitario (si pensi alle aree di programmazione e finanziamento, all'area giuridica, amministrativa, tecnico-sanitaria, vigilanza e controllo, agli osservatori epidemiologici), i dati provenienti dalle SDO consentono di orientare gli interventi in queste aree come non è mai stato possibile in precedenza.

Il passaggio importante è però rappresentato dalla possibilità di aumentare gli spazi di azione delle realtà periferiche, ovvero dalla capacità che sapranno esercitare le istituzioni nel dare risposta ad una delle esigenze maggiormente invocate a livello politico, amministrativo e sociale: il decentramento delle azioni di governo.

Questa finalità corrisponde tecnicamente all'adozione della metodologia di lavoro in rete, in particolare alla modalità distribuita, meccanismo che garantisce al tempo stesso un attento controllo di gestione a livello centrale ed una maggiore consapevolezza da parte dei centri produttori di servizi. Tali centri, che in sanità con la nuova riforma sono costituiti da imprese operanti sia nel settore pubblico che nel settore privato, debbono infatti poter assicurare livelli quali-quantitativi in grado di coniugare al tempo stesso necessità di economicità del servizio sanitario e bisogni sanitari della popolazione.

In effetti, contrariamente ad ogni ottimistica presentazione tecnica, la modalità di raccolta ed analisi dei dati a livello regionale resta ancora l'accesso in rete telematica tradizionale o addirittura la consegna dei *floppy disk*, con un accentramento su un nodo centrale dei dati ospedalieri che ordinariamente

porta ad analisi a scadenze periodiche, generalmente annuali.

Il modello sviluppato presso l'USSI parte dal principio della condivisione per proporre un nuovo modello di rete distribuita che in sanità può portare grossi vantaggi, soprattutto se inserito in un contesto integrato tra i vari servizi ed i vari aspetti, secondo una modalità *multilivello, polifunzionale, statistica, condivisa*.

Il primo prototipo realizzato, *DRG Analyst*, ha rappresentato un modello di interazione tra operatori del settore sanitario, statistici ed epidemiologi, ed ha condotto alla realizzazione di un libro ed altre pubblicazioni originali<sup>4,5,6</sup> nelle quali sono state formulate alcune importanti ipotesi di lavoro per progetti futuri. Inoltre, alcune innovazioni in statistica sono state proposte sia nel campo degli indici comparativi di attività ospedaliera che nel campo delle implementazioni grafiche. Alcuni esempi sono presentati nei paragrafi seguenti.

Recentemente, sono state presentate nuove idee circa implementazioni sperimentali, basate su reti distribuite, che partono dai concetti precedentemente sviluppati estendendoli a meccanismi di *meta-apprendimento* su grandi data-base (*data-mining*). Gli ambiti in cui sviluppare tali sistemi per la ricerca sui servizi sanitari costituiscono l'oggetto del presente articolo.

#### **Rete Informativa Condivisa Atto Primo: la creazione di DRG Analyst**

Al fine di poter consentire una analisi ed una rappresentazione delle attività di un sistema sanitario regionale, nel 1996 il DFCE (*F. Carinci, V. Lepore, F. Vitullo*) ha realizzato un primo prototipo di sistema informativo, scritto in linguaggio SAS<sup>®</sup>, moduli BASE, macro linguaggio, STAT, GRAPH ed IML, basato su un piccolo insieme di caratteristiche estratte dalla SDO.

Il progetto ha visto la partecipazione dell'Assessorato Sanità della Regione Abruzzo e la costituzione del Gruppo di Coordinamento Regionale per l'implementazione dell'Osservatorio Epidemiologico, alla cui attività ha partecipato un folto gruppo di operatori appartenenti a 21 presidi ospedalieri abruzzesi.

Per evidenziare l'ambito nel quale il sistema doveva operare, il software è stato denominato semplicemente "*DRG Analyst*": un appellativo che al contempo chiariva i limiti delle finalità poste dal progetto in questa prima iniziativa, sottolineandone comunque il contesto attuale e gli specifici utilizzi.

La raccolta dei dati relativi al primo semestre del 1995, per complessive 111.089 schede provenienti da 21 ospedali, ha portato ad esaminare numerosi aspetti del sistema sanitario regionale, sebbene sia

stato analizzato un insieme di dati composto da sole sette caratteristiche presenti nella SDO.

*Il "magazzino" di base per i dati sanitari di popolazione*

La figura 1 rappresenta le caratteristiche dei dati raccolti e gli archivi utilizzati per completare il sistema in base alle analisi progettate.

Un primo insieme di dati è rappresentato da archivi contenenti disposizioni "ministeriali" relativamente a DRG (descrizione, MDC, tipo, soglia, peso relativo, degenza media del progetto ROD e relativo sistema di tariffe), discipline e codici di reparto (codice, descrizione).

Un secondo nucleo comprendeva dati territoriali, riguardanti strutture (ASL, ex-USL, ospedali e reparti con relativi codici e numero posti letto) e comuni (coordinate di digitalizzazione, popolazione, ASL ed ex-USL di appartenenza).

