

F. Carinci

Il modello GIPSY-RECPAM: un versatile approccio metodologico per l'analisi integrata del paziente cardiologico

Sereatrix snc, Pescara, Italy

RIASSUNTO. L'analisi per partizioni ricorsive effettuata per il progetto GISSI-PSICOLOGIA risulta essere un modello estremamente attuale da potersi applicare in diverse situazioni nell'ambito di cura cardiologico. La metodologia consiste nell'applicazione congiunta di strumenti di misurazione multifattoriali e nell'impiego di sofisticate tecniche statistiche in grado di consentire una efficace e sintetica rappresentazione clinico-epidemiologica del paziente infartuato. Tali caratteristiche offrono la possibilità di applicare il modello RECPAM in maniera versatile nel contesto di soluzioni online ad ampio contenuto tecnologico. Per il nostro studio, si sono utilizzate le scale psicometriche appartenenti al questionario CBA-H. Le prospettive di utilizzo nel panorama cardiologico italiano odierno e futuro sono pertanto numerose e particolarmente rilevanti nel contesto di cura e di valutazione integrata del paziente cardiologico.

Parole chiave: RECPAM, GISSI-PSICOLOGIA, partizioni ricorsive, analisi multivariata, alberi di regressione, psicomatria.

ABSTRACT. *THE GIPSY-RECPAM MODEL: A VERSATILE APPROACH FOR INTEGRATED EVALUATION IN CARDIOLOGIC CARE. Tree-structured methodology applied for the GISSI-PSICOLOGIA project, although performed in the framework of earliest GISSI studies, represents a powerful tool to analyze different aspects of cardiologic care. The GISSI-PSICOLOGIA project has delivered a novel methodology based on the joint application of psychometric tools and sophisticated statistical techniques. Its prospective use could allow building effective epidemiological models relevant to the prognosis of the cardiologic patient. The various features of the RECPAM method allow a versatile use in the framework of modern e-health projects. The study used the Cognitive Behavioral Assessment H Form (CBA-H) psychometrics scales. The potential for its future application in the framework of Italian cardiology is relevant and particularly indicated to assist planning of systems for integrated care and routine evaluation of the cardiologic patient.*

Key words: RECPAM, GISSI-PSICOLOGIA, recursive partitioning, multivariate analysis, regression trees, psychometrics.

Premessa

Le accresciute possibilità di collezionare ed elaborare database di enormi dimensioni hanno determinato una drammatica crescita nella gamma di applicazioni per i più vari scopi sia di ricerca che applicativi. Come noto, in questo ambito la cardiologia italiana rappresenta un modello internazionale di assoluto rilievo.

Anche grazie a questi contributi, l'epidemiologia clinica ha subito una notevole evoluzione e diversificazione, comprendendo oggi ambiti che spaziano dalla ricerca sugli esiti, alla ricerca sui servizi sanitari, passando per molte altre specializzazioni nelle quali si utilizzano strumenti e metodologie sofisticate accomunate da una notevole solidità scientifica ed una accresciuta efficienza operativa.

Nonostante ciò, occorre sottolineare come ancora oggi, particolarmente in ambito clinico cardiologico, sia ancora forte la necessità di strumenti che possano consentire di utilizzare al meglio le conoscenze epidemiologiche acquisite dal punto di vista pratico.

Il problema diventa particolarmente rilevante in contesti complessi, quali quelli tipici degli interventi di riabilitazione, in cui numerosi fattori, oltre a quelli clinici, concorrono a definire la prognosi del paziente.

In tale ambito, il sottoprogetto GISSI-2 denominato GISSI-PSICOLOGIA (GIPSY) (1, 2) ha rappresentato un esempio importante di interazione multidisciplinare il cui rilievo metodologico merita ulteriore approfondimento.

Il presente articolo rivisita la metodologia utilizzata nel GIPSY per poterne facilitare un suo utilizzo in nuovi studi cardiologici, ed in generale in tutti i casi in cui la gamma delle dimensioni esaminate risulta particolarmente complessa e difficile da analizzare.

In particolare, nell'articolo si illustreranno i principi e le modalità d'uso di tecniche statistiche finalizzate alla valutazione integrata dell'impatto combinato di caratteristiche cliniche e psicologiche.

Infine, sulla base delle recenti innovazioni tecnologiche, se ne indicheranno le prospettive in un'ottica di telemedicina ed in generale nel campo dell'*e-health*.

Lo studio GIPSY

Lo studio in oggetto (1, 3) mirava ad identificare il ruolo dei fattori psicologici nella sopravvivenza a 6 mesi di una coorte di pazienti infartuati.

La rilevazione psicométrica è stata effettuata tramite il questionario CBA-H (4-6).

L'analisi prevedeva in prima istanza l'applicazione di tecniche multivariate standard quali il modello di Cox, ed in particolare l'applicazione dei modelli di Cox del tipo univariato e multivariato con correzione per le variabili cliniche.

A tale riguardo, uno dei problemi maggiori consisteva nella ricerca simultanea dei cut-off appropriati per le variabili incluse in modelli ad effetti indipendenti, e quindi la ricerca delle interazioni significative, a più livelli, tra le variabili, con relativa identificazione di tutti i cut-offs relativi a tali interazioni.

Alcuni dei problemi evidenziati non erano stati risolti utilizzando le variabili ed i relativi cut-off di rischio cardiovascolare (3) e psicologico (4, 5).

Le tabelle I-II riassumono i risultati delle analisi univariate e multivariate. In entrambi i casi non si evidenziava alcun ruolo predittivo per le variabili psicologiche esaminate.

