



MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITA' E DELLA RICERCA

Programmi di ricerca cofinanziati - Modello E Relazione scientifica conclusiva sui risultati di ricerca ottenuti - ANNO 2007 prot. 2007AYHZWC

1. Area Scientifico Disciplinare principale	<i>13: Scienze economiche e statistiche</i>
2. Coordinatore Scientifico del programma di ricerca	<i>GASPARINI Mauro</i>
- Università	<i>Politecnico di TORINO</i>
- Facoltà	<i>Facoltà di INGEGNERIA</i>
- Dipartimento/Istituto	<i>Dip. MATEMATICA</i>
3. Titolo del programma di ricerca	<i>Metodi statistici per l'apprendimento nella ricerca clinica</i>
4. Settore principale del Programma di Ricerca:	<i>SECS-S/01</i>
5. Costo originale del Programma:	<i>136.200 €</i>
6. Quota Cofinanziamento MIUR:	<i>41.126 €</i>
7. Quota Cofinanziamento Ateneo:	<i>27.905 €</i>
8. Finanziamento totale:	<i>69.031 €</i>
9. Durata:	<i>24 mesi</i>

10. Obiettivo della ricerca eseguita

Le prove cliniche sono un successo della civiltà contemporanea poiché "Rappresentano una delle poche aree in cui gli effetti di nuova tecnologia e di nuovi programmi su esseri umani vengono studiati usando principi statistici moderni di disegno sperimentale" (Simon 2006). Questo è il punto di partenza di questo progetto, che si concentra su alcuni aspetti metodologici usati nelle prove cliniche e più in generale negli studi biomedici.

Gli stessi aspetti sono presenti anche in altri campi della moderna sperimentazione, come la gestione della qualità in campo aziendale o la genomica in campo biologico. Tuttavia, poiché le prove cliniche sono un campo maturo e ampiamente partecipato, i fondamenti concettuali e le pratiche statistiche di base possono essere date per scontate, così che la proposta attuale può concentrarsi sugli aspetti più strettamente metodologici.

Le prove cliniche sono i mattoni necessari per lo sviluppo clinico di una nuova terapia, e le prove cliniche comparative di fase III sono le più centrali. Nel caso di un'azienda farmaceutica, le prove di fase III sono fatte per ottenere un'autorizzazione dalle Autorità Sanitarie e immettere sul mercato la nuova sostanza, mentre nel caso di prove cliniche indipendenti un simile esercizio di "fase III" è condotto per ottenere riconoscimento condiviso del metodo e raccomandazione di uso per la pratica clinica.

Per rendere possibile le prove cliniche comparative e dimostrative su larga scala, durante lo sviluppo clinico lo sponsor di una nuova terapia deve comunque prendere importanti decisioni e risolvere delicati problemi inferenziali. In questo modo, nuovi risultati vengono incorporati in un processo sequenziale e le migliori decisioni cliniche e scientifiche vengono prese in modo da massimizzare la probabilità di ottenere buone medicine e in modo da utilizzare le risorse disponibili in maniera ottimale.

Esiste quindi una molteplicità di tipi di prove cliniche, e più in generale di studi biomedici, volti ad esplorare alcuni importanti quesiti epistemologici.

La ricerca biomedica di sviluppo clinico presenta dunque allo statistico una serie di problemi che si possono collettivamente descrivere come la costruzione dell'evidenza clinica su base sperimentale, ovvero come l'apprendimento nella ricerca clinica. La statistica diventa in questo modo parte fondamentale del moderno movimento chiamato "evidence-based medicine".

Il processo di apprendimento descritto avviene sia all'interno di un singolo studio sia sul collegamento di più studi.

Esempi di apprendimento all'interno di un singolo studio sono:

- la ricerca della dose ottimale;
- i disegni sperimentali adattativi,

mentre esempi di apprendimento su più studi sono

- il passaggio dalle prove cliniche di fase II a quelle di fase III,
- la valutazione della riproducibilità dei risultati,
- l'analisi dei sottogruppi e l'individuazione di bioprofili per la generazione di ipotesi di lavoro per gli studi futuri
- le analisi multivariate di risposte multiple.

Obiettivi finali della ricerca proposta sono lo studio approfondito di ciascuno di questi argomenti, la produzione di ricerca scientifica a livello di frontiera e la concreta collaborazione con centri assistenza e ricerca clinica e con centri di produzione biomedica.

11. Descrizione della Ricerca eseguita e dei risultati ottenuti

UNITA' DI TORINO POLITECNICO

L'unità di TORINO POLITECNICO ha organizzato una riunione scientifica di apertura il 10 Ottobre 2008 che ha visto la partecipazione e la presentazione del programma da parte di tutte le unità. Ha quindi partecipato attivamente a tutte le riunioni successive, come documentato dal sito web <http://www.smlcr.altervista.org/>, mantenendo un ruolo di stimolo e di coordinamento.

Il compito TO1, la ricerca della dose per terapie combinate, ha visto l'impegno di Mauro Gasparini ed Elisa Bianchini. La critica di un metodo basato sulle copule è dovuto a Yin e Yang, dell'MD Anderson Cancer Center di Houston, è stata redatta da Gasparini congiuntamente con ricercatori della Novartis di Basilea ed è stata accettata da Applied Statistics (JRSS C). Vedasi Gasparini et al. (2010). Un secondo articolo di approfondimento delle implicazioni teoriche e pratiche dei diversi modelli usati per descrivere l'interazione tra gli effetti tossici di due terapie è in corso di redazione (Gasparini et al.). La dottoressa Elisa Bianchini sta concludendo un dottorato di ricerca in "Statistica Biomedica" presso la Scuola di Dottorato in Scienze biomediche cliniche e sperimentali dell'Università degli Studi di Milano. Questo PRIN ha contribuito in maniera significativa alle spese di viaggio e di studio per il conseguimento del dottorato, previsto per i primi mesi del 2011. Tale dottorato e più in generale le attività di ricerca di Bianchini nell'ambito del compito I sono incentrati su diversi temi: un inquadramento generale degli studi di Fase I in ambito Oncologico, l'analisi del cambiamento nel disegno degli studi di Fase I derivanti dall'introduzione di nuovi farmaci con meccanismo di azione citostatico (Terapie Biologiche o Target), lo sviluppo di modelli Bayesiani per misure ripetute di tossicità, basati sul metodo del Continual Reassessment di O'Quigley.

