



## IL MODELLO PREDITTIVO

A cura di L Cianfanelli, C Piccinelli, G Carreras, G Como,  
F Fagnani, E Pagano, N Segnan, C Senore.



## Indice

Introduzione

1. Gli stati del modello

1.1 Stati di salute

1.2 Stati per l'abitudine al fumo

1.3 Stati per l'attività fisica

2 Inizializzazione popolazione

2.1 Inizializzazione stato di salute

3 Evoluzione della popolazione

3.1 Transizioni relative al fumo

3.2 Transizioni relative all'attività fisica

3.3 Transizioni relative agli stati di salute

3.4 Output del modello

3.5 Intervento di prevenzione

Bibliografia

Appendice A: distribuzione ex fumatori

## Introduzione

L'esposizione al fumo e alla sedentarietà rende gli individui più suscettibili a problemi di salute, impattando negativamente l'aspettativa e la qualità di vita degli individui stessi. L'obiettivo di questo studio è quantificare gli effetti di interventi di prevenzioni su fumo e sedentarietà nella popolazione italiana. Gli effetti sono misurati in termini di DALYs, definiti come la somma tra gli anni di vita persi a causa di morte prematura (YLL) e gli anni di vita vissuti con disabilità (YLD). Il guadagno netto di DALYs dovuto ad un intervento di prevenzione è valutato comparando uno scenario baseline con uno scenario in cui è implementato l'intervento di prevenzione in esame. Al fine di simulare l'andamento della popolazione e stimare YLL e YLD nei diversi scenari, il modello predittivo descrive l'evoluzione degli individui tramite catene di Markov indipendenti, il cui spazio degli stati tiene conto delle possibili esposizione ai fattori di rischio e lo stato di salute dell'individuo. I parametri del modello sono calibrati sulla base della letteratura, di dati Istat e Global Burden of Disease Study (GBD).

L'utilizzo di modelli Markoviani per stimare l'effetto di interventi di prevenzione in una popolazione non è nuovo in letteratura [8, 5, 7]. Tra le varie applicazioni si trovano in letteratura lavori sull'esposizione al fumo [5, 7, 6], sulla sedentarietà [1], e sull'obesità [2, 3]. Il nostro modello, oltre a generalizzare in parte la letteratura esistente per quanto concerne i singoli fattori di rischio (come nel caso di [1] per quanto riguarda la sedentarietà), tiene in considerazione in contemporanea di due diversi fattori di rischio, fumo e sedentarietà.

## 1. Gli stati del modello

Si descrive ogni individuo tramite una catena di Markov indipendente. La catena di Markov di ogni individuo è a tempo discreto, con passo temporale di un anno.

Si considera una popolazione di individui con età maggiore o uguale a 25 anni, stratificata per età e genere, con l'assunzione che tutti gli individui con età maggiore o uguale a 90 siano inclusi in un'unica fascia di età, chiamata 90+ (queste assunzioni sono arbitrarie e rilassabili senza perdita di generalità). Di conseguenza, l'età  $e$  prende valori nell'insieme  $\{25, 26, \dots, 89, 90+\}$ . Ogni individuo è anche descritto dal genere  $g$ , che prende valore nell'insieme  $\{m, f\}$  (maschio o femmina).

Per ciascun individuo si considerano due fattori di rischio (fumo e sedentarietà) e 5 patologie traccianti ad essi correlate: malattia coronarica (IMA), cancro ai polmoni/trachea (CP), ictus (ICT), bronco pneumopatia cronico-ostruttiva (BP) e diabete di tipo 2 (DIA). Si indica con  $\mathcal{M}$  l'insieme delle patologie.

La condizione di ogni individuo è completamente determinata da età, genere, fumo, attività fisica e salute, si può descrivere cioè tramite cinque indici, nella forma

$$(e, g, f, a, sm).$$

Chiamiamo  $e$  e  $g$  i parametri dell'individuo, che evolvono deterministicamente, ed identifichiamo come stato della catena la tripla  $(f, a, sm)$ , la cui evoluzione è stocastica. Gli stati possibili sono quindi il prodotto di tutte le combinazioni di esposizione al fumo, esposizione alla sedentarietà e stato di salute.

I dettagli sugli stati di fumo, attività fisica e stati di salute sono spiegati nelle prossime sottosezioni. L'inizializzazione della popolazione, e le probabilità di transizione della catena sono dettagliati rispettivamente in Sezione 2 e 3.

## 1.1 Stati di salute

Si considerano i seguenti stati:

- sano (S);
- malati di cancro ai polmoni/trachea (CP);
- malati di bronco pneumopatia cronico-ostruttiva (BP);
- individuo che ha avuto un episodio di infarto al miocardio (IMA);
- individuo che ha avuto un episodio di ictus (ICT);
- malati di diabete di tipo 2 (DIA);
- tutte le combinazioni di due patologie;
- tutte le combinazioni di tre patologie;
- tutte le combinazioni di quattro patologie;
- lo stato con tutte le patologie.

