



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in
Metodologia Statistica per la Ricerca Scientifica
XX ciclo

Valutazione della persistenza
in trattamento antiipertensivo:
un'analisi multilivello paziente-medico
attraverso modelli con effetti casuali

Mirko Di Martino

Dipartimento di Scienze Statistiche "P. Fortunati"
Marzo 2008



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in
Metodologia Statistica per la Ricerca Scientifica
XX ciclo

Valutazione della persistenza
in trattamento antiipertensivo:
un'analisi multilivello paziente-medico
attraverso modelli con effetti casuali

Mirko Di Martino

Coordinatore
Prof. Daniela Cocchi

Tutor
Prof. Paola Monari

Co-tutor
Prof. Rossella Miglio

Settore Disciplinare
MED/01

Dipartimento di Scienze Statistiche "P. Fortunati"
Marzo 2008

Sommario

Introduzione	1
Obiettivi	4
1. Il governo clinico delle patologie croniche.....	5
1.1. Introduzione	5
1.2. Efficacia teorica ed efficacia pratica	6
1.3. I sistemi informativi per il governo clinico.....	7
1.3.1. Archivi amministrativi aziendali	8
1.3.2. Banche dati cliniche.....	10
1.4. Gli strumenti per il governo clinico	11
2. Analisi multilevel: contesti applicativi e metodologia	13
2.1. Introduzione alle analisi multilevel.....	13
2.2. La correlazione intraclasse.....	14
2.3. Aspetti metodologici dei modelli multilevel.....	16
2.4. Il modello multilevel a sola intercetta.....	20
2.5. Stima dei parametri nei modelli multilevel.....	21
2.5.1. Il metodo della massima verosimiglianza.....	21
2.5.2. L'approccio bayesiano alle analisi multilevel	23
2.5.3. Altri metodi per la stima dei parametri.....	29
2.6. Verifica di ipotesi nei modelli multilevel.....	32
2.7. Modelli lineari generalizzati multilevel	35
2.7.1. Stima dei parametri nei modelli lineari generalizzati multilevel.....	38
2.7.2. Integrazione di effetti fissi ed effetti casuali.....	40
3. Modelli per l'analisi della sopravvivenza con effetti casuali	43
3.1. L'analisi della sopravvivenza.....	43
3.2. Il modello a rischi proporzionali di Cox	44
3.3. Il modello di Cox con effetti casuali.....	47
3.3.1. Stima dei parametri nel modello di Cox con effetti casuali.....	51
4. Analisi della persistenza in trattamento antiipertensivo	54
4.1. Fonte dei dati.....	54
4.2. Criteri di arruolamento	54
4.3. Caratterizzazione dei pazienti	55
4.4. Periodo di follow-up	55
4.5. Criteri di esclusione.....	56
4.6. Valutazione della persistenza in trattamento.....	56
4.7. Struttura gerarchica paziente-medico.....	57
5. Risultati.....	58
5.1. Discussione	68
Bibliografia	71

Introduzione

Fai attenzione quando leggi libri di medicina.

Potresti morire per un errore di stampa.

Mark Twain

I grandi trial clinici hanno dimostrato con estrema chiarezza come un corretto trattamento farmacologico di alcuni fattori di rischio come l'ipertensione, l'ipercolesterolemia o l'iperglicemia possa ridurre significativamente la mortalità e la morbilità nella popolazione (Dahlof *et al.*, 1991; Pignone *et al.*, 2000).

Tuttavia, in pratica clinica, l'utilizzo delle terapie farmacologiche nelle patologie croniche risulta spesso inadeguato sia in termini di numerosità dei pazienti trattati sia in termini di aderenza e persistenza al trattamento (Degli Esposti *et al.*, 2002; Di Martino *et al.*, 2005). Ne consegue che, nella popolazione generale ma anche nei pazienti trattati, il rischio di mortalità e morbilità non viene adeguatamente controllato dalla terapia farmacologica.

Gli studi clinici controllati e randomizzati costituiscono il *gold standard* dell'evidenza scientifica e sono insostituibili nella valutazione dell'efficacia di una nuova terapia. Eppure, i risultati derivanti da questi studi non possono essere generalizzati a tutti i soggetti che utilizzano il farmaco nella reale pratica clinica (Dieppe *et al.*, 2004) per due motivi principali: in primo luogo, negli studi clinici randomizzati l'aderenza e la persistenza al trattamento farmacologico sono garantite da un monitoraggio continuo e sistematico; in secondo luogo, i gruppi minoritari, le persone anziane e i pazienti a rischio di eventi avversi sono spesso deliberatamente esclusi dallo studio.

Emerge, quindi, la necessità di conoscere le modalità con cui le evidenze scientifiche trovano reale applicazione nella pratica clinica routinaria, identificando quei fattori di disturbo che, nei contesti reali, limitano l'efficacia e l'appropriatezza clinica. Nell'ambito delle terapie farmacologiche croniche, uno

di questi fattori è costituito dalla ridotta aderenza al trattamento (Joint National Committee VII, 2003). Su questo tema, recenti studi osservazionali hanno individuato le caratteristiche del paziente associate a bassi livelli di *compliance* (Di Martino *et al.*, 2007); altri hanno focalizzato l'attenzione sulle caratteristiche del medico e sulla sua capacità di comunicare ai propri assistiti l'importanza del trattamento farmacologico (Teutsch, 2003). Dalle attuali evidenze scientifiche, tuttavia, non emerge con chiarezza il peso relativo dei due diversi attori, paziente e medico, nel determinare i livelli di *compliance* nel trattamento delle patologie croniche.

In questo contesto, i pazienti e i medici di medicina generale possono essere concettualizzati come un sistema gerarchico osservabile su differenti livelli della gerarchia; le variabili oggetto di interesse possono essere definite su ciascuno di questi livelli. La struttura dei dati nella popolazione è, dunque, gerarchica e i dati campionari sono visti come risultato di un campionamento a due stadi: nella prima fase si assume un campione di medici di medicina generale, nella seconda fase si estrae un campione di pazienti da ciascuno dei medici selezionati.

Storicamente, le analisi di dati gerarchicamente organizzati sono state spesso realizzate mediante tecniche standard come l'analisi della varianza o la regressione multipla dopo aver spostato tutte le variabili su un solo livello di interesse. Ciò avveniva mediante due distinte procedure: aggregazione e disaggregazione. L'aggregazione è lo spostamento di variabili originariamente osservate su un basso livello della gerarchia verso un livello superiore. Al contrario, la disaggregazione è lo spostamento di variabili verso un livello più basso della gerarchia.

Tuttavia, analizzare variabili che appartengono a differenti livelli della gerarchia su un singolo e comune livello è inadeguato. Da un lato, l'aggregazione comporta una sostanziale perdita di informazioni e, di conseguenza, l'analisi statistica perde potenza. Dall'altro, se i dati vengono disaggregati, i test statistici ordinari considerano i valori disaggregati come informazioni indipendenti provenienti dal vasto insieme di unità di basso livello. Eppure, la reale numerosità

campionaria per le variabili disaggregate è costituita dal più basso numero di unità di livello superiore. Pertanto, l'utilizzo dell'ampio insieme di casi disaggregati come numerosità campionaria porta ad ottenere risultati statisticamente significativi che sono totalmente spuri.

In generale, nelle situazioni in cui i dati sono gerarchicamente organizzati, le osservazioni individuali non sono del tutto indipendenti. I pazienti iscritti allo stesso medico di medicina generale tenderanno ad essere tra loro simili a causa dei processi di selezione e a causa della “storia comune” che gli individui condividono afferendo allo stesso medico. I test statistici tradizionali sono fortemente basati sull'assunto di indipendenza tra le osservazioni. Se questa ipotesi risulta violata, le stime degli errori standard prodotte dai test statistici convenzionali sono troppo piccole e, di conseguenza, i risultati che si ottengono appaiono “impropriamente” significativi. Emerge, quindi, la necessità di utilizzare un modello statistico che tenga conto della non indipendenza delle osservazioni e che consenta, allo stesso tempo, di analizzare simultaneamente variabili che “provengono” da diversi livelli della gerarchia.

Obiettivi

Questa tesi, partendo da un'esigenza conoscitiva concreta, si propone di illustrare i principali aspetti metodologici dei modelli di regressione con effetti casuali enfatizzandone le applicazioni su dati gerarchicamente organizzati.

Gli obiettivi specifici di questo lavoro possono essere sintetizzati come segue:

- valutare l'utilizzo dei farmaci antiipertensivi in pratica clinica attraverso un sistema informativo che prevede l'integrazione tra banche dati amministrative e dati raccolti dai Medici di Medicina Generale;
- analizzare la persistenza in trattamento antiipertensivo con tecniche proprie dell'analisi della sopravvivenza;
- illustrare i principali aspetti metodologici dei modelli di regressione multilivello, con particolare riferimento ai modelli per l'analisi della sopravvivenza con effetti casuali (*shared frailties model*);
- valutare quanta parte della variabilità totale è attribuibile al livello-paziente, quanta parte della variabilità è attribuibile al livello-medico;
- spiegare la variabilità totale in funzione delle caratteristiche del paziente e in funzione delle caratteristiche del medico.

1. Il governo clinico delle patologie croniche

1.1. Introduzione

Il governo clinico riguarda l'insieme delle attività con cui i servizi sanitari si rendono responsabili del miglioramento continuo della qualità dell'assistenza nei limiti delle risorse disponibili. Il governo clinico risulta quindi costituito da una serie organica di scelte di politica sanitaria effettuate alla luce di un utilizzo sistematico di tutti gli strumenti a disposizione per favorire l'efficacia e l'appropriatezza clinica. Il governo clinico delle patologie croniche rappresenta indubbiamente un aspetto di grande rilevanza strategica per la gestione sanitaria a causa dell'elevata prevalenza di queste malattie e dell'elevato volume di spesa che generano. In Italia (ISTAT, 2007), il 42.8% della popolazione soffre di almeno una patologia cronica ed il 13.3% presenta tre o più malattie croniche contemporaneamente. Per quel che riguarda il volume di spesa, il solo apparato cardiovascolare assorbe il 35.1% della spesa complessiva per farmaci (OsMed, 2005). I grandi trial clinici hanno dimostrato con estrema chiarezza come un corretto trattamento farmacologico di alcuni fattori di rischio come l'ipertensione, l'ipercolesterolemia o l'iperglicemia possa ridurre significativamente la mortalità e la morbilità nella popolazione (Dahlof *et al.*, 1991; Pignone *et al.*, 2000). Tuttavia, in pratica clinica, l'utilizzo delle terapie farmacologiche nelle patologie croniche risulta spesso inadeguato sia in termini di numerosità dei pazienti trattati sia in termini di aderenza e persistenza al trattamento (Degli Esposti *et al.*, 2002; Di Martino *et al.*, 2005). Ne consegue che, nella popolazione generale ma anche nei pazienti trattati, il rischio di mortalità e morbilità non viene adeguatamente controllato dalla terapia farmacologica.

Su questo tema, recenti studi osservazionali hanno individuato le caratteristiche del paziente associate a bassi livelli di *compliance* (Di Martino *et al.*, 2007); altri hanno focalizzato l'attenzione sulle caratteristiche del medico e sulla sua capacità di comunicare ai propri assistiti l'importanza del trattamento farmacologico (Teutsch, 2003). Dalle attuali evidenze scientifiche, tuttavia, non

emerge con chiarezza il peso relativo dei due diversi attori, paziente e medico, nel determinare i livelli di *compliance* nel trattamento delle patologie croniche.

1.2. Efficacia teorica ed efficacia pratica

Gli studi clinici randomizzati (studi sperimentali) costituiscono il *gold standard* dell'evidenza scientifica e sono insostituibili nella valutazione dell'efficacia di una nuova terapia farmacologica. Tuttavia, i risultati derivanti da questi studi non possono essere generalizzati a tutti i soggetti che utilizzano il farmaco nella reale pratica clinica (Dieppe *et al.*, 2004) per due motivi principali: in primo luogo, negli studi clinici randomizzati l'aderenza e la persistenza al trattamento farmacologico sono garantite da un monitoraggio continuo e sistematico; in secondo luogo, i gruppi minoritari, le persone anziane e i pazienti a rischio di eventi avversi sono spesso deliberatamente esclusi dallo studio. In molti casi, infatti, la scarsa efficacia di una terapia farmacologica in pratica clinica dipende proprio dalla ridotta aderenza al trattamento (Joint National Committee VII, 2003). Inoltre, l'esclusione "da protocollo" di alcune categorie di pazienti dagli studi sperimentali determina una situazione in cui l'utilizzo del farmaco su alcune tipologie di pazienti (generalmente quelli caratterizzati da una situazione clinica più complessa) non è supportato da solide evidenze scientifiche sull'efficacia, sulla tollerabilità e sulla sicurezza della terapia. Il termine "efficacia" necessita alcune precisazioni. Secondo la definizione classica, l'efficacia è la capacità di raggiungere il risultato desiderato. In ambito sanitario, l'efficacia è riferita agli esiti dell'intervento terapeutico ed è tanto maggiore quanto minore è lo scarto tra i risultati ottenuti (*outcome*) e gli obiettivi attesi. Tuttavia, una approfondita comprensione di tale termine richiede la distinzione tra efficacia teorica (*efficacy*) ed efficacia pratica (*effectiveness*). L'*efficacy* è la capacità di raggiungere il risultato desiderato nelle condizioni "ideali" tipiche degli studi sperimentali, con casistiche selezionate, variabili di contesto controllate e condizioni organizzative migliori di quelle presenti nella pratica routinaria. L'*effectiveness* è la capacità di raggiungere il risultato desiderato in condizioni operative reali, con casistiche non

selezionate e, di solito, con variabili di contesto e condizioni organizzative peggiori di quelle proprie degli studi sperimentali (Black, 1996). Pertanto, il governo clinico delle patologie croniche non può basarsi esclusivamente sulle informazioni che derivano dagli studi clinici controllati e randomizzati. Emerge, infatti, la necessità di conoscere le modalità con cui le evidenze scientifiche trovano reale applicazione nella pratica clinica routinaria, identificando quei fattori di disturbo che, nei contesti reali, limitano l'efficacia e l'appropriatezza clinica. Sarebbe, inoltre, opportuno disporre di informazioni che descrivano localmente le diverse realtà sanitarie, al fine di governare i processi assistenziali secondo le caratteristiche e secondo le esigenze delle specifiche aree geografiche.

Dal momento che “non si può pensare di governare ciò che non si conosce” (Grilli e Taroni, 2004), la più immediata implicazione pratica del governo clinico è indubbiamente rappresentata dalla realizzazione delle condizioni organizzative necessarie a sviluppare, in modo sistematico e continuativo, la sorveglianza ed il monitoraggio dei processi assistenziali.

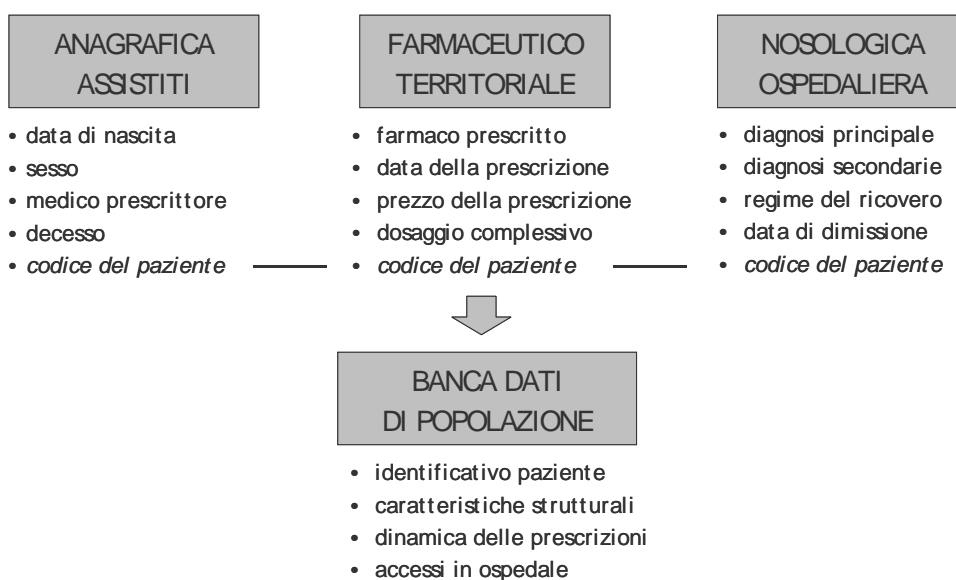
1.3. I sistemi informativi per il governo clinico

Un sistema informativo è l'interconnessione di attrezzature (hardware, software) per l'archiviazione, la gestione e l'elaborazione dei dati. Un sistema informativo è caratterizzabile in funzione degli obiettivi per cui viene utilizzato, dei criteri secondo i quali viene strutturato, delle procedure operative di funzionamento (Greenes e Shortliffe, 1990). Un sistema informativo non deve essere strutturato in modo perfetto ma semplicemente adeguato rispetto alle finalità informative dei soggetti che ne esprimono la necessità e ne promuovono la realizzazione (Black, 1997). In relazione al governo clinico delle patologie croniche, gli obiettivi conoscitivi, attualmente non soddisfatti in modo sistematico, sono rappresentati dalla valutazione epidemiologica delle caratteristiche dei pazienti in trattamento e delle modalità di utilizzo della terapia farmacologica.

1.3.1. Archivi amministrativi aziendali

Una possibile fonte idonea alla valutazione dell'aderenza di un intervento terapeutico agli standard di efficacia – valutazione delle modalità di utilizzo del trattamento farmacologico (farmacoutilizzazione) – sono gli archivi amministrativi aziendali (Farmer, 1999). Gli archivi amministrativi aziendali, sovrapponibili per alcuni versi ai *claims database* di derivazione nordamericana utilizzati prevalentemente nei sistemi sanitari su base assicurativa, rappresentano un insieme di fonti informative che vengono originariamente realizzate in modo autonomo per specifici scopi dipartimentali ma che spesso dispongono dei requisiti necessari per poter essere integrate e per poter condividere le informazioni in esse contenute. L'integrazione degli archivi amministrativi attraverso l'identificativo del paziente offre, limitatamente alle informazioni contenute negli archivi originali, la possibilità di creare una banca dati amministrativa di popolazione.