L'archivio fondamentale, denominato SDO, era rappresentato dall'insieme dei files di data-base ospedalieri, comprendente un sottoinsieme di informazioni sui pazienti dimessi in termini di comune di residenza, ospedale di dimissione, reparto di dimissione, DRG, durata degenza, regime ricovero e modalità di dimissione.

*Prodotti per il controllo di gestione: la combinazione di reports aziendali e territoriali, le rappresentazioni grafiche per il supporto decisionale*

Le funzionalità del software realizzato risentivano delle necessità maggiormente avvertite al momento della raccolta dei dati (ricordiamo che in nes-

suna regione meridionale erano all'epoca state predisposte le SDO). Quali erano le urgenze all'epoca di questo progetto, ovvero alla fine del 1995?

La prima necessità di ogni istituzione era quella di disporre di un *report* riguardo l'attività ospedaliera. Questa esigenza, che in un primo momento non era stata soddisfatta da tutti i presidi, è oggi entrata nelle operazioni di routine per la quasi totalità dei centri partecipanti, con un leggero ritardo peraltro prevedibile.

Malgrado ciò, non siamo a conoscenza di un sistema informativo in grado di produrre *in tempo reale* un prospetto *cumulativo* regionale della attività ospedaliera.

DRG Analyst è stato progettato per la produzione di prospetti di questo tipo che possano essere costruiti "su misura", ovvero selezionando, con delle opzioni appropriate, un sottoinsieme di pazienti, di reparti, di ospedali, di comuni di provenienza e quant'altro secondo le intenzioni dell'utente. In tal modo viene a configurarsi un sistema di monitoraggio dell'intera casistica regionale "per problemi" che possono essere volta per volta esaminati.

Nel realizzare i supporti di lavoro, è stato fondamentale utilizzare SAS®, in particolare i moduli BASE, STAT e GRAPH, e un tipo di implementazione capace di sfruttare al massimo le peculiarità del linguaggio macro.

Per ogni unità (o insieme di unità) selezionata è quindi possibile produrre un tabulato di base relativo ai singoli DRG riguardo a numero assoluto e percentuale di pazienti dimessi per regime di ricovero, rango del DRG, percentuale cumulativa, degenza media, degenza media troncata, coefficiente di variazione (CV), numero assoluto e percentuale di di-

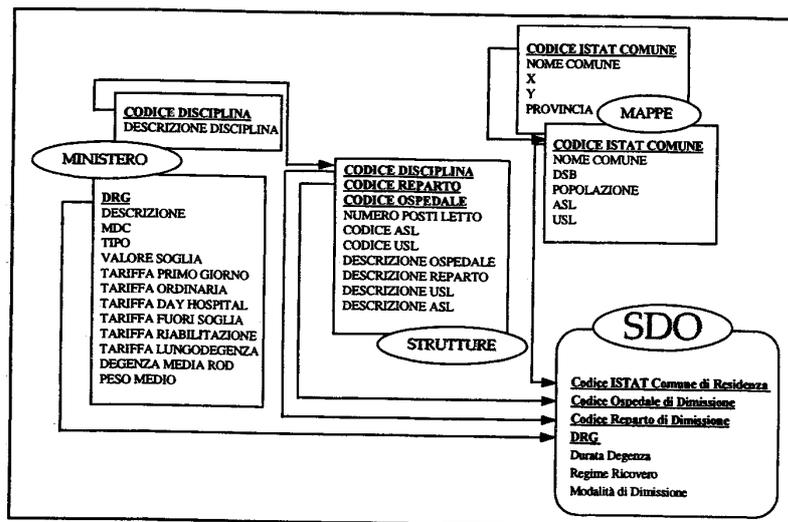


Figura 1 - DRG Analyst: contenuto degli archivi e relazioni.

missioni fuori soglia ed al di sotto di soglie pre-specificate (2-3 giorni), fatturato, etc. Tra le molte elaborazioni proposte, è di particolare interesse oggi quella riguardante il numero dei ricoveri ripetuti, ovvero un nuovo parametro denominato *fattore di unicità*, misura sulla quale implicitamente numerose regioni hanno improntato diverse strategie di tariffazione.

Oltre alla copiosa produzione di reports possibili, il software ha consentito l'utilizzo combinato di rappresentazioni grafiche particolarmente utili in questo campo, quali gli *istogrammi a pila asimmetrici*, i *boxplots* confrontati con uno standard, i *grafici a bolla* e le *mappe geografiche* di offerta di prestazioni.

*Prodotti per l'epidemiologia e la ricerca sui servizi sanitari: cartografia, grafici di dispersione, starplots*

Le potenzialità di questi risultati per finalità di riorganizzazione delle strutture dal punto di vista dell'efficienza aziendale sono evidenti per quanti hanno analizzato la nuova riforma ed affrontato testi specialistici. Meno evidente è come impiegare le sintesi descrittive fornite dalla statistica per finalità maggiormente vicine alla definizione dei bisogni sanitari, secondo quelli che sono gli interessi dell'epidemiologia.

