

DIPARTIMENTO DI
ECONOMIA E FINANZA

**METODI E ANALISI
STATISTICHE
2021**



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

**DIPARTIMENTO DI
ECONOMIA E FINANZA**

**METODI E ANALISI
STATISTICHE
2021**



**UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO**

Tutti i diritti di traduzione, riproduzione e adattamento, totale o parziale, con qualsiasi mezzo (comprese le copie fotostatiche e i microfilm) sono riservati

Toma E., (a cura di) edizione 2021. *Metodi e Analisi Statistiche*, Dipartimento di Economia e Finanza, Università degli studi di Bari *Aldo Moro*.

© Copyright 2021 by Università degli Studi di Bari Aldo Moro
www.uniba.it

Prima edizione: dicembre 2021

ISBN 978-88-6629-066-7

Gli articoli qui presentati sono stati tutti oggetto di valutazione interna.

Editing finale: F.D. d'Ovidio, E. Toma

Sommario

Ernesto Toma Presentazione	pag. 5
Massimo Bilancia, Fabio Manca, Gianvito Pio A Brief Review on Compositional Inference of Microbiome Data	« 7
Claudia Marin, Fabio Manca Application of text mining in the biomedical field	« 21
Ilaria Pepe, Ernesto Toma Alcune considerazioni sulla presenza in Italia dei Minori Stranieri Non Accompagnati (MSNA) in relazione alla pandemia Covid-19	« 35
Aurélie P. Scarafino, Antonella Massari, Claudia Marin Big Data, Intelligenza Artificiale e turismo sostenibile	« 59
Nunziata Ribecco, Angela Maria D’Uggento, Arianna Ruggiero L’esperienza della didattica a distanza in alcune scuole pugliesi	« 71
Carlo Cusatelli, Crescenzo Gallo, Alessandro Rinaldi Distributive adaption of the Lorenz curve to a Squirecle	« 87
Angela Maria D’Uggento, Marina Tarantini L’approccio modellistico per una programmazione più efficace dei processi amministrativi: il caso della mortalità da Covid-19	« 101
Andrea Tanga, Angela Maria D’Uggento, Danilo Di Bona Efficacia del Benraluzimab nel trattamento dell’asma severo eosinofilo. Identificazione di subfenotipi di risposta	« 125
Agata Maria Madia Carucci, Giovanni Vannella La valutazione delle azioni di sostenibilità ambientale delle imprese attraverso i censimenti permanenti	« 135

Sabrina Diomede, Giovanni Tagliatela	
<i>A generalization of the logistic function</i>	« 155
Anna Lucia Muserra, Francesco Grimaldi	
<i>Crisi finanziaria e politiche di bilancio. Alcune evidenze empiriche</i>	« 167

Efficacia del Benraluzimab nel trattamento dell'asma severo eosinofilico. Identificazione di subfenotipi di risposta

Andrea Tanga¹, Angela Maria D'Uggento^{2*}, Danilo Di Bona³

¹Dottore in Scienze statistiche, ²Dipartimento di Economia e Finanza, Università degli Studi di Bari Aldo Moro ³Dipartimento dell'Emergenza e dei Trapianti di organi (DETO), Università degli Studi di Bari Aldo Moro

Riassunto: Il presente contributo presenta i principali risultati dell'analisi statistica condotta sugli effetti di un farmaco biologico, il Benralizumab, utilizzato nella terapia dell'asma severo. Si tratta di un anticorpo monoclonale umanizzato di tipo IgG1 κ in grado di legarsi in modo selettivo alla catena α del recettore dell'interleuchina 5 (IL-5R) bloccando il pathway di attivazione di IL-5 / IL5R nonché delle cellule effettrici. Con riferimento alla popolazione di pazienti sottoposti a tale terapia, lo studio retrospettivo si pone l'obiettivo di individuare le caratteristiche dei pazienti che sono hanno mostrato una migliore risposta alla terapia con Benralizumab al fine di creare un modello predittivo, presupposto per l'attivazione di percorsi di "medicina personalizzata". Le analisi svolte hanno consentito di individuare 4 sub-fenotipi di pazienti associati a un differente grado di risposta alla terapia. Le caratteristiche dei pazienti che meglio rispondono alla terapia.

