

MODELLI DI BAYES NELLE COMPLICANZE DEL DIABETE

Vittorio Colagrande, Marta Di Nicola, Rocco Di Mascio, Fabio Capani¹, Piero Vellante, Enzo Ballone

Dipartimento di Medicina e Scienze dell'Invecchiamento, Laboratorio di Biostatistica, Università degli Studi "G. d'Annunzio", Chieti

¹Unità di Diabetologia, Università degli Studi "G. d'Annunzio", Chieti

Riassunto

Le due più comuni complicanze che i pazienti con diabete di tipo 1 e 2 sviluppano sono la Retinopatia Diabetica (RD) e la Nefropatia Diabetica (ND). In tale contesto, riveste particolare importanza lo studio del tempo di sviluppo della ND per un paziente che ha già sviluppato la RD. Nel lavoro tale problematica è stata affrontata utilizzando, in primo luogo, la formula di Bayes per determinare la probabilità condizionata del tempo di sviluppo di una seconda complicanza, dato il tempo di sviluppo di una prima complicanza. Successivamente, è stato introdotto il fattore di Bayes per valutare l'evidenza statistica fornita dall'osservazione di un dato tempo di RD sul tempo di sviluppo della ND. Vengono riportati i risultati ottenuti dall'applicazione dei due modelli ad una popolazione di 5.473 pazienti con diabete di tipo 2 seguiti dal Servizio di Diabetologia dell'Ospedale Civile di Pescara (Italy). I risultati evidenziano che l'acquisire informazione sul tempo di sviluppo della RD apporta un contributo importante alla conoscenza, in termini probabilistici, del tempo di sviluppo della ND.

PREMESSA

Avere un predittore precoce delle complicanze del diabete può essere d'aiuto nella prevenzione e/o nel management delle complicanze [1]. Risulta importante, ad esempio, conoscere se lo sviluppo di una complicanza modifica il tempo di sviluppo di una seconda complicanza.

In questo lavoro viene affrontata l'analisi delle relazioni tra i tempi di sviluppo di due complicanze con un approccio che prevede l'applicazione della "formula di Bayes" e del "fattore di Bayes" [3, 4, 6]. La popolazione studiata attiene i pazienti con diabete di tipo 2 del Registro delle Prestazioni del Servizio di Diabetologia dell'Ospedale Civile di Pescara [2].

OBIETTIVI

Il lavoro si propone di valutare, in termini probabilistici, il tempo di sviluppo di un evento (complicanza, progressione, ricaduta, ecc), dato un tempo antecedente di sviluppo di un altro evento. Nello specifico si vuole determinare il valore predittivo della Retinopatia Diabetica (RD) per l'evento Nefropatia Diabetica (ND) attraverso un'analisi statistica di tipo non parametrico.

MATERIALI E METODI

Sono stati individuati 7.083 pazienti seguiti dal Servizio di Diabetologia dal 1993 al 2002. Di questi 1.610 (22,7%) avevano avuto solo un contatto occasionale con il Servizio, prevalentemente per motivi legali (certificazioni inerenti rilascio di patente di guida, esenzione ticket farmaceutici, screening per diabete, ecc) e sono quindi stati esclusi non essendo parte integrante della coorte seguita dal Servizio. I restanti 5.473 pazienti con diabete di tipo 2 sono stati arruolati nello studio ed hanno costituito il database di riferimento.

Gli eventi (le complicanze) presi in esame sono la RD e la ND, così definiti: RD pazienti con retinopatia background o proliferativa; ND pazienti con albuminuria >300 mg/24h, incremento della creatinina o in dialisi.

Di ogni paziente diabetico arruolato sono state rilevate la data di diagnosi di diabete (inizio del follow-up), quella di registrazione della RD e della ND e, se tali eventi non si sono verificati, quella dell'ultimo controllo presso il Servizio.

Stabilito un periodo di osservazione di n anni, le variabili temporali di sviluppo delle due complicanze T_R e T_N definiscono il tempo (in anni) tra la data di registrazione dell'evento verificato o quella dell'ultimo controllo (dati troncati) e la data della diagnosi di diabete (tempo 0). Gli eventi RD e ND sono stati analizzati, rispettivamente, negli intervalli di tempo $[0, t_R)$ e $[0, t_N)$, con $0 < t_R \leq t_N \leq n$.