Il Compito TO2 era l'Analisi di sottogruppi mediante STEPP. A consolidamento del metodo e della sua diffusione, si è giunti alla pubblicazione di Bonetti et al. (2009b). E' stata inoltre portata avanti un interessante scambio di comunicazioni con altri colleghi (Royston e Sauerbrei) sulle pagine del Journal of Clinical Oncology per discutere i meriti relativi del nostro approccio STEPP e di uno dei metodi alternativi basati sui fractional polynomials. Vedasi Bonetti et al. (2008). Infine, si è portato avanti lo sviluppo di estensioni ulteriori del metodo STEPP, in particolare all'ambito dei competing risks. L'articolo in questione è stato accettato sulla sezione di metodi statistici del prestigioso Journal of Clinical Oncology (Lazar et al. 2010). Una ulteriore pubblicazione è stata preparata per descrivere un'applicazione dei metodi ad uno studio danese di tumore al seno: (Gunnarsdottira et al.). Infine, si è preparata e distribuita la release del software R che implementa l'algoritmo mediante permutazioni descritto in Bonetti et al. (2009) ed in Lazar et al. (2010). Il pacchetto in questione può essere scaricato da <https://sites.google.com/site/stepprpackage/> Questi lavori sono stati svolti in collaborazione con colleghi della Harvard University e del Dana Farber Cancer Institute di Boston, USA.

Si è invece affrontato il tema TO3, quello dell'eterogeneità dal punto di vista dello studio della concentrazione dei tempi di sopravvivenza, con particolare attenzione ai modelli di cura, nella pubblicazione Bonetti et al. (2009a). che descrive i metodi in questione. E' attualmente in corso uno sviluppo aggiuntivo che studi la performance del metodo nei casi di sbilanciamento nella numerosità campionaria dei due gruppi.

Per quanto riguarda il compito TO4, un articolo di confronto tra diversi stimatori nonparametrici della funzione di sopravvivenza in presenza di dati censurati a destra e troncati a sinistra è stato sottoposto alla rivista Statistica (Gasparini e Gandini). Tale lavoro è scritto congiuntamente da Gasparini e Gandini e affronta il problema

dal punto di vista applicativo all'interno di una consulenza con il Centro Microcitomie dell'Università di Torino sulla sopravvivenza dei malati di talassemia maggiore.

Nel 2009 è stato inserito nell'Unità locale di TORINO POLITECNICO il dottor Luca Pozzi, che sta attivamente contribuendo al filone di ricerca sul compito TO5: metodi statistici alternativi per la fase II. In particolare, il dottor Pozzi ha fatto uno stage di studio di qualche mese alla Novartis di Basilea che ha visto sfociare in un articolo in fase di redazione (Pozzi et al.) alcune proposte alternative di simulazione della selezione della dose ottimale nel passaggio dalla fase II alla fase III della ricerca clinica. Gasparini ha partecipato alla ricerca come coautore e come supervisore di Pozzi. L'addestramento e la formazione del dottor Pozzi sono da considerarsi un successo del gruppo, visto che sta intraprendendo con successo il suo secondo anno di dottorato di Biostatistica dell'Università di Berkeley (California) su temi molto vicini a quelli del PRIN.

UNITA' DI BOLOGNA: INDIVIDUAZIONE DELLA STRUTTURA DI INTERAZIONE DELLE VARIABILI

La ricerca si è concentrata principalmente sull'analisi di dati da microarray con attenzione all'identificazione delle strutture di interazione esistente tra le variabili, e quindi tra geni.

Un primo argomento di ricerca riguarda l'identificazione della rete di regolazione genica sottostante i dati. In questo ambito, da un punto di vista statistico, la principale difficoltà è causata dalla sproporzione esistente tra dimensione campionaria e numero di variabili coinvolte nell'analisi. In questo tipo di applicazioni infatti il numero di punti sperimentali è solitamente molto inferiore al numero di geni oggetto di studio. Una possibile soluzione al problema è stata identificata nell'utilizzo dei coefficienti di correlazione parziale di ordine ridotto che, in presenza di strutture di rete sparse, si sono dimostrati strumenti efficaci per la ricostruzione della rete. La ricerca si è indirizzata quindi verso l'utilizzo di modelli grafici gaussiani ed ha ottenuto risultati sia in ambito metodologico che applicato. Si è inoltre studiata una possibile estensione del metodo proposto che possa essere applicata in un contesto di meta-analisi, ossia nel caso in cui si dispone di più insiemi di dati, ottenuti mediante esperimenti che coinvolgono lo stesso tipo di cellula ma sotto condizioni sperimentali differenti tra loro. Le tecniche sviluppate sono state implementate nel pacchetto statistico "qgraph", pubblicato all'interno della versione 2.4 di Bioconductor, che include efficienti routine scritte nel linguaggio C che consentono l'applicazione anche a problemi di dimensione medio/grande.