L'epidemiologia dimostra molta attenzione per tutte le componenti che influenzano il profilo di cura del paziente. Da questo punto di vista i tassi di ospedalizzazione, la variabilità delle degenze e delle procedure e tutti gli aspetti legati alla pratica assistenziale possono portare a conclusioni utili tramite l'esplorazione della scheda di dimissione ospedaliera.

Lo studio della variabilità geografica delle caratteristiche per DRG può evidenziare problemi e suggerire soluzioni altrimenti non individuabili: in questo caso abbiamo ripartito semplicemente i dati dei pazienti dimessi in base ai comuni di provenienza dei pazienti stessi ed analizzato le singole componenti separatamente.

Il percorso svolto risponde ad una esigenza estremamente attuale nel campo della ricerca sui servizi sanitari: un ambito di ricerca in graduale espansione è quello della analisi della variabilità geografica per piccole aree (*small area variation analysis*)<sup>7,8</sup>: sotto questa etichetta si riuniscono una serie di tecniche statistiche che, fondendo informazioni socio-economico-sanitarie per ogni cellula areale (contee, comuni, aree di codice postale) con i dati di dimissione ospedaliera ed in alcuni casi di altre realtà assistenziali (medicina di base, assistenza domiciliare), mira alla descrizione geografica, alla ricerca di eventuali *patterns* che possano evidenziare dei problemi per particolari aree, allo stu-

dio delle associazioni tra le caratteristiche di ogni area e le componenti della domanda e dell'offerta di servizi sanitari.

Tali dati sono ottenibili in DRG Analyst scegliendo un insieme di pazienti dimessi nella regione e prendendo come unità di osservazione i comuni scelti singolarmente o raggruppati.

Questo è stato fatto ad esempio per i *tassi di ospedalizzazione* per area (numero di dimissioni, in un determinato periodo di tempo, relative a pazienti provenienti da una determinata zona sul totale degli abitanti residenti in quel comune), in quanto esiste tutta una letteratura sul tema che dimostra come il valore di questi tassi può essere collegato alla pratica clinica; questo fenomeno può inevitabilmente portare alla modificazione della spesa sanitaria, come anche associarsi a comportamenti eterogenei che possono trovare diverse giustificazioni.

Provvedendo alla digitalizzazione in formato SAS<sup>®</sup> dei profili comunali della Regione Abruzzo, è stato possibile rappresentare sotto forma di mappe geografiche i dati relativi alle schede di dimissione ospedaliera. Le peculiarità di SAS<sup>®</sup>/GRAPH consentono quindi di ottenere mappe epidemiologiche per qualsiasi ripartizione territoriale passibile di collegamento con un database di questo tipo.

Variabilità globale e concentrazione geografica hanno offerto indicazioni circa i problemi di salute o eventualmente di assistenza in un territorio, anche se alcune cautele debbono essere osservate nella interpretazione dei dati.

In ogni mappa, le colorazioni utilizzate sono state riferite a categorie create sulla base dei tassi, nel nostro caso calcolati a livello comunale, raggruppati secondo gradazioni che permettono di individuare zone chiare, a bassa intensità, e zone più scure, ad alta intensità.

La scelta delle classi può essere fatta fondamentalmente in due modi: soggettivamente, qualora si conosca approfonditamente il fenomeno esaminato, ed oggettivamente, quando è un metodo statistico a guidare la classificazione.

Le opzioni implementate nella nostra macro EPIMAP, inclusa in DRG Analyst, permettono di scegliere tra: 1) definire le classi numericamente *a priori*; 2) creare le classi per *percentili* della distribuzione della caratteristica in esame; 3) lanciare una procedura di *cluster analysis* aggregativa come quella precedentemente descritta (metodo dei centroidi) specificando il numero finale di classi da ottenere. Le mappe presentate nel libro da noi redatto<sup>4</sup> sono state ottenute seguendo il terzo approccio (per motivi di spazio si rimanda al volume per le figure).

Dal punto di vista della valutazione della attività comparativa di reparti ed ospedali, è risultato particolarmente importante riassumere i risultati in termini di indici classici quali la *degenza media*, la *presenza media*, il *tasso di saturazione*, l'*indice di*

rotazione, l'intervallo di turnover e l'ammontare di lire per dimesso e lire per punto DRG<sup>4</sup>. Tali indici sono stati rappresentati sia in forma tabulata che sotto forma di grafico di dispersione, nel quale sono evidenziate medie ed intervalli di confidenza del sistema regionale posto come standard.

L'impatto maggiore per l'indagine sui servizi sanitari è stato però raggiunto dagli *starplot*, figure create per rappresentare a livello multidimensionale il confronto tra reparti o ospedali. Questa implementazione si è rivelata principalmente per due ordini di motivi:

a) la capacità di rappresentare indici basati su diverse composizioni di pazienti per diverse strutture. In questo ambito è infatti fondamentale tener conto, nei confronti, delle diverse tipologie di pazienti in termini di distribuzione di frequenza per DRG nelle varie strutture. Questo aspetto, che sembra sottovalutato in letteratura, è diventato necessario in una implementazione in rete in

quanto la molteplicità dei confronti esige la correttezza del confronto tra realtà eterogenee.

b) la capacità di rappresentare dinamicamente le informazioni, proponendo forme che vanno ad animarsi sulla base dei diversi standard identificati.