Il risultato destava alcuni interrogativi metodologici.

In particolare, un possibile effetto di mascheramento puramente "statistico" in cui le variabili cliniche, con il loro peso rilevante in fase acuta, coprisse l'impatto di profili psicologici ad alto rischio.

Tale situazione è in effetti comune in situazioni in cui l'analisi si limita all'osservazione degli effetti principali, penalizzando di fatto il ruolo di variabili che anziché avere un ruolo predittivo su tutta la popolazione, potrebbero giocare un ruolo solo in alcuni strati specifici.

Una strategia alternativa consiste nell'analisi 'per patterns', ovvero nella valutazione dei rischi per sottogruppi di variabili e per sottogruppi di osservazioni, tramite un utilizzo combinato di analisi esplorativa ed analisi multivariata.

Il metodo RECPAM

La tecnica RECPAM (RECURSIVE Partitioning and AMalgamation) (7,8) è una generalizzazione dei metodi del tipo CART (9) che consente di applicare modelli di regressione ad albero in situazioni estremamente frequenti in biostatistica ed epidemiologia clinica. Tale strumento è estremamente utile nell'attuare una stratificazione prognostica (10), ovvero per segmentare popolazioni di pazienti sulla base di semplici criteri clinici, sulla base di outputs facilmente leggibili in quanto interpretabili secondo la stessa logica di tecniche standard quali il modello di Cox e la regressione logistica.

Il caso dello studio GIPSY offre una brillante presentazione del funzionamento e dei vantaggi di RECPAM nelle sue varie modalità.

L'algoritmo parte analizzando l'intera popolazione di pazienti GIPSY, cercando di ripartire il campione in due gruppi in modo tale da massimizzare la differenza in ter-

Tabella I. Variabili psicologiche: modelli univariati

Variabile	HR	P> χ^2	I.C. 95%	
Ansia	1.289	0.32	0.784	2.119
Paure Sanitarie	1.161	0.56	0.701	1.923
Depressione	1.705	0.08	0.940	3.092
Depressione nei 3 mesi precedenti	1.228	0.42	0.748	2.016
Benessere nei 3 mesi precedenti	1.420	0.17	0.865	2.331
Stress nei 3 mesi precedenti	1.255	0.39	0.751	2.096
Nevroticismo	1.202	0.47	0.732	1.973
Introversione	0.653	0.10	0.394	1.082
Ansia Sociale	1.178	0.65	0.582	2.387
Sensi di Colpa	1.089	0.74	0.661	1.793
Comportamento di Tipo A	1.415	0.18	0.847	2.363

Tabella II. Modello multivariato ristretto (backward elimination)

Variabile	HR	P> χ^2	I.C. 95%	
Test da sforzo controindicato	3.759	0.0001	2.167	6.523
Scompenso cardiaco tardivo	3.391	0.0001	1.847	6.224
Scompenso cardiaco precoce	2.203	0.0129	1.182	4.106
Fattori di Rischio Cardiovascolare	2.037	0.0132	1.161	3.573

mini del criterio indicato: in questo caso il rischio di mortalità a 6 mesi secondo il modello di Cox.

Specificando il modello di Cox, l'input prevede quindi la scelta di una variabile temporale, di una variabile indicatrice di censoring, le variabili di aggiustamento, o *globali*, e l'elenco delle *covariate di albero*. Per il nostro studio, queste ultime erano quindi le scale psicométriche appartenenti al nucleo centrale (sezioni A,B e C) del questionario CBA-H.

L'algoritmo parte con l'analizzare la prima covariata per ogni ipotetica suddivisione in due classi, tramite quella che viene definita come "split defining question".

Ad esempio, la variabile ansia, che è di tipo ordinale, viene testata per una partizione binaria che prevede da una parte il sottogruppo con livello 0, dall'altra quello con livello maggiore di zero. Quindi, si produce il modello di Cox e si utilizza come *misura di informazione* di questo *cut-off* la statistica del *likelihood ratio* (LR) (7) ottenuta dal tradizionale test del confronto di ipotesi:

$H_0: \beta_1=0$, ovvero i due gruppi non differiscono per l'evento mortalità a 6 mesi

$H_1: \beta_1 \neq 0$, ovvero i due gruppi differiscono per l'evento mortalità a 6 mesi

Da un punto di vista matematico, il LR viene calcolato come differenza del *log likelihood* del modello che include la nuova variabile binaria (a valori 0,1) derivata dalla dimensione ansia, e del modello ristretto che non la include.

Tabella III. Large Tree Partitioning: ricerca del primo split

Searching First Split through 1-2449	
Depressione 3 mesi precedenti	3.56
Benessere 3 mesi precedenti	3.50
Depressione	3.29
Stress 3 mesi precedenti	2.88
Estroversione	2.80
Sensi di colpa	2.10
Comportamento di Tipo A	1.88
Nevroticismo	1.74
Ansia	0.99
Paure	0.92
Ansia sociale	0.19

Il processo viene iterato per tutte le possibili partizioni binarie, e si memorizza il livello della variabile ansia che ha fatto registrare il valore più alto di LR, per poi passare ad esaminare la seconda variabile.

Nel caso del GIPSY, RECPAM analizzava al primo passaggio il contenuto informativo delle variabili indicate come covariate di albero nella fase di input, portando alla graduatoria in Tabella III.