Volendo mettere in relazione i tempi di sviluppo di ND con quelli di RD, si è ritenuto opportuno far ricorso ad una formula delle probabilità condizionate, in particolare la formula di Bayes, per ricavare informazioni sulla capacità predittiva dell'evento $\{T_R < t_R\}$ sull'evento $\{T_N < t_N\}$.

La formula di Bayes richiede anzitutto il calcolo delle probabilità a priori ($P(T_N < t_N)$) che, ai fini del presente lavoro, rappresentano le probabilità che un generico paziente diabetico verifichi la ND in un tempo inferiore a t_N anni, indipendentemente dallo sviluppo della RD. Tali probabilità sono state determinate facendo riferimento ad una popolazione di 6.828 pazienti con diabete di tipo 2 afferiti ai Servizi di Diabetologia degli Ospedali di Pescara e Chieti e seguita retrospettivamente per 25 anni di diabete.

Le probabilità a priori vengono, poi, aggiornate sulla base delle informazioni relative ai tempi di registrazione dei due eventi (probabilità $P(T_R < t_R | T_N < t_N)$) derivanti dal campione di 5.473 pazienti seguiti dal Servizio di Diabetologia di Pescara. Si perviene, quindi, al calcolo delle probabilità a posteriori o probabilità di Bayes relative al tempo di ND. Queste danno una misura della probabilità che un paziente con diabete di tipo 2 sviluppi la ND in un tempo inferiore a t_N anni, dato un tempo di RD inferiore a t_R anni.

Nello specifico, le probabilità a posteriori sono state determinate attraverso la seguente formula

$$P(T_N < t_N | T_R < t_R) = \frac{P(T_N < t_N) \cdot P(T_R < t_R | T_N < t_N)}{P(T_R < t_R)} \quad (1)$$

Le probabilità $P(T_R < t_R)$, $P(T_N < t_N)$ e $P(T_R < t_R | T_N < t_N)$ sono state ottenute come complemento ad uno delle funzioni di sopravvivenza libera da evento (EFS), definite come $S_N(t_N) = P(T_N \geq t_N)$, $S_R(t_R) = P(T_R \geq t_R)$ e $S_R(t_R | T_N < t_N) = P(T_R \geq t_R | T_N < t_N)$ e calcolate applicando il metodo delle Tavole di Sopravvivenza.

La (1) evidenzia come l'“opinione” a priori circa l'evento $\{T_N < t_N\}$ viene trasformata in una “opinione” a posteriori attraverso l'informazione apportata dalla osservazione dell'evento $\{T_R < t_R\}$; pertanto la trasformazione è indicativa del contributo dell'osservazione sperimentale all'analisi condotta.

È opportuno far rilevare che, nelle usuali applicazioni della formula di Bayes, gli eventi condizionanti hanno il significato di possibili “cause” dell'evento osservato (visto come “effetto”). Nel caso della formula (1) proposta nel presente lavoro, invece, l'evento condizionante $\{T_N < t_N\}$ è da intendersi come uno stato di informazione e non deve essere visto in funzione causale. L'osservazione di $\{T_R < t_R\}$, inoltre, può essere presa come un evento in qualche modo “predittivo” dell'evento $\{T_N < t_N\}$. I risultati dell'applicazione della formula (1) alla casistica in esame sono stati riportati e discussi anche in un precedente lavoro [2].

Un'altra relazione di natura probabilistica, derivante da una espressione alternativa della formula di Bayes, consente di mettere maggiormente in luce l'informazione aggiuntiva prodotta dall'osservazione sperimentale, distinguendo tale informazione da quella a priori.

Si tratta di introdurre gli ODDS ($ODD = \text{Prob}/(1-\text{Prob})$) sia a priori che a posteriori e considerare la seguente relazione

$$\text{Odds a posteriori} = \text{Odds a priori} \times \text{fattore di Bayes} \quad (2)$$

Il fattore di Bayes definito dalla (2) permette di ottenere una misura della “evidenza statistica” fornita dall’osservazione di un tempo di RD inferiore ad un dato valore t_R , a favore dell’evento tempo di ND inferiore a t_N .