Figura 1. Dagli archivi amministrativi aziendali ad una banca dati di popolazione



Per alcuni anni, in occasione dei primi utilizzi di questa fonte per gli studi di farmacoutilizzazione, numerose sono state le critiche mosse verso questa soluzione. In generale, alcuni ritenevano gli archivi amministrativi aziendali non idonei per scopi diversi da quelli amministrativi e contabili volti esclusivamente alla quantificazione delle risorse consumate. Due considerazioni vanno espresse a questo riguardo. In primo luogo, queste fonti informative sono utilizzate per il calcolo di indicatori di processo. Gli archivi amministrativi aziendali sono infatti utilizzati per ricostruire il profilo analitico e cronologico dell'assistibile (caratteristiche del paziente, prescrizioni farmacologiche) e per valutare eventuali scostamenti rispetto agli standard di efficacia terapeutica (continuità delle prescrizioni e durata del trattamento). In questa prospettiva, gli archivi amministrativi aziendali risultano idonei in quanto permettono la ricostruzione della dinamica degli interventi (necessaria per una valutazione di processo) e non viziano la valutazione in quanto rilevano ciò che avviene nella pratica clinica sulla base dell'arbitrio degli operatori sanitari (medici prescrittori). In secondo luogo, nonostante il valore dell'informazione in Sanità sia universalmente riconosciuto, lo sviluppo di sistemi informativi adeguati è notevolmente ritardato da problemi organizzativi (difficoltà di motivare il personale dedicato al corretto e costante utilizzo degli strumenti) e finanziari (costi di realizzazione elevati ed ancora percepiti come superiori ai potenziali risparmi). In questa situazione, quindi, il "riciclo" degli archivi amministrativi aziendali per gli scopi del governo clinico appare giustificato dall'assenza di una reale alternativa e dall'idoneità di questo strumento in ragione della rappresentatività delle coorti osservate, della durata del periodo di osservazione (potenzialmente senza fine), della qualità dei dati inseriti e dei limitati costi di ri-utilizzo (Birnbaum *et al.*, 1999).

Gli archivi amministrativi, nella loro destinazione originaria, sono utilizzati per scopi contabili (conteggio rimborsi, consumo di risorse). Tuttavia, poiché le informazioni sono rilevate con indicazione del paziente beneficiario e della data di riferimento dell'intervento, è possibile ribaltare sul paziente l'insieme delle informazioni tracciando, per ogni assistibile, un profilo analitico e cronologico degli interventi effettuati e delle risorse assorbite.

Il ribaltamento dell'informazione dal fattore (consumo di farmaco) al paziente (destinazione del farmaco) rappresenta un criterio di attribuzione della spesa che non modifica la spesa complessiva (100%) ma risulta essenziale per valutare l'aderenza degli interventi terapeutici agli standard di efficacia poiché consente di analizzare il modo in cui il paziente ha effettuato gli interventi e, quindi, il modo in cui il paziente ha assorbito le risorse destinategli (Motheral e Fairman, 1997).

Figura 2. L'attribuzione del dato attraverso la matrice fattore-paziente

	<i>Farmaco a</i>	<i>Farmaco b</i>	...	<i>Farmaco n</i>	<i>Totale</i>
<i>Paziente 1</i>	35%	0%	...	35%	70%
<i>Paziente 2</i>	0%	0%	...	0%	0%
...
<i>Paziente m</i>	10%	15%	...	5%	30%
<i>Totale</i>	45%	15%	...	40%	100%

1.3.2. Banche dati cliniche

Il principale difetto delle banche dati amministrative è la mancanza di dati clinici.

Le banche dati amministrative, infatti, nascendo per scopi amministrativi e contabili tralasciano completamente dati inerenti alle abitudini di vita del paziente (fumatore o non fumatore), ai sintomi ed alle diagnosi (diagnosi di ipertensione arteriosa, ipercolesterolemia o diabete), alla stadiazione della patologia (rischio cardiovascolare assoluto), agli indicatori di esito intermedio (valore della pressione arteriosa, del colesterolo, della glicemia). Allo scopo di completare il contributo informativo garantito da una banca dati amministrativa, è possibile attivare una banca dati "clinica". Questo strumento, realizzabile, in funzione delle specifiche esigenze di analisi, presso la medicina generale, la medicina

specialistica, i vari dipartimenti o le farmacie, offre la possibilità di raccogliere le rimanenti informazioni per la completezza della caratterizzazione del paziente, della definizione degli interventi, della valutazione degli esiti. Le banche dati cliniche comportano, però, alcuni problemi di carattere gestionale come la necessità di collaborazione degli utenti per la qualità e la completezza dei dati, le difficoltà di addestramento al corretto e costante utilizzo degli strumenti per la raccolta dei dati, i costi di realizzazione e mantenimento (Wyatt, 1994).

In conclusione, la strategia di realizzazione di un sistema informativo per la gestione clinica ed economica dei servizi sanitari può essere graduale (prima la banca dati amministrativa poi la banca dati clinica) oppure parziale (una sola tra le due banche dati). È, comunque, necessario che la strategia sia coerente con gli obiettivi di valutazione prefissati e che preveda una prospettiva di integrazione delle diverse fonti informative e di condivisione dei dati raccolti.

1.4. Gli strumenti per il governo clinico

Un importante strumento per la valutazione delle terapie farmacologiche croniche in pratica clinica è costituito dagli studi osservazionali. Gli studi osservazionali (studi clinici non interventistici) sono studi in cui *“i medicinali sono prescritti nel modo consueto, conformemente alle condizioni fissate nell’autorizzazione all’immissione in commercio. L’inclusione del paziente in una determinata strategia terapeutica non è decisa in anticipo da un protocollo di sperimentazione ma rientra nella normale pratica clinica e la decisione di prescrivere il medicinale è del tutto indipendente da quella di includere il paziente nello studio”* (Ministero della Salute, 2002). Dalla definizione proposta, si evince immediatamente come gli studi osservazionali costituiscano lo strumento privilegiato per valutare l'*effectiveness* (efficacia pratica) delle terapie croniche.

Gli studi osservazionali non si pongono in contrapposizione con gli studi sperimentali anzi, sotto alcuni punti di vista, ne costituiscono il completamento. Le due metodologie di analisi, infatti, presentano obiettivi e connotati profondamente diversi. Gli studi sperimentali sono disegnati per analizzare la

relazione tra processo terapeutico ed esiti clinici in condizioni ideali, strettamente controllate dagli sperimentatori che, nelle visite periodiche di follow-up, monitorizzano e incentivano sistematicamente l'aderenza del paziente alla terapia farmacologica, riducendo al minimo tutte le possibili deviazioni dal protocollo.

Queste particolari condizioni sono necessarie per eliminare l'effetto di possibili "fattori di disturbo" e per definire gli standard terapeutici necessari per ottenere il risultato desiderato. Gli studi osservazionali, invece, sono condotti per valutare l'efficacia degli interventi terapeutici in pratica clinica, cioè in condizioni reali. Questi studi sono utili alla misurazione e valutazione dei processi terapeutici, degli esiti clinici e dell'allocazione delle risorse nelle popolazioni reali. Secondo queste premesse, gli studi di efficacia teorica e di efficacia pratica dovrebbero essere sviluppati entrambi in quanto complementari: gli uni per fissare le regole relative a "come dovremmo trattare", gli altri per verificare la conformità tra standard terapeutici e processi applicati (Black, 1996; Revicki e Frank, 1999).

Tra le diverse tipologie di studi osservazionali, rivestono un ruolo particolarmente rilevante per il governo clinico delle patologie croniche gli studi di farmacoutilizzazione, finalizzati ad esprimere quantitativamente le modalità con cui i farmaci vengono prescritti e assunti in pratica clinica. Attraverso questi studi le terapie farmacologiche vengono valutate e confrontate con gli standard di trattamento sulla base di alcune caratteristiche particolarmente rilevanti nelle patologie croniche, quali la persistenza (durata della terapia), l'aderenza (dose media giornaliera assunta), la ciclicità del trattamento.

Gli studi di farmacoutilizzazione consentono, pertanto, di identificare eventuali forme di non appropriatezza, di quantificare l'allontanamento dagli standard terapeutici, di esprimere giudizi e operare confronti con altre realtà sanitarie, consentono di valutare e, se necessario, ridefinire i criteri di allocazione delle risorse.

2. Analisi multilevel: contesti applicativi e metodologia

2.1. Introduzione alle analisi multilevel

Negli studi di epidemiologia e farmacoepidemiologia si analizza frequentemente la relazione tra individuo e contesto socio-sanitario. Il concetto generale sottostante questa tipologia di analisi è che gli individui interagiscono con il contesto sociale cui appartengono: da un lato, gli individui sono influenzati dai gruppi sociali di appartenenza, dall'altro, le caratteristiche di questi gruppi dipendono, a loro volta, dagli individui che ne fanno parte. Generalmente gli individui e i gruppi sociali sono concettualizzati come un sistema gerarchico in cui gli individui e i gruppi sono definiti su livelli differenti di questo sistema gerarchico. Un simile sistema può essere osservato su differenti livelli della gerarchia e le variabili oggetto di interesse possono essere definite su ciascuno di questi livelli. Nelle analisi *multilevel* la struttura dei dati nella popolazione è gerarchica e i dati campionari sono visti come risultato di un campionamento a più stadi da questa popolazione gerarchicamente strutturata. Pertanto, nel campo della ricerca osservazionale sull'utilizzo dei farmaci, la popolazione può essere costituita dai medici di medicina generale e dai pazienti afferenti ai medici stessi, e la procedura di campionamento si sviluppa in due stadi: nella prima fase si assume un campione di medici, nella seconda fase si estrae un campione di pazienti da ciascuno dei medici selezionati. In questo contesto, i pazienti sono "annidati" all'interno dei propri medici di medicina generale e l'analisi *multilevel* riflette una struttura gerarchica realmente esistente. In altri contesti, tuttavia, l'annidamento può anche essere rappresentato da una serie di misure ripetute all'interno di singoli individui (analisi di dati longitudinali) o da un collettivo di soggetti all'interno di differenti studi scientifici (meta-analisi).

Storicamente, le analisi di dati gerarchicamente organizzati sono state spesso realizzate mediante tecniche standard come l'analisi della varianza o la regressione multipla dopo aver spostato tutte le variabili su un solo livello di interesse. Ciò avveniva mediante due distinte procedure: aggregazione e

disaggregazione. L'aggregazione è lo spostamento di variabili originariamente osservate su un basso livello della gerarchia verso un livello superiore (nelle valutazioni medico-paziente, l'aggregazione sarebbe ottenuta assegnando a ciascun medico le medie di variabili osservate sui rispettivi pazienti). Al contrario, la disaggregazione è lo spostamento di variabili verso un livello più basso della gerarchia (ad esempio assegnando a tutti i pazienti variabili che esprimono le caratteristiche dei medici di medicina generale cui afferiscono).

Tuttavia, analizzare variabili che appartengono a differenti livelli della gerarchia su un singolo e comune livello è inadeguato per due ragioni.

La prima è di natura statistica. Da un lato, l'aggregazione comporta una sostanziale perdita di informazioni e, di conseguenza, l'analisi statistica perde potenza. Dall'altro, se i dati vengono disaggregati, i test statistici ordinari considerano i valori disaggregati come informazioni indipendenti provenienti dal vasto insieme di unità di basso livello. Eppure, la reale numerosità campionaria per le variabili disaggregate è costituita dal più basso numero di unità di livello superiore. Pertanto, l'utilizzo dell'ampio insieme di casi disaggregati come numerosità campionaria porta ad ottenere risultati statisticamente significativi che sono totalmente spuri.

La seconda ragione è di natura concettuale, consiste nell'analizzare i dati ad un certo livello e formulare le conclusioni ad un altro livello (fallacia del livello sbagliato). Questo tipo di errore interpretativo può assumere sostanzialmente due forme: la "*ecological fallacy*" che consiste nell'interpretare dati aggregati come se fossero dati individuali e la "*atomistic fallacy*", problema in cui si incorre quando si formulano inferenze su un livello elevato della gerarchia basandosi su analisi realizzate a un livello inferiore (Alker, 1969).

2.2. La correlazione intraclassa

La caratteristica principale delle analisi multilevel è costituita dal fatto che, trattandosi di dati gerarchicamente organizzati, in genere le osservazioni individuali non sono del tutto indipendenti. Nel contesto dell'utilizzo dei farmaci

per le patologie croniche, i pazienti che afferiscono al medesimo medico di medicina generale tenderanno ad essere tra loro simili a causa di eventuali processi di selezione e a causa della “storia comune” che i pazienti condividono afferendo allo stesso medico. Ne consegue che la correlazione media tra pazienti iscritti allo stesso medico di medicina generale risulterà più elevata di quella tra pazienti che afferiscono a medici differenti. Questa particolare forma di correlazione prende il nome di *correlazione intraclassa*. I test statistici tradizionali sono fortemente basati sull’assunto di indipendenza tra le osservazioni. Se questa ipotesi risulta violata, le stime degli errori standard prodotte dai test statistici convenzionali sono troppo piccole e, di conseguenza, i risultati che si ottengono appaiono “impropriamente” significativi. La dipendenza tra le osservazioni individuali può essere considerata come un fattore che “riduce” la numerosità campionaria effettiva. Considerando un campionamento a due stadi in cui tutti i gruppi sono costituiti dallo stesso numero di unità elementari, la numerosità campionaria effettiva n_{eff} può essere calcolata come segue (Kish, 1965):

$$n_{eff} = n / [1 + (n_{clus} - 1)\rho] \quad (2.1)$$

dove n è la numerosità campionaria totale, n_{clus} è la dimensione di ciascun gruppo e ρ è la correlazione intraclassa. Tuttavia, le correzioni per gli effetti da disegno come quella proposta da Kish presentano due pesanti limiti. In primo luogo, la correlazione intraclassa varia al variare della variabile di interesse. In secondo luogo, i problemi relativi all’analisi di strutture gerarchiche sono in genere resi più complessi dalla presenza di variabili misurate su tutti i livelli della gerarchia.

Emerge, quindi, la necessità di utilizzare un modello statistico che tenga conto della non indipendenza delle osservazioni e che consenta, allo stesso tempo, di analizzare simultaneamente variabili che “provengono” da diversi livelli della gerarchia.

2.3. Aspetti metodologici dei modelli multilevel

I modelli di regressione multilevel sono noti in letteratura sotto una varietà di nomi come *modelli a coefficienti casuali*, *modelli a componenti della varianza*, *modelli lineari gerarchici*, *modelli a effetti misti*. Questi modelli, pur non essendo esattamente identici, risultano tra loro molto simili: tutti assumono che esiste una struttura gerarchica nei dati, con una singola variabile risposta misurata al livello inferiore e variabili esplicative potenzialmente misurate su tutti i livelli esistenti nella gerarchia. Concettualmente, il modello di regressione multilevel può essere visto come un sistema gerarchico di equazioni di regressione. Sarà di seguito presentata la struttura dei modelli multilevel per dati su due livelli.

Sia J il numero dei gruppi, con un numero differente n_j di individui all'interno di ciascun gruppo. Sia Y la variabile risposta misurata al livello più basso della gerarchia (rilevata, quindi, per ciascun individuo), sia X una variabile esplicativa misurata sul livello degli individui e sia Z una variabile esplicativa misurata sul livello dei gruppi. Il livello degli individui costituisce il primo livello, quello dei gruppi il secondo livello. Sarà utilizzato l'indice j per i gruppi ($j = 1, \dots, J$) e l'indice i per gli individui ($i = 1, \dots, n_j$).

La stima dei parametri di un modello multilevel che si adatti alla situazione appena descritta può essere concettualmente distinta in due fasi successive.

Nella prima fase, vengono realizzati, all'interno di ciascun gruppo, modelli di regressione separati, al fine di predire la variabile risposta Y in funzione della variabile esplicativa X come segue:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (2.2)$$

In questa equazione di regressione β_{0j} è la usuale intercetta, β_{1j} è il classico coefficiente di regressione per la variabile esplicativa X misurata sul livello degli individui mentre e_{ij} costituisce il solito termine d'errore. Si assume che, all'interno di ciascun gruppo, gli errori al livello individuale siano indipendenti e normalmente distribuiti con media nulla e varianza σ_e^2 .

$$e_{ij} \sim N(0, \sigma_e^2)$$

La differenza rispetto al modello di regressione non gerarchico sta nell'assunzione che ogni gruppo ha una differente intercetta β_{0j} e un differente coefficiente di regressione β_{1j} . Dal momento che variano tra le unità di livello superiore, questi coefficienti sono spesso denominati coefficienti *casuali*. Le unità di secondo livello sono quindi viste come un campione proveniente da una più vasta popolazione di gruppi. Il passo successivo nel modello di regressione gerarchico è quello di spiegare la variazione dei coefficienti β_{0j} e β_{1j} introducendo le variabili esplicative misurate sul livello superiore, come segue:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (2.3)$$

e

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (2.4)$$

Si assume che i termini di errore nelle equazioni di regressione a livello di gruppo u_{0j} e u_{1j} , spesso denominati macro-errori, siano normalmente distribuiti con media nulla e varianze σ_{u0}^2 e σ_{u1}^2 , rispettivamente. Si assume inoltre che i macro-errori siano indipendenti tra i gruppi e indipendenti dagli errori di livello individuale e_{ij} . σ_{u01} rappresenta la covarianza tra i macro-errori u_{0j} e u_{1j} .

$$u_{0j} \sim N(0, \sigma_{u0}^2); \quad u_{1j} \sim N(0, \sigma_{u1}^2); \quad \text{cov}(u_{0j}, u_{1j}) = \sigma_{u01}$$

I coefficienti γ_{00} , γ_{10} , γ_{01} e γ_{11} , non avendo la caratteristica di variare tra le unità di un livello superiore, sono detti coefficienti *fissi*.