Keywords: Benralizumab; studio retrospettivo; regressione logistica, cluster analysis.

* Autore corrispondente: angelamaria.duggento@uniba.it.

Il lavoro qui descritto è frutto di un progetto comune, tuttavia Tanga ha scritto l'introduzione, il paragrafo 2 e curato le elaborazioni e il commento dei risultati nel paragrafo 3; D'Uggento ha curato il paragrafo 3.1 e contribuito alla individuazione delle metodologie e Di Bona ha contribuito alla redazione del paragrafo 3 e redatto il paragrafo 4.

1. Introduzione

L'ASMA è una malattia caratterizzata da difficoltà respiratoria dovuta a riduzione del calibro dei bronchi per infiammazione della parete bronchiale, contrazione della muscolatura liscia bronchiale e caratterizzata da un'infiammazione cronica delle vie aeree che determina un'ostruzione che può variare notevolmente da soggetto a soggetto.

L'etiologia, al momento, è ignota e questo fa sì che sia oggetto di studio da parte di molti ricercatori di ambito medico. Si distinguono una forma estrinseca, i cui fattori scatenanti sono gli allergeni inalatori, ed una intrinseca, cioè non allergica, i cui trigger sono aspecifici. Il trattamento di fondo prevede l'utilizzo di farmaci anti-infiammatori (corticosteroidi orale o inalatorio), frequentemente associati ad un farmaco che dilata i bronchi, entrambi somministrati per via inalatoria.

La diagnosi di asma viene fatta considerando la storia del paziente, le caratteristiche dei sintomi (sibili, respiro corto e tosse) e i risultati degli esami finalizzati ad accertare il livello di limitazione del flusso aereo. Quest'ultimo aspetto viene osservato attraverso la risposta che si ottiene somministrando un farmaco broncodilatatore. La risposta, ovvero la capacità di espellere l'aria dai polmoni, può essere valutata con specifiche prove di funzionalità respiratoria, tra cui la più importante è quella denominata FEV1. FEV1 è l'acronimo di Forced Expiratory Volume in the 1st second, procedura che misura il volume di aria che il paziente riesce ad espirare in 1 secondo.

Per quanto riguarda la severità della patologia, si distinguono forme lievi, moderate e severe. Si definisce asma severo quella condizione che necessita della terapia massimale inalatoria per ottenere il controllo (no esacerbazioni, no sintomi moderati/severi). L'Asma grave è una patologia con specificità che la rendono differente dalle più diffuse forme di asma e, in Italia, colpisce tra il 5 e il 10% della totalità degli asmatici per un totale di circa 300.000 persone tra adulti, adolescenti e bambini.

In Italia, come in altre nazioni europee, la prevalenza dell'asma bronchiale nella popolazione è del 4-5 % con un tasso di incidenza di 3-4 casi ogni 1000 abitanti all'anno. Per quanto riguarda l'età, l'asma insorge con notevole frequenza in età infantile rappresentando la più comune malattia cronica e la principale causa di ricoveri ospedalieri nell'infanzia. Si può poi osservare un'attenuazione dei sintomi asmatici, o anche la scomparsa, in un 20-30% dei casi, all'epoca della pubertà soprattutto nei soggetti di sesso maschile. Un secondo picco di alta frequenza

dell'asma bronchiale si riscontra in età adulta (prevalenza 2,5%, incidenza 2 casi ogni mille all'anno).

Il sesso maschile è nettamente più colpito nei casi di asma infantile (nel rapporto di 2 a 1 tra maschi e femmine) mentre nell'adolescenza e nell'età adulta non c'è differenza tra i due sessi. La familiarità sembra giocare un ruolo importante nella precocità di comparsa dell'affezione. Per quanto riguarda i fattori ambientali la prevalenza è maggiore nelle aree urbane e nelle nazioni industrializzate.

L'Asma grave condiziona notevolmente la vita quotidiana delle persone che ne sono affette: scuola, lavoro e vita sociale. Il processo di diagnosi e cura di questa patologia ha ancora molta strada da percorrere per migliorare il livello di cura ed assistenza da garantire ai pazienti.