Molti autori [5, 6] collocano il fattore di Bayes nell’ambito del confronto tra una ipotesi e la sua complementare. In questo lavoro, stabilite le due ipotesi $H_1 = \{T_N < t_N\}$ e $H_2 = \{T_N \geq t_N\}$, il fattore di Bayes permette di ricavare informazioni, in termini probabilistici, sull’“evidenza statistica” fornita dall’osservazione $\{T_R < t_R\}$ all’evento H_1 . In tale contesto utilizzando la (1) e la (2) si ottiene, per il fattore di Bayes, la seguente espressione

$$B_{12} = \frac{P(T_R < t_R | H_1)}{P(T_R < t_R | H_2)} = \frac{P(T_N < t_N | T_R < t_R)}{[1 - P(T_N < t_N | T_R < t_R)]} \cdot \frac{[1 - P(T_N < t_N)]}{P(T_N < t_N)} \quad (3)$$

dove $P(T_N < t_N)$ e $P(T_N < t_N | T_R < t_R)$ sono rispettivamente le probabilità a priori e a posteriori presenti nella formula di Bayes (1).

Nell’uso standard del fattore di Bayes, l’evento di cui si valuta l’evidenza sperimentale è un’ipotesi non osservabile, mentre in questo lavoro è un evento osservabile nel futuro ($\{T_N \geq t_N\}$). Il fattore di Bayes, allora, può essere assunto come un indicatore diretto della “capacità predittiva” rispetto all’evento $\{T_N < t_N\}$ dell’osservazione $\{T_R < t_R\}$.

Per quanto attiene la valutazione dell’ “evidenza statistica” sulla base dei valori del fattore B_{12} , sono stati proposti diversi intervalli con relative classi di evidenze. In realtà, in considerazione del fatto che delle probabilità, specie se piccole, interessa l’ordine di grandezza, risulta “naturale” che gli intervalli di valori di B_{12} vengano espressi in termini del logaritmo del rapporto di probabilità. Infatti tale rapporto risulta diverso da zero se le probabilità del numeratore e del denominatore sono diverse, invece tende a zero quando tali probabilità sono quasi uguali. In considerazione, poi, che la (2) in termini di logaritmo si trasforma in $\log(\text{Odds a posteriori}) = \log(\text{Odds a priori}) + \log(B_{12}) = \log(P(H_1)/P(H_2)) + \log(B_{12})$, si evidenzia come il logaritmo del rapporto di “credibilità” dei due eventi H_1 e H_2 viene modificato di un incremento pari al logaritmo del fattore di Bayes.

Jeffreys [8] propone di esprimere il fattore di Bayes in termini di $\log_{10}(B_{12})$. Kass e Raftery [5], pur riportando l’interpretazione di Jeffreys, ritengono che può essere più utile calcolare il doppio del logaritmo naturale del fattore di Bayes, in quanto permette di riportarsi alla medesima scala del “likelihood ratio test statistics”. Successivamente Jackman [7] suggerisce nuovamente il calcolo del doppio del logaritmo naturale di B_{12} per determinare le classi dell’evidenza statistica a favore dell’ipotesi H_1 , come riportato in Tabella 1.

Tab. 1: Intervalli di evidenza statistica per il fattore di Bayes.

$2 \log_e (B_{12})$	B_{12}	Evidenza a favore di H_1
< 0	< 1	Negativa
0 – 2	1 – 3	Meritevole di menzione
2 – 5	3 – 12	Positiva
5 – 10	12 – 150	Forte
> 10	> 150	Molto forte

RISULTATI

Nel presente lavoro si è inteso porre l'attenzione sull'applicazione del fattore di Bayes alla casistica presa in esame, a partire dai risultati forniti dalla formula di Bayes (1) presentati in Tabella 2. Tale tabella riporta nella seconda colonna le probabilità a priori per i tempi di ND, nell'ultima riga le probabilità per i tempi di RD e, all'interno, le probabilità a posteriori calcolate sul database dei 5.473 pazienti diabetici presi in esame.

Tab. 2: Probabilità a posteriori di un tempo di ND < t_N anni per un paziente diabetico con un tempo di RD < t_R anni.