La figura 2 evidenzia un esempio di questo tipo di confronti; ogni asse va interpretato secondo la legenda della figura 3.

**Dalla analisi cumulativa delle SDO alla condivisione ed alla meta-analisi: RISS-H, il Sistema Informativo Condiviso, l'approccio Touch+Pile-Up+meta-Analyze (TPA)**

Nella prima versione del programma di rete, DRG Analyst poteva solo operare in modalità "simulata", ovvero esisteva la possibilità "virtuale" di elaborare i dati da un sito periferico, ma in realtà i dati erano centralizzati, ovvero risiedevano in una area della

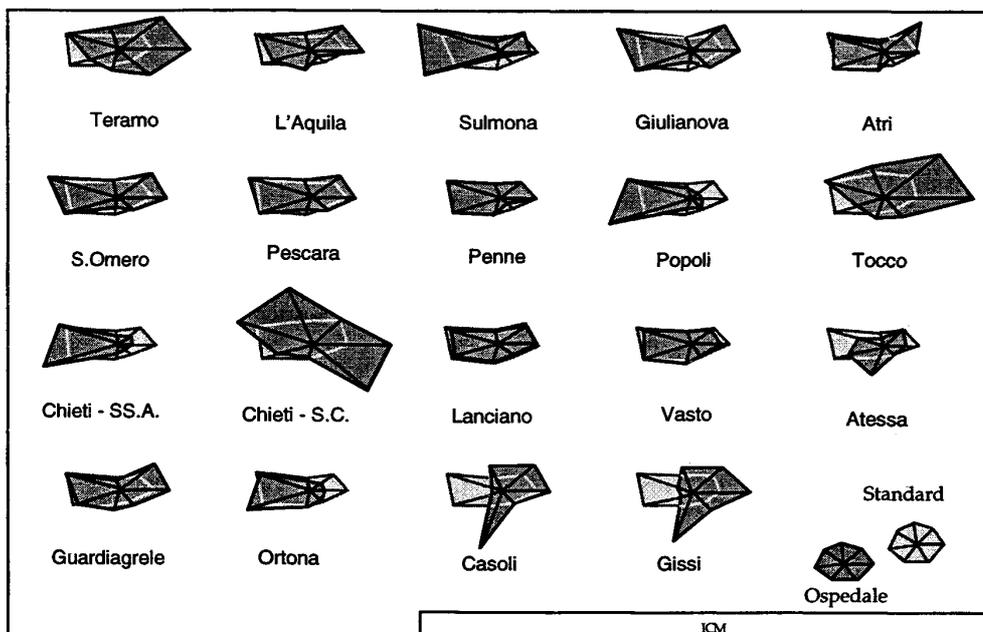


Figura 2. Esempio di applicazione dello starplot

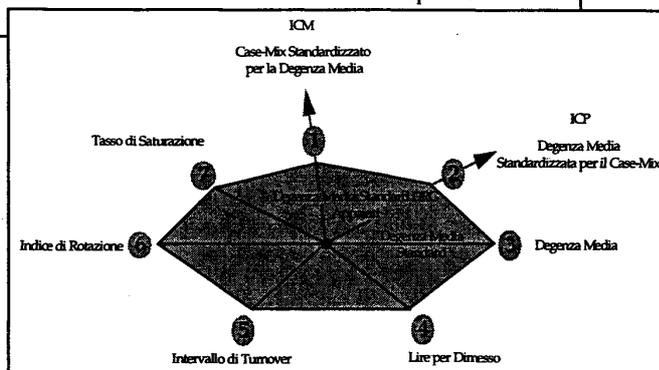


Figura 3. Legenda dello starplot

memoria di massa della macchina AXP residente presso il DFCE. Il passaggio dalla logica di DRG Analyst alla logica condivisa esige la formulazione di un metodo "distribuito", che potesse appunto permettere la allocazione dei files a livello periferico e quindi la analisi dei dati "frammentata".

Il lavoro del periodo successivo alla prima elaborazione dei dati ci ha quindi visti impegnati a realizzare un sistema alternativo in grado di soddisfare le esigenze di un nuovo progetto, successivamente denominato RISS-H (*F. Carinci*). La tecnica impiegata per la realizzazione del nuovo software è stata così definita approccio "Touch+Pile-Up+meta-Analyze (TPA)", in quanto i passi da compiere per completare il processo corrispondono a tre ben distinte procedure (la tecnica è descritta in figura 4).