Un secondo argomento di ricerca riguarda l'identificazione di gruppi di geni che corrispondono a moduli funzionali di specifico interesse all'interno del genoma. In particolare si è studiata l'applicazione di tecniche di clustering per l'identificazione di moduli trascrizionali, e si è evidenziato come le procedure tradizionalmente utilizzate di clustering gerarchico basato sulla regola del legame medio forniscano, in questo ambito, performance insoddisfacenti. Si è proposto quindi di utilizzare una misura di dissimilarità basata sul Lambda di Wilks che, combinata con la metodologia delle correlazioni canoniche sparse, può essere applicata anche nel caso in cui il numero delle variabili eccede la numerosità campionaria. L'utilità della procedura proposta è avallata sia da risultati di tipo metodologico sia da applicazioni a dati simulati e reali.

UNITA' DI BOLOGNA: ESPERIMENTI ADATTATIVI PER LE PROVE CLINICHE

L'attività di ricerca svolta ha riguardato principalmente il tema degli esperimenti reali e simulati per le prove cliniche, incentrandosi in particolare sulla stesura di lavori scientifici, di natura sia metodologica che applicativa, e sull'organizzazione e partecipazione a convegni e seminari. Negli studi clinici gli esperimenti sequenziali adattativi vengono incontro a varie esigenze e, tra queste, una di fondamentale rilevanza consiste nel cercare di conciliare l'etica verso il singolo paziente ed il guadagno atteso dall'informazione acquisita con il metodo sperimentale. Per quanto riguarda gli esperimenti clinici comparativi a risposta binaria, è stata proposta ed analizzata una metodologia in grado di combinare requisiti di efficienza inferenziale e l'esigenza di massimizzare la percentuale di pazienti che ricevono il trattamento migliore attraverso un opportuno criterio compromesso. Tale criterio si basa sull'utilizzo di un sistema pesi, essi stessi funzione degli effetti dei trattamenti oggetto di studio, il che consente di derivare allocazioni target dei trattamenti in grado di riflettere l'importanza relativa dei due obiettivi in questione. Inoltre, è stato proposto un disegno sequenziale in grado di far convergere le assegnazioni dei trattamenti verso il target ottimo così derivato (1). Tale metodologia è stata estesa anche nel caso in cui l'outcome sperimentale sia una variabile continua ed inoltre si rendano disponibili informazioni riguardanti fattori prognostici di particolare rilievo.

Un'altra tematica di ricerca che è stata approfondita riguarda la sperimentazione virtuale nell'ambito delle scienze biomediche e

sanitarie. In particolare, il lavoro (Giovagnoli, M. Zagoraïou (2010) contiene una ricca rassegna dei problemi statistici derivanti dai trials simulati in ambito clinico dove abbiamo: * evidenziato la potenzialità di questa metodologia attraverso applicazioni reali; * presentato le finalità e peculiarità dei modelli di simulazione utilizzati in questo contesto; * illustrato le tecniche di validazione dei suddetti modelli; * analizzato numerose problematiche che sorgono nella progettazione degli esperimenti simulati, con particolare attenzione alle esigenze di pianificazione ottimale degli stessi, proponendo alcune possibili soluzioni.

UNITA' DI MILANO

Nell'ambito del progetto PRIN 2007 "Metodi Statistici per l'Apprendimento in Biomedicina": l'Unità di ricerca Milanese guidata dal Prof. Elia Mario Biganzoli (MED01/Statistica Medica) ha sviluppato le attività relative alla "valutazione dell'impatto diagnostico, prognostico e di risposta alle terapie in ambito oncologico dei bioprofili ottenuti da metodologie di indagine molecolare ad elevata resa in presenza di eventi multipli durante il follow-up, tramite l'applicazione di tecniche statistiche multivariate innovative"

Il progetto PRIN ha rappresentato la naturale continuazione del Progetto NoE BIOPATTERN "Computational Intelligence for Biopattern Analysis in Support of e-Healthcare" FP6/2002/IST/1 N° IST-2002-508803 dove lo stesso gruppo di ricerca è stato leader della Task Force 5: Evaluation and Benchmarking, con funzioni di supporto biostatistico alla ricerca Informatica Biomedica per l'integrazioni di dati clinici e biologici e l'analisi dei bioprofili nelle patologie maggiori: oncologiche, neurologiche e cardiovascolari. Il parere finale della Commissione Europea riguardo ai risultati del progetto BIOPATTERN, ha riconosciuto il valore fondamentale dell'azione della biostatistica congiunta all'informatica biomedica nei progetti di ricerca di trasferimento e di supporto alla decisione in biomedicina. Nello stesso contesto, la Commissione ha auspicato la continuazione di BIOPATTERN tramite alleanze sancite da accordi quadro interistituzionali a livello nazionale ed internazionale.

Il Progetto PRIN 2007 ha avuto quindi un ruolo fondamentale per la successiva ricaduta nazionale dei risultati delle azioni di integrazione multidisciplinare nella rete di ricerca internazionale sviluppata nel corso del progetto BIOPATTERN. Queste azioni sono state determinanti per la crescita della capacità di ricerca di tipo metodologico del gruppo Milanese, al fine di affrontare in modo sempre più efficace la valutazione di problemi biostatistici complessi alla luce delle metodiche biotecnologiche che generano insiemi di dati u2018omici' di elevate dimensioni e la successiva valutazione dell'impatto sulla decisione medica individualizzata. Gli sviluppi metodologici connessi al progetto e l'evoluzione contingente dell'Unità di ricerca Milanese nel contesto della ricerca accademica e sanitaria, hanno permesso la differenziazione e crescita nella produzione scientifica biomedica orientata all'analisi quantitativa per il supporto alla decisione medica nella ricerca su diverse patologie maggiori estendosi dall'oncologia verso cardiologia, reumatologia ed ematologia.