La severità dell'asma viene valutata in base alle caratteristiche dei sintomi ed ai trattamenti necessari per controllarli, nonché alla presenza di riacutizzazioni. L'asma severo viene trattato con l'associazione di più farmaci a dosaggi massimali.

Con riferimento alla terapia, essa utilizza differenti tipologie di farmaci:

- farmaci di fondo, per il trattamento di mantenimento da seguire in maniera continuativa; riducono l'infiammazione delle vie aeree, controllano i sintomi e riducono i rischi futuri, come le riacutizzazioni e il declino della funzione polmonare;
- farmaci sintomatici (o di emergenza): offrono sollievo immediato dai sintomi e possono, quindi, essere utilizzati al bisogno nelle fasi di riacutizzazione o di peggioramento dell'asma. Sono consigliati anche per la prevenzione a breve termine della possibile chiusura dei bronchi (bronco-costrizione) indotta dall'attività fisica;
- farmaci aggiuntivi: vengono utilizzati soprattutto nell'asma severo, quando i sintomi persistono o si verificano riacutizzazioni nonostante il trattamento ottimizzato con alte dosi dei farmaci di fondo e sono stati attuati, quando possibile, i provvedimenti per intervenire sui fattori di rischio modificabili.

Non sempre la terapia standard è sufficiente per tenere sotto controllo la malattia. Laddove queste terapie non abbiano successo, si può ricorrere ai farmaci biologici, il cui target è uno specifico pathway infiammatorio. La nuova terapia in uso in pazienti adulti affetti da asma grave di tipo eosinofilo non controllato, in aggiunta ai corticosteroidi inalatori e beta-agonisti a lunga durata d'azione, è basata sull'utilizzo del Benralizumab, un anticorpo monoclonale umanizzato di tipo IgG1 κ , in grado di legarsi in modo selettivo alla catena α del recettore dell'interleuchina 5 (IL-5R) bloccando il pathway di attivazione di IL-5 / IL5R nonché delle cellule effettrici.

Il contributo è organizzato nel modo seguente: dopo l'introduzione, nel paragrafo 2 sono presentati i dati e metodi statistici utilizzati, nel paragrafo 3 vengono illustrati i principali risultati ottenuti e, infine, brevi conclusioni sono discusse nel paragrafo 4.

2. Materiali e metodi

Il campione analizzato nello studio è composto da 160 pazienti affetti da asma severo, residenti in varie zone del sud Italia e in cura presso del Dipartimento dell'Emergenza e dei Trapianti di organi (DETO) del Policlinico di Bari. A ciascun paziente è stato somministrato il trattamento a base di Benralizumab ad un tempo T0 (considerato la baseline), al tempo T6 (dopo 6 mesi) per verificare quali siano state le risposte a tale trattamento in relazione alle caratteristiche dei pazienti di interesse per il clinico.

I dati sono stati rilevati dalle schede dei pazienti, gestite in formato elettronico dall'equipe medica di ricerca. Le variabili di interesse per l'analisi statistica sono state le seguenti:

- età
- genere
- BMI (body mass index)
- status di fumatore (classificato in tre categorie: sì, no, ex fumatore)
- età di insorgenza della malattia
- anni di presenza della malattia
- ereditarietà (atopia familiarità)
- numero di riacutizzazioni
- somministrazione corticosteroidi orali – OCS (dicotomica: sì/no)
- dose media prednisone
- esito dell'ACT - Asthma Control Test
- FEV1% (flusso massimo espiratorio massimo al 1° secondo)
- livello di eosinofili ematici
- livello di Ige totali

Ai fini del monitoraggio dell'efficacia della terapia, sono state utilizzate le seguenti variabili, rilevate rispettivamente ai tempi T0 e T6:

- numero di riacutizzazioni
- dose media prednisone
- ACT

- FEV1% (flusso massimo espiratorio massimo al 1° secondo)

La variabile numero di riacutizzazioni (esacerbazioni) è stata ricodificata sulla base del periodo di tempo in cui i sintomi hanno mostrato un peggioramento, con conseguente intensificarsi della terapia, espressa come variazione percentuale: se $\Delta=100\%$ è stata attribuita l'etichetta "Super risposta", se $50\% < \Delta < 100\%$ l'etichetta "Risposta" e se $\Delta < 50\%$, quella di "Non risposta".