L'analista (rappresentato da un cuore), identifica l'obiettivo del problema nella realizzazione di una determinata struttura, che definiamo "oggetto di sintesi". Tale oggetto di sintesi, ad esempio, può essere la mediana di una determinata misura, diciamo la mediana delle giornate di degenza ( $m$ ). Ora ipotizziamo che tale mediana, data la mole dei database utilizzati a livello periferico, non sia calcolabile immediatamente utilizzando tutti i dati dei differenti archivi, ma debba essere ottenuta tramite accessi separati per ogni sito. Dopo aver realizzato un contatto (*Touch*) per ogni singolo nodo, a partire da quello con numerosità  $N_1$ , per concludere l'elaborazione deve essere individuata una costruzione logico-matematica in grado di portare ad una analisi tendente alla misura obiettivo, la mediana appunto. Dobbiamo quindi passare alla procedura *Pile-Up*, che consiste nel trovare il metodo migliore per giungere ad una mediana globale e che quindi cumula i risultati del contatto corrente con i contatti precedenti; in questo caso, molto semplice, tale procedura consiste nel calcolo di distribuzioni di frequenze per giornate di degenza (PROC SORT+PROC FREQ in SAS®).

Tale processo si itera per il nodo di numerosità  $N_2$ , finchè si giunge al ritorno al sito di partenza, nel

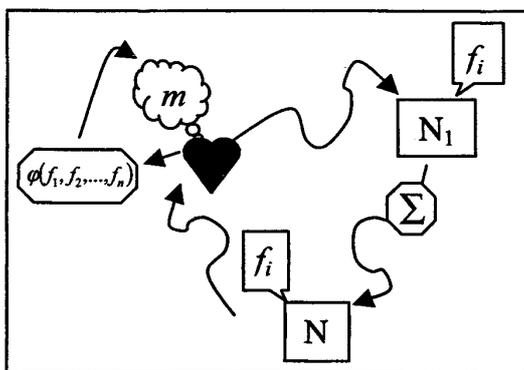


Figura 4. Tecnica Touch+Pile-Up+Meta-Analyze (TPA)

quale si utilizza la procedura di meta-analisi (*meta-Analyze*).

Nei casi più semplici non subentra una approssimazione stocastica in tale processo, ovvero non si eseguono procedure che portano a stime probabilistiche circa la esattezza dell'oggetto di sintesi ricercato. Nel caso della mediana, cioè, si ripropone il modello di calcolo tradizionale (50% dei pazienti nella distribuzione cumulata, ricostruita a valle come somma delle tabelle dai singoli siti).

Un aspetto ancor più interessante è la possibilità di attuare strategie di contatto che prevedano, per oggetti di sintesi a carattere probabilistico, l'utilizzo di analisi statistiche quali l'analisi multivariata (nel caso di indagini epidemiologiche la regressione logistica, ad esempio).

Si pensi all'interesse rivestito dall'esplicitazione *in tempo reale* del rischio relativo (*odds ratio*), per un determinato paziente con determinate caratteristiche, di incorrere in un intervento chirurgico in un sistema sanitario di riferimento prescelto (ASL, tipo di ospedale, regione), rispetto ad un altro paziente nello standard. La analogia con il campo di ricerca della meta-analisi è quindi evidente.

Ancora, si pensi alla rilevanza della possibilità di individuare *in tempo reale* nuove tipologie di pazienti e nuove classificazioni in generale (DRG corretti, nuovi sistemi di tariffe regionali). Per tali scopi sono applicabili, tra le varie tecniche disponibili, i metodi multivariati ad alberi di regressione (l'approccio RECPAM<sup>9,10</sup>, in particolare) approcci dotati di grandi potenzialità nell'ambito epidemiologico, i cui possibili sviluppi sono tuttora in fase di prova e saranno proposti come ulteriore modulo di approfondimento del sistema RISS-H.

In definitiva, nell'ambito dei nuovi progetti presentati dall'USSI, si stanno proponendo orizzonti di sviluppo notevoli che certamente influiranno sia sulla implementazione del software necessario a svolgere le analisi in questione, sia sulla individuazione di nuove tecniche costruite *ad hoc* per risolvere i problemi derivanti dalla messa a disposizione di enormi masse di dati e dal collegamento in rete.

Ad alleviare il peso della complessità per l'utente di strumenti statistici sempre più sofisticati, una parte rilevante del progetto RISS-H è stata programmata per lo studio di interfacce utente semplificate, oltre che per l'impiego di modalità di collegamento ottimali. Tali operazioni verranno svolte attraverso l'utilizzo dei moduli SAS®/AF e CONNECT su piattaforme miste, ovvero workstation SUN/SOLARIS® e PC.

Un'altra caratteristica interessante del progetto RISS-H è la compatibilità con i sistemi pre-esistenti all'interno delle strutture sanitarie. I siti periferici, infatti, conosceranno i vantaggi di un protocollo comune con il centro di coordinamento, basato sull'impiego dei moduli SAS® visuali e di connessione,

non vincolandosi necessariamente ad alcun software specifico per la gestione dei dati.