Nel periodo del progetto PRIN 2007, l'Unità di Milano ha quindi realizzato l'integrazione con il progetto di rete nazionale per la Bioinformatica Oncologica RNBIO nel contesto dell'Istituto Superiore di Sanità ed Alleanza Contro il Cancro, con il contributo e la diretta organizzazione dei corsi di Bioinformatica per la Proteomica e Biostatistica per la Bioinformatica. L'attività internazionale ha poi riguardato lo sviluppo del gruppo di interesse speciale nell'ambito IEEE-INNS International Neural Network Society e la realizzazione di un corso triennale Erasmus Life Long Learning Programme sull'analisi dei dati genomici con la collaborazione con le Università di Salerno, Napoli, Helsinki, Warwick e Milano.

Missione primaria del gruppo biostatistico Milanese è infatti il supporto quantitativo alla crescita della rilevanza sociale ed economica della ricerca biomedica di trasferimento sulle patologie maggiori, tramite la valutazione biostatistica applicata su fronti di ricerca locali e globali con le reti di collaborazione esistenti ed in fase di sviluppo, creando nuove opportunità di lavoro per i giovani Biostatistici e Bioinformatici.

In questo ambito, grazie anche alle attività svolte durante il progetto, il gruppo ha ricevuto riconoscimento a livello mondiale nella Biostatistica per l'analisi di dati di follow-up in presenza di rischi che competono, per lo studio delle dinamiche di malattia, nella valutazione dei profili molecolari e del rischio tramite tecniche multivariate per la medicina individualizzata. Il gruppo di fatto rappresenta l'eccellenza mondiale per lo sviluppo di metodi di intelligenza artificiale nell'analisi dei dati di follow-up.

Per quanto riguarda la descrizione della ricerca ed dei risultati:

1. Analisi in-silico di dati microarray su piattaforma Affimetrix, orientata allo studio dell'espressione di geni relativi alla polarità cellulare in linee cellulari di tumori epiteliali, selezionando e

combinando l'informazione relativa all'espressione di singoli geni. Sono state utilizzate conoscenze pregresse della biologia molecolare, permettendo la strutturazione di conoscenza a priori ed ipotesi mirate sulla regolazione dell'espressione.

2. Per quanto riguarda l'attività di ricerca sui metodi statistici multivariati per l'analisi dei bioprofili, l'interesse si è orientato alla valutazione delle diverse metodologie di clustering confrontate con modelli grafici e metodi di classificazione, per identificare il corrispondente profilo biologico e la sua relazione con caratteristiche riconosciute di aggressività del tumore. Particolare attenzione è stata dedicata alla caratterizzazione dell'algoritmo di Affinity Propagation e delle sue relazioni con i problemi di ottimizzazione in fisica meccanica statistica.

3. La collaborazione con le Università di Bologna e Firenze, relative all'uso dei modelli grafici per lo studio delle reti metaboliche e di co-espressione hanno riguardato la possibilità di valutare le strutture di indipendenza condizionale tra le variabili misurate e di poterle interpretare ai fini della caratterizzazione dei bioprofili.

4. Progettazione ed il supporto biostatistico alla realizzazione di un database per la raccolta (retrospettiva e prospettica) e l'integrazione di informazioni cliniche e biologiche derivanti da pazienti affette da carcinoma ovarico presso la Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, con specifico interesse per l'archiviazione dei dati da tecnologie genomiche. Definizione di un progetto per l'analisi della struttura di variabilità in esperimenti di gene expression profiling e per il confronto tra diverse piattaforme sperimentali.

5. Consulenza metodologica biostatistica per il confronto della classificazione clinico u2013 patologica, molecolare e genotipica del carcinoma mammario in progetti collaborativi tra Italia e Cina, ed Italia - Sudan. Supporto per la "sistematizzazione" del database e la selezione delle pazienti.

6. Supporto alla pianificazione degli esperimenti proteomici per la misura di valori di diverse citochine in campioni clinici di tumore al polmone con sistemi multiplex (Bioplex - Biorad). Analisi dei dati e supporto all'interpretazione dei risultati tramite tecniche statistiche multivariate con applicazione innovativa al filtraggio del bias di misura indotto dal sistema multiplex. In presenza di elevata variabilità analitica e di bias è stato possibile isolare alcune citochine riprodotte in due esperimenti indipendenti, validate successivamente dal punto di vista biologico tramite sistemi tradizionali ELISA.

7. Supporto alla pianificazione dello studio di valutazioni di nuovi marcatori patologici per la prognosi del melanoma mucoso. L'obiettivo dello studio è di valutare la rilevanza e l'impatto di fattori clinici e patologici sulla prognosi, e sul decorso successivo all'intervento, in soggetti affetti da melanoma mucoso maligno di testa e collo. Il melanoma suddetto è noto per essere una patologia rara e a cattiva prognosi; a tale riguardo, la casistica dell'Istituto Nazionale Tumori, formata da 93 pazienti che sono stati ricevuti in un arco di 30 anni, costituisce a tutt'oggi la casistica più numerosa riportata da un singolo istituto. L'end-point valutato in questo studio è la sopravvivenza libera da malattia; il tempo di sopravvivenza è calcolato dalla data di primo trattamento alla data di prima ripresa, definita come la prima evidenza di recidiva locale (comprese quelle sui linfonodi regionali) o comparsa di metastasi distanti. Analisi dei dati tramite metodologie avanzate di selezione dei modelli di regressione e aggregazione mediante model averaging e boosting.

8. Supporto alla pianificazione di esperimenti volti ad integrare il contributo diagnostico/prognostico del linfonodo sentinella nel melanoma cutaneo utilizzando informazioni da esperimenti di carattere genomico, con misure di espressione congiunta su piattaforma c-DNA microarray Illumina.

9. Sono state sviluppate tecniche di regressione flessibile per l'analisi di relazioni di dipendenza complesse, con l'obiettivo di definire schemi prognostici elevandone la capacità predittiva e valutandone l'accuratezza in termini di calibrazione e discriminazione. In particolare le applicazioni hanno riguardato lo sviluppo e validazione di uno strumento (nomogramma) e di modelli predittivi per la prognosi dei pazienti affetti da Sarcomi Retroperitoneali dei tessuti molli. I modelli potranno fornire una misura del rischio di mortalità. Nomogrammi per la previsione della sopravvivenza specifica per causa sono stati proposti per tutti i sarcomi. Il lavoro è stato seguito da una validazione internazionale proposta in collaborazione con gruppi di ricerca negli USA che hanno sviluppato strumenti simili.