Sostituendo le equazioni (2.3) e (2.4) nella equazione (2.2) il modello di regressione multilevel può essere esplicitato come una singola equazione di regressione, come segue:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} + e_{ij} \quad (2.5)$$

Il termine $X_{ij}Z_j$ è denominato interazione *cross-level* poiché coinvolge variabili esplicative misurate su differenti livelli della gerarchia.

La parte $[\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j]$ nell'equazione (2.5) contiene i coefficienti fissi ed è, pertanto, denominata parte fissa del modello. Il segmento $[u_{1j}X_{ij} + u_{0j} + e_{ij}]$ nell'equazione (2.5) contiene invece i termini di errore casuali e viene denominato parte casuale del modello. Questo segmento costituisce una struttura complessa di errore e, come si può notare dalla formula, gli errori per le osservazioni all'interno dei gruppi sono correlati poiché u_{0j} e u_{1j} risultano comuni per le osservazioni che appartengono al medesimo gruppo.

Dall'equazione di regressione (2.5) si nota, inoltre, che il termine d'errore u_{1j} è connesso con X_{ij} . Dal momento che la variabile esplicativa X_{ij} e il termine di errore u_{1j} sono fra loro moltiplicati, l'errore totale che ne risulta sarà differente per differenti valori di X_{ij} , una situazione che, nei modelli di regressione ordinari, prende il nome di "eteroschedasticità". Risultano, pertanto, violate le assunzioni di indipendenza e di omoschedasticità degli errori su cui si basano i modelli di regressione ordinari.

Attraverso l'equazione di regressione (2.5) è, dunque, possibile stimare, attraverso i coefficienti a effetti fissi, gli effetti indipendenti delle variabili di secondo livello, delle variabili di primo livello e la loro interazione. Il modello multilevel, inoltre, permette di quantificare la variabilità nei diversi livelli della gerarchia: variabilità entro il gruppo, espressa dalla varianza σ^2_e ; variabilità tra i gruppi, espressa dalle varianze degli effetti casuali σ^2_{u0} e σ^2_{u1} . Gli effetti stimati dal modello possono essere sintetizzati come segue:

parte fissa

γ_{00} : risposta media in Y (tra tutti i gruppi e tutti gli individui) nel caso in cui sia X che Z presentano valore zero;

γ_{01} : effetto di Z sulla risposta media in Y ;

γ_{10} : effetto di X su Y nel caso in cui Z presenta valore zero;

γ_{11} : effetto di Z sull'effetto di X su Y ;

parte casuale

σ_e^2 : varianza intra-classe (tra le unità di livello inferiore) controllando per l'effetto di X ;

σ_{u0}^2 : varianza condizionata dell'intercetta di primo livello, esprime la variabilità della risposta media tra le unità di secondo livello, controllando per X e Z ;

σ_{u1}^2 : varianza condizionata del coefficiente di regressione di primo livello, esprime la variabilità dell'effetto di X su Y tra le unità di secondo livello, controllando per Z ;

σ_{u01} : covarianza condizionata tra intercetta e coefficiente di regressione di primo livello.

Nei casi in cui, dopo l'aggiustamento per le covariate nel modello, la variabilità residua tra le unità di secondo livello delle intercette e dei coefficienti di regressione risulta trascurabile, i macro-errori nelle equazioni (2.3) e (2.4) si approssimano tutti allo zero e tendono a zero anche le stime delle varianze degli effetti casuali σ_{u0}^2 e σ_{u1}^2 . In una simile circostanza, il modello di regressione multilevel (2.5) si riduce ad una regressione multipla tradizionale che include variabili indipendenti misurate sia sul primo che sul secondo livello, in cui tutti i coefficienti sono fissi, senza componenti casuali al secondo livello. In questo caso gli individui all'interno dei gruppi possono essere considerati indipendenti,

condizionatamente alle covariate di primo e di secondo livello incluse nel modello. Al contrario, la persistenza di una variabilità significativa tra le intercette o tra i coefficienti di regressione dopo l'inclusione di variabili di secondo livello, potrebbe esprimere la necessità di analizzare altri fattori di secondo livello potenzialmente responsabili di questa variabilità residua, assumendo che tutte le covariate di primo livello che possono comportare differenze tra i gruppi siano già state incluse nell'equazione (2.2).

2.4. Il modello multilevel a sola intercetta

Nel caso in cui si analizzi una struttura dati gerarchica senza alcuna variabile esplicativa, il modello di regressione multilevel può essere utilizzato per produrre una stima della correlazione intraclasse. Per queste finalità si utilizza il modello a sola intercetta. Se non ci sono variabili esplicative X , l'equazione (2.2) si riduce a:

$$Y_{ij} = \beta_{0j} + e_{ij} \quad (2.6)$$

Allo stesso modo, se non esistono variabili esplicative di secondo livello Z , l'equazione (2.3) si riduce a:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2.7)$$

Sostituendo la (2.7) nella (2.6) si ottiene:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (2.8)$$

Il modello a sola intercetta (2.8) non spiega alcuna varianza in Y . Semplicemente scompone la varianza in due componenti indipendenti: σ_e^2 (la varianza degli errori di primo livello e_{ij}) e σ_{u0}^2 (la varianza degli errori di secondo livello u_{0j}).

Utilizzando questo modello è possibile definire la correlazione intraclasse ρ mediante la seguente equazione:

$$\rho = \sigma_{u0}^2 / (\sigma_{u0}^2 + \sigma_e^2) \quad (2.9)$$

Il denominatore del secondo membro della equazione (2.9) costituisce la varianza totale. La correlazione intraclasse ρ indica, quindi, la proporzione di varianza spiegata dalla struttura in gruppi della popolazione. La correlazione intraclasse ρ può anche essere interpretata come la correlazione attesa tra due unità campionate casualmente che si trovano nello stesso gruppo.

2.5. Stima dei parametri nei modelli multilevel

La stima dei parametri nei modelli multilevel (intercetta, coefficienti di regressione e componenti della varianza) viene generalmente realizzata attraverso il metodo della Massima Verosimiglianza. Questo metodo produce stime asintoticamente efficienti e consistenti. Inoltre, con grandi campioni, le stime di Massima Verosimiglianza sono generalmente robuste rispetto a leggere violazioni dell'assunzione di normalità distributiva degli errori. Tuttavia, in alcuni contesti di analisi, il metodo della Massima Verosimiglianza non è ottimale e risulta più opportuno utilizzare metodi di stima alternativi quali l'approccio bayesiano, il metodo dei minimi quadrati generalizzati, le equazioni di stima generalizzate, lo stimatore di Huber/White.

2.5.1. Il metodo della massima verosimiglianza

Il metodo di stima della Massima Verosimiglianza (*Maximum Likelihood* - ML) consiste nel massimizzare la Funzione di Verosimiglianza, definita in base alla probabilità di osservare una data realizzazione campionaria condizionatamente ai valori assunti dai parametri oggetto di stima. Nell'ambito dei modelli di regressione multilevel esistono due differenti Funzioni di Verosimiglianza. Una viene chiamata Funzione di Verosimiglianza Completa (*Full Maximum Likelihood* - FML) nella quale sono inclusi sia i coefficienti di regressione che le componenti

della varianza. L'altra viene chiamata Funzione di Verosimiglianza Ristretta (*Restricted Maximum Likelihood* - RML); nel metodo RML soltanto le componenti della varianza sono incluse nella funzione di verosimiglianza e i coefficienti di regressione vengono stimati in una fase successiva. Entrambi i metodi producono le stime puntuali dei parametri, gli errori standard e la devianza complessiva del modello. Il metodo FML, nel momento in cui vengono stimate le componenti della varianza, tratta i coefficienti di regressione come quantità fisse e incognite ma non tiene in considerazione i gradi di libertà persi stimando gli effetti fissi. Il metodo RML, invece, stima le componenti della varianza dopo la rimozione degli effetti fissi dal modello. Ne consegue che le stime FML delle componenti della varianza sono distorte e risultano generalmente troppo piccole. Le stime RML sono meno distorte e hanno la proprietà, nelle situazioni in cui i gruppi siano perfettamente bilanciati, di coincidere con le stime, ottimali, ottenute mediante l'analisi della varianza. Pertanto, sotto il punto di vista teorico, il metodo RML produce stime migliori rispetto al metodo FML, soprattutto nei casi in cui il numero dei gruppi è piccolo. In pratica, tuttavia, le differenze tra i due metodi di stima sono piuttosto contenute e il metodo FML continua ad essere frequentemente utilizzato poiché presenta due vantaggi rispetto al metodo RML. In primo luogo, i processi computazionali sono generalmente più rapidi. In secondo luogo, dal momento che i coefficienti di regressione sono inclusi nella funzione di verosimiglianza, è possibile utilizzare un test chi-quadrato basato sulle verosimiglianze per confrontare modelli annidati che differiscono sia nella parte fissa che nella parte casuale. Al contrario, col metodo RML, il test chi-quadrato basato sulle verosimiglianze può essere utilizzato per confrontare modelli che differiscono esclusivamente nella parte casuale. Per una accurata trattazione dei metodi FML e RML si rimanda a Searle, Casella, McCulloch, 1992 e Longford, 1993.

Il calcolo delle stime di massima verosimiglianza richiede una procedura iterativa. Al primo passo della procedura vengono generati valori iniziali per i diversi parametri, generalmente ottenuti stimando i parametri incogniti come se la regressione riguardasse un singolo livello gerarchico. Nei passaggi successivi, le

stime iniziali vengono iterativamente migliorate fino al momento in cui la procedura di stima converge. Tuttavia, una ridotta numerosità campionaria o una non corretta specificazione del modello potrebbero causare problemi di convergenza impedendo l'arresto della procedura di stima. Nelle analisi multilivello, inoltre, la mancata convergenza dell'algoritmo è spesso attribuibile alla presenza, nel modello, di molte componenti della varianza prossime allo zero.

In generale, oltre ai problemi di convergenza, i metodi di stima della Massima Verosimiglianza non sono adatti nelle situazioni in cui la numerosità campionaria è molto piccola o nelle situazioni in cui risultano violati gli assunti di normalità distributiva degli errori.

2.5.2. L'approccio bayesiano alle analisi multilevel

Nei casi in cui la numerosità campionaria risulta particolarmente esigua o nei casi in cui gli errori non sono normalmente distribuiti, i metodi di stima bayesiani costituiscono una efficace alternativa ai metodi di stima basati sulla Massima Verosimiglianza.

Nella statistica Bayesiana, l'incertezza sui valori in popolazione dei parametri del modello multilevel (coefficienti di regressione e componenti della varianza) viene espressa assegnando ai parametri stessi una distribuzione di probabilità. Tale distribuzione viene generalmente chiamata distribuzione *a priori*, dal momento che viene specificata indipendentemente dai dati campionari. Dopo l'osservazione dei risultati campionari, la distribuzione *a priori* viene combinata con la Verosimiglianza dei dati al fine di produrre una distribuzione *a posteriori* che esprime l'incertezza sui parametri di popolazione dopo l'osservazione dei dati campionari. Generalmente, la varianza della distribuzione *a posteriori* è inferiore rispetto alla varianza della distribuzione *a priori*, in quanto l'osservazione dei risultati campionari riduce l'incertezza sui parametri da stimare. Per la scelta della distribuzione *a priori*, esistono due principali alternative: *a priori* informative e *a priori* non informative. Le distribuzioni *a priori* informative sono caratterizzate da

una piccola varianza, esprimono una forte credenza a priori sui parametri non noti della popolazione e influenzano notevolmente la distribuzione a posteriori.

Le distribuzioni a priori non informative hanno, per contro, una scarsissima influenza sulla distribuzione a posteriori ed hanno la sola funzione di produrre, combinate con la verosimiglianza, la distribuzione a posteriori. Un esempio di a priori non informativa è costituito dalla distribuzione uniforme definita su tutto l'asse reale, con la quale si afferma semplicemente che il parametro non noto della popolazione può variare tra meno infinito e più infinito, con tutti i valori ugualmente probabili. Un altro esempio di a priori non informativa è costituito dalla distribuzioni normale con varianza infinita.

Una volta determinata la distribuzione a posteriori, nei casi in cui questa abbia una forma matematicamente semplice, è possibile utilizzare le caratteristiche della distribuzione stessa per calcolare una stima puntuale e un intervallo di confidenza (bayesiano) per il parametro di popolazione sconosciuto. Nel caso di una distribuzione a posteriori normale sarà, pertanto, possibile utilizzare la media come stima puntuale del parametro incognito e basare la costruzione dell'intervallo di confidenza sulla deviazione standard della distribuzione a posteriori.

Si consideri il tradizionale modello lineare

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

con

\mathbf{Y} vettore ($n \times 1$) aleatorio di variabili osservabili

\mathbf{X} matrice ($n \times k$) non stocastica di rango k

$\boldsymbol{\beta}$ vettore ($k \times 1$) di parametri aleatori

$\boldsymbol{\varepsilon}$ vettore ($n \times 1$) di errori aleatori.

In generale, si assume che $\boldsymbol{\varepsilon}$ si distribuisca normalmente con

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} ; E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_n$$

Si avrà dunque

$$\mathbf{Y}|\mathbf{X},\boldsymbol{\beta},\sigma^2 \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

con funzione di verosimiglianza data da:

$$p(\mathbf{Y}|\mathbf{X},\boldsymbol{\beta},\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

Per poter procedere all'inferenza sui parametri $\boldsymbol{\beta}$ e σ^2 è necessario assegnare loro una distribuzione iniziale e, mediante la formula di Bayes, determinare la distribuzione finale di $(\boldsymbol{\beta}, \sigma^2)$.

La presenza di una distribuzione di probabilità a priori assegnata ai coefficienti di regressione rende i modelli lineari bayesiani concettualmente simili ai modelli a coefficienti casuali.

Le distribuzioni a priori possono essere non informative alla Jeffreys, informative coniugate, informative non coniugate.

Nel caso in cui σ^2 sia nota,

- assegnando una distribuzione non informativa alla Jeffreys ai parametri $\boldsymbol{\beta}$, si ottiene, come distribuzione finale, una normale multivariata k -dimensionale;
- assumendo come distribuzione iniziale coniugata al modello per $\boldsymbol{\beta}$ una normale multivariata si ottiene, come distribuzione finale, ancora una normale multivariata.

Nel caso in cui σ^2 sia non nota,

- assegnando una distribuzione non informativa alla Jeffreys ai parametri (β , σ^2) si ottiene, come distribuzione finale marginale per σ^2 , una gamma inversa, come distribuzione finale marginale per β , una t di Student k-dimensionale con $n - k$ gradi di libertà;
- assumendo come distribuzione iniziale congiunta per β e σ^2 coniugata al modello una normal-gamma si ottiene, come distribuzione finale marginale per β , una t di Student k-dimensionale.

In molte situazioni è avvertita la necessità di incorporare nell'analisi opinioni "a priori" riguardo ad alcuni coefficienti del modello lineare, opinioni che mal si esprimono con le distribuzioni iniziali precedentemente presentate.

D'altra parte, il ricorso a distribuzioni informative non coniugate farebbe sorgere grossi problemi computazionali. Una interessante mediazione tra il problematico utilizzo di informative non coniugate e l'esigenza di incorporare nell'analisi informazioni a priori è resa possibile dalle informazioni iniziali di tipo gerarchico.

Nel modello lineare l'idea è la seguente: sempre nell'ambito del modello normale (primo stadio) si assegnano distribuzioni normali ai coefficienti (secondo stadio), queste distribuzioni sono specificate a meno di parametri aleatori a cui ancora una volta si assegna una distribuzione normale (terzo stadio) e così via.

Con una simile struttura, si utilizza una a priori di tipo gerarchico, poiché la distribuzione a priori dei coefficienti del modello dipenderà da parametri, definiti *iperparametri*, ai quali si assegnerà una ulteriore distribuzione a priori, chiamata *iper a priori*.

Questo procedimento potrebbe seguire un numero arbitrario di volte, tuttavia, le informazioni che possono essere incorporate in successivi stadi vanno riducendosi sempre più. Per questa ragione è spesso conveniente assegnare una distribuzione iniziale non informativa al terzo stadio.

Nei modelli lineari gerarchici bayesiani, al contrario dei modelli multilevel tradizionali, la gerarchia è logica più che strutturale. Non è, infatti, necessario che i dati siano realmente organizzati secondo un annidamento gerarchico. Questa caratteristica conferisce ai modelli gerarchici bayesiani una maggiore flessibilità di utilizzo.

Il modello a due stadi

Dato il modello $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, sia \mathbf{Y} , dato $\boldsymbol{\beta}_1$, distribuito normalmente con valore atteso $\mathbf{X}_1\boldsymbol{\beta}_1$ e matrice di varianze e covarianze \mathbf{C}_1 :

$$\mathbf{Y}|\boldsymbol{\beta}_1 \sim N(\mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{C}_1) \text{ (Primo stadio)}$$

Si assuma poi che $\boldsymbol{\beta}_1$, dato $\boldsymbol{\beta}_2$, sia distribuito normalmente:

$$\boldsymbol{\beta}_1|\boldsymbol{\beta}_2 \sim N(\mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{C}_2) \text{ (Secondo stadio)}$$

Si assume che tutte le matrici di varianze e covarianze \mathbf{C}_i siano definite positive e di dimensioni appropriate. Si noti come, in questa struttura, la distribuzione a priori per $\boldsymbol{\beta}_1$ dipenda dall'iperparametro $\boldsymbol{\beta}_2$.