La variabile dose media prednisone, che rappresenta la dose media di corticosteroide somministrato per via orale al paziente per combattere le infezioni causate dall'asma, è stata anch'essa oggetto di ricodifica col medesimo criterio della variazione percentuale usato per il numero di riacutizzazioni.

La variabile ACT riporta gli esiti del test per il controllo dell'asma ed esprime, attraverso una scala numerica (da 0=peggiore a 25=migliore), lo stato di malattia. Dal punteggio di 20 in su, la malattia è considerata controllata. Ai fini della presente analisi, la valutazione è effettuata attraverso una differenza tra T0 e T6 attribuendo le seguenti etichette: "Super risposta" se $20 \leq \Delta < 25$; "Risposta" se $3 \leq \Delta < 20$; "Non risposta" se $\Delta < 3$.

La variabile FEV1, espressa in litri o in percentuale rispetto alla norma, indica il grado di ostruzione delle vie aeree; è il parametro più oggettivo, ma non correla con la sintomatologia in modo chiaro e, dunque, viene considerato come una misura di beneficio surrogato, poiché, per il clinico è più importante controllare il sintomo, piuttosto che il miglioramento di un parametro strumentale. La variazione è stata valutata in termini percentuali con il seguente criterio: una variazione $\Delta > 80\%$ è stata ricodificata come "Super risposta", se $12\% \leq \Delta \leq 80\%$, "Risposta", se $\Delta < 12\%$ "Non risposta".

Una ulteriore variabile è stata ottenuta "incrociando" gli esiti "assenza esacerbazioni" e "assenza corticosteroide", (denominata ESA0andOCS0), ottenendo una variabile risposta politomica a tre categorie in cui:

- viene attribuita la categoria "super risposta" se le classificazioni delle variazioni percentuali attribuite alle variabili "numero riacutizzazioni" e "dose media di corticosteroide" per ciascun paziente sono entrambe "super risposta";
- se le precedenti classificazioni sono una "risposta" e l'altra "super risposta", la categoria sarà "risposta";
- se una delle due è "non rispondente", sarà attribuita la categoria "non risposta".

Prima di iniziare l'analisi è stata fatta una ulteriore ricodifica della variabile risposta ESA0andOCS0 trasformandola in binaria con i seguenti esiti: "Super risposta" e "Non super risposta".

Le analisi statistiche sono state realizzate con il software IBM SPSS Statistical Package for Social Sciences. In particolare, dopo il test di Shapiro-Wilk per la verifica della normalità delle distribuzioni, per capire quali variabili fossero più rilevanti per individuare le caratteristiche del paziente che ha sviluppato una super risposta al trattamento, sono state utilizzate la regressione logistica multivariata, l'analisi dei cluster. Queste ultime tecniche sono state usate per porre le basi della c.d. medicina personalizzata, che si pone l'obiettivo di dare al clinico gli strumenti conoscitivi per individuare la giusta dose di farmaco da somministrare al paziente, dati alcuni predittori.

3. Risultati

3.1 *Analisi descrittiva delle caratteristiche dei pazienti*

I 160 pazienti coinvolti in questo studio preliminare sono di genere prevalentemente femminile (68,1%), hanno età media di 53,8 anni ($\pm 13,9$ anni), che sale a 56,0 anni per le donne e si attesta sui 53 anni per gli uomini. La tabella 1 mostra i dati rilevati al tempo 0 sui pazienti, distinti per genere.