A livello dell'intero campione, la variabile "depressione nei 3 mesi precedenti", per il particolare livello (>2, ≤ 2) indicato da RECPAM, risultava essere il fattore con il maggiore peso predittivo, ovvero in grado di suddividere il campione in due gruppi contraddistinti dalla più marcata differenza in termini di prognosi infausta nei 6 mesi successivi l'infarto.

Quindi, la prima ripartizione viene fatta in base alla semplice domanda: "il punteggio nella scala di depressione nei 3 mesi precedenti l'infarto è superiore o eguale a 2?".

La procedura ad albero crea ciclicamente nuove partizioni, identificandole in base ad una sola condizione per volta.

La procedura si conclude quindi tramite una regola conclusiva, o "stopping rule" (SR). Nella analisi della sopravvivenza questa consiste nel "event size", ovvero la numerosità minima di eventi presenti nel nodo, e la node size, ovvero la numerosità minima in termini di totale di osservazioni presenti nel nodo.

Lo schema finale assume una caratteristica forma di albero binario (Figura 1).

Affinché un modello con troppi parametri non rischi di perdere in accuratezza (overfitting), alla fase precedente se ne aggiungono altre.

Una seconda fase si occupa del processo di 'semplificazio-

ne' del grande albero, applicando la cosiddetta 'potatura' (pruning). In tal modo si cerca di limitare la ripartizione a gruppi più consistenti, e quindi ad interazioni di ordine inferiore, al fine di garantire una maggiore predittività sui campioni futuri.

Nella 'potatura', due nodi 'figli' provenienti dallo stesso 'padre' vengono consolidati in una sola foglia, secondo il criterio di significatività statistica evidenziato dal principio di Akaike (Akaike Information Criterion, o AIC) (7). L'AIC è in pratica una penalità che viene pagata sulla significatività del LR in ragione del numero di parametri immessi nel modello:

$$AIC = -2 (\log \text{likelihood}) + 2 * (\text{numero di parametri})$$

Allo stesso modo, una terza fase effettua la cosiddetta procedura di amalgamazione (amalgamation), al fine di semplificare il modello ad albero unendo i sottogruppi che si originano da "rami" diversi.

Per le varie opzioni software a disposizione per l'applicazione di RECPAM, riferirsi alla pagina web segnalata in bibliografia (8).

Il modello GIPSY-RECPAM

L'applicazione di RECPAM ai dati GIPSY suggeriva che le misurazioni dello stato di benessere nei 3 mesi precedenti avevano peso e direzioni sovrapponibili nel predire il rischio di sopravvivenza a 6 mesi del paziente infartuato.

In altri termini, era possibile condensare questa informazione in una sola variabile (somma algebrica delle tre condizioni) che esprimesse il "malessere nei 3 mesi precedenti l'infarto".

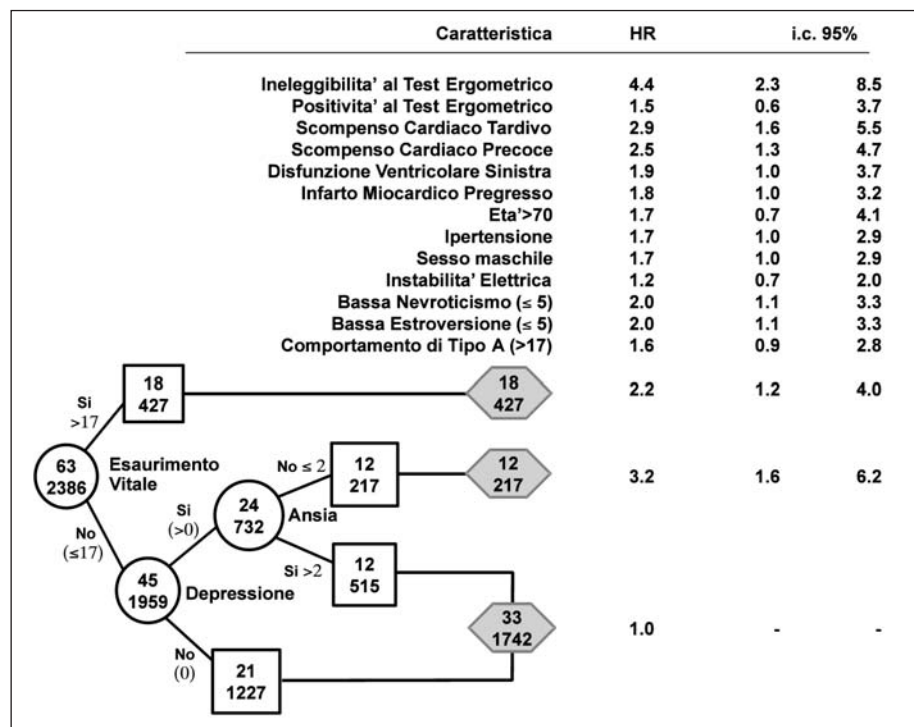


Figura 1. RECPAM-GISSI-PSI: il modello EF-IP. I numeri nei cerchi, quadrati ed esagoni sono il numero di eventi e non eventi a 6 mesi rispettivamente