Applicando il teorema di Bayes:

$$p(\boldsymbol{\beta}_1|\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\beta}_1) p(\boldsymbol{\beta}_1)$$

con \mathbf{X}_1 , \mathbf{C}_1 , \mathbf{X}_2 , \mathbf{C}_2 , $\boldsymbol{\beta}_2$ noti, si ottiene la distribuzione finale:

$$\boldsymbol{\beta}_1|\mathbf{Y} \sim N(\mathbf{V}_0\mathbf{b}_0, \mathbf{V}_0)$$

dove

$$\mathbf{V}_0 = (\mathbf{C}_2^{-1} + \mathbf{X}_1' \mathbf{C}_1^{-1} \mathbf{X}_1)^{-1}$$

$$\mathbf{b}_0 = (\mathbf{X}_1' \mathbf{C}_1^{-1} \mathbf{y} + \mathbf{C}_2^{-1} \mathbf{X}_2 \boldsymbol{\beta}_2)$$

Finora è stato assunto che le matrici di varianze e covarianze nei diversi stadi fossero note. Accade spesso che questa ipotesi non sia accettabile. In tal caso è necessario assegnare una distribuzione iniziale anche a tali componenti.

Il modello a tre stadi

Dato il modello $\mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, sia \mathbf{Y} , dato $\boldsymbol{\beta}_1$, distribuito normalmente con valore atteso $\mathbf{X}_1 \boldsymbol{\beta}_1$ e matrice di varianze e covarianze \mathbf{C}_1 :

$$\mathbf{Y} | \boldsymbol{\beta}_1 \sim N(\mathbf{X}_1 \boldsymbol{\beta}_1, \mathbf{C}_1) \text{ (Primo stadio)}$$

$$\boldsymbol{\beta}_1 | \boldsymbol{\beta}_2 \sim N(\mathbf{X}_2 \boldsymbol{\beta}_2, \mathbf{C}_2) \text{ (Secondo stadio)}$$

$$\boldsymbol{\beta}_2 | \boldsymbol{\beta}_3 \sim N(\mathbf{X}_3 \boldsymbol{\beta}_3, \mathbf{C}_3) \text{ (Terzo stadio)}$$

con \mathbf{X}_i e \mathbf{C}_i ($i = 1, 2, 3$) note, $\boldsymbol{\beta}_3$ noto.

Dove

$$\boldsymbol{\beta}_1 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{u}_1 \quad ; \quad \mathbf{u}_1 \sim N(0, \mathbf{C}_2)$$

$$\boldsymbol{\beta}_2 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \mathbf{u}_2 \quad ; \quad \mathbf{u}_2 \sim N(0, \mathbf{C}_3)$$

Una approfondita trattazione del modello a tre stadi è presente nel Cifarelli e Muliere, 1989, p. 245.

Nei contesti reali di analisi, soprattutto nei casi in cui risultano violati gli assunti di normalità distributiva, non è sempre possibile ricondursi a una distribuzione a posteriori dalla forma matematicamente semplice. In queste situazioni la distribuzione a posteriori, generalmente multivariata, può essere approssimata utilizzando le procedure di simulazione *Markov Chain Monte Carlo* (MCMC). Queste tecniche, piuttosto impegnative sotto il punto di vista computazionale, generano campioni casuali tratti da complesse distribuzioni a posteriori. Attraverso un elevato numero di campioni casuali è, infatti, possibile approssimare la reale forma delle distribuzioni. Le distribuzioni a posteriori simulate vengono successivamente utilizzate per calcolare le stime puntuali e gli intervalli di confidenza. Solitamente, viene impiegata la distribuzione marginale di ciascun parametro. La moda della distribuzione a posteriori marginale viene frequentemente utilizzata come stima puntuale del parametro incognito. Questa soluzione costituisce l'equivalente bayesiano dello stimatore di massima verosimiglianza. In alcune situazioni, tuttavia, si preferisce utilizzare la media della distribuzione a posteriori poiché più semplice da calcolare, o la mediana, nel caso di distribuzioni a posteriori asimmetriche.

2.5.3. Altri metodi per la stima dei parametri

Minimi quadrati generalizzati

Le stime dei minimi quadrati generalizzati (*Generalized Least Squares* - GLS) possono essere ottenute da una procedura di massima verosimiglianza restringendo il numero di iterazioni a uno. Le stime ottenute col metodo dei minimi quadrati generalizzati sono asintoticamente equivalenti alle stime di massima verosimiglianza, pertanto, su grandi campioni, le due procedure producono stime praticamente indistinguibili. Dal momento che le stime GLS

sono molto più rapide da calcolare rispetto alle stime FML, queste possono essere utilizzate come stime provvisorie nel caso in cui l'algoritmo di massima verosimiglianza risulti particolarmente lento, come nel caso di *data set* estremamente grandi. Inoltre, il metodo dei minimi quadrati generalizzati risulta particolarmente utile nei casi in cui le procedure di massima verosimiglianza non convergono. L'analisi dei risultati GLS può essere molto utile al fine di individuare il problema che determina la non convergenza dell'algoritmo di massima verosimiglianza.

Tuttavia, ricerche basate su simulazioni, hanno evidenziato come gli stimatori GLS siano meno efficienti rispetto agli stimatori ML, pertanto, qualora non sussistano problemi di rapidità computazionale o di convergenza, il metodo della massima verosimiglianza risulta certamente preferibile.

Equazioni di stima generalizzate

Il metodo delle equazioni di stima generalizzate (*Generalized Estimating Equations* - GEE) stima le varianze e le covarianze della parte casuale dei modelli multilivello direttamente dai residui. Questo determina una maggiore rapidità computazionale rispetto al metodo FML. Una volta stimate le componenti della varianza attraverso il metodo GEE, i coefficienti di regressione della parte fissa del modello vengono stimati attraverso il metodo dei minimi quadrati generalizzati.

Nei casi in cui la numerosità campionaria sia ragionevolmente elevata, gli stimatori GEE risultano molto robusti rispetto alla violazione dell'assunto di normalità distributiva degli errori.

Tuttavia, se gli errori sono normalmente distribuiti, gli stimatori GEE sono meno efficienti rispetto agli stimatori di massima verosimiglianza. Dal momento che anche gli stimatori di massima verosimiglianza sono piuttosto robusti rispetto a leggere violazioni dell'assunto di normalità distributiva, il metodo GEE non presenta evidenti vantaggi, se non la maggiore rapidità computazionale, rispetto ai metodi basati sulla massima verosimiglianza.

Stimatore di Huber/White

Nei casi in cui la variabile dipendente non si distribuisce normalmente le stime puntuali prodotte dal metodo della massima verosimiglianza sono ancora consistenti e asintoticamente non distorte. Le stime degli errori standard, tuttavia, non sono attendibili e non possono essere utilizzate per la verifica di ipotesi né per la costruzione di intervalli di confidenza. Questo problema non si attenua all'aumentare della numerosità campionaria.

Nelle situazioni in cui non è conveniente trasformare la variabile risposta, una possibile alternativa è costituita dallo stimatore di Huber/White, un metodo di correzione che produce errori standard robusti. Nel metodo della massima verosimiglianza lo stimatore delle varianze e delle covarianze campionarie è basato sulla matrice di informazione o, più in generale, sull'inversa della matrice Hessiana. Gli errori standard sono, pertanto, costituiti dalla radice quadrata delle varianze campionarie poste sulla diagonale principale di tale inversa. Utilizzando la notazione matriciale, la matrice di varianze e covarianze delle stime dei coefficienti di regressione può essere scritta come segue:

$$\mathbf{V}_{ML}(\hat{\boldsymbol{\beta}}) = \mathbf{H}^{-1}$$

dove \mathbf{H} è la matrice Hessiana.

Lo stimatore di Huber/White è dato da:

$$\mathbf{V}_R(\hat{\boldsymbol{\beta}}) = \mathbf{H}^{-1}\mathbf{C}\mathbf{H}^{-1} \quad (2.10)$$

dove \mathbf{V}_R è la matrice di varianze e covarianze robuste dei coefficienti di regressione mentre \mathbf{C} è una matrice di correzione.

Vista la posizione centrale della matrice di correzione \mathbf{C} rispetto alle due matrici inverse \mathbf{H}^{-1} , lo stimatore di Huber/White espresso dall'equazione (2.10) viene spesso denominato stimatore *sandwich*.

Il termine di correzione si basa sui residui osservati. Se i residui sono normalmente distribuiti, V_{ML} e V_R sono entrambi stimatori consistenti delle varianze e covarianze dei coefficienti di regressione ma V_{ML} risulta più efficiente di V_R . Quando i residui non seguono una distribuzione normale, V_{ML} non è più attendibile mentre V_R è ancora stimatore consistente delle varianze e covarianze dei coefficienti di regressione.

Per una approfondita trattazione tecnica dello stimatore di Huber/White si rimanda a Greene, 1997. L'utilizzo dello stimatore di Huber/White per correggere gli errori standard delle componenti della varianza è trattato nel Goldstein, 1995.

La breve esposizione dei metodi di stima nei modelli multilevel presentata in questo lavoro non esaurisce tutte le soluzioni disponibili in letteratura. Tra i metodi non presentati in questo paragrafo, rivestono particolare interesse il *profile likelihood method* e la tecnica del *bootstrapping*, per la cui trattazione si rimanda a Hox, 2002.

2.6. Verifica di ipotesi nei modelli multilevel

Test di Wald

Uno dei test più utilizzati per la verifica di ipotesi nei modelli di regressione multilevel è il *test di Wald* in cui la statistica test Z viene calcolata rapportando la stima puntuale del parametro di interesse all'errore standard della stima stessa. La distribuzione di riferimento per la statistica Z è la normale standardizzata.

Il test di Wald si basa sull'assunto che i parametri sottoposti a verifica di ipotesi abbiano una distribuzione campionaria normale, con una varianza campionaria che può essere stimata a partire dalla matrice di informazione. Come discusso da Fears et al. (1996), in situazioni particolari, la statistica di Wald non risulta adatta a testare le componenti della varianza, soprattutto nei casi in cui

queste siano prossime allo zero o nei casi in cui la numerosità campionaria sia molto ridotta. Si precisa, inoltre, che gli errori standard utilizzati per la costruzione del test sono di natura asintotica, pertanto sono validi per grandi campioni. Non si conosce con precisione quale sia la numerosità campionaria sufficiente affinché gli errori standard possano essere considerati accurati; per gli approfondimenti su questo tema si rimanda agli studi di simulazione condotti da Van der Leeden et al. (1997).

E', comunque, opportuno precisare che, nelle regressioni multilivello, la numerosità campionaria rilevante per i coefficienti di regressione e le componenti della varianza di secondo livello è costituita dal numero dei gruppi che, generalmente, non è molto elevato.

Confronto tra modelli annidati

Dalla funzione di verosimiglianza è possibile calcolare una statistica, la *devianza*, che costituisce un indicatore della bontà di adattamento del modello ai dati. La devianza è definita come: $-2 \times \ln(\text{verosimiglianza})$, dove per verosimiglianza si intende il valore della funzione di verosimiglianza alla convergenza. I modelli con devianza inferiore presentano un miglior adattamento ai dati.

Se due modelli sono annidati, ovvero se un modello specifico può essere derivato da un modello più generale rimuovendo uno o più parametri dal modello generale, è possibile confrontare statisticamente i due modelli utilizzando le loro devianze. Infatti, la differenza tra le devianze di due modelli annidati, sotto l'ipotesi nulla di equivalenza tra i due modelli, si distribuisce come un Chi-quadrato con gradi di libertà pari alla differenza nel numero dei parametri stimati dai due modelli. Questa proprietà può essere utilizzata per verificare l'ipotesi che l'adattamento ai dati del modello più generale sia significativamente superiore all'adattamento ai dati del modello specifico. Come già anticipato nel paragrafo 2.5.1, si precisa che, qualora il test includa uno o più coefficienti di regressione, i parametri del modello devono essere necessariamente stati stimati col metodo FML.

Il test Chi-quadrato delle devianze, anche noto come test del rapporto delle verosimiglianze, è asintoticamente equivalente al test di Wald. Tuttavia, il test del rapporto delle verosimiglianze risulta più efficace del test di Wald per sottoporre a verifica di ipotesi le componenti della varianza. Ciò non si verifica nei casi in cui il modello venga stimato utilizzando un'approssimazione della funzione di verosimiglianza, come può avvenire nel caso dei modelli lineari generalizzati con effetti casuali, la cui trattazione è rimandata al paragrafo 2.7.1.

Si puntualizza, infine, che i test eseguiti sulle componenti della varianza, sia nel caso del test di Wald che nel caso del test del rapporto delle verosimiglianze, sono “a una coda” dal momento che le varianze non possono essere negative. Per una approfondita trattazione della verifica di ipotesi sulle componenti della varianza si rimanda a Berkhof e Snijders, 2001.

Confronto tra modelli non annidati

Un indice di adattamento finalizzato al confronto di modelli non annidati è costituito dal criterio di informazione di Akaike (*Akaike's Information Criterion - AIC*). Questo indice è basato sul principio di parsimonia, secondo il quale i modelli semplici sono da preferire rispetto ai modelli complessi, e aggiusta il confronto statistico tra i modelli per il numero dei parametri stimati. Nel caso dei modelli di regressione multilivello il criterio di informazione di Akaike assume la seguente forma:

$$AIC = d + 2q \quad (2.11)$$

dove d è la devianza del modello e q è il numero dei parametri stimati. L'utilizzo di questo indice presuppone che i modelli posti a confronto facciano riferimento allo stesso data set, utilizzando lo stesso metodo per la stima dei parametri.

Un indice di adattamento piuttosto simile all'*AIC* è costituito dal criterio di informazione di Schwarz (*Schwartz's Bayesian Information Criterion - BIC*) che è dato da:

$$BIC = d + q \cdot \ln(N) \quad (2.12)$$

dove N rappresenta il numero di osservazioni.

Dal momento che i dati gerarchicamente organizzati presentano, ai vari livelli, differenti numerosità campionarie, nelle analisi multilivello si preferisce utilizzare l'indice AIC poiché è molto più semplice da calcolare.

In generale, dato un certo numero di modelli da confrontare, saranno da preferire quelli che presentano i valori più bassi di AIC o BIC .

2.7. Modelli lineari generalizzati multilevel

Sono numerosi i contesti di analisi in cui la variabile risposta non si distribuisce in modo simmetrico. Questo non consente l'utilizzo diretto delle metodologie statistiche basate sull'ipotesi di normalità.

L'approccio moderno a questo tipo di situazione consiste nell'includere esplicitamente nel modello statistico la trasformazione necessaria e un'opportuna distribuzione d'errore. Questi modelli statistici appartengono alla classe dei modelli lineari generalizzati e possono essere definiti attraverso tre componenti:

- una variabile risposta y con una specifica distribuzione d'errore con media μ e varianza θ^2 ;
- un'equazione di regressione lineare e additiva che costituisce il predittore η della variabile risposta y ;
- una funzione legame che collega i valori attesi della variabile risposta y ai valori predetti per η : $\eta = f(\mu)$.

Per una approfondita trattazione dei modelli lineari generalizzati si rimanda a McCullagh e Nelder, 1989.

In questo paragrafo saranno presentate due estensioni multilivello dei modelli lineari generalizzati: il modello di regressione logistica e il modello di regressione di Poisson.

Sia J il numero dei gruppi, con un numero differente n_j di individui all'interno di ciascun gruppo. Sia Y una variabile risposta di natura dicotomica (codificata come 0 o 1) misurata al livello più basso della gerarchia (rilevata, quindi, per ciascun individuo), sia X una variabile esplicativa misurata sul livello degli individui e sia Z una variabile esplicativa misurata sul livello dei gruppi. Sarà utilizzato l'indice j per i gruppi ($j = 1, \dots, J$) e l'indice i per gli individui ($i = 1, \dots, n_j$). Utilizzando la funzione legame *logit* e una distribuzione d'errore binomiale per la probabilità di successo P_{ij} , il modello multilevel di regressione logistica assume la seguente forma:

$$\text{logit}(P_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} \quad (2.13)$$

dove

$$P \sim \text{Binomiale}(\mu)$$

con

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

Nell'equazione (2.13) si nota come la struttura multilivello sia contenuta dal predittore lineare. Inoltre, il termine d'errore di primo livello e_{ij} non fa parte dell'equazione di regressione perché è implicitamente contenuto nella specificazione della distribuzione d'errore. Tuttavia, è possibile stimare un fattore di scala per la varianza di primo livello. Se questo fattore di scala risulta pari a uno, gli errori osservati seguono esattamente la distribuzione d'errore teorica binomiale. Se il fattore di scala risulta significativamente maggiore o minore di uno, si è in presenza di sovradisersione o sottodispersione. I fenomeni di variabilità extra-binomiale possono essere causati da numerosi fattori tra i quali l'omissione di covariate rilevanti o di termini di interazione significativi, la

presenza di osservazioni anomale, un'errata specificazione della funzione legame.

Nel caso specifico dei modelli multilevel, si aggiungono due fattori che possono indurre fenomeni di sovradisersione: una ridotta dimensione dei gruppi e l'omissione di un intero livello nella struttura gerarchica del modello.

Al fine di calcolare il coefficiente di correlazione intraclasse, alcuni autori suggeriscono di utilizzare la quantità $\pi^2/3$ come stima della varianza di primo livello. Tale quantità corrisponde, infatti, alla varianza di una distribuzione logistica con fattore di scala pari a uno. Questo metodo approssimato risulta particolarmente appropriato nei casi in cui la variabile risposta deriva dalla dicotomizzazione di una variabile dipendente quantitativa (Hox, 2002, pag. 117; Snijders e Bosker, 1999, pag. 223).