Tempo di ND	$1-S_N(t_N)^a$	Tempo di RD					
		$T_R < 2$	$T_R < 4$	$T_R < 6$	$T_R < 8$	$T_R < 10$	$T_R < 12$
$T_N < 2$	0,0260	0,1331	-	-	-	-	-
$T_N < 4$	0,0624	0,1434	0,1655	-	-	-	-
$T_N < 6$	0,0948	0,1449	0,1794	0,2752	-	-	-
$T_N < 8$	0,1255	0,1517	0,1893	0,2873	0,3447	-	-
$T_N < 10$	0,1475	0,1592	0,1937	0,3082	0,3502	0,3639	-
$T_N < 12$	0,1667	0,1723	0,2054	0,3086	0,3509	0,3722	0,4300
$1-S_R(t_R)$		0,0289	0,0465	0,0773	0,1231	0,1659	0,2540

^a Probabilità a priori stimate su una popolazione diabetica di 6.828 pazienti con diabete di tipo 2 seguita retrospettivamente per 25 anni.

FONTE: Ballone E, Colagrande V, Di Nicola M, Di Mascio R, Di Mascio C e Capani F. Probabilistic approach to developing nephropathy in diabetic patients with retinopathy.

Statistics in Medicine 2003; 22: 3889-3897.

Le probabilità a posteriori, presenti nella Tabella 2, vanno analizzate per colonna ed indicano le probabilità di un tempo di sviluppo della ND prima di 2, 4, 6, 8, 10 o 12 anni dalla diagnosi di diabete, dato un tempo di sviluppo della RD

inferiore a 2, 4, 6, 8, 10 o 12 anni dalla diagnosi di diabete. Si osservi, ad esempio, che il valore di probabilità del 43,00% ottenuto per $t_R = t_N = 12$ anni, se confrontato con il 16,67% della probabilità a priori di una ND entro 12 anni di diabete, evidenzia il peso delle osservazioni sperimentali sul campione di 5.473 pazienti diabetici.

I risultati dell'applicazione del fattore di Bayes (3) alle probabilità di Tabella 2 sono riportati nella Tabella 3.

Tab. 3: Fattore di Bayes di un tempo di ND < t_N anni rispetto ad un tempo di ND $\geq t_N$ per un paziente diabetico con un tempo di RD < t_R anni.

Tempo di ND	Tempo di RD					
	$T_R < 2$	$T_R < 4$	$T_R < 6$	$T_R < 8$	$T_R < 10$	$T_R < 12$
$T_N < 2$	5,75	–	–	–	–	–
$T_N < 4$	2,52	2,98	–	–	–	–
$T_N < 6$	1,62	2,09	3,63	–	–	–
$T_N < 8$	1,25	1,63	2,81	3,67	–	–
$T_N < 10$	1,09	1,39	2,57	3,12	3,31	–
$T_N < 12$	1,04	1,29	2,23	2,70	2,96	3,77

Si può osservare, anzitutto, che tutti i valori di B_{12} risultano maggiori di 1, ciò è indicativo del fatto che l'osservazione di un tempo di RD minore di t_R anni $\{T_R < t_R\}$ fornisce all'ipotesi $\{T_N < t_N\}$ un supporto sperimentale maggiore rispetto all'ipotesi $\{T_N \geq t_N\}$. D'altra parte, fissato un tempo t_R , essendo B_{12} decrescente al crescere di t_N , l'evidenza statistica risulta tanto più "debole" quanto più aumenta la distanza temporale $t_N - t_R$.

Valutando poi, i risultati della Tabella 3 secondo l'interpretazione di Tabella 1, si osserva che si ha "evidenza positiva" a favore di H_1 in corrispondenza della diagonale principale, dato che per $t_N = t_R$ il fattore di Bayes assume valori compresi tra 3 e 4. In particolare il valore $B_{12} = 5,75$, ottenuto per $t_N = t_R = 2$, è indicativo del fatto che, l'aver osservato l'evento $\{T_R < 2\}$ attribuisce una evidenza statistica decisamente positiva all'ipotesi $\{T_N < 2\}$ rispetto all'ipotesi $\{T_N \geq 2\}$. Importanti risultano anche i valori del fattore di Bayes prossimi alla diagonale principale (in particolare con $t_N = t_R + 2$); invece quando t_N risulta molto distante da t_R , il fattore di Bayes è prossimo a 1.

Si può dedurre, in sostanza, che si ha evidenza sperimentale apprezzabile di sviluppo di ND e RD negli stessi intervalli di tempo o intervalli "vicini", mentre tale evidenza risulta sempre meno apprezzabile, ma "meritevole di menzione", quanto più l'intervallo temporale $[0, t_N)$ risulta maggiore dell'intervallo $[0, t_R)$.