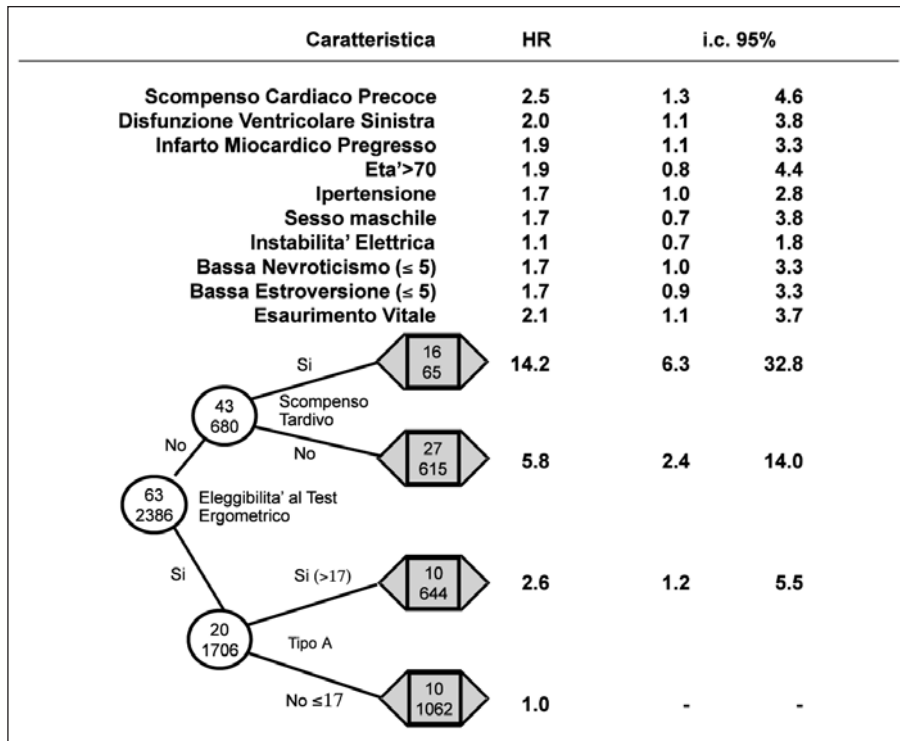


Figura 2. RECPAM-GISSI-PSI: il modello EF-IC

Questo nuovo costrutto, definito come esaurimento vitale ("Vital Exhaustion"), risponde all'ipotesi che la popolazione infartuale possa presentare tratti caratteristici che individuano un fattore di rischio unico, a differenza della popolazione normale sulla quale probabilmente i diversi fattori risultano scarsamente correlati ed esprimono scale psicometriche ben distinte.

RECPAM rende pertanto possibile formulare modelli statistici in grado di integrare indagini psicometriche ed analisi clinico-cardiologica. Nel caso del GIPSY, questa tecnica ha consentito di rispondere alle seguenti domande di rilevanza clinica:

- a parità di condizioni cliniche, quali profili psicologici sono associati ad un più elevato rischio di evento fatale a 6 mesi dall'infarto?
- quali sottogruppi di pazienti sono immediatamente definibili sul piano clinico-psicologico come possibili target di intervento?
- quali fattori "leva" di carattere psicologico possono contribuire ad attenuare il peso delle complicanze cliniche sulla prognosi del paziente infartuato?

L'analisi evidenziava tre modelli finali, di seguito riportati in dettaglio.

Modello 1: Effetti Fissi + Interazioni Psicometriche (EF-IP)

In questo modello le variabili cliniche appaiono come variabili fisse, in quanto esprimono fattori di confondimento noti, che sappiamo essere rilevanti nei campioni GISSI (3); nella terminologia RECPAM, tali effetti fissi sono anche detti variabili globali.

La figura 1 rappresenta il modello EF-IP. Come atteso, la maggioranza delle variabili cliniche esercita un ruolo importante a livello di tutto il campione nel predire un esi-

to negativo, ma a queste si aggiungono importanti interazioni di ordine psicologico.

Inoltre, l'analisi evidenzia ulteriori variabili indipendenti di carattere psicometrico: nevroticismo, estroversione e comportamento di tipo A.

In sintesi, ecco il messaggio clinico risultante dal modello EF-IP:

"a prescindere da tutte le altre caratteristiche cliniche e psicologiche, la prognosi del paziente con infarto miocardico acuto risulta associata all'appartenenza a tre profili.

Il profilo 'a più basso rischio' corrisponde ad un tipo di paziente che presenta bassi livelli di esaurimento vitale CON nessun sintomo di depressione UNITAMENTE a stato di ansia. Rispetto a questa classe, coloro i quali presentano un alto livello

di esaurimento vitale riportano un rischio significativo, più che doppio, di prognosi infausta a 6 mesi.

La categoria maggiormente a rischio (più che triplo rispetto alla prima) è quella equivalente ai pazienti contraddistinti da bassi livelli di esaurimento vitale ED uno stato di depressione NON COLLEGATO ad uno stato di ansia".

Il primo risultato del GIPSY sottolineava quindi il ruolo autonomo della componente psicologica, suggerendone un suo possibile utilizzo di routine, parallelamente all'esame clinico.

Il modello EF-IP evidenzia come le componenti psicologiche, seppure offuscate da indicatori clinici di indubbia rilevanza, costituiscano un patrimonio informativo aggiuntivo in grado di migliorare la prognosi di circa un terzo dei pazienti, riducendo quindi la relativa parte di eventi registrati a 6 mesi.

Modello 2: Effetti Fissi + Interazioni Clinico-psicometriche (EF-IC)

Una ulteriore analisi ha permesso di identificare quale fosse in generale la rilevanza di profili misti clinico-psicologici associati ad un aumento del rischio di mortalità a 6 mesi.

La figura 2 presenta il modello ad "Effetti Fissi + Interazioni Clinico-psicometriche" (EF-IC).