Nei contesti di analisi in cui Y_{ij} denota un conteggio di eventi rari per l'unità di primo livello i nel gruppo j , utilizzando la funzione legame *logaritmo naturale* e una distribuzione d'errore di Poisson per il conteggio atteso D_{ij} , si ottiene il modello multilevel di regressione di Poisson:

$$\log(D_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} + \log(T_{ij}) \quad (2.14)$$

dove

$$D \sim \text{Poisson}(\lambda)$$

con

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

Nell'equazione (2.14) il termine $\log(T_{ij})$ è inserito nel modello come *offset* e permette di considerare nella regressione i differenti periodi individuali di esposizione a rischio T_{ij} .

La possibilità di utilizzare il termine *offset* rende la regressione di Poisson particolarmente adatta alle situazioni in cui la variabile risposta è costituita da un

tasso di incidenza, calcolato come numero di eventi rapportato agli anni-persona. In questi casi, la regressione di Poisson assume che il tasso di incidenza sia costante nel tempo di follow-up.

Nelle situazioni in cui il conteggio sia riferito a eventi non rari e non sia comunque possibile far riferimento alla distribuzione normale, viene generalmente utilizzata la funzione legame *inversa* associata a una distribuzione d'errore gamma.

2.7.1. Stima dei parametri nei modelli lineari generalizzati multilevel

L'approccio prevalentemente utilizzato per la stima dei parametri nei modelli lineari generalizzati multilivello consiste nell'approssimare il legame non lineare attraverso una funzione quasi lineare. In questo metodo, noto come metodo della quasi-verosimiglianza, la funzione non lineare viene linearizzata attraverso un'approssimazione basata sull'espansione in serie di Taylor. Questa tecnica approssima una funzione non lineare attraverso una serie infinita di termini. In genere viene utilizzato esclusivamente il primo termine della serie (approssimazione di Taylor di primo ordine) tuttavia, quando si utilizza anche il secondo termine (approssimazione di Taylor di secondo ordine) si ottengono risultati generalmente migliori. Il metodo di stima della massima verosimiglianza procede iterativamente, partendo da valori approssimati dei parametri che vengono migliorati in ciascuna delle iterazioni successive. Pertanto, i valori stimati dei parametri cambiano durante le iterazioni e l'espansione in serie di Taylor deve essere ripetuta dopo ogni passo della procedura di stima.

In questo contesto, la procedura di stima può essere realizzata attraverso il metodo della quasi-verosimiglianza marginale (*marginal quasi-likelihood* - MQL), in cui l'espansione in serie di Taylor può utilizzare solo i valori correnti della parte fissa del modello, oppure attraverso il metodo della quasi-verosimiglianza penalizzata (*penalized quasi-likelihood* - PQL), in cui l'espansione in serie di Taylor può

essere migliorata utilizzando sia i valori correnti della parte fissa che le stime correnti dei residui di livello superiore.

Il metodo della quasi-verosimiglianza penalizzata, basato su approssimazioni di Taylor di secondo ordine fornisce le stime più accurate ma potrebbe causare problemi di convergenza nei casi in cui la numerosità campionaria sia molto ridotta o nei casi in cui il modello contenga un elevato numero di parametri. In queste situazioni, il metodo della quasi-verosimiglianza marginale basato su approssimazioni di Taylor di primo ordine costituisce, comunque, una valida alternativa. Un confronto approfondito tra i due metodi di stima è presente in Goldstein e Rasbash, 1996.

Nei metodi di stima basati sullo sviluppo in serie di Taylor la verosimiglianza che viene massimizzata non è la reale verosimiglianza. Per questo motivo il test del rapporto delle verosimiglianze, basato sul confronto tra le devianze dei modelli, non risulta molto accurato. Questo problema, inoltre, caratterizza tutti gli indici basati sulla verosimiglianza, tra questi il criterio di informazione di Akaike. Pertanto, nei casi in cui si ricorre alla linearizzazione mediante serie di Taylor, per la verifica di ipotesi risulta preferibile il test di Wald.

Un metodo di stima alternativo alla quasi-verosimiglianza che consente di utilizzare i test statistici basati sulle devianze è costituito dall'integrazione numerica della complessa funzione di verosimiglianza. Infatti, al contrario delle tecniche basate sulle serie di Taylor che massimizzano una verosimiglianza approssimata, attraverso l'integrazione numerica è possibile massimizzare la corretta verosimiglianza. Per questo motivo, utilizzando il metodo della massima verosimiglianza con l'integrazione numerica, restano appropriati tutti i test statistici e gli indici di adattamento basati sulla verosimiglianza.

Inoltre, alcune ricerche basate sulle simulazioni (Hartford & Davidian, 2000; Rodriguez & Goldman, 2001) suggeriscono che i metodi di integrazione numerica producono stime più precise rispetto al metodo della quasi-verosimiglianza.

Anche i metodi di integrazione numerica possono, tuttavia, comportare problemi di convergenza simili a quelli riscontrati per il metodo della quasi-verosimiglianza penalizzata con approssimazioni di secondo ordine.

Per una approfondita trattazione dei metodi di integrazione numerica si rimanda a Wolfinger, 1993.

2.7.2. Integrazione di effetti fissi ed effetti casuali

Nell'ambito dei modelli lineari generalizzati con effetti casuali si è avvertita la necessità di quantificare l'eterogeneità tra i gruppi attraverso una misura di effetto che avesse la stessa interpretazione degli effetti fissi espressi, nel caso della regressione logistica, attraverso gli *odds ratio* (OR), nel caso della regressione di Poisson attraverso i *rate ratio*.

“... the intraclass correlation is not comparable with the fixed effects, which have odds ratio interpretations. This is unfortunate, because random effects are not very different from fixed effects in nature; random effects are fixed effects with an additional distributional assumption ...” (Larsen e Merlo, 2005).

Nel caso in cui si disponga di una variabile dipendente dicotomica e i dati presentino una struttura gerarchica su due livelli, si supponga di confrontare due individui con covariate identiche che appartengono a due gruppi differenti.

L'*odds ratio* tra i due soggetti posti a confronto è costituito dall'esponenziale della differenza tra gli effetti casuali dei due gruppi: $OR = \exp(u_{01} - u_{02})$. Pertanto, l'*odds ratio* tra il soggetto con la più alta propensione e il soggetto con la più bassa propensione è dato da: $OR = \exp(|u_{01} - u_{02}|)$.

Ripetendo i confronti selezionando casualmente i soggetti da gruppi diversi si ottiene una serie di *odds ratio* di cui è possibile studiare la distribuzione.

La mediana di questa distribuzione può essere utilizzata per quantificare l'eterogeneità tra i gruppi.

Assumendo l'indipendenza e la normalità distributiva degli effetti casuali, $u_{0j} \sim N(0, \sigma^2_{u0})$, la distribuzione della variabile casuale odds ratio può essere caratterizzata attraverso la seguente funzione di ripartizione:

$$\begin{aligned}
 F(z) &= P(\exp(|u_{01} - u_{02}|) \leq z) \\
 &= P\left(\frac{|u_{01} - u_{02}|}{\sqrt{2 \times \sigma^2_{u0}}} \leq \frac{\log(z)}{\sqrt{2 \times \sigma^2_{u0}}}\right) = 2\Phi\left(\frac{\log(z)}{\sqrt{2 \times \sigma^2_{u0}}}\right) - 1
 \end{aligned} \tag{2.15}$$

dove $\Phi(\cdot)$ indica la funzione di ripartizione della distribuzione normale standardizzata.

Pertanto, la funzione densità di probabilità per la distribuzione di $\exp(|u_{01} - u_{02}|)$ può essere espressa come segue:

$$f(z) = \frac{\partial}{\partial z} F(z) = \frac{1}{z} \sqrt{\frac{2}{\sigma^2_{u0}}} \varphi\left(\frac{\log(z)}{\sqrt{2 \times \sigma^2_{u0}}}\right) \tag{2.16}$$

Ponendo $F(z) = 0.5$ si ottiene:

$$z = \exp\left(\sqrt{2 \times \sigma^2_{u0}} \times \Phi^{-1}(0.75)\right) \tag{2.17}$$

dove $\Phi^{-1}(0.75)$ è il 75° percentile della distribuzione normale standardizzata.

Seguono le caratteristiche principali dell'odds ratio mediano (*Median Odds Ratio* - MOR) espresso dalla (2.17):

- è una funzione della varianza di secondo livello σ^2_{u0} ;

- può assumere esclusivamente valori maggiori o uguali all'unità: un odds ratio mediano pari a uno indica assenza di variabilità tra i gruppi mentre, nel caso di una considerevole variabilità tra i gruppi, l'odds ratio mediano risulterà sostanzialmente maggiore di uno;
- è direttamente confrontabile con gli *odds ratio* per gli effetti fissi;
- può essere esteso a modelli che includono tre o più livelli gerarchici e a modelli con coefficienti di regressione casuali (*random slope*);
- è immediatamente generalizzabile al modello di regressione di Poisson con effetti casuali (*median rate ratio*) e al modello di regressione di Cox con effetti casuali (*median hazard ratio*) la cui trattazione è rimandata al prossimo capitolo.

3. Modelli per l'analisi della sopravvivenza con effetti casuali

3.1. L'analisi della sopravvivenza

L'analisi statistica dei tempi di sopravvivenza è stata oggetto, negli ultimi anni, di un rinnovato interesse da parte di vari settori della ricerca.

L'elemento essenziale per cui un problema è ammissibile nell'ambito di tale analisi è costituito dalla presenza di una variabile aleatoria a valori non negativi che descrive il tempo di accadimento di uno o più eventi di interesse. E' proprio la genericità estrema con cui tale evento è definito che consente di includere situazioni della più diversa natura, interessando vari campi di studio.

Alcune particolarità dei dati di durata rendono le tecniche per l'analisi della sopravvivenza non riconducibili alle procedure statistiche standard utilizzate per l'analisi di altri dati. La prima caratteristica è che i dati di sopravvivenza, generalmente, non sono distribuiti in modo simmetrico: ciò non consente l'utilizzo delle metodologie basate sull'ipotesi di normalità. La seconda caratteristica è che i tempi di sopravvivenza sono spesso censurati. In numerosi contesti di analisi, infatti, l'effettivo tempo di fallimento F non è osservabile per tutti gli individui entrati a far parte dello studio perché la limitata durata del periodo di osservazione implica la presenza di soggetti che, al termine del follow-up, non hanno ancora sperimentato l'evento di interesse. Inoltre, è possibile che il paziente esca dall'osservazione per cause diverse dall'evento in studio oppure che l'evento stesso non si verifichi anche ipotizzando di poter seguire il soggetto per l'intero arco della sua esistenza.

Indicando con t_i il tempo osservato per l'unità i e con δ_i l'indicatore di evento per l'unità i si ha che:

per $\delta_i = 1$, t_i = tempo dell'evento;

per $\delta_i = 0$, t_i = tempo della censura.

Nel riassumere i dati di sopravvivenza, due funzioni rivestono un interesse centrale: la funzione di sopravvivenza e la funzione di rischio.

Sia F una variabile aleatoria a valori non negativi che esprime il tempo di fallimento di un individuo.

La *funzione di sopravvivenza* $S(t)$, al tempo t , è la probabilità che F sia maggiore o uguale al valore t :

$$S(t) = P(F \geq t) \quad \forall 0 \leq t < +\infty$$

Pertanto, la funzione S si riferisce alla coda destra della distribuzione di F poiché raccoglie cumulativamente tutte le probabilità di sopravvivenza oltre un certo tempo fissato.

La *funzione di rischio* $\lambda(t)$ rappresenta, invece, il tasso istantaneo di evento per un soggetto che sia sopravvissuto fino a quell'istante. Si consideri la probabilità che la variabile aleatoria F , tempo di sopravvivenza, sia compresa tra t e $t + \delta t$, con la condizione che F sia maggiore o uguale a t , ovvero $P(t \leq F < t + \delta t | F \geq t)$. La funzione di rischio $\lambda(t)$ è il limite di questa probabilità divisa per l'intervallo di tempo δt , al tendere di δt a zero:

$$\lambda(t) = \lim_{\delta t \rightarrow 0^+} \left\{ \frac{P(t \leq F < t + \delta t | F \geq t)}{\delta t} \right\}$$

3.2. Il modello a rischi proporzionali di Cox

Nei casi in cui si disponga, per ciascuna unità i , dei dati sul tempo (t_i, δ_i) e di un vettore di covariate \mathbf{x}_i , sia $\lambda_0(t)$ la funzione di rischio di base, ovvero la funzione di rischio del generico soggetto caratterizzato dal fatto di avere tutti i valori delle

covariate identicamente nulli ($\mathbf{x} = \mathbf{0}$). Il modello di regressione di Cox assume la forma seguente:

$$\lambda(t | \mathbf{x}) = \lambda_0(t) \cdot \exp(\mathbf{x}'\boldsymbol{\beta}) \quad (3.1)$$

Dove $\boldsymbol{\beta}$ è il vettore dei coefficienti di regressione (β_1, \dots, β_p).

La caratteristica fondamentale del modello di Cox è quella di non formulare alcuna ipotesi iniziale sull'andamento nel tempo della funzione di rischio $\lambda_0(t)$, lasciando che siano i dati stessi a suggerirlo.

L'equazione (3.1) può essere scritta nella forma equivalente:

$$\frac{\lambda(t | \mathbf{x})}{\lambda_0(t)} = \exp(\mathbf{x}'\boldsymbol{\beta}) \quad (3.2)$$

per evidenziare come alla base del modello di Cox sia l'ipotesi che il rischio relativo non dipenda dal tempo, rimanendo costante per l'intera durata del follow-up. Questa ipotesi, pur consentendo notevoli semplificazioni per ciò che concerne gli aspetti inferenziali di stima e verifica di ipotesi, risulta a volte in contrasto con l'andamento reale del fenomeno. Non è raro infatti che l'effetto dovuto ad un particolare fattore di rischio o l'effetto protettivo di una particolare terapia tenda a diminuire nel tempo.

La stima del vettore di parametri incogniti $\boldsymbol{\beta}$ e le relative procedure di verifica di ipotesi sono basate sulla funzione di verosimiglianza parziale:

$$L(\boldsymbol{\beta}) = \prod_{i \in D} \frac{\lambda_i(t_{(k)})}{\sum_{j \in R_k} \lambda_j(t_{(k)})} = \prod_{i \in D} \frac{\exp[\boldsymbol{\beta}'\mathbf{x}_i]}{\sum_{j \in R_k} \exp[\boldsymbol{\beta}'\mathbf{x}_j]} \quad (3.3)$$

che lo stesso Cox propone di considerare come funzione di verosimiglianza per il vettore β .

In questa espressione $t_{(k)}$ indica il k -esimo tempo di fallimento osservato, i identifica l'individuo che ha sperimentato l'evento in $t_{(k)}$, R_k l'insieme di individui ancora a rischio un istante prima di $t_{(k)}$ e D l'insieme degli individui che hanno sperimentato l'evento durante il periodo di osservazione.

Questa funzione, sebbene non possa essere considerata una funzione di verosimiglianza associata ad un reale esperimento statistico, mantiene tutte le proprietà che caratterizzano una funzione di verosimiglianza ordinaria.

Nella (3.3) si nota come i tempi di troncamento non diano un contributo diretto alla verosimiglianza, essendo la produttoria limitata agli effettivi tempi di fallimento osservati. Inoltre, ciascun termine della produttoria stessa può essere interpretato come la probabilità che, osservato un fallimento in $t_{(k)}$, sia stato proprio l' i -esimo individuo tra coloro ancora esposti a rischio a sperimentare l'evento. Il metodo della massima verosimiglianza parziale permette di stimare, con una modesta perdita di efficienza, i coefficienti di regressione senza formulare alcuna ipotesi iniziale sul rischio di base.

Per una approfondita trattazione del modello di regressione di Cox e delle sue estensioni si rimanda a Therneau e Grambsch, 2000. Di particolare interesse (pagine 79-85) la definizione di alcune tipologie di residuo sviluppate nel contesto specifico del modello a rischi proporzionali di Cox.

Il tema dei residui è di estrema attualità perché l'estensione della definizione di residuo all'analisi della sopravvivenza comporta due problemi sostanziali. Il primo è legato alla presenza di osservazioni incomplete, caratteristica peculiare dell'analisi della sopravvivenza. Benché, in generale, si tenda a mantenere inalterata la definizione di residuo in corrispondenza dei dati troncati, questa soluzione non è priva di ripercussioni negative sulla loro capacità diagnostica. Inoltre, mentre la variabile risposta è un tempo, cioè un valore sulla semiretta positiva, la quantità ricostruita dal modello, ed in particolare dal modello

di Cox, è generalmente una funzione del tempo, sia essa la funzione di sopravvivenza o la funzione di rischio stimata.

3.3. Il modello di Cox con effetti casuali

L'utilizzo di effetti casuali nei modelli di regressione è aumentato molto rapidamente negli ultimi anni. Tuttavia, l'estensione di questa particolare modellistica all'ambito dell'analisi della sopravvivenza è stata molto più lenta.

Uno dei problemi principali per cui si è avvertita la necessità di includere nei modelli di durata un effetto casuale è quello dell'omissione di covariate. Sono, infatti, numerosi i contesti di analisi in cui non tutte le covariate rilevanti vengono incluse nella specificazione del modello poiché il ricercatore potrebbe non conoscere o non essere in grado di misurare tutte le variabili esplicative rilevanti. In queste situazioni, possono essere individuate due differenti fonti di variabilità nei tempi di sopravvivenza: la variabilità spiegata dalle covariate incluse nel modello e l'eterogeneità causata dalle variabili non considerate nelle analisi.

Nei modelli di durata l'eterogeneità dovuta all'omissione di covariate o, più in generale, ad una errata specificazione del modello determina distorsioni nella stima della funzione di rischio. Il motivo di questa distorsione è attribuibile ad una peculiarità dei modelli di durata: la composizione del campione si modifica con il passare del tempo. Infatti, col trascorrere del tempo di follow-up, nel campione (*risk set*) aumenta la proporzione delle unità a basso rischio.