Tabella 1. *Principali caratteristiche dei pazienti arruolati nello studio secondo il genere (medie e percentuali)*

Caratteristiche (medie)	F	M	Caratteristiche (%)	F	M
Età (anni compiuti)	54	53	Fumo		
Peso (kg)	67,12	81,81	No	76,2	70,6
Altezza (cm)	159,63	172,71	Si	7,3	5,9
BMI	26,08	27,11	ex fumatore	16,5	23,5
Anni di malattia	21	20	Familiarità atopia (=si)	56,0	29,4
Numero riscutizzazioni/anno	5	5	OCS (si)	59,6	58,8
Dose prednisone	7	8	Poliposi nasale (=si)	52,7	67,4
ACT	13	14	Bronchiectasie (=si)	22,9	19,6
FEV1pre (litri)	1,64	2,24			
FEV1pre (%)	66,9	62,8	Esito trattamento		
Eosinofili ematici	825	998	NonSuperRisposta	32,2	40,5
IgE totali	407	423	SuperRisposta	67,8	59,5

3.2 *Analisi multivariate: modello logistico e analisi dei cluster*

Per individuare le variabili che hanno maggior influenza sulla risposta del paziente alla terapia sperimentale è stato stimato un modello logistico multivariato, in cui la variabile risposta binaria è stata codificata attribuendo il valore 1 a tutti i soggetti che, dopo un determinato periodo dalla somministrazione del farmaco, hanno azzerato i propri livelli di ESA e OCS, etichettati come Super-rispondenti, e 0 nei restanti casi.

I predittori inseriti nel modello sono stati: età, età di insorgenza della malattia, anni di malattia, atopia, esinofili, Ige, FEV1%, numero di esacerbazioni, poliposi nasale e bronchiectasie.

Utilizzando la procedura stepwise backward con il metodo Wald, sono risultate significative (con $\alpha=0,05$) solo l'età del paziente ($p\text{-value}=0,020$), l'età di insorgenza della malattia ($p\text{-value}=0,050$), la poliposi nasale ($p\text{-value}=0,034$), e la somministrazione di azitromicina ($p\text{-value}=0,002$). In particolare, l'età di insorgenza dell'asma e l'assunzione di azitromicina sono risultati fattori di protezione, come evidenziato dagli odds ratio (Exp(B)) in Tabella 2.

Tabella 2. *Modello di regressione logistica binaria per i pazienti con super risposta*

Predittori	B	S.E.	Wald	df	p-value	Exp(B)
Età	0,039	0,017	5,426	1	0,001	1,040
Età insorgenza asma	-0,440	0,022	3,854	1	0,049	0,957
Azitromicina	-1,818	0,597	9,276	1	0,002	0,162
Poliposi nasale	0,858	0,404	4,519	1	0,034	2,359
Costante	2,159	0,995	4,707	1	0,030	8,666

La cluster analysis è stata implementata considerando le medesime variabili usate come predittori nel modello logistico.

È stato effettuato un clustering gerarchico con il metodo Ward e la distanza euclidea al quadrato, ottenendo 4 gruppi: il primo composto da 23 pazienti, il secondo da 19, il terzo da 35, il quarto da 28. Il numero dei pazienti considerati nella analisi dei gruppi è sceso da 160 a 105 a causa di alcuni dati mancanti nel dataset.

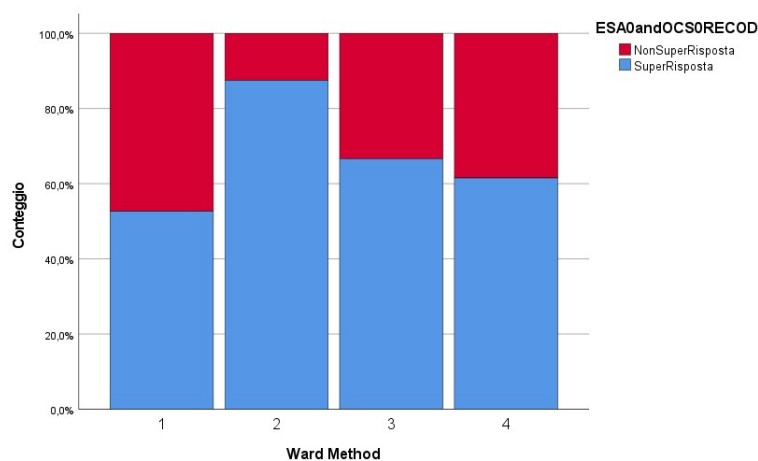
Ai fini della valutazione del clinico, il cluster più significativo è risultato essere il secondo, in cui si trova il maggior numero di super-rispondenti (Fig. 1).