La logica di questo modello è differente, ovvero del tipo "inserisci tutte le variabili come covariate di albero nell' algoritmo RECPAM, quindi usa quelle identificate come globalmente utili - non entrate nell'albero - come possibili fattori di confondimento".

Il risultato offre un esame "paritetico" tra variabili cliniche e psicologiche: il profilo psicologico conta nel contesto dei pazienti eleggibili al test ergometrico, ed è relati-

vo al comportamento di Tipo A. In questa sottoclasse si registra un innalzamento del rischio di due volte e mezza maggiore rispetto agli eleggibili NON di tipo A.

Inoltre, un basso livello di estroversione ed uno stato di esaurimento vitale risultavano importanti a livello dell'intera popolazione (aumento del rischio quasi doppio).

Modello 3: Interazione Psicologica modificatrice di Effetti Fissi (IP-EF)

Una particolare modalità d'uso di RECPAM consente l'analisi per sottogruppi, tramite una o più variabili considerate come *determinanti*. In questo caso il metodo viene costruito sulla base di un criterio diverso: anziché identificare sottogruppi associati ad una diversa prognosi, RECPAM identifica sottogruppi in cui l'effetto di uno o più determinanti cambia in maniera significativa. In questo caso quindi il criterio non consiste nel livello di rischio in sé, ma nel differenziale di rischio tra due livelli diversi del determinante.

Ad esempio, nel caso di un trattamento farmacologico, in questo modo RECPAM consente di individuare sottogruppi di pazienti che riportano risposte mediamente diverse.

Nel caso del GIPSY, si è adottato una logica tipica della analisi di sopravvivenza (figura 3). Le curve nella parte sinistra evidenziano il diverso rischio di tre categorie cliniche con diversa mediana di sopravvivenza a 6 mesi.

Compito del modello di Cox tradizionale è quello di stimare l'importanza di queste componenti a livello dell'intera popolazione, riassumendole con un differenziale espresso da una costante γ , che ad esempio potrebbe dirci "se un soggetto passa dalla categoria A alla categoria B il suo rischio aumenta esponenzialmente, in ragione di e^γ " (proportional hazards).

La domanda posta nella analisi per sottogruppi in RECPAM è la seguente: se noi assumiamo questa proporzionalità, esistono dei sottogruppi in cui questo fattore proporzionale cambia, ovvero omogenei rispetto ad una ripartizione di γ in γ_1 e γ_2 (parte destra figura 3)?

La rilevanza clinica di questa analisi sta nel fatto che se la 'distanza' tra gli effetti clinici diminuisce, allora potremmo avere trovato una associazione che identifica nel fattore *determinante* un modificatore di effetto, ovvero una possibile leva clinica che può essere oggetto di particolare attenzione, identificando come target gruppi ben definiti dall'applicazione di strumenti di misura.

Nell'applicazione GIPSY, il criterio in questione diventa la stratificazione clinica, mentre le

caratteristiche psicologiche vengono immesse come covariate di albero.

Pertanto, l'albero definitivo ripartirà la popolazione in ragione di un particolare livello di una scala del questionario che identifica sotto-modelli in cui i parametri predittivi di tipo clinico saranno *significativamente diversi*.

Il risultante modello RECPAM del tipo IP-EF è presentato in figura 4.

L'analisi per *determinante* si legge quindi così:

"a prescindere dalla presenza o meno di scompenso precoce, nei pazienti con esaurimento vitale, il passaggio da una categoria clinica di eleggibilità al test ergometrico e/o scompenso tardivo all'altra (fasce A, B e C in figura 3) risulta associato ad un innalzamento del rischio medio pari a $e^{2.13}=8.4$ volte; lo stesso aumento è pari a $e^{1.06}=2.9$ volte nei pazienti nei quali NON si riscontra esaurimento vitale".

Il risultato dimostra che il ruolo della situazione psicologica precedente l'infarto merita di essere esplorato in fase acuta, al fine di attenuare l'impatto delle stesse variabili cliniche che il cardiologo ben conosce e che considera come fondamentali per la valutazione del rischio del paziente infartuato.

Uso del modello GIPSY-RECPAM nella pratica clinica

I modelli precedenti esplicitano la logica di uno studio estremamente sofisticato e complesso dal punto di vista interpretativo. Lo studio GIPSY, oltre che produrre un mo-