In conclusione è importante rilevare che l'applicazione della formula di Bayes proposta nel lavoro ha permesso di valutare, per una popolazione di pazienti diabetici, in che misura la probabilità non condizionata (a priori) del tempo di sviluppo di una complicanza (ND) è stata incrementata dalla conoscenza del tempo di sviluppo dell'altra complicanza (RD).

Il fattore di Bayes, poi, ha consentito di dare una misura della "capacità predittiva" dell'osservazione di un tempo di RD inferiore a t_R in riferimento all'evento tempo di ND inferiore a t_N . I valori ottenuti per esso indicano che l'ipotesi $\{T_N < t_N\}$ ha ottenuto, dall'osservazione dell'evento $\{T_R < t_R\}$, un "supporto sperimentale" comunque maggiore rispetto all'ipotesi $\{T_N \geq t_N\}$.

Va osservato, infine, che l'interesse del fattore di Bayes proposto risiede nella constatazione che esso risulta essere indipendente dalla "massa" di probabilità inizialmente collocata in H_1 e H_2 , anche se dipende dal modo in cui la probabilità iniziale è ripartita all'interno delle ipotesi stesse. "Questa dipendenza, relativamente limitata dalla distribuzione iniziale, rende il fattore di Bayes uno strumento utile quando si vuole confrontare l'evidenza sperimentale a favore di due ipotesi disgiunte" [6] e quindi fornisce un indicatore diretto della capacità predittiva dell'osservazione $\{T_R < t_R\}$ sul tempo di sviluppo della Nefropatia Diabetica.

BIBLIOGRAFIA

1. AHMED M., ABU EL-ASRAR, ET ALL. (2002). Retinopathy as a predictor of other diabetic complications. *International Ophthalmology*, 24, 1-11.
2. BALLONE E., COLAGRANDE V., DI NICOLA M., DI MASCIOR., DI MASCIOR., CAPANI F. (2003). Probabilistic approach to developing nephropathy in diabetic patients with retinopathy. *Statistics in Medicine*, 22, 3889-3897.
3. CRAIG BA., FRYBACK DG., KLEIN R., KLEIN BEK. (1999). A Bayesian approach to modelling the natural history of a chronic condition from observations with intervention. *Statistics in Medicine*, 18(11), 1355-1371.
4. GOODMAN SN. (1999). Toward Evidence-Based Medical Statistics 2: The Bayes Factor. *Annals of Internal Medicine*, 130, 1005-1013.
5. KASS RE., RAFTERY AE. (1995). Bayes Factor. *Journal of the American Statistical Association*, 90(430), 773-795.
6. PICCINATO L. (1997). Il Fattore di Bayes come strumento pratico di Statistica Applicata. Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma "La Sapienza" (Serie B n.1), Roma.
7. JACKMAN S. (2003). Bayes Factor. *Encyclopedia of Research Methods for the Social Sciences* (Volume 3), Sage Publication, Thousand Oaks.
8. JEFFREYS H. (1961). *Theory of Probability* (3rd ed.), Oxford University Press, Oxford U.K.

BAYES' MODELS IN THE DIABETIC COMPLICATIONS

Summary

The two most common chronic complications that patients with type I or II Diabetes Mellitus develop are Diabetic Retinopathy (DR) and Diabetic Nephropathy (DN). An interesting question is whether a diabetic patient that had developed a Retinopathy will develop a Nephropathy sooner. This problem was approached by initially applying Bayes' formula to determine the conditional probability of the time of development of a second complication, given the time of development of a first complication. Then, we introduced Bayes factor to evaluate the statistical evidence provided by the observation of a given time of DR on the time of development of DN. In this article we discussed the results of application of two Bayesian models (Bayes' formula and Bayes' factor) obtained using a study population of 5,473 patients with type 2 Diabetes treated at the Servizio di Diabetologia dell' Ospedale Civile di Pescara (Italy). The results indicate that acquiring information on the time of development of DR provides a positive contribution to the knowledge, in probabilistic terms, of the time of development of DN.

Particularly, Bayes' formula permitted the evaluation the extent to which the knowledge concerning the time of developing DR increased the probability of developing DN with respect to an estimate performed without this knowledge. While, Bayes' factor permitted the measurement of the predictive capacity with respect to the time of DN of the observation of the time of DR.