Ne consegue che anche le eventuali deviazioni dall'assunto di proporzionalità dei rischi potrebbero essere spiegate dall'eterogeneità causata dall'omissione di covariate.

Per tenere in considerazione nel modello tale eterogeneità e descrivere l'influenza delle covariate non osservate, Vaupel introdusse nel 1979 il concetto di *frailty*, una variabile aleatoria U continua con densità $f(u; \theta)$ indipendente dalle covariate \mathbf{X} , che agisce sulla funzione di rischio di base. L'idea sottostante è che

gli individui posseggono differenti “fragilità” e che i soggetti più fragili svilupperanno l’evento in studio prima degli altri.

Nell’ambito della regressione di Cox, l’inclusione nel modello della variabile aleatoria U può avvenire in due modi differenti:

$$\lambda(t | \mathbf{x}, u) = \lambda_{00}(t) \cdot \exp(\mathbf{x}'\boldsymbol{\beta}) \cdot u \quad (3.4)$$

dove $\lambda_{00}(t)$ è la funzione di rischio di base ($\mathbf{x} = \mathbf{0}$, $u = 1$), U ha distribuzione asimmetrica con $E(U) = 1$ (in genere si fa riferimento a una distribuzione gamma o log-normale);

oppure

$$\lambda(t | \mathbf{x}, u) = \lambda_{00}(t) \cdot \exp(\mathbf{x}'\boldsymbol{\beta} + u) = \lambda_{00}(t) \cdot \exp(\mathbf{x}'\boldsymbol{\beta}) \cdot \exp(u) \quad (3.5)$$

dove $\lambda_{00}(t)$ è la funzione di rischio di base ($\mathbf{x} = \mathbf{0}$, $u = 0$), U ha distribuzione simmetrica con $E(U) = 0$ (in genere si fa riferimento alla distribuzione normale).

In entrambi i casi, si considera una realizzazione di U per ogni unità. Per una approfondita trattazione di questi modelli si rimanda a Keiding et al., 1997.

Nei casi in cui i dati presentino una struttura gerarchica, un’interessante estensione del modello di Cox con frailty è costituito dal modello di Cox con frailty condivise (*shared frailties Cox regression model*) in cui si assume che tutti gli individui all’interno dello stesso gruppo condividano la stessa frailty u .

Infatti, se si considera una realizzazione di U per ogni gruppo, il modello di Cox con frailty può essere visto come un modello di regressione multilivello.

La differenza tra i modelli con frailty individuali e i modelli con frailty condivise è sostanziale: i primi hanno l’obiettivo di descrivere l’influenza di covariate non

osservate nei modelli a rischi proporzionali; i secondi quello di spiegare le correlazioni tra soggetti che appartengono allo stesso gruppo nei casi in cui i dati siano gerarchicamente organizzati e i tempi di sopravvivenza non possano essere considerati indipendenti.

Nei modelli con frailty condivise si assume, infatti, che i tempi di sopravvivenza possano essere considerati indipendenti condizionatamente all'effetto casuale condiviso.

Il modello di regressione di Cox con frailty condivise a livello di gruppo, nell'ipotesi di una struttura gerarchica su due livelli, può essere formalizzato come segue:

$$\lambda(t | \mathbf{x}_{ij}, \mathbf{u}_j) = \lambda_{00}(t) \cdot \exp(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{u}_j) \quad (3.6)$$

dove:

\mathbf{x}_{ij} è il vettore di covariate (sia di primo che di secondo livello) dell'unità i del gruppo j ;

\mathbf{z}_{ij} è il vettore di covariate (di primo livello) con effetto casuale: $\mathbf{z}_{ij} \in \{\mathbf{x}_{ij}, \text{intercetta}\}$;

\mathbf{u}_j è il vettore di effetti casuali (shared frailties) del gruppo j .

Generalmente, si assume che il vettore di effetti casuali \mathbf{u}_j segua una distribuzione normale multivariata con media zero e matrice non nota di varianze e covarianze: $\mathbf{u}_j \sim N(\mathbf{0}, \boldsymbol{\Omega}_u)$.

Il modello (3.6), proposto da Vaida e Xu nel 2000, consente una specificazione non parametrica della funzione di rischio di base ed è in grado di incorporare la variabilità casuale sia nella funzione di rischio di base che nei coefficienti di regressione delle covariate di primo livello. Nei casi in cui $\boldsymbol{\Omega}_u$ sia pari a zero, le risposte degli individui all'interno dei gruppi possono essere considerate

indipendenti e l'equazione di regressione (3.6) si riduce all'usuale modello di regressione di Cox a rischi proporzionali.

Il modello di regressione (3.6) gode, inoltre, della proprietà di invarianza di scala, infatti

se $u_{kj} \sim N(0, \sigma_k^2)$ e $z_{kij}^* = cz_{kij}$
 si ha $z_{kij}u_{kj} = cz_{kij}(u_{kj} / c) = z_{kij}^*u_{kj}^*$
 dove $u_{kj}^* \sim N(0, (\sigma_k / c)^2)$.

Questo assicura che gli aspetti inferenziali non dipendano dall'unità di misura delle covariate e, conseguentemente, degli effetti casuali.

Al fine di poter includere nel modello di regressione covariate tempo dipendenti sia di primo che di secondo livello, il modello (3.6) è stato successivamente esteso e formalizzato come segue (Maples et al., 2002):

$$\lambda_{ij}(t) = \lambda_{00}(t) \cdot \exp[r_{0j} + \mathbf{r}_{1j}'\mathbf{x}_{ij}(t) + \boldsymbol{\beta}_x'\mathbf{x}_{ij}(t) + \boldsymbol{\beta}_z'\mathbf{z}_j(t)] \quad (3.7)$$

dove $\mathbf{z}_j(t)$ denota il vettore di covariate di secondo livello al tempo t , $\mathbf{x}_{ij}(t)$ il vettore p -dimensionale di covariate di primo livello al tempo t ed $\mathbf{r}_j = (r_{0j}, \mathbf{r}_{1j})$ il vettore $(p + 1)$ -dimensionale di effetti casuali non osservati a livello di gruppo.

Anche nel modello (3.7) si assume che il vettore di effetti casuali \mathbf{r}_j segua una distribuzione normale multivariata con media pari a zero e matrice di varianze e covarianze non nota.

3.3.1. Stima dei parametri nel modello di Cox con effetti casuali

Sia $\boldsymbol{\gamma}$ il vettore dei parametri contenuti nella matrice di varianze e covarianze $\boldsymbol{\Omega}_u$ e sia $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0)$ il “vettore” dei parametri di interesse.

Ciascuna osservazione può essere scritta come $y_{ij} = (t_{ij}, \delta_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij})$, dove

- t_{ij} è il tempo di sopravvivenza dell'unità i del gruppo j ;
- δ_{ij} è l'indicatore di evento (1 se si osserva l'evento di interesse, 0 altrimenti);
- \mathbf{x}_{ij} è il vettore di covariate, di primo e di secondo livello;
- \mathbf{z}_{ij} è il vettore di covariate di primo livello con effetto casuale;
- \mathbf{u}_j è il vettore di effetti casuali del gruppo j .

Sia $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,n_j})$ l'insieme dei dati relativi al gruppo j ($j = 1, \dots, n$). Vaida e Xu mostrano come, per il gruppo j , condizionatamente agli effetti casuali, la log-verosimiglianza abbia la seguente forma:

$$l_j = l_j(\boldsymbol{\beta}, \lambda_0; \mathbf{y}_j | \mathbf{u}_j) = \sum_{i=1}^{n_j} [\delta_{ij} \log \lambda_0(t_{ij}) + \delta_{ij}(\boldsymbol{\beta}' \mathbf{x}_{ij} + \mathbf{z}_{ij}' \mathbf{u}_j) - \Lambda_0(t_{ij}) \exp(\boldsymbol{\beta}' \mathbf{x}_{ij} + \mathbf{z}_{ij}' \mathbf{u}_j)] \quad (3.8)$$

dove $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ è la funzione di rischio cumulato di base.

La verosimiglianza dei dati osservati è dunque:

$$L(\boldsymbol{\theta}) = \int L(\boldsymbol{\theta} | \mathbf{u}) p(\mathbf{u} | \boldsymbol{\gamma}) d\mathbf{u} = \prod_{j=1}^n \int \exp(l_j) p(\mathbf{u}_j | \boldsymbol{\gamma}) d\mathbf{u}_j \quad (3.9)$$

dove $L(\boldsymbol{\theta} | \mathbf{u}) = \prod_j \exp(l_j)$ è la verosimiglianza condizionata agli effetti casuali $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$.

Come evidenziato da Vaida e Xu, la verosimiglianza dei dati osservati $L(\boldsymbol{\theta})$ non esiste in forma chiusa poiché l'integrale non ha soluzione analitica. Pertanto, la stima dei parametri nel modello di regressione di Cox con frailty condivise deve necessariamente avvenire attraverso metodi approssimati: tra questi, uno dei più diffusi è l'algoritmo EM (*Expectation-Maximization*). In questa soluzione, gli effetti casuali vengono trattati come dati mancanti.

In generale, indicando con \mathbf{Y} la matrice dei dati che si avrebbe in assenza di mancata risposta, con \mathbf{Y}_{obs} l'insieme dei valori osservati e con \mathbf{Y}_{mis} l'insieme dei valori mancanti, si ha: $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$.

L'algoritmo EM è un metodo che consente, attraverso un procedimento iterativo, di effettuare le stime di massima verosimiglianza dei parametri in presenza di dati incompleti, riconducendo il problema ad un problema standard di stima per dati completi. Partendo da una stima iniziale $\boldsymbol{\theta}^{(0)}$ dei parametri (*starting guess*), l'algoritmo consiste, ad ogni iterazione t del procedimento, nella applicazione dei seguenti due passi:

E-step. Calcolo del valore atteso $H(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ della verosimiglianza $L(\boldsymbol{\theta} | \mathbf{Y}_{obs})$ rispetto alla distribuzione dei dati mancanti condizionatamente ai dati osservati e alle stime correnti dei parametri $\boldsymbol{\theta}^{(t)}$:

$$H(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) \equiv \int L(\boldsymbol{\theta} | \mathbf{Y}_{obs}, \mathbf{Y}_{mis}) f(\mathbf{Y}_{mis} | \mathbf{Y}_{obs}, \boldsymbol{\theta}^{(t)}) d\mathbf{Y}_{mis}$$

M-step. Massimizzazione di $H(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ rispetto a $\boldsymbol{\theta}$.

L'algoritmo genera una successione $\{\boldsymbol{\theta}^{(t)}\}_{t=1,2,\dots}$ che, sotto alcune ipotesi di regolarità, si dimostra (Dempster et al., 1977) convergere alla stima di massima verosimiglianza di $\boldsymbol{\theta}$.

Nel caso specifico del modello di regressione Cox con frailty condivise, l'*M-step* tratta le stime correnti degli effetti casuali \mathbf{u} come valori fissi e aggiorna $\boldsymbol{\beta}$ e λ_0 come nell'usuale modello di Cox mentre l'*E-step* calcola \mathbf{u} come valore atteso condizionatamente ai dati e ai valori correnti di $\boldsymbol{\beta}$ e di λ_0 (Parner, 1997).

In relazione alla stima dei parametri nel modello di Cox con effetti casuali, costituisce una valida alternativa all'algoritmo *EM* il metodo della verosimiglianza parziale penalizzata (*Penalized Partial Likelihood* - PPL), per la cui trattazione si rimanda a Therneau e Grambsch, 2000 e Ripatti e Palmgren, 2000.

4. Analisi della persistenza in trattamento antiipertensivo

4.1. Fonte dei dati

I dati utilizzati per questo studio si inseriscono nel contesto del progetto PANDORA (Degli Esposti *et al.*, 1999), uno studio prospettico, iniziato nel 1996 con il supporto di Pfizer Italia Srl, che prevede la creazione di un archivio elettronico contenente dati che derivano dalla Medicina Generale. Il progetto si propone di migliorare la pratica clinica, con particolare riferimento alle patologie croniche, attraverso la conduzione e la diffusione di studi epidemiologici e farmacoepidemiologici. Un gruppo di Medici di Medicina Generale (MMG) che dipendono dall'Azienda Unità Sanitaria Locale di Ravenna si è volontariamente proposto di partecipare allo studio. I MMG dispongono di un Personal Computer collegato al misuratore automatico della pressione arteriosa, modello Dinamap 1846SX (Critikon, Tampa, Florida, USA), e utilizzano un programma di gestione paziente, denominato So.Ge.Pa.®, per inserire le informazioni sui loro assistiti che includono l'anamnesi clinica, i farmaci prescritti e i risultati di tutti gli esami strumentali o di laboratorio effettuati dai loro pazienti ipertesi. Ogni singolo atto prescrittivo compiuto dai medici è, quindi, registrato e contribuisce a costruire una memoria longitudinale individuale. Il software trasferisce periodicamente i dati via modem ad una stazione remota per la creazione di un database generale.

Le informazioni raccolte sono state integrate con i dati di ricovero attraverso procedure di *record linkage* con l'archivio amministrativo nosologico ospedaliero dell'Azienda Unità Sanitaria Locale di Ravenna (diagnosi principale e secondarie, data e reparto di ammissione/dimissione).

4.2. Criteri di arruolamento

Sono stati arruolati nello studio tutti i pazienti che presentavano almeno una misurazione di pressione arteriosa nel periodo compreso fra il 01/01/1997 e il 31/12/2002. Nel caso di misurazioni ripetute, si è fatto riferimento alla rilevazione

più vicina al 01/01/1997. La data di questa rilevazione costituisce la data di arruolamento. Sono stati, in secondo luogo, individuati tutti i soggetti che presentavano prescrizioni di farmaci antiipertensivi nel periodo successivo alla data di arruolamento.

Sono state considerate le seguenti famiglie di farmaci antiipertensivi: Diuretici (codice ATC C03), Beta Bloccanti non associati (codice ATC C07A), Calcio Antagonisti (codice ATC C08), Sostanze ad azione sul sistema renina-angiotensina (ACE-Inibitori non associati e Antagonisti dell'Angiotensina II non associati, codici ATC C09A e C09C, rispettivamente) e altri antiipertensivi (codice ATC C02).

4.3. Caratterizzazione dei pazienti

Nei 365 giorni antecedenti alla data di arruolamento, per ciascun paziente è stato valutato l'utilizzo cronico (presenza di almeno 2 prescrizioni) delle seguenti terapie farmacologiche: antidiabetici (codice ATC A10), farmaci per la terapia cardiaca (codice ATC C01), farmaci per il trattamento delle affezioni respiratorie (codice ATC R03).

In base alla presenza/assenza di prescrizioni per farmaci antiipertensivi nei 365 giorni antecedenti alla data di arruolamento, i soggetti sono stati classificati in già trattati e nuovi trattati, rispettivamente. Inoltre, nei sei anni antecedenti alla data di arruolamento sono stati ricercati i seguenti progressi ricoveri cardio/cerebrovascolari, identificati mediante la codifica ICD-9: infarto miocardico acuto, angina pectoris, scompenso cardiaco, ictus, attacco ischemico transitorio.

4.4. Periodo di follow-up

A partire dalla prima prescrizione di farmaci antiipertensivi successiva o coincidente alla data di arruolamento, gli assistiti sono stati osservati per 365 giorni al fine di misurare la persistenza in trattamento. Nel periodo di follow-up è

stato valutato l'utilizzo concomitante (presenza di almeno 2 prescrizioni) di statine (codice ATC C10AA) e antiaggreganti (codice ATC B01AC). Tutte le informazioni relative al periodo di follow-up sono state raccolte fino al 31/12/2003.

4.5. Criteri di esclusione

Sono stati esclusi dalle analisi i pazienti che, nei 365 giorni antecedenti alla data di arruolamento, non risultavano sempre presenti nella lista degli assistibili del medico di medicina generale che ha effettuato la misurazione di pressione arteriosa, i pazienti con età (calcolata alla data di arruolamento) minore di 20 anni e i pazienti che presentavano, nella prima prescrizione di antiipertensivi, un'associazione di differenti classi farmacologiche.

Sono stati, inoltre, esclusi tutti i soggetti che presentavano un periodo di follow-up inferiore a 365 giorni. Rientrano in questa categoria di pazienti i soggetti che hanno ricevuto la prima prescrizione di antiipertensivi dopo il 31/12/2002, i soggetti che hanno cambiato medico di medicina generale durante il periodo di follow-up e i soggetti deceduti durante il periodo di follow-up.

4.6. Valutazione della persistenza in trattamento

La durata del trattamento antiipertensivo è stata calcolata come segue: (giorni intercorsi tra la prima e l'ultima prescrizione) + (proiezione, in giorni, dell'ultima prescrizione).

La proiezione dell'ultima prescrizione è stata stimata sulla base delle Dosi Definite Giornaliere in questa contenute (WHO - Collaborating Centre for Drug Statistics Methodology, 2002). Sono stati definiti persistenti i soggetti che presentavano una durata del trattamento maggiore di 273 giorni.

4.7. Struttura gerarchica paziente-medico

L'analisi della persistenza in trattamento è stata realizzata seguendo un approccio multilivello. I pazienti costituiscono le unità di primo livello, i medici di medicina generale le unità di secondo livello. La struttura gerarchica di riferimento considera, quindi, l'annidamento dei pazienti all'interno del proprio medico di medicina generale.

Le variabili esplicative di primo livello (misurate sul paziente) sono state: età (calcolata alla data di arruolamento), sesso, presenza/assenza di pregressi ricoveri cardio/cerebrovascolari, presenza/assenza di terapie croniche pregresse (antidiabetici, farmaci per la terapia cardiaca, farmaci per il trattamento delle affezioni respiratorie), incidenza al trattamento antiipertensivo (già trattato/nuovo trattato), valore di pressione arteriosa sistolica e diastolica all'arruolamento, classe di antiipertensivo alla prima prescrizione¹, presenza/assenza di terapie croniche concomitanti (statine, antiaggreganti). Le variabili esplicative di secondo livello (misurate sul medico di medicina generale) sono state: età (calcolata al 01/01/1997), sesso, numero di assistibili (numero medio di assistibili nel periodo di arruolamento).