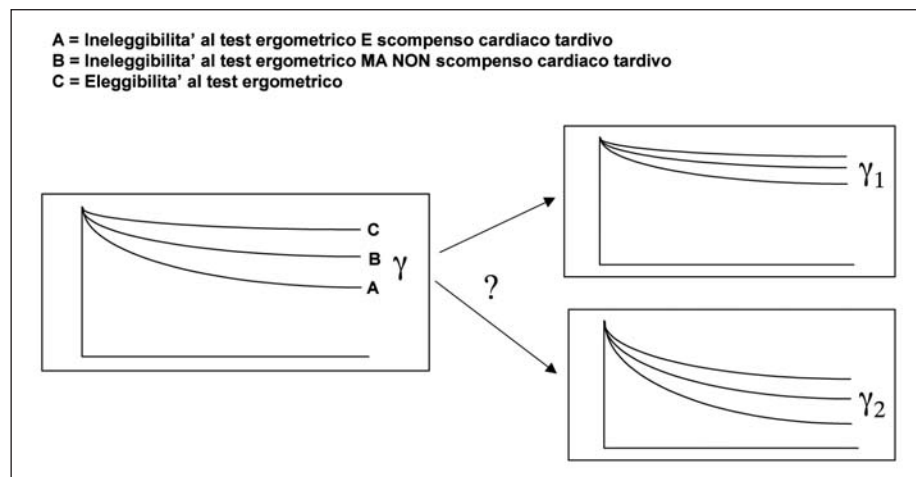


Figura 3. RECPAM-GISSI-PSI: analisi con determinante

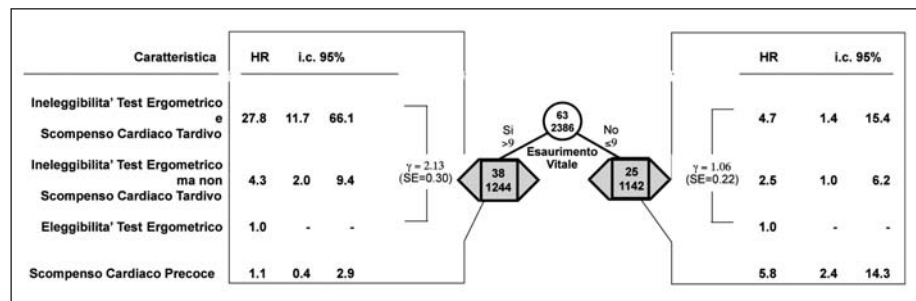


Figura 4. RECPAM-GISSI-PSI: il modello IP-EF

dello epidemiologico, ha però anche contribuito a costruire uno strumento di rilievo per la valutazione di routine in ambito cardiologico.

L'indubbio vantaggio di uno strumento come RECPAM è quello di consentire una efficace sintesi di aspetti multidisciplinari in contesti operativi particolarmente complessi.

Nel caso della valutazione psicometrica, la lunga sequenza delle domande inserite in un questionario quale il CBA-H potrebbe risultare infatti inappropriata in ambito ospedaliero.

Al tempo stesso risulta però importante contemplare tutte le dimensioni di rilievo per la prognosi del paziente tra quelle a disposizione del cardiologo.

RECPAM rappresenta un modo conveniente per utilizzare modelli statistici in un contesto clinico. Nell'ambito della valutazione psicologica, il suo utilizzo prospettico può così schematizzarsi:

a) *Uso del modello GIPSY come scoring system*

Il modello RECPAM può essere utilizzato per allocare il 'prossimo' paziente: sulla base delle risposte date e dei parametri clinici, esiste uno ed un solo profilo di rischio che può essere assegnato ad ogni singolo soggetto.

Spetta al clinico scegliere tra i diversi modelli proposti a seconda delle particolari condizioni e considerazioni personali.

Se egli ha già una chiara idea del profilo clinico del paziente, allora la componente psicometrica 'da aggiungere' a questa traccia è quella delineata dal modello EF-IP. Nel caso dello studio GIPSY, questo prevede l'allocazione del paziente in uno dei nodi terminali dell'albero, con conseguente possibile aumento di rischio, da aggiungere eventualmente ad un profilo di rischio psicologico globale a seconda dei livelli di estroversione, nevroticismo e comportamento di Tipo A.

Se invece il clinico si affida ad una ripartizione 'equa' tra variabili cliniche e psicologiche, allora il modello di riferimento è quello EF-IC, ovvero un modello che integra le due componenti ed assegna il paziente ad una classe di rischio psicologica solo se appartenente alla categoria con test ergometrico svolto. In questo caso le componenti globali sono non solo tutte le caratteristiche cliniche, ma anche quelle psicologiche, che questa volta includono anche l'esaurimento vitale. Concretamente, il modello EF-IC è pertanto coerente con il precedente, ma si riferisce ad un altro punto di vista dell'utente finale.

Il terzo modello, IP-EF, è sostanzialmente utile per la pianificazione dei servizi sanitari, in quanto pone l'attenzione su un possibile intervento in termini preventivi per una attenuazione dell'impatto su pazienti che hanno vissuto una particolare difficoltà generale in più dimensioni psicologiche concomitanti. Anche in questo caso, la scala dell'esaurimento vitale può allocare un singolo paziente in una classe specifica.

In tutti i casi precedenti, è possibile coadiuvare il clinico con meccanismi automatici, quali ad esempio semplici software installati su portatili o palmari, che possono immediatamente presentare la situazione clinica di riferimento, come in una sorta di 'pannello di controllo' esperto.

b) *Validazione esterna: applicazione del modello GIPSY su nuovi campioni*

Allo stesso tempo, è possibile produrre delle nuove stime su altri campioni, costruendo gli stessi modelli per diverse situazioni di riferimento. Per fare questo non occorre utilizzare RECPAM, in quanto il modello è già disponibile. Praticamente, è necessario utilizzare lo stesso modello statistico, cosicché gli stessi risultati diventano confrontabili nello spazio e nel tempo.

Lo stesso tipo di procedura è stata applicata per validare un sistema di classificazione capace di stratificare i pazienti con infarto miocardico acuto sulla base delle poche informazioni disponibili all'ammissione (11). In questa ulteriore applicazione, le stime degli *odds ratio* ottenuti per la struttura ad albero nel GISSI-2 sono state poi confrontate con quelle ottenute applicando l'albero stesso nel GISSI-1 e GISSI-3.

Pertanto, è possibile utilizzare il modello di Cox finale, ad esempio quello in figura 1, dove le classi terminali dell'albero diventano covariate aggiuntive nel modello per un nuovo campione, per poi analizzare le eventuali divergenze dalle stime ottenute.