Gli eventuali aderenti al modello RISS-H, in definitiva, manterranno una architettura client-server complessiva che riguarderà più la gestione del codice e le elaborazioni statistiche, che il controllo dei data-base: questo perché risulta indispensabile, nel contesto della gestione dei dati sanitari, rispettare conoscenze e strutture esistenti, per non incorrere in difficoltà di apprendimento e di approvazione di nuovi supporti che porterebbero ad un rigetto immediato.

Al fine di poter esplicitare un piccolo esempio dell'utilizzo della modalità di lavoro TPA, presentiamo una piccola macro, allegata in appendice (%INDEXES), che mostra le procedure di calcolo necessarie a produrre il listato di RISS-H riguardante due indici fondamentali nella valutazione comparativa della attività ospedaliera: *l'indice di case-mix e l'indice comparativo di performance*. Per le formule e l'interpretazione dei risultati si consulti il testo di riferimento già citato<sup>4</sup>.

Il codice riportato consente di esercitarsi direttamente con la presente macro su alcune analisi comparative dell'attività ospedaliera. Per citarne un possibile utilizzo, nell'ambito di uno studio recente<sup>11</sup>, si sono valutate le performance dei singoli reparti di chirurgia di 5 ospedali abruzzesi, si sono elaborati gli indici e sono state valutate le performance in termini di giornate di degenza medie per ognuno dei singoli reparti contro lo standard costituito dall'unione dei singoli reparti. Risulta chiaro che in linea teorica è possibile identificare tanti standard quante sono le possibili combinazioni: è quindi interesse del singolo selezionare solo quelle dotate di un minimo di interpretabilità rispetto agli obiettivi e, soprattutto, alle ipotesi preposte.

## Conclusioni

Le tecnologie telematiche mettono a disposizione della statistica importanti strumenti per riassumere enormi quantità di dati. Lo sviluppo continuo, se da un lato offre delle opportunità, propone alla statistica una continua sfida alla creazione di metodi che possano essere in grado di sintetizzare in maniera efficiente le informazioni disponibili.

Questa condizione impone però alle scienze statistiche l'esigenza di una rifondazione se non totale, sicuramente sostanziale, nel senso indicato dalla intrinseca *operatività in network* dei metodi stessi. In questo caso non si tratta di una semplice "analisi in modalità web", ma di una imponente ridefinizione dei metodi, da connettere sinergicamente alle tecniche del *data-mining*, del *knowledge discovery*, dei *neural networks*.

In tale ambito, il software SAS<sup>®</sup> è di sicuro aiuto come piattaforma di implementazione, per motivi

rilevanti: possiede procedure statistiche sofisticate, funzionalità di data-base e di connessione in rete, è multiplatforma.

Al fine di ottenere un successo tangibile nella applicazione pratica di tali innovazioni nel campo sanitario, è importante sottolineare l'esigenza di un grosso sforzo di coesione tra *software house*, forze della ricerca scientifica, figure esperte nei singoli ambiti di applicazione, e soprattutto utenti finali, veri protagonisti in questa epoca che per la prima volta riconoscerà un ruolo importante al livello periferico.

L'ambito del sistema sanitario riveste un livello di priorità tale da porlo sicuramente tra i primi settori destinati ad usufruire del cambiamento in atto, ed in questo scenario l'introduzione di sistemi quali RISS-H può costituire una valida opportunità per lo sviluppo di nuovi modelli operativi.

## Bibliografia

- 1) Fetter R, Shin Y, Freeman J, Averill R, Thompson J. Case-Mix definition by diagnosis related groups. *Med Care* 1980; 18 (Suppl.): 2.
- 2) Fetter RB, Brand DA, Gamache D. DRGs their design and development, Michigan: Ann Arbor, 1991.
- 3) Casale C., Labbrozzi D., Nicolucci A., Carinci F., Avanzi C., Dell'Aquila R., Forcella M., Montemurno C., Procaccini D., Ruscitto F., Spada S., Stella I. Evaluation of the Efficiency of a Nephrology Department by the DRGs and a Barber Nomogram: the role of comorbidity; *Contributions to Nephrology*, Basel, Karger, Vol. 109, 84-89, 1994.
- 4) Vitullo F., Carinci F., Lepore V., Tognoni G., *Aziende Sanitarie e Modelli di Uso dei DRG*, Il Pensiero Scientifico Editore 1997.
- 5) Vitullo F., Carinci F., Valerio M., *L'Oncologia in Abruzzo. Atlante delle ospedalizzazioni interne e delle migrazioni sanitarie. Primo semestre 1995*, Ed. Consorzio Mario Negri Sud - Regione Abruzzo
- 6) Carinci F., SDO Abruzzo, *Convegno 'Dai data base ospedalieri all'osservatorio epidemiologico: opportunità per il Sistema Sanitario Regionale'*, Pescara, 5 luglio 1996.
- 7) Diehr P, Cain K, Kreuter W, Rosenkranz S. Can small-area analysis detect variation in surgery rates? The power of small-area variation analysis. *Med Care* 1992; 30: 6.
- 8) Diehr P, Cain K, Ye Z, Abdul-Salam. Small area variation analysis, Methods for comparing several diagnosis related groups. *Med Care* 1993; 31(5): YS45-YS53.
- 9) Carinci F., Nicolucci A., Ciampi A., Labbrozzi D., Bettinardi O., Zotti A.M., Tognoni G. on behalf of the GISSI Investigators, Role of Interactions between Psychological and Clinical Factors in determining 6-Month Mortality among patients with Acute Myocardial Infarction. Application of recursive partitioning techniques to the GISSI-2 Data-Base, *European Heart Journal*, 18:835-845, 1997.
- 10) Ciampi A. Constructing Predictions Trees from Data: the RECPAM approach, in *Computational Aspects of Model Choice*, Physica-Verlag., 1994.
- 11) Menna A. I DRGs come strumento di valutazione comparativa della attività ospedaliera. Studio su alcuni database di ospedali in Abruzzo, *Tesi di Laurea*, Facoltà di Scienze Statistiche Demografiche e Sociali, Università La Sapienza, Roma, Maggio 1997.