10. Sviluppo di metodi e modelli flessibili per l'analisi di dati longitudinali in presenza di censura, in particolare per lo studio dell'"event-history" nella patologia oncologica in generale, per la valutazione prognostica e predittiva di eventi (recidive locali, comparsa di metastasi a distanza). Obiettivo principale è il supporto

biostatistico allo studio di ipotesi di tipo biologico sulla dinamica di malattia, e l'ipotesi di Tumour Dormancy. Studio della biologia e dinamica di recidiva dei tumori mammari con recettori per estrogeni, progesteronici e c-erbB2 negativi (triplo negativi).

UNITA' DI NOVARA

Abbiamo messo a punto un metodo generale basato sul bootstrap per la stima conservativa della potenza e della dimensione campionaria dei test statistici. Dalle simulazioni effettuate prendendo ad esempio il test di Wilcoxon per due campioni è emerso che tale soluzione bootstrap fornisce risultati migliori di altre tecniche note in letteratura e basate sulla distribuzione asintotica della statistica test. L'articolo è stato sottoposto a Journal of Biopharmaceutical Statistics, ed è stato accettato per la pubblicazione. Si è in attesa di correggere le bozze.

Per quanto riguarda la stima della dimensione campionaria di fase III sulla base dei dati di fase II, abbiamo messo a punto una strategia basata sull'ottimizzazione della potenza globale delle due fasi, basata a sua volta su un criterio di scelta mobile della quantità di conservatività da applicare. Tale strategia si è rivelata molto efficace. Infatti, fornisce potenza globale maggiore delle strategie a conservatività fissa e stime della dimensione campionaria più precise. Una relazione è stata esposta alla scorsa riunione intermedia di Bologna, il 19 Giugno 2009. L'articolo è stato pubblicato sulla prestigiosa rivista Pharmaceutical Statistics. Si è in attesa di conoscere la esatta collocazione sul numero della rivista.

Abbiamo inoltre lavorato intensamente sulla stima della probabilità di riproducibilità della significatività statistica del test di Wilcoxon a due campioni, utile alle agenzie regolatrici per stabilire se effettuare o meno il secondo dei due trial pivotali. Sotto il profilo metodologico, abbiamo trovato che esiste la possibilità teorica di effettuare RP-testing anche per questo test di Wilcoxon, in analogia al risultato in De Martini (2008), Stat.Prob.Lett. Tale risultato si basa su una modellizzazione delle distribuzioni sotto l'ipotesi alternativa da cui deriva l'ordinamento stocastico in senso stretto della statistica test e la conseguente monotonicità della funzione di potenza. Abbiamo quindi sviluppato tali risultati in senso pratico confrontando alcuni stimatori della RP per il test di Wilcoxon, con il fine di stimare la potenza e di effettuare RP-testing. Per quest'ultimo fine si sono tutti rivelati molto attendibili, con il "disagreement" rispetto al test classico che tende rapidamente a zero. Per quanto concerne lo scopo di stima, l'approssimazione della potenza basata sulla normalità asintotica con limiti inferiori della varianza ha fornito i risultati globalmente migliori. Anche i metodi nonparametrici basati sul semplice plug-in non hanno sfigurato, specialmente per potenze effettive su valori intorno all'80-90%. Ne sono scaturiti quindi due diversi articoli: il primo, teorico, è attualmente sottoposto a Stat.Prob.Lett, mentre il secondo, più pratico, è stato sottoposto a Communication in Statistics - Simulation and Computation.

Infine, abbiamo studiato soluzioni al problema della stima della potenza per test sugli end-points primari multipli. Pur trattandosi di un problema parametrico, si incontrano notevoli difficoltà tecniche a causa della multidimensionalità del problema e della distorsione del test Intersection Union (IUT). Pertanto, abbiamo sviluppato alcune tecniche parametriche approssimate e le abbiamo confrontate con la soluzione bootstrap generale da noi messa a punto recentemente (vedi sopra). In un primo lavoro di simulazione è emerso che sia le tecniche parametriche che quella bootstrap sono distorte. In ogni caso la tecnica bootstrap ha mostrato i risultati migliori e la convergenza più rapida. Si è quindi pensato di migliorare la tecnica bootstrap di base con la calibrazione, ed in una seconda fase di simulazioni sono emersi risultati più che buoni. Da tale lavoro ne è scaturito un articolo, formalizzato in un quaderno di dipartimento, che sarà presto sottoposto ad una rivista internazionale.

12. Problemi riscontrati nel corso della ricerca

Non ci sono stati problemi degni di nota, se non piccole variazioni nella composizione dei gruppi di ricerca dovuti a motivi personali.