L'operazione richiede la riduzione dei modelli RECPAM ad equazioni matematiche corrispondenti ai tradizionali modelli epidemiologici. Ad esempio, nel modello GIPSY tutte le variabili sono dicotomiche, ed in questo caso basta disporre i dati in maniera tale da avere due serie di variabili numeriche $C_1+...C_n$, cliniche e $P_1+...P_n$, psicologiche, del tipo *dummy variables*, ovvero codificate con 1=si e 0=no.

Le equazioni RECPAM contengono quindi covariate fisse e termini di interazione per la struttura ad albero. Questi ultimi assumono una struttura caratteristica, che in termini informatici viene realizzata nella maniera più semplice con sequenze IF..THEN..ELSE, lasciando quindi un termine al di fuori che agisce come termine di riferimento.

Le formule matematiche relative ai modelli ottenuti sono riportate in appendice tecnica.

La replicazione del modello RECPAM su altri campioni consente quindi l'utilizzo come sistema esperto, ovvero capace di 'aggiustare' le stime imparando dall'esperienza.

c) *Applicazione soggettiva del metodo RECPAM: partizione 'guidata'*

In alcuni casi una selezione automatica del modello potrebbe sembrare inappropriata: occorre quindi introdurre criteri "clinici" sia nella scelta degli *split* che nella eliminazione dei sottogruppi ritenuti marginali (*pruning* e *amalgamation*). RECPAM consente di enucleare tali direttive in vari modi. Occorre però applicare di nuovo la procedura, ad esempio specificando diversi insiemi di covariate globali e sottogruppi psicometrici di partenza.

d) *Applicazione 'manuale' della logica RECPAM: modelli statistici innestati*

La logica del metodo RECPAM può essere utilizzata nel contesto generale dei modelli di regressione. Si possono pertanto utilizzare tecniche miste, alternando procedure più tradizionali alla costruzione di modelli che ricorsi-

vamente includono covariate ed interazioni secondo uno stile prossimo a quello di RECPAM. Ad esempio, è possibile 'dicotomizzare' le dimensioni psicometriche sulla base del criterio di massimo LR ad ogni *step*, come effettuato nel *recursive partitioning*, o amalgamare sottogruppi applicando il criterio del minimo AIC.

e) *Traduzione del modello da generatore di ipotesi a generatore di evidenze*

Gli alberi di regressione sono spesso inclusi genericamente nell'insieme di tecniche più o meno propriamente definite come "data mining". In biostatistica, tali metodi sono più correttamente citati come "analisi esplorativa", intesi come "generatori di ipotesi" in grado di suggerire *patterns* possibili, piuttosto che validare relazioni causa-effetto.

Ciò nonostante, questi metodi consentono poi di verificare l'ipotesi tramite studi *ad hoc*.

Ad esempio, RECPAM può suggerire sperimentazioni cliniche controllate nelle quali testare con assegnazione casuale l'uso del modello ad albero come strumento di supporto alla decisione clinica. I disegni di tipo 'clusterizzato', basandosi sulla randomizzazione dei centri di cura, offrono delle possibilità notevoli di impiego di questi modelli in studi di implementazione ed intervento nel servizio sanitario.

Nel caso del GIPSY, questo può essere semplicemente usato nella sua forma estesa oppure offerto al clinico nella riduzione ottenuta dal modello RECPAM, pianificando quindi interventi diversi per le diverse sottocategorie ottenute, valutandone le differenze sulla base degli esiti prognostici.

Prospettive

Oggi esistono notevoli possibilità di sviluppo nell'uso congiunto di sistemi che utilizzano dimensioni diverse della pratica clinica.

L'integrazione di tali strumenti consentirà di estendere la gamma di modelli a situazioni in grado di includere aspetti più direttamente connessi all'analisi dei servizi sanitari.

In particolare, il metodo RECPAM consente di costruire e testare modelli per i quali la prognosi dell'individuo è influenzata non solo dalle caratteristiche del soggetto, ma anche da quelle organizzative e strutturali (12). L'inclusione di modelli multilivello consentirà di incorporare relazioni che in molti casi assumono forme 'gerarchiche' ben più complesse di quanto possibilmente catturabili tramite i metodi tradizionali.

La regressione ad alberi può quindi fornire una lettura sintetica a più livelli, ad esempio considerando anche fattori di ordine psicologico-comportamentale rilevanti per l'operatore sanitario in senso generale, in rapporto agli esiti della cura (è pensabile un CBA-H anche per il cardiologo, lo psicologo, l'infermiere, e l'amministratore?).

Sulla base di queste idee, ed anche per garantire una maggiore efficienza e stabilità degli algoritmi RECPAM, sono allo studio soluzioni per la produzione e l'utilizzo di alberi in rete telematica (13).

Le possibilità di un utilizzo clinico di questi strumenti sono oggi innumerevoli. Tali prospettive sono estremamente rilevanti per un contesto di cure integrate in cui team clinici e cittadini sono destinati a condividere strutture e processi di cura.

Gli attuali sviluppi nel campo dell'*e-health* consentono di immaginare cicli virtuosi in cui raccolta dei dati interattiva e modelli statistici sono strettamente connessi, e quindi aspetti psicologici e clinici possono insieme contribuire alla costruzione di sofisticate soluzioni di telemedicina.

In quest'ottica, l'esperienza concreta dell'analisi GIPSY rappresenta non solo un utile riferimento metodologico, ma anche un positivo esempio di interazione epi-psicometrica da considerarsi come possibile punto di partenza per nuove esperienze multidisciplinari ad alto contenuto innovativo.