## APPENDICE

```

/*****
Macro: INDEXES
Autori: Fabrizio Carinci, Angelo Menna
Data: Ottobre 1997
-----

```

Commento: Programma SAS che elabora gli indici ICM e ICP per il confronto fra le attivita' di uno o piu' reparti (ospedali) di interesse e uno o piu' reparti (ospedali) di riferimento (standard).  
La costruzione degli indicatori e' basata sulla distribuzione di pazienti e giorni di degenza DRG-specifici.

(in: Aziende Sanitarie e Modelli di Uso dei DRG, F.Vitullo, F.Carinci, V.Lepore, G.Tognoni, Il Pensiero Scientifico Editore, Roma, Febbraio 1997)

```

-----
Unita' di Statistica e Sistemi Informativi
Consorzio Mario Negri Sud
S.Maria Imbaro (CH) - ITALY
*****
OPTIONS LS=80 PS=500 NOMPRINT NOMTRACE
NOMLOGIC NOSYMBOLGEN NOSOURCE NOSOURCE2 NONUMBER NODATE
NOMOTES FORMDLIM=' ' ;

```

```

%MACRO INDEXES(
libdata = /* Dir. Dati */
base = /* Lista Ospedali */
conditio= /* Condizione */
base_rep= /* Scelta Reparti */
std= /* Lista Ospedali Standard */
std_rep= /* Scelta Reparti Standard */
n_cod=4 /* Num. cifre signif. reparti */
n_codstd=4 /* Num. cifre signif. rep. Standard */
ind= /* Indici da calcolare */
labelind= /* Labels per indici */
title= /* Titolo Listato */
);

```

```

%MACRO WORDS(string);
%local count word;
%let count=1;
%let word=%qscan(&string,&count,%str( ));
%do %while(%word ne);
%let count=%eval(&count+1);
%let word=%qscan(&string,&count,%str( ));
%end;
%eval(&count-1)
%MEND WORDS;

%local i inbase instd n_base nbaserep n_std nstdrep;
%if %bquote(&base)^= %then %let n_base=%WORDS(&base);
%if %bquote(&std)^= %then %let n_std=%WORDS(&std);
%if %bquote(&base_rep)^= %then %let nbaserep=%WORDS(&base_rep);
%if %bquote(&std_rep)^= %then %let nstdrep=%WORDS(&std_rep);
%if %bquote(&ind)^= %then %let n_ind=%WORDS(&ind);
%do i=1 %to %nbaserep;
%if %i>1 %then
%let inbase=%inbase,"%SCAN(&base_rep,%i)";
%else %let inbase=%SCAN(&base_rep,%i);
%end;
%do i=1 %to %nstdrep;
%if %i>1 %then %let instd=%instd,"%SCAN(&std_rep,%i)";
%else %let instd=%SCAN(&std_rep,%i);
%end;
%do i=1 %to %n_base;
FILENAME SDO "%libdata%SCAN(&base,%i)_SDO.CSV";
DATA SDO (DROP=LOSPRE LOSPOST TIPO);
LENGTH DIM_REP $4 LOSPRE 3 LOS 3
LOSPOST 3 DRG 3 TIPO $1;
INFILE SDO DELIMITER=';' DSD MISSEVER;
INPUT DIM_REP LOSPRE LOS LOSPOST DRG TIPO;
IF SUBSTR(TRIM(LEFT(DIM_REP)),1,&n_cod) IN (&inbase);
%if %bquote(&conditio)^= %then %STR(IF &conditio);
RUN;
PROC SORT;BY DRG;RUN;
PROC FREQ DATA=SDO NOPRINT;
TABLES LOS / OUT=%if %i=1 %then TABNEW;%else TABLOS;;
BY DRG;RUN;
PROC DATASETS NOLIST;DELETE SDO / MEMTYPE=DATA;
QUIT;RUN;
%if %i>1 %then %do;
PROC APPEND DATA=TABLOS BASE=TABNEW;RUN;
%end;
%end;
PROC SORT DATA=TABNEW;BY DRG LOS;RUN;
PROC SUMMARY DATA=TABNEW NOPRINT;
FREQ COUNT;
OUTPUT OUT=BASELOS (DROP=_TYPE_ RENAME=( _FREQ_=COUNT));
BY DRG LOS;RUN;
%do i=1 %to %n_std;
FILENAME SDO "[PROG.DATA]%SCAN(&std,%i)_SDO.CSV";
DATA SDO (DROP=LOSPRE LOSPOST TIPO);
LENGTH DIM_REP $4 LOSPRE 3
LOS 3 LOSPOST 3 DRG 3 TIPO $1;