¹ Sulla base del principio dell'intenzione al trattamento (*intention to treat analysis*), i pazienti sono stati classificati esclusivamente in relazione alla classe di antiipertensivo alla prima prescrizione, trascurando le eventuali modifiche della terapia farmacologica (associazioni o cambi di terapia) avvenute durante il periodo di follow-up. Per gli approfondimenti sull'*intention to treat analysis* si rimanda a Pocock, 1983.

5. Risultati

Applicando i criteri di arruolamento e i criteri di esclusione precedentemente descritti, sono stati individuati 38 Medici di Medicina Generale per un totale di 4584 pazienti.

Dopo una prima lettura dei dati, si è ritenuto opportuno escludere 100 nuovi trattati con antiipertensivi che presentavano valori di pressione arteriosa sistolica e diastolica normali, ovvero inferiori a 120/80 mmHg (Joint National Committee VII, 2003). E', infatti, verosimile che questi soggetti abbiano assunto i farmaci antiipertensivi per motivi diversi dall'ipertensione arteriosa e, pertanto, sono stati esclusi dalle valutazioni di persistenza in trattamento. Inoltre, per evitare i fenomeni di sovradisersione trattati nel paragrafo 2.7, sono stati esclusi dalle analisi 4 Medici di Medicina Generale che presentavano un numero di pazienti inferiore a 5. Ciò ha comportato una ulteriore riduzione, pari a 10 soggetti, delle unità di primo livello. Infine, sono stati esclusi 4 pazienti che presentavano valori di pressione arteriosa sistolica inferiori a 30 mmHg.

Il collettivo su cui sono state effettuate le valutazioni di persistenza in trattamento ammonta, quindi, a 4470 pazienti afferenti a 34 Medici di Medicina Generale.

Le caratteristiche delle unità di primo e secondo livello incluse nelle analisi sono presentate nelle tabelle 5-1 e 5-2. Nella tabella 5-3 sono, inoltre, riportate le percentuali di utilizzo delle diverse classi di farmaci antiipertensivi alla prima prescrizione, e l'eventuale trattamento concomitante con statine o antiaggreganti.

Tabella 5-1. Caratteristiche dei Medici di Medicina Generale

Medici	N	34
Età	<i>range</i>	35 - 51
Età	media \pm DS	42 \pm 4
Sesso	% maschi	58.8
Numero assistibili	media \pm DS	1172 \pm 446
Pazienti inclusi	media \pm DS	132 \pm 89

Tabella 5-2. Caratteristiche dei pazienti

		Nuovi trattati	Già trattati	Totale
Pazienti	N	1936	2534	4470
Età	<i>range</i>	21 - 96	23 - 95	21 - 96
Età	media ± DS	62 ± 12	66 ± 12	64 ± 12
Sesso	% maschi	46.9	41.6	43.9
Pregressi eventi CV	%	12.5	27.5	21.0
Diabete	%	2.5	6.3	4.7
Cardiopatìa	%	1.2	8.8	5.5
Asma/Bpco	%	2.5	5.4	4.1
PAS	media ± DS	153 ± 19	146 ± 21	149 ± 20
PAD	media ± DS	86 ± 11	81 ± 10	83 ± 11

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica.

Tabella 5-3. Classe di antiipertensivo alla prima prescrizione e trattamenti concomitanti

		Nuovi trattati	Già trattati	Totale
Classe di antiipertensivo				
Azione su renina-angiotensina	%	56.4	50.2	52.9
Calcio Antagonisti	%	16.4	20.7	18.9
Beta Bloccanti	%	15.3	16.9	16.2
Diuretici	%	9.8	9.2	9.5
Altri antiipertensivi	%	2.0	2.9	2.5
Trattamenti concomitanti				
Statine	%	11.5	15.0	13.5
Antiaggreganti	%	14.6	21.6	18.5

Complessivamente, 835 pazienti (pari al 18.7%) sono risultati non persistenti in trattamento, poiché hanno interrotto la terapia farmacologica antiipertensiva durante i 365 giorni di follow-up. Nei soggetti esposti per la prima volta al trattamento antiipertensivo la percentuale di interruzione terapeutica è risultata pari al 28.0%.

La tabella 5-4 mostra i risultati di una regressione logistica “tradizionale” che non tiene in alcuna considerazione la struttura gerarchica paziente-medico. I dati sono stati disaggregati, pertanto, le variabili esplicative misurate sul medico di medicina generale sono state “spostate” verso il livello più basso della gerarchia, costituito dai pazienti in trattamento antiipertensivo.

Tabella 5-4. Modello di regressione logistica su dati disaggregati: predittori di discontinuità

		Odds Ratio	ES	P-value*	IC 95%
Età del medico	5 anni	1.31	0.07	< 0.001	1.18-1.46
Sesso del medico	(femmina)	0.71	0.07	< 0.001	0.59-0.85
Numero di assistibili	100 soggetti	1.02	0.01	0.132	0.99-1.05
Età del paziente	10 anni	0.83	0.03	< 0.001	0.77-0.89
Sesso del paziente	(femmina)	0.94	0.08	0.437	0.79-1.11
Pregressi eventi CV	(assenza)	0.92	0.11	0.490	0.74-1.16
Diabete	(assenza)	1.24	0.25	0.292	0.83-1.84
Cardiopatìa	(assenza)	0.95	0.22	0.826	0.61-1.48
Asma/Bpco	(assenza)	0.98	0.22	0.927	0.64-1.51
Inizio del trattamento	(recente)	0.31	0.03	< 0.001	0.26-0.37
PAS	5 mm/Hg	0.99	0.01	0.520	0.97-1.02
PAD	5 mm/Hg	0.91	0.02	< 0.001	0.87-0.95
Classe di FAI	-	-	-	< 0.001	-
Altri antiipertensivi	(renina-angiotensina)	1.40	0.36	0.198	0.84-2.33
Diuretici	(renina-angiotensina)	2.15	0.28	< 0.001	1.67-2.78
Beta Bloccanti	(renina-angiotensina)	1.38	0.15	0.003	1.11-1.72
Calcio Antagonisti	(renina-angiotensina)	1.11	0.13	0.381	0.88-1.39
Statine	(assenza)	0.65	0.09	0.002	0.50-0.86
Antiaggreganti	(assenza)	0.64	0.08	0.001	0.49-0.83

Tra parentesi è indicata la categoria di riferimento (Odds Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco Antiipertensivo. *: test di Wald.

Come ampiamente trattato nei paragrafi 2.1 e 2.2, i risultati prodotti da questo modello di regressione non sono attendibili per due motivi: in primo luogo, l’ipotesi di indipendenza tra le osservazioni risulta violata, pertanto le stime degli errori standard prodotte dai test statistici convenzionali sono troppo piccole e i risultati che si ottengono appaiono “impropriamente” significativi; in secondo luogo, alcune delle variabili esplicative sono misurate sul medico di medicina

generale ma i test statistici ordinari considerano i valori disaggregati come informazioni indipendenti provenienti dal vasto insieme delle unità di primo livello, i pazienti. La reale numerosità campionaria per le variabili disaggregate, tuttavia, è costituita dal ben più esiguo numero di medici di medicina generale.

Il modello di regressione logistica a intercetta casuale, senza covariate di primo o secondo livello, ha stimato un valore pari a 0.71 per la varianza di secondo livello σ^2_{u0} con un errore standard pari a 0.22 (test di Wald: $p < 0.001$; test del rapporto delle verosimiglianze: $\chi^2_{(1)} = 207.32$, $p < 0.001$)². Il coefficiente di correlazione intraclasse ρ ammontava al 17.68%. L'odds ratio mediano risultava pari a 2.23.

Dal momento che parte della variabilità tra i gruppi potrebbe essere attribuibile alla diversa composizione degli assistiti tra i diversi medici di medicina generale, sono state inserite nel modello a intercetta casuale tutte le covariate di primo livello. La stima della varianza condizionata $\sigma^2_{u0|\mathbf{x}}$ è risultata pari a 0.48 con un errore standard di 0.16 (test di Wald: $p = 0.001$; test del rapporto delle verosimiglianze: $\chi^2_{(1)} = 142.63$, $p < 0.001$). Il coefficiente di correlazione intraclasse residuo $\rho|\mathbf{x}$ ammontava al 12.73%. L'odds ratio mediano risultava pari a 1.94.

² Come discusso nel paragrafo 2.6, il test del rapporto delle verosimiglianze è più efficace del test di Wald qualora si sottopongano a verifica di ipotesi le componenti della varianza. Tuttavia, il test di Wald è piuttosto diffuso grazie ad una maggiore rapidità computazionale. In questa tesi, per completezza di esposizione e per agevolare il confronto tra le due metodologie, sono stati presentati i risultati di entrambi i test. Si precisa, inoltre, che tutti i test sulle componenti della varianza sono stati realizzati “a una coda” dal momento che le varianze non possono essere negative.

Inserendo nel modello a intercetta casuale anche le covariate di secondo livello sono stati ottenuti i risultati mostrati in tabella 5-5.

Tabella 5-5. Modello di regressione logistica con intercetta casuale: predittori di primo e secondo livello

		Odds Ratio	ES	P-value*	IC 95%
Età del medico	5 anni	1.31	0.22	0.097	0.95-1.81
Sesso del medico	(femmina)	0.83	0.22	0.481	0.49-1.40
Numero di assistibili	100 soggetti	0.99	0.03	0.724	0.92-1.06
Età del paziente	10 anni	0.84	0.03	< 0.001	0.78-0.91
Sesso del paziente	(femmina)	0.91	0.08	0.302	0.77-1.08
Pregressi eventi CV	(assenza)	0.91	0.11	0.454	0.72-1.15
Diabete	(assenza)	1.25	0.26	0.278	0.83-1.89
Cardiopatia	(assenza)	0.96	0.22	0.877	0.61-1.52
Asma/Bpco	(assenza)	1.07	0.24	0.774	0.69-1.66
Inizio del trattamento	(recente)	0.34	0.03	< 0.001	0.28-0.41
PAS	5 mm/Hg	0.98	0.01	0.102	0.95-1.01
PAD	5 mm/Hg	0.95	0.02	0.028	0.90-0.99
Classe di FAI	-	-	-	< 0.001	-
Altri antiipertensivi	(renina-angiotensina)	1.49	0.40	0.136	0.88-2.53
Diuretici	(renina-angiotensina)	2.37	0.32	< 0.001	1.82-3.09
Beta Bloccanti	(renina-angiotensina)	1.39	0.16	0.004	1.11-1.75
Calcio Antagonisti	(renina-angiotensina)	1.12	0.14	0.365	0.88-1.42
Statine	(assenza)	0.69	0.10	0.010	0.52-0.91
Antiaggreganti	(assenza)	0.70	0.09	0.009	0.54-0.92
Componenti della varianza		σ^2			
σ_{u0}^2	-	0.42	0.14	0.002	-

Tra parentesi è indicata la categoria di riferimento (Odds Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco Antiipertensivo. *: test di Wald.

Test del rapporto delle verosimiglianze per σ_{u0}^2 :
 $\chi^2_{(1)} = 109.12, p < 0.001.$

Nelle tabelle che seguono sono presentate le stime ottenute mediante il modello di regressione di Poisson con intercetta casuale (tabella 5-6) e mediante il modello di regressione di Cox con *frailty* condivise a livello di gruppo (tabella 5-7).

Tabella 5-6. Modello di regressione di Poisson con intercetta casuale: predittori di primo e secondo livello

		Rate Ratio	ES	P-value*	IC 95%
Età del medico	5 anni	1.30	0.20	0.095	0.96-1.76
Sesso del medico	(femmina)	0.85	0.22	0.516	0.52-1.39
Numero di assistibili	100 soggetti	0.99	0.03	0.678	0.93-1.05
Età del paziente	10 anni	0.85	0.03	< 0.001	0.79-0.91
Sesso del paziente	(femmina)	0.89	0.07	0.118	0.77-1.03
Pregressi eventi CV	(assenza)	0.95	0.10	0.625	0.78-1.16
Diabete	(assenza)	1.17	0.21	0.381	0.82-1.67
Cardiopatìa	(assenza)	0.97	0.20	0.874	0.64-1.45
Asma/Bpco	(assenza)	1.08	0.21	0.697	0.73-1.59
Inizio del trattamento	(recente)	0.34	0.03	< 0.001	0.29-0.41
PAS	5 mm/Hg	0.98	0.01	0.029	0.95-0.99
PAD	5 mm/Hg	0.95	0.02	0.027	0.91-0.99
Classe di FAI	-	-	-	< 0.001	-
Altri antiipertensivi	(renina-angiotensina)	1.44	0.33	0.106	0.92-2.25
Diuretici	(renina-angiotensina)	2.36	0.26	< 0.001	1.90-2.92
Beta Bloccanti	(renina-angiotensina)	1.32	0.13	0.004	1.10-1.60
Calcio Antagonisti	(renina-angiotensina)	1.11	0.12	0.333	0.90-1.36
Statine	(assenza)	0.72	0.09	0.012	0.56-0.93
Antiaggreganti	(assenza)	0.72	0.09	0.005	0.57-0.90
Componenti della varianza		σ^2			
σ^2_{u0}	-	0.39	0.12	< 0.001	-

Tra parentesi è indicata la categoria di riferimento (Rate Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco AntiIpertensivo. *: test di Wald.

Test del rapporto delle verosimiglianze per σ^2_{u0} :
 $\chi^2_{(1)} = 154.13, p < 0.001$.

Tabella 5-7. Modello di regressione di Cox con frailty gamma-distribuite condivise a livello di gruppo: predittori di primo e secondo livello

		Hazard Ratio	ES	P-value*	IC 95%
Età del medico	5 anni	1.32	0.20	0.067	0.98-1.77
Sesso del medico	(femmina)	0.91	0.21	0.678	0.58-1.42
Numero di assistibili	100 soggetti	0.98	0.03	0.438	0.92-1.04
Età del paziente	10 anni	0.86	0.03	< 0.001	0.80-0.91
Sesso del paziente	(femmina)	0.91	0.07	0.202	0.79-1.05
Pregressi eventi CV	(assenza)	0.95	0.10	0.650	0.78-1.17
Diabete	(assenza)	1.17	0.21	0.381	0.82-1.66
Cardiopatìa	(assenza)	0.97	0.20	0.889	0.65-1.46
Asma/Bpco	(assenza)	1.06	0.21	0.769	0.72-1.56
Inizio del trattamento	(recente)	0.37	0.03	< 0.001	0.31-0.44
PAS	5 mm/Hg	0.98	0.01	0.052	0.96-1.01
PAD	5 mm/Hg	0.96	0.02	0.036	0.92-0.99
Classe di FAI	-	-	-	< 0.001	-
Altri antiipertensivi	(renina-angiotensina)	1.43	0.33	0.114	0.92-2.23
Diuretici	(renina-angiotensina)	2.23	0.24	< 0.001	1.80-2.76
Beta Bloccanti	(renina-angiotensina)	1.30	0.13	0.007	1.07-1.57
Calcio Antagonisti	(renina-angiotensina)	1.11	0.12	0.321	0.90-1.36
Statine	(assenza)	0.73	0.09	0.012	0.56-0.93
Antiaggreganti	(assenza)	0.72	0.09	0.006	0.57-0.91
Componenti della varianza		σ^2			
θ (<i>frailty variance</i>)	-	0.31	0.10	< 0.001	-

Tra parentesi è indicata la categoria di riferimento (Hazard Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco AntiIpertensivo. *: test di Wald.

Test del rapporto delle verosimiglianze per θ :
 $\chi^2_{(1)} = 104.15, p < 0.001.$

In relazione alla variabile che distingue i già trattati dai nuovi trattati, l'analisi grafica dei residui parziali di Schoenfeld ha evidenziato una leggera violazione dell'assunto di proporzionalità dei rischi. Probabilmente, col passare del tempo di follow-up, l'atteggiamento dei nuovi trattati rispetto alla terapia farmacologica antiipertensiva si avvicina a quello dei soggetti già in trattamento e la differenza tra i tassi istantanei di interruzione delle due categorie di pazienti diviene progressivamente meno marcata.

Confrontando i risultati presentati nelle tabelle 5-5, 5-6, e 5-7 si nota come i tre modelli di regressione con effetti casuali producano, in relazione alla parte fissa del modello, stime estremamente simili. La tabella 5-8 offre un confronto diretto tra le misure di effetto stimate dai diversi modelli di regressione.

Tabella 5-8. Effetti fissi: confronto tra i diversi modelli di regressione

		Odds Ratio	Rate Ratio	Hazard Ratio
Età del medico	5 anni	1.31	1.30	1.32
Sesso del medico	(femmina)	0.83	0.85	0.91
Numero di assistibili	100 soggetti	0.99	0.99	0.98
Età del paziente	10 anni	0.84*	0.85*	0.86*
Sesso del paziente	(femmina)	0.91	0.89	0.91
Pregressi eventi CV	(assenza)	0.91	0.95	0.95
Diabete	(assenza)	1.25	1.17	1.17
Cardiopatìa	(assenza)	0.96	0.97	0.97
Asma/Bpco	(assenza)	1.07	1.08	1.06
Inizio del trattamento	(recente)	0.34*	0.34*	0.37*
PAS	5 mm/Hg	0.98	0.98*	0.98
PAD	5 mm/Hg	0.95*	0.95*	0.96*
Classe di FAI	-	-	-	-
Altri antiipertensivi	(renina-angiotensina)	1.49	1.44	1.43
Diuretici	(renina-angiotensina)	2.37*	2.36*	2.23*
Beta Bloccanti	(renina-angiotensina)	1.39*	1.32*	1.30*
Calcio Antagonisti	(renina-angiotensina)	1.12	1.11	1.11
Statine	(assenza)	0.69*	0.72*	0.73*
Antiaggreganti	(assenza)	0.70*	0.72*	0.72*

Tra parentesi è indicata la categoria di riferimento (Hazard Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco Antiipertensivo. *: $p < 0.05$.