Bibliografia

- 1) Labbrozzi D, Carinci F, Nicolucci A, Bettinardi O, Zotti AM, Tognoni G a nome del gruppo italiano per lo studio della sopravvivenza nell'infarto del miocardio (GISSI). Le caratteristiche psicologiche del paziente infartuato: risultati del GISSI-2, *G Ital Cardiol* 1996; 26: 85-106.
- 2) Carinci F, Nicolucci A, Ciampi A, Labbrozzi D, Bettinardi O, Zotti AM, Tognoni G. 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 database. *Eur Heart J*, 1997; 18: 835-45.
- 3) Volpi A, De Vita C, Franzosi MG et al. Determinants of 6-month mortality in survivors of myocardial infarction after thrombolysis. Results of the GISSI-2 database. The Ad hoc Working Group of the Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico (GISSI), *Circulation* 1993; 88(2): 416-29.
- 4) Sanavio E. Le Scale CBA. Milano, Raffaello Cortina, 2002.
- 5) Bianconi G, Poggioli E, Merelli E, Razzaboni E, Comelli D. Psychological issues related to multiple sclerosis. *G Ital Med Lav Ergon*. 2006; 28(1 Suppl 1): 22-8.
- 6) Zotti AM, Bertolotti G, Michielin P, Sanavio E, Vidotto G. Il CBA Forma HOSPITAL. Manuale. OS: Firenze 2009.
- 7) Ciampi A, Negassa A, Lou. Z. Tree-structured prediction for censored survival data and the Cox model, *Journal of Clinical Epidemiology*, 1995; 48, 5, 675-68.
- 8) Carinci F. Generalised regression trees by RECURSIVE Partitioning and AMalgamation using statistical software, URL: <http://www.serecetrix.eu/staff/carinci/recpam/index.html>
- 9) <http://www.med.monash.edu.au/healthservices/chsr/FabrizioCarinci/2000/software/recpam/>
- 10) Breiman, L, Friedman J, Olshen R. and Stone C. Classification and Regression Trees, CRC Press, Berkeley, 1984.
- 11) Nicolucci A, Carinci F, Ciampi A. Stratifying patients at risk of diabetic complications. An integrated look at clinical, socio-economic and care-related factors. *Diabetes Care* 1998; 21: 1439-44
- 12) Fresco C, Carinci F, Maggioni AP, Ciampi A, Nicolucci A, Santoro E, Tavazzi L, Tognonia G. Very early assessment of the risk of in-hospital death in 11,483 patients with acute myocardial infarction, *Am Heart J* 1999; 138(6 Pt 1), 1058-64.
- 13) Carinci F. RECPAM Information Generation Network (REIGN): Parallel Regression Trees for Multilevel Information Systems and Data Mining, URL: <http://www.serecetrix.eu/staff/carinci/recpam/index.html>
- 14) Carinci F. RECPAM Information Generation Network (REIGN): Parallel Regression Trees for Multilevel Information Systems and Data Mining, Victorian Partnership for Advanced Computing (VPAC), Expertise Program Grants 2002, September 2001, Monash University, URL: http://www.med.monash.edu.au/healthservices/chsr/projects/chsr_reign.html.

APPENDICE TECNICA

Le formule attengono all'esponente del modello di Cox, in cui ITE=Ineleggibilità al test ergometrico, EV=Esaurimento Vitale, D=Depressione, A=Ansia, ST=Scompenso tardivo SP=Scompenso precoce, TA=Tipo A
 I modelli matematici ottenibili dal modello GIPSY-RECPAM sono i seguenti:

– *Modello EF-IP:*

$$(\beta_1 C_1 + \dots + \beta_{10} C_{10}) + (\beta_{11} P_1 + \dots + \beta_{13} P_3) + \frac{\beta_{14} EV}{C_1} + \frac{\beta_{15} (1-EV) * D * (1-A)}{C_2}$$

Il modello ottenuto include 13 parametri globali e 2 classi RECPAM, equivalenti ad un ulteriore parametro globale e ad un termine di interazione di secondo ordine (pari alla profondità dell'albero). La terza classe rimane fuori dal modello in quanto agisce come classe di riferimento.

– *Modello EF-IC:*

$$(\beta_1 C_1 + \dots + \beta_7 C_7) + (\beta_8 P_1 + \dots + \beta_{10} P_3) + \frac{\beta_{11} (ITE * ST)}{C_1} + \frac{\beta_{12} (ITE) * (1-ST)}{C_2} + \frac{\beta_{13} TA * (1-ITE)}{C_3}$$

In questo modello figurano 13 parametri globali e 3 classi RECPAM, che equivalgono a tre termini di interazione di primo ordine. Anche in questo caso, la classe con rischio inferiore rimane fuori dal modello in quanto classe di riferimento.

– *Modello IP-EF:*

$$(EV) * \frac{[\beta_1 (ITE * ST) + \beta_2 ITE * (1-ST) + \beta_3 SP]}{S_1} + (1-EV) * \frac{[\beta_4 (ITE * ST) + \beta_5 ITE * (1-ST) + \beta_6 SP]}{S_2}$$

Da quest'ultima equazione risulta evidente la caratteristica di questo modello rispetto ai precedenti, che è anche la ragione per cui l'analisi per *determinante* equivale ad una analisi per sottogruppi. Infatti, si otterrebbero gli stessi risultati se noi separassimo il campione secondo EV, stimando separatamente i due modelli S₁ e S₂.

Richiesta estratti: Prof. Fabrizio Carinci, Senior Biostatistician, Sereatrix snc, Pescara, Italy, E-mail: research@fabcarinci.net