13. Risorse umane complessivamente ed effettivamente impegnate (da consuntivo)

	(mesi uomo)
TOTALE	106
da personale universitario	72
altro personale	0
Personale a contratto a carico del PRIN 2007	0

14. Modalità di svolgimento (dati complessivi)

Partecipazioni a convegni:

	Già svolti (numero)	Da svolgere (numero)	Descrizione
in Italia	38	0	19 interventi nazionali con una media di due persone per intervento
all'estero	34	1	17 interventi internazionali con una media di due persone per intervento Bonetti parteciperà all'IBC in Dicembre 2010
TOTALE	72	1	

Per ogni campo di testo max 8.000 caratteri spazi inclusi

Articoli pertinenti pubblicati:

	Numero	Descrizione
su riviste italiane con referee	1	<i>Gasparini M. e Gandini, M. A comparison of nonparametric estimators of survival under left-truncation and right-censoring motivated by a case study. Accettato dalla rivista Statistica.</i>
su riviste straniere con referee	48	<p><i>GASPARINI JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS, vol. 59; p. 543-544, ISSN: 0035-9254</i></p> <p><i>DI GAETANO EUROPEAN JOURNAL OF HUMAN GENETICS, vol. 17; p. 91-99</i></p> <p><i>RAINERO HEADACHE, vol. 48(7); p. 1108-1114, ISSN: 0017-8748</i></p> <p><i>GASPARINI THOUSAND OAKS: SAGE publications, ISBN/ISSN: 9781412927147</i></p> <p><i>Bonetti Journal of Clinical Oncology; Volume: 26; pp.: 3813-14</i></p> <p><i>Bonetti Lifetime Data Analysis; Volume: 15; pp.: 493-518</i></p> <p><i>Bonetti Statistics in Medicine; Volume: 28; pp.: 1255-68</i></p> <p><i>Gunnarsd'ottira ... Breast Cancer Research and Treatment, In press.</i></p> <p><i>Lazar Journal of Clinical Oncology (Statistics in Oncology series)</i></p> <p><i>Baldi Biometrika, to appear.</i></p> <p><i>Forcina Statistical Modelling, to appear.</i></p> <p><i>Baldi Antognini Advances in Model Oriented Design and Analysis pp. 17-24, ISBN: 978-3-7908-2409-4, Springer.</i></p> <p><i>Baldi Antognini Journal of Statistical Planning and Inference, 140, 2607-2617.</i></p> <p><i>Baldi Antognini Complex data modeling and computationally intensive statistical methods, Series: Contributions to Statistics (P. Mantovan, P. Secchi Eds), 27-40, ISBN: 978-88-470-1385-8, Springer.</i></p> <p><i>F. Pizza Journal of Clinical Sleep Medicine, 6, 41-45.</i></p> <p><i>A Giovagnoli Optimal Design and Related Areas in Optimization and Statistics (L. Pronzato, A. Zhigljavsky eds) 81-96, Springer, ISBN: 978-0-387-79935-3</i></p> <p><i>Roverato Biometrics, to appear.</i></p> <p><i>Studený Kybernetika. 45, 2, 208-248.</i></p> <p><i>Castelo Journal of Computational Biology. 16, 2, 213-227.</i></p> <p><i>Castelo Methods in Molecular Biology Series (Wang, J., Aik-choon, T. and Tian, T. Eds), Humana Press, in corso di pubblicazione.</i></p> <p><i>Demicheli International Journal of Radiation Biology, 86 (7), pp. 542-547.</i></p> <p><i>Biganzoli Current Breast Cancer Reports, 2 (2), pp. 96-106.</i></p> <p><i>Ardoino Cancer, 116 (10), pp. 2429-2436.</i></p> <p><i>Soria Computers in Biology and Medicine, 40 (3), pp. 318-330.</i></p> <p><i>Demicheli Cancer Science, 101 (3), pp. 826-830.</i></p> <p><i>Lisboa IEEE Computational Intelligence Magazine, 5 (1), art. no. 5386112, pp. 14-18.</i></p> <p><i>Lisboa Frontiers in Artificial Intelligence and Applications, 196 (1), pp. 81-93.</i></p>