Le caratteristiche dei cluster possono essere ricostruite attraverso i dati esposti nella Tabella 3. I pazienti che hanno risposto meglio alla terapia risultano essere coloro che hanno un'età media e un'età di insorgenza media inferiore agli altri, che

hanno un flusso espiratorio massimo al primo secondo in termini percentuali più alto, con un alto indice di massa corporea, a cui si associa solitamente uno stato infiammatorio più elevato, coloro che presentano un livello di eosinofili più alto, anch'essi sintomatici di livelli maggiori di infiammazione, che presentano poche bronchiectasie e un minor numero di polipi nasali.

In altri termini, la cluster analysis ha evidenziato che il paziente che ha una infiammazione più elevata, ma il cui flusso espiratorio è maggiormente conservato grazie alla terapia massimale inalatoria utilizzata, risponde meglio degli altri.

Figura 1. Distribuzione percentuale dei pazienti nei quattro cluster secondo la variabile super-risposta si/no.



Al contrario, la peggiore risposta alla terapia si riscontra per il cluster 1, caratterizzato da una più elevata presenza di patologia allergica come evidenziato dal prick test positivo in tutti i casi e dall'esecuzione in anamnesi di immunoterapia.

Tabella 3. Caratteristiche dei pazienti secondo i cluster

Variabili	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Età (media)	58	35	55	62
Età insorgenza asma (media)	46	24	27	45
FEV1 pre (%)	55,7	79,3	54,1	62,2
Eosinofili ematici (medio)	613	1363	992	891
BMI (medio)	26,8	28,3	26,3	26,3
Poliposi nasale (% si)	34,8	47,4	65,7	60,7
Bronchiectasie	17,4	21,1	42,9	17,9
Azitromicina (assunzione=si)	13,0	0,0	22,9	21,4

4. Conclusioni

Le tecniche statistiche trovano un impiego sempre crescente nelle analisi di dati clinici, soprattutto in quella che è la nuova sfida della “medicina personalizzata”.

Impiegate negli studi retrospettivi, consentono di individuare i fattori di rischio che hanno maggiormente influito sulla risposta ai protocolli sperimentali, orientando le scelte dei medici e rendendo le diagnosi più accurate.

Tenuto conto della variabilità dei pazienti, le analisi di cluster sembrano essere molto informative, poiché consentono di identificare diversi gruppi di pazienti simili tra loro per numerose caratteristiche. Questo ha consentito di indentificare 4 sottogruppi di pazienti all'interno del sottogruppo T2. Dunque, contrariamente a quanto ritenuto, cioè che il fenotipo T2 sia abbastanza omogeneo e, quindi, target ideale di terapia con il Benralizumab, in realtà si osserva una notevole eterogeneità tra i pazienti, che condiziona la risposta clinica.

La sperimentazione, condotta dall'equipe del Dipartimento dell'Emergenza e dei Trapianti di organi (DETO) del Policlinico di Bari, di cui fa parte il prof. Di Bona, proseguirà lo studio con l'integrazione di database multicentrici, rilevati sull'intero territorio nazionale. Ciò consentirà di disporre di popolazioni più numerose e database più ricchi di informazioni sui pazienti tali da giustificare in ricorso anche ad altre tecniche di analisi, tra cui quelle di machine learning.

Le due tecniche multivariate utilizzate, regressione logistica e analisi dei cluster, hanno consentito di delineare il profilo dei pazienti che meglio hanno risposto alla terapia con Benralizumab in base alla risposta infiammatoria dei biomarker e alle caratteristiche cliniche. Questi risultati sono utili per individuare percorsi terapeutici differenti e modulabili.

Riferimenti bibliografici

- Agache I, Song Y, Rocha C, et al. (2020). Efficacy and safety of treatment with dupilumab for severe asthma: a systematic review of the EAACI guidelines-Recommendations on the use of biologicals in severe asthma. *Allergy*. 75(5):1058-1068.
- Bleecker ER, FitzGerald JM, Chanez P, et al. (2016). SIROCCO study investigators. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β 2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial. *Lancet*. 388(10056):2115-2127.