```

```

INFILE SDO DELIMITER=';' DSD MISSEVER;
INPUT DIM_REP LOSPRE LOS LOSPOST DRG TIPO;
IF SUBSTR(TRIM(LEFT(DIM_REP)),1,&n_codstd) IN (&n_codstd);
RUN;
PROC SORT;BY DRG;RUN;
PROC FREQ DATA=SDO NOPRINT;
TABLES LOS /OUT=%if %i=1 %then TABNEW;%else TABLOS;;
BY DRG;RUN;
PROC DATASETS NOLIST;DELETE SDO / MEMTYPE=DATA;
QUIT;RUN;
%if %i>1 %then %do;
PROC APPEND DATA=TABLOS BASE=TABNEW;RUN;
%end;
%end;
PROC SORT DATA=TABNEW;BY DRG LOS;RUN;
PROC SUMMARY DATA=TABNEW NOPRINT;
FREQ COUNT;
OUTPUT OUT=STDLOS DROP=_TYPE_ RENAME=( _FREQ_=COUNT));
BY DRG LOS;RUN;
DATA INDEXES;
MERGE BASELOS (RENAME=(COUNT=BASE_RIC))
STDLOS (RENAME=(COUNT=STD_RIC));
BY DRG LOS;
IF BASE_RIC=. THEN BASE_RIC=0;
IF STD_RIC=. THEN STD_RIC=0;
IF FIRST.DRG THEN DO;
TOTBASE=0;TOTSTD=0;TOTRICB=0;TOTRICS=0;
END;
/* Totale Giorni di Degenza Base per DRG */
IF BASE_RIC>0 THEN TOTBASE+(LOS*BASE_RIC);
/* Totale Giorni di Degenza Standard per DRG */
IF STD_RIC>0 THEN TOTSTD+(LOS*STD_RIC);
TOTRICB=BASE_RIC; /* Tot.Dimessi Base per DRG */
TOTRICS=STD_RIC; /* Tot.Dimessi Standard per DRG */
/* Analisi per DRG Comuni */
IF LAST.DRG AND TOTRICB>0 AND TOTRICS>0;
RUN;
DATA INDEXES(KEEP=TOTDRG &ind);
SET INDEXES END=LAST;
TOTDRG+1; /* Numero DRG in comune */
TB+TOTRICB; /* Totale Dimessi nella Base */
TS+TOTRICS; /* Totale Dimessi nello Standard */
DEN+TOTSTD; /* Totale Giorni Degenza Standard */
%do i=1 %to %n_ind;
%if %UPCASE(%SCAN(&ind,%i))=ICM %then %do;
/* Tot. Giorni Degenza Stand. per Performance Standard */
NUM_ICM+((TOTSTD/TOTRICS)*TOTRICB);
/* Media GG. Degenza Stand. per Performance Standard */
CMIXSDM=NUM_ICM/TB;
ICM=(NUM_ICM/DEN)*(TS/TB); /* Indice di Case-Mix */
%end;
%else %if %UPCASE(%SCAN(&ind,%i))=ICP %then %do;
NUM_ICP+((TOTBASE/TOTRICB)*TOTRICS);
DMSCMX=NUM_ICP/TS;
ICP=NUM_ICP/DEN;
%end;
%end;
IF LAST;
RUN;
TITLE1 "RISS-H - Rete Informativa
Sistema Sanitario Ospedaliero";
TITLE3 " &title ";
TITLE2 "BASE DATASETS:&base";
PROC PRINT NOOBS UNIFORM LABEL;
VAR TOTDRG &IND;
LABEL TOTDRG="Totale DRG"
%do i=1 %to %n_ind;
%SCAN(&ind,%i)="&SCAN(&labelind,%i,)"
%end;
;
RUN;
%MEND INDEXES;

%MACRO RISS_H;
%local osp i;
%let osp=LA AQ CH PO AT;
%do i=1 %to 5;
%INDEXES(libdata=[RISS.PROG.DATA],
base=%SCAN(&osp,%i),
conditio=TRIM(LEFT(TIPO))="C",
base_rep=09,
std=LA AQ CH PO AT,
std_rep=09,
n_cod=2,
n_codstd=2,
ind=ICM ICP,
labelind=Indice di Casemix /
Indice di Performance,
title=DRG Chirurgici);
%end;
%MEND RISS_H;
%RISS_H;

```