		<p>Stefanini <i>BMC Bioinformatics</i>, 10 (SUPPL. 12), art. no. 1471 .</p> <p>Ferraro <i>Clinical Chemistry and Laboratory Medicine</i>, 47 (10), pp. 1297-1303.</p> <p>Fernandes <i>Lecture Notes in Computer Science 5517 LNCS (PART 1)</i>, pp. 989-996.</p> <p>Lisboa <i>IEEE Transactions on Neural Networks</i>, 20 (9), pp. 1403-1416.</p> <p>Bagnoli <i>Endocrine-Related Cancer</i>, 16 (2), pp. 443-453.</p> <p>Ambrogi <i>Computational Statistics and Data Analysis</i>, 53 (7), pp. 2767-2779.</p> <p>Lama <i>Computational Statistics and Data Analysis</i>, 53 (5), pp. 1906-1922.</p> <p>Ambrogi <i>Statistics in medicine</i>, 27 (30), pp. 6407-6425.</p> <p>Bacciu <i>Lecture Notes in Computer Science 5178 LNAI (PART 2)</i>, pp. 181-188.</p> <p>Demicheli <i>Breast cancer research : BCR</i>, 10 (5).</p> <p>Boracchi <i>International Journal of Biological Markers</i>, 23 (4), pp. 199-206.</p> <p>Bacciu <i>KES (2) 2008: 181-188</i></p> <p>Gree <i>European Journal of Cancer. Suppl 6 (7) 2008. p18</i></p> <p>Green <i>Breast Cancer Research 10 2008.</i></p> <p>Green <i>Modern Pathology 21 2008. 34a.</i></p> <p>D. Soria <i>International Journal of Knowledge Engineering and Soft Data Paradigms (IJKESDP)</i>, 1(3):195-215, 2009</p> <p>Fernandes <i>Int. J. Knowledge Engineering and Soft Data Paradigm. In press, 2009</i></p> <p>Ardoino <i>Int. J. Biomedical Engineering and Technology, In press, 2009.</i></p> <p>Lisboa <i>Computational Intelligence and Bioengineering. IOS press, Amsterdam, 2009.</i></p> <p>DE MARTINI D. <i>JOURNAL OF BIOPHARMACEUTICAL STATISTICS</i>, vol. 21 n.1, ISSN: 1054-3406</p> <p>DE MARTINI D. (2010). <i>Adapting by calibration the sample size of a phase III trial on the basis of phase II data. PHARMACEUTICAL STATISTICS</i>, ISSN: 1539-1604</p>
su altre riviste italiane	0	
su altre riviste straniere	0	
comunicazioni a convegni/congressi internazionali	17	<p>Gasparini: March 18, 2010, University of Texas at San Antonio, San Antonio, Texas</p> <p>Gasparini: March 15, 2010, Alcon Research, LTD, Fort Worth, Texas</p> <p>Taccetti G., Bianchini E., Zavataro L., Campana S., Defilippi G., Ravenni N., Trevisan F., Boni V., Braccini G., Braggion C., Cariani L.</p> <p><i>Pseudomonas aeruginosa eradication in cystic fibrosis: preliminary data from a randomized multicenter study of two different early antibiotic treatment protocols.</i></p> <p>Abstract accettato alla Conferenza Sulla Fibrosi Cistica Nord America 2010. 21/23 Ottobre 2010 Baltimore.</p> <p>F.M.L. Di Lascio, A. Roverato (2010) "Comparison of correlation-based dissimilarity measures for clustering genes with DAG structure". <i>Proceedings del CIBB 2010, USB stick con ISBN: 978-88-95272-87-0, p.1-9.</i></p> <p>Roverato, A. e Castelo, R. (2010). "Learning undirected graphical models from multiple datasets with the generalized non-rejection rate" <i>Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (P. Myllymäki, T. Roos and T. Jaakkola Eds), 249-256, HIIT Publications.</i></p> <p>A. Giovagnoli, "Ethics and Inference in Clinical Trials: Admissible Allocations and Response Adaptive Randomization", <i>IWSM 2009, Second International Workshop in Sequential Methodologies, University of Technology of Troyes, June 15-17, 2009 relazione invitata</i></p> <p>A. Giovagnoli, "Actual and Simulated Trials in Clinical Research", <i>ENBIS9 Conference, Goteborg, September 21-23, 2009 relazione invitata</i></p> <p>A. Giovagnoli, <i>Randomized Group Up-And-Down (U&D) Experiments, Queen Mary - University of London, 5 November 2009 seminario invitato</i></p> <p>A. Giovagnoli, "Design for Clinical trials" <i>Graduate School in STATISTICS and ACTUARIAL SCIENCES of the Belgian French Speaking Community, Institut de Statistique - Université Catholique de Louvain (UCL) 6-8 maggio 2009 seminario invitato</i></p> <p>M. Zagoraiou, "The Covariate Adjustable Biased Coin Design for balancing clinical trials in the presence of covariates", <i>8th International Conference of Numerical Analysis and Applied Mathematics, 19-25 Settembre 2010, Rodi, Grecia relazione invitata</i></p> <p>M. Zagoraiou, "Simulated clinical trials: some design issues" e "Covariate adjusted designs for combining efficiency, ethics and randomness in normal response trials", <i>mODa9-Model Oriented Data Analysis, 14-18 Giugno 2010,</i></p>