## 1.2 Stati per l'abitudine al fumo

Riguardo all'esposizione al fumo, supponiamo che ogni individuo possa essere non fumatore, fumatore o ex-fumatore. Nel caso degli ex fumatori teniamo inoltre traccia di quanti anni sono passati dalla cessazione. Di conseguenza, dividiamo gli ex fumatori in ex fumatori da 1 anno, da 2 anni, etc, fino ad accorpate in unica categoria gli ex fumatori da più di 15 anni. Per l'esposizione al fumo abbiamo quindi 18 stati:

- non fumatore (NF)
- ex fumatore da 1, 2, ..., 15 anni ( $ex_1, ex_2, \dots, ex_{15}$ )
- ex fumatore da più di 15 anni ( $ex_+$ ).

### 1.3 Stati per l'attività fisica

Si considerano solo due stati, attivo (att) e sedentario (sed). Tale scelta è concorde con la letteratura esistente (si veda ad esempio [1]).

## 2 Inizializzazione popolazione

Per inizializzare la popolazione utilizziamo le seguenti fonti:

- la distribuzione iniziale degli individui per età e genere è ricavata da dati Istat del 2019;
- i dati sulle prevalenze delle cinque patologie traccianti per età e genere sono ricavati da GBD e sono relativi all'Italia nel 2019. Il GBD non fornisce prevalenze congiunte, cioè la frazione di individui che ha due o più delle patologie traccianti. In mancanza di tali informazioni, si assume indipendenza nelle prevalenze delle diverse patologie (si veda Sezione 2.1);
- La distribuzione congiunta di esposizione al fumo e attività fisica è ricavata da dati Istat del 2019, che però forniscono, per quanto concerne il fumo, una divisione solo in fumatori, non fumatori, ed ex fumatori. La distribuzione degli ex fumatori nelle varie sottoclassi è discussa in Appendice A.

Sia  $P_{f,a,g}^{e,s,m}$  il numero di individui di età  $e$ , genere  $g$  e stato  $(a, f, sm)$ . Laddove ci sia un sottoinsieme di indici ad indicare lo stato è sottintesa una marginalizzazione della distribuzione,

ad esempio  $P_{sm}^{e,g} = \sum_{f,a} P_{f,a,g}^{e,s,m}$ . In mancanza di distribuzioni congiunte di esposizione ai fattori di rischio e stato di salute si assume che le distribuzioni siano indipendenti, cioè

$$P_{f,a,g}^{e,s,m} \propto P_{f,a}^{e,g} \cdot P_{sm}^{e,g}. \tag{1}$$

### 2.1 Inizializzazione stato di salute

Si indica con  $N_m^{e,g}$  il numero di individui di età  $e$ , genere  $g$  e patologia  $m$ , ricavati da GBD. In mancanza di ulteriori dati assumiamo che la probabilità di avere due (o più) patologie sia proporzionale alla probabilità di avere le due (o più) separatamente. Vale quindi la seguente relazione, dove gli indici  $m$  e  $n$  scorrono sulle patologie,  $m \in sm$  significa che lo stato  $sm$  è uno stato in cui è presente la patologia  $m$  e  $P^{e,g}$  indica il numero totale di individui con età  $e$  e genere  $g$  :

$$P_{sm}^{e,g} \propto \prod_{\substack{m \in M \\ m \in sm}} N_m^{e,g} \prod_{\substack{n \in M \\ n \in sm}} (P_{n,sm}^{e,g} - N_n^{e,g}). \quad (2)$$

### 3 Evoluzione della popolazione

Le simulazioni avvengono entro un orizzonte temporale arbitrario (tipicamente 30 anni). Si lavora con una coorte aperta, ovvero ogni anno inseriamo dei nuovi 25enni, la cui distribuzione iniziale è la stessa della distribuzione dei 25enni utilizzata in sede di inizializzazione della popolazione.

La catena di Markov di ogni individuo è eterogenea, nella misura in cui le probabilità di transizione dipendono dall'età dell'individuo, e quindi dal tempo trascorso. Facciamo la seguente assunzione sulle transizioni.

**Assunzione 1** La probabilità di passare da uno stato  $(f, a, sm) \rightarrow (f', a', sm')$  per un individuo di età e genere  $(e, g)$ , indicata da  $P^{e,g} \{(f, a, sm) \rightarrow (f', a', sm')\}$ , si fattorizza nel seguente modo:

$$P^{e,g} \{(f, a, sm) \rightarrow (f', a', sm')\} = P_{f,f'}^{e,g} \cdot P_{a,a'}^{e,g} \cdot P_{sm,sm'}^{e,g}(f, a),$$

dove  $P_{f,f'}^{e,g}$  è la probabilità che un individuo passi da stato di fumo  $f$  a stato di fumo  $f'$ ,  $P_{a,a'}^{e,g}$  è la probabilità che un individuo passi da stato di attività  $a$  a stato di attività  $a'$ , e  $P_