La tabella 5-9 pone, invece, a confronto le componenti della varianza stimate nella parte casuale dei tre modelli.

Tabella 5-9. Componenti della varianza: confronto tra i diversi modelli di regressione

	σ^2	ES	P-value*	P-value [^]
Regressione logistica con intercetta casuale				
$\sigma^2_{u0 \mathbf{x}}$	0.48	0.16	0.0011	< 0.0001
$\sigma^2_{u0 \mathbf{x}, \mathbf{z}}$	0.42	0.14	0.0015	< 0.0001
Regressione di Poisson con intercetta casuale				
$\sigma^2_{u0 \mathbf{x}}$	0.43	0.13	0.0003	< 0.0001
$\sigma^2_{u0 \mathbf{x}, \mathbf{z}}$	0.39	0.12	0.0004	< 0.0001
Regressione di Cox con frailty condivise				
$\theta \mathbf{x}$	0.35	0.11	0.0005	< 0.0001
$\theta \mathbf{x}, \mathbf{z}$	0.31	0.10	0.0008	< 0.0001

\mathbf{x} : vettore di covariate di primo livello; \mathbf{z} : vettore di covariate di secondo livello; θ : frailty variance. *: test di Wald; ^: test del rapporto delle verosimiglianze.

Si puntualizza che le varianze condizionate di secondo livello $\sigma^2_{u0|\mathbf{x}}$ e $\sigma^2_{u0|\mathbf{x}, \mathbf{z}}$ non sono direttamente confrontabili con le *frailty variance* $\theta|\mathbf{x}$ e $\theta|\mathbf{x}, \mathbf{z}$. Infatti la frailty corrisponde all'esponentiale dell'intercetta casuale poiché è il logaritmo della frailty ad entrare nel predittore lineare come effetto casuale.

Nella tabella successiva sono presentate le stime ottenute restringendo le analisi ai nuovi trattati e utilizzando un modello di regressione di Cox con frailty gamma-distribuite condivise a livello di gruppo.

Al fine di semplificare il modello, sono state escluse le seguenti covariate: numero di assistibili, pregressi eventi CV, diabete, cardiopatia, asma/bpco. Queste variabili, infatti, non risultavano statisticamente significative, non erano fattori confondenti e aumentavano la variabilità campionaria delle altre stime.

Tabella 5-10. Modello di regressione di Cox con frailty gamma-distribuite condivise a livello di gruppo: analisi ristrette ai nuovi trattati, predittori di primo e secondo livello

		Hazard Ratio	ES	P-value*	IC 95%
Età del medico	5 anni	1.39	0.18	0.011	1.08-1.80
Sesso del medico	(femmina)	0.72	0.15	0.118	0.48-1.09
Età del paziente	10 anni	0.83	0.03	< 0.001	0.77-0.90
Sesso del paziente	(femmina)	0.88	0.08	0.167	0.73-1.05
PAS	5 mm/Hg	0.98	0.01	0.178	0.95-1.01
PAD	5 mm/Hg	0.92	0.02	0.001	0.87-0.97
Classe di FAI	-	-	-	< 0.001	-
Altri antiipertensivi	(renina-angiotensina)	1.38	0.41	0.283	0.77-2.49
Diuretici	(renina-angiotensina)	2.37	0.31	< 0.001	1.84-3.06
Beta Bloccanti	(renina-angiotensina)	1.23	0.15	0.098	0.96-1.56
Calcio Antagonisti	(renina-angiotensina)	1.07	0.15	0.632	0.82-1.40
Statine	(assenza)	0.83	0.13	0.246	0.61-1.13
Antiaggreganti	(assenza)	0.70	0.11	0.022	0.52-0.95
Componenti della varianza		σ^2			
θ (<i>frailty variance</i>)	-	0.22	0.08	0.003	-

Tra parentesi è indicata la categoria di riferimento (Hazard Ratio = 1).

CV: Cardio/cerebrovascolare; PAS: Pressione Arteriosa Sistolica; PAD: Pressione Arteriosa Diastolica; FAI: Farmaco Antiipertensivo. *: test di Wald.

Test del rapporto delle verosimiglianze per θ :
 $\chi^2_{(1)} = 47.39, p < 0.001.$

5.1. Discussione

L'importanza clinica, sociale ed economica del trattamento dell'ipertensione arteriosa richiede la messa in opera di strumenti di monitoraggio dell'uso dei farmaci antiipertensivi che consentano di verificare la trasferibilità alla pratica clinica dei dati ottenuti nelle sperimentazioni.

L'attuazione di una adatta strategia terapeutica non è un fenomeno semplice da realizzare perché le condizioni in cui opera il medico di medicina generale sono profondamente differenti da quelle che si predispongono nell'esecuzione degli studi randomizzati e controllati. Un requisito indispensabile dei *clinical trials* è che la terapia antiipertensiva sia somministrata a dose adeguata, in modo continuativo e per lunghi periodi per risultare efficace nel ridurre e/o normalizzare i valori pressori.

I risultati di questo studio mostrano che il 19% dei trattati con antiipertensivi ha interrotto la terapia farmacologica durante i 365 giorni di follow-up. Nei nuovi trattati, la percentuale di interruzione terapeutica ammontava al 28%.

Le caratteristiche-paziente individuate dall'analisi multilivello indicano come la probabilità di interrompere il trattamento sia più elevata nei soggetti che presentano una situazione clinica generale migliore (giovane età, assenza di trattamenti concomitanti, bassi livelli di pressione arteriosa diastolica). Questi soggetti, oltre a non essere abituati ad assumere altre terapie croniche, percepiscono in minor misura i potenziali benefici del trattamento antiipertensivo e tenderanno a interrompere la terapia farmacologica alla comparsa dei primi effetti collaterali. Il modello ha inoltre evidenziato come i nuovi trattati presentino una più elevata probabilità di interruzione terapeutica, verosimilmente spiegata dalla difficoltà di abituarsi all'assunzione cronica del farmaco in una fase di assestamento della terapia in cui i principi attivi di prima scelta potrebbero non adattarsi pienamente, in termini di tollerabilità, alle caratteristiche del paziente.

Anche la classe di farmaco di prima scelta riveste un ruolo essenziale nella determinazione dei livelli di *compliance*. Il fenomeno è probabilmente legato ai

diversi profili di tollerabilità delle numerose alternative terapeutiche (Degli Esposti, Di Martino, Saragoni et al., 2004).

L'appropriato riconoscimento dei predittori-paziente di discontinuità (*risk profiling*) e la loro valutazione globale nella pratica clinica quotidiana potrebbe contribuire a migliorare il rapporto medico-paziente e incrementare i livelli di compliance al trattamento.

L'analisi delle componenti della varianza ha evidenziato come il 18% della variabilità nella persistenza in trattamento antiipertensivo sia attribuibile al livello Medico di Medicina Generale. Controllando per le differenze demografiche e cliniche tra gli assistiti dei diversi medici, la quota di variabilità attribuibile al livello medico risultava pari al 13%.

La capacità empatica dei prescrittori nel comunicare ai propri pazienti l'importanza della terapia farmacologica riveste un ruolo importante nel determinare i livelli di compliance al trattamento. La crescente presenza, nella formazione dei medici, di corsi di carattere psicologico finalizzati a migliorare il rapporto medico-paziente potrebbe, inoltre, spiegare la relazione inversa, particolarmente evidente nella sottoanalisi effettuata sui nuovi trattati, tra età del medico e persistenza in trattamento.

La proporzione non trascurabile di variabilità spiegata dalla struttura in gruppi degli assistiti evidenzia l'opportunità e la necessità di investire nella formazione dei Medici di Medicina Generale con l'obiettivo di sensibilizzare ed "educare" i medici alla motivazione ma anche al monitoraggio dei soggetti trattati, alla sistematica valutazione in pratica clinica dei predittori-paziente di discontinuità e a un appropriato utilizzo della classe di farmaco di prima scelta.

Limiti dello studio

Uno dei possibili limiti di questo studio risiede nella ridotta rappresentatività del campione di medici (la partecipazione al progetto era su base volontaria) e di pazienti (la presenza di almeno una misurazione di pressione arteriosa, dettata dai

criteri di arruolamento, potrebbe aver distorto il campione analizzato, selezionando i pazienti che si recano dal proprio medico con maggior frequenza). Questo potrebbe spiegare la minore incidenza di interruzioni terapeutiche rispetto a studi condotti, nella stessa area geografica, mediante database amministrativi di popolazione (Degli Esposti, Di Martino, Saragoni et al., 2004).

Conclusioni

L'analisi dei dati contenuti nei database della medicina generale ha consentito di valutare l'impiego dei farmaci antiipertensivi nella pratica clinica e di stabilire la necessità di porre una maggiore attenzione nella pianificazione e nell'ottenimento dell'obiettivo che il trattamento si prefigge. Alla luce dei risultati emersi da questa valutazione, sarebbe di grande utilità la conduzione di ulteriori studi osservazionali volti a sostenere il progressivo miglioramento della gestione e del trattamento dei pazienti a rischio cardiovascolare nell'ambito della medicina generale.

Bibliografia

- Alker, H.R. (1969). A typology of fallacies, in M. Dogan & S. Rokkan (Eds.), *Quantitative ecological analysis in the social sciences*. Cambridge, Ma.: M.I.T. Press.
- Berkhof, J., Snijders, T.A.B. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics*, 26, 133-152.
- Birnbaum, H.G., Cremieux, P.Y., Greenberg, P.E. (1999). Using healthcare claims data for outcome research and pharmacoeconomic analyses. *Pharmacoeconomics*, 16, 1-8.
- Black, N. (1996). Why do we need observational studies to evaluate the effectiveness of health care? *British Medical Journal*, 312, 1215-1218.
- Black, N. (1997). Developing high quality clinical databases. The key to a new research paradigm. *British Medical Journal*, 315, 381-2.
- Breslow, N.E., Clayton, D.G. (1993). Approximate inference in generalised linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.
- Cifarelli, D.M., Muliere, P. (1989). *Statistica Bayesiana*. Pavia, Gianni Iuculano Editore.
- Dahlof, B., Lindholm, L.H., Hansson, L. et al. (1991). Morbidity and mortality in the Swedish Trial in Old Patients with Hypertension (STOP-Hypertension). *Lancet*, 338, 1281-1285.
- Degli Esposti, E., Berto, P., Buda, S. et al. (1999). The PANDORA project: results of the pilot study. *American Journal of Hypertension*, 12, 790-796.
- Degli Esposti, E., Sturani, A., Di Martino, M. et al. (2002). Long-term persistence with antihypertensive drugs in new patients. *Journal of Human Hypertension*, 16, 439-444.
- Degli Esposti, L., Di Martino, M., Saragoni, S. et al. (2004) Pharmacoeconomics of antihypertensive drug treatment: an analysis of

how long patients remain on various antihypertensive therapies. *Journal of Clinical Hypertension*, 6(2), 76-84.

- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.
- Di Martino, M., Degli Esposti, L., Ruffo, P. et al. (2005). Underuse of lipid-lowering drugs and factors associated with poor adherence: a real practice analysis in Italy. *European Journal of Clinical Pharmacology*, 61, 225-230.
- Di Martino, M., Veronesi, C., Degli Esposti, L. et al. (2008). Adherence to antihypertensive drug treatment and blood pressure control: a real practice analysis in Italy. *Journal of Human Hypertension*, 22(1), 51-3.
- Dieppe, P., Bartlett, C., Davey, P. et al. (2004). Balancing benefits and harms: the example of non-steroidal anti-inflammatory drugs. *British Medical Journal*, 329, 31-4.
- Diez-Roux, A.V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, 21, 171-92.
- Farmer, K.C. (1999). Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. *Clinical Therapeutics*, 21, 1074-90.
- Fears, T.R., Benichou, J., Gail, M.H. (1996). A reminder of the fallibility of the Wald statistic. *American Statistician*, 50(3), 226-227.
- Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold/New York: Halsted.
- Goldstein, H., Browne, W., Rasbash, J. (2002). Multilevel Modelling of medical data. *Statistics in Medicine*, 21, 3291-315.
- Goldstein, H., Rasbash, J. (1996). Improved approximations to multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 159, 505-513.
- Greene, W.H. (1997). *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.

- Greenes, R.A., Shortliffe, E.H. (1990). Medical Informatics. An emerging academic discipline and institutional priority. *Journal of the American Medical Association*, 263(8), 1114-20.
- Grilli, R., Taroni, F. (2004). *Governo clinico. Governo delle organizzazioni sanitarie e qualità dell'assistenza*. Roma: Il Pensiero Scientifico Editore.
- Gutierrez, R.G. (2002). Parametric frailty and shared frailty survival models. *The Stata Journal*, 2(1), 22-44.
- Hartford, A., Davidian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 34, 139-164.
- Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, 1, 255-273.
- Hox, J. (2002). *Multilevel Analysis: techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- ISTAT (2007). *Condizioni di salute, fattori di rischio e ricorso ai servizi sanitari*. Indagine Multiscopo sulle famiglie “Condizioni di salute e ricorso ai servizi sanitari”, anno 2005.
- Joint National Committee VII (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *JAMA*, 289(19), 2560-72.
- Keiding, N., Andersen, P.K., Klein, J.P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16, 215-224.
- Kish, L. (1965). *Survey Sampling*. NY: Wiley.
- Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, 48, 795-806.
- Larsen, K., Merlo, J. (2005). Appropriate Assessment of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression. *American Journal of Epidemiology*, 161, 81-88.

- Lee, Y., Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, 58, 619-678.
- Lindsey, J.K. (1993). *Models for Repeated Measurements*. Oxford: Clarendon Press.
- Longford, N.T. (1993). *Random coefficient models*. Oxford: Clarendon Press.
- Maples, J.J., Murphy, S.A., Axinn, W.G. (2002). Two-level proportional hazards models. *Biometrics*, 58, 754-763.
- McCullagh, P., Nelder, J.A. (1989). *Generalized linear models*. Second Edition. London: Chapman & Hall.
- McGilchrist, C.A. (1993). REML estimation for survival models with frailty. *Biometrics*, 49, 221-225.
- McGilchrist, C.A. (1994). Estimation in generalised mixed models. *Journal of the Royal Statistical Society, Series B*, 56, 61-69.
- McGilchrist, C.A., Yau, K.K.W. (1995). The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models. *Communications in Statistics – Theory and Methods*, 24, 2963-2980.
- Ministero Della Salute (2002). Circolare Ministeriale n. 6 del 2 settembre 2002, emanata dal Ministero della Salute, pubblicata sulla *G.U. n. 214 del 12 settembre 2002*.
- Moskowitz, D.S., Hershberger Scott, L. (2001). *Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications*. Lawrence Erlbaum Associates.
- Motheral, B.R., Fairman, K.A. (1997). The use of claims databases for outcome research: rational challenges and strategies. *Clinical Therapeutics*, 19(2), 346-66.
- Nielsen, G.G., Gill, R.D., Andersen, P.K. et al. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19, 25-44.
- OSMED - Osservatorio Nazionale sull'Impiego dei Medicinali (2005). *L'uso dei farmaci in Italia*. Rapporto nazionale anno 2004.

- Parner, E. (1997). *Inference in semiparametric frailty models*. PhD thesis, University of Aarhus, Denmark.
- Pignone, M., Phillips, C., Mulrow, C. (2000). Use of lipid lowering drugs for primary prevention of coronary heart disease: meta-analysis of randomised trials. *British Medical Journal*, 321, 983-6.
- Pocock, S.J. (1983). *Clinical Trials: A Practical Approach*. Chichester: Wiley.
- Revicki, D.A., Frank, L. (1999). Pharmacoeconomic evaluation in the real world. *Pharmacoeconomics*, 15(5), 423-34.
- Ripatti, S., Palmgren, J. (2000). Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood. *Biometrics*, 56(4), 1016-1022.
- Rodriguez, G., Goldman, N. (2001). Improved estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 164, part 2, 339-355.
- Searle, S.R., Casella, G., McCulloch, C.E. (1992). *Variance components*. New York, Wiley.
- Sinha, D., Tanner, M.A., Hall, W.J. (1994). Maximization of the marginal likelihood of grouped survival data. *Biometrika*, 81, 53-60.
- Snijders, T.A.B., Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modelling*. Thousand Oaks, CA: Sage.
- Teutsch, C. (2003). Patient-doctor communication. *The Medical Clinics of North America*, 87(5), 1115-45.
- Therneau, T.M., Grambsch, P.M. (2000). *Modeling survival data: extending the Cox model*. New York, Springer-Verlag.
- Vaida, F., Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, 19, 3309-3324.
- Van Der Leeden, R., Busing, F., Meijer, E. (1997). Applications of bootstrap methods for two-level models. Paper, Multilevel Conference, Amsterdam, April 1-2, 1997.

- Vaupel, J.W., Manton, K.G., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439-454.
- WHO, Collaborating Centre for Drug Statistics Methodology. (2002). *Guidelines for ATC classification and DDD assignment*, 5th edition.
- Wolfinger, R.W. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80, 791-795.
- Wyatt, J.C. (1994). Clinical data systems, part 1: data and medical records. *Lancet*, 344, 1543-7.
- Wyatt, J.C. (1994). Clinical data systems, part 2: components and techniques. *Lancet*, 344, 1609-14.
- Wyatt, J.C. (1994). Clinical data systems, part 3: development and evaluation. *Lancet*, 344, 1682-8.
- Yau, K.K.W., Mcgilchrist, C.A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine*, 17, 1201-1213.