	<p>Bertinoro, Italia.</p> <p>D. Soria, J.M. Garibaldi, F. Ambrogi, P.J.G. Lisboa, P. Boracchi, E. Biganzoli, Clustering breast cancer data by consensus of different validity indices, in <i>Proceedings of the 4th International Conference on Advances in Medical, Signal and Information Processing (MEDSIP)</i>, 2008. Italy IET Conference Publications, (540 CP)</p> <p>F. Ambrogi, E. Raimondi, D. Soria, P. Boracchi, E. Biganzoli, Cancer profiles by Affinity Propagation, <i>Proceedings of the Seventh International Conference on Machine Learning and Applications (ICMLA08)</i>, IEEE Computer Society, 650-655, 2008</p> <p>D. Soria, J.M. Garibaldi, E. Biganzoli, I.O. Ellis, A Comparison of Three Different Methods for Classification of Breast Cancer Data, <i>Proceedings of the Seventh International Conference on Machine Learning and Applications (ICMLA08)</i>, IEEE Computer Society, 619-624, 2008.</p> <p>Fernandes, A.S., Jarman, I.H., Etehell, T.A., Fonseca, J.M., Biganzoli, E., Bajdik, C., Lisboa, P.J.G. Missing data imputation in longitudinal cohort studies - Application of PLANN-ARD in breast cancer survival (2008) <i>Proceedings - 7th International Conference on Machine Learning and Applications, ICMLA 2008</i>, pp. 644-649.</p> <p>D. Soria, J.M. Garibaldi, F. Ambrogi, P.J.G. Lisboa, P. Boracchi, E. Biganzoli, Application of Affinity Propagation on breast cancer data sets, <i>23rd European Conference on Operational Research (EURO XXIII)</i>, Bonn, Germany, 06-07-2009</p> <p>DE MARTINI D., DE CAPITANI L (2010). Reproducibility Probability Testing for the Wilcoxon rank-sum test. In: <i>Atti del convegno. Padova, 14-15 Giugno 2010</i></p>
<p>comunicazioni a convegni/congressi nazionali</p>	<p>19</p> <p>Gasparini: October 1, 2010 University of Bologna, Bologna, Italy</p> <p>Gasparini: September 23, 2010, Greco-Italian meeting on Statistics, Sardegna, Italy</p> <p>F.M.L. Di Lascio, A. Roverato (2010) "Partial ordering of dissimilarity measures for gene clustering". <i>Proceedings del 45mo Meeting Scientifico della Societa' Italiana di Statistica</i>, USB stick con ISBN 978-88-6129-566-7, p.1-8.</p> <p>Baldi Antognini, A. Giovagnoli (2009) Adaptive randomization for ethics and inference in clinical trials with covariate information, <i>Atti del VII Congresso della Societa' Italiana di Biometria</i>, 10-12 giugno 2009 Ponte di legno (Bs), 1-4.</p> <p>A. Roverato, F.M.L. Di Lascio (2009) "Clustering variables with DAG structure", <i>Proceedings del Workshop S.Co. 2009 - Maggioli Editore</i>, p.369-374, ISBN: 978-88387-4385-1.</p> <p>Roverato, A., Castelo, R. (2009). "An application of the non-rejection rate in meta-analysis". <i>Statistical Methods for the analysis of large data-sets -- Book of short papers</i>, Cleup, pp. 187-190.</p> <p>A. Roverato, "Wilks' Lambda Dissimilarity Measures For Gene Clustering: An Approach Based On The Identification Of Transcription Modules", <i>Greco Italian Meeting</i>, 23-25 settembre 2010, Porto San Paolo, Sardegna.</p> <p>F.M.L. Di Lascio, "Comparison of correlation-based dissimilarity measures for clustering genes with DAG structure", <i>CIBB 2010, Sessione Specializzata su "Data Clustering and Bioinformatics"</i>, 16-18 settembre 2010, Palermo, Italia.</p> <p>F.M.L. Di Lascio, "Clustering variables with DAG structure", <i>S.Co. 2009</i>, 14-16 settembre 2009, Milano, relazione invitata.</p> <p>F.M.L. Di Lascio, "Partial ordering of dissimilarity measures for gene clustering", <i>45mo Meeting Scientifico della SIS</i>, 14-16 giugno 2010, Padova.</p> <p>A. Baldi Antognini, "Adaptive randomization for ethics and inference in clinical trials with covariate information", <i>VII Congresso della Societa' Italiana di Biometria</i>, 10-12 giugno 2009, Ponte di legno (Bs).</p> <p>M. Zagoraiou, "On the Optimal Designs for the Universal Kriging", <i>S.Co. 2009</i>, 14-16 September 2009, Milano.</p> <p>Soria, J.M. Garibaldi, F. Ambrogi, P.J.G. Lisboa, P. Boracchi, E. Biganzoli, Classification techniques for breast cancer data, <i>Mini EURO Conference on Computational Biology, Bioinformatics and Medicine (Mini EURO CCBBM 2008)</i>, Rome, Italy, 15 - 17 September 2008</p> <p>Vellido A, Biganzoli E, Lisboa P.J.G. Machine learning in cancer research: implications for personalised medicine. <i>ESANN 2008</i>: 55-64</p> <p>D. Soria, F. Ambrogi, P. Boracchi, J.M. Garibaldi, E. Biganzoli, Application of Affinity Propagation on a large breast cancer data set, <i>Proceedings of the Conference on Statistical methods for the analysis of large data sets (SIS 2009)</i>, 531-534, 2009.</p> <p>N. Bassani, F. Ambrogi, R. Bosotti, M. Bertolotti, A. Isacchi, E. Biganzoli. Non-Parametric MANOVA Methods for Detecting Differentially Expressed Genes in Real-Time RT-PCR experiments, <i>CIBB 2009 Sixth International Meeting on Computational Intelligence Methods for Bioinformatics And Biostatistics 15-17 October 2009 - Genova (Italy)</i></p> <p>R.Tagliaferri, E. Biganzoli, P. Lisboa, F. Masulli, L. Peterson, R. Panel and discussion: <i>INNS and IEEE-CIS Perspectives in Bioinformatics and Biostatistics. CIBB 2009 Sixth International Meeting on Computational Intelligence Methods for Bioinformatics And Biostatistics 15-17 October 2009 - Genova (Italy)</i></p> <p>F. Ambrogi, E. Biganzoli. Some issues in clustering breast cancer biomarker data. Merging statistical, biological and clinical perspectives. <i>2° Workshop Bioinformatica, Biostatistica e Machine Learning</i>, 6 ottobre 2009, Università di Salerno, Fisciano.</p> <p>Fernandes AS, Jarman IH, Etehell TA, Fonseca JM, Biganzoli E, Bajdik C, Lisboa PJG Stratification Methodologies for Neural Networks Models of Survival. <i>IWANN (1) 2009</i>: 989-996.</p>

rapporti interni	7	<p><i>Di Gaetano C., Matullo G., Ursino M., Piazza A., Gasparini M.</i> <i>A proximity-based method to identify genomic regions correlated with latitude.</i></p> <p><i>Gasparini M., Neuenschwander B. e Bailey S.</i> <i>Modelling drug-drug interactions in dose-finding combination trials.</i></p> <p><i>Pozzi L., Schmidli H., Gasparini M., Racine A.</i> <i>A Bayesian adaptive dose selection trial with a count endpoint.</i></p> <p><i>Giovagnoli, M. Zagoraïou (2010) "Simulated Clinical Trias: some design issues" Quaderni del Dipartimento di Scienze Statistiche, Università di Bologna, serie Ricerche, 2010, N.1.</i></p> <p><i>DE CAPITANI L, DE MARTINI D. (2010). On stochastic orderings of the Wilcoxon Rank Sum test statistic - with applications to reproducibility probability estimation testing, quaderno di dipartimento SEMEQ</i></p> <p><i>A. LUCADAMO, N. ACCOTO, DE MARTINI D. (2010). Power Estimation for Multiple Co-Primary Endpoints: a comparison among conservative solutions, quaderno di dipartimento DIMEQUANT</i></p> <p><i>DE CAPITANI L, DE MARTINI D. (2010). Reproducibility Probability Estimation and Testing for the Wilcoxon rank-sum test, quaderno di dipartimento DIMEQUANT</i></p>
brevetti depositati	0	
TOTALE	92	

Per ogni campo di testo max 8.000 caratteri spazi inclusi

Data 16/12/2010 18:07

Firma

Si autorizza alla elaborazione e diffusione delle informazioni riguardanti i programmi di ricerca presentati ai sensi del D. Lgs. n. 196/2003 del 30.6.2003 sulla "Tutela dei dati personali". La copia debitamente firmata deve essere depositata presso l'Ufficio competente dell'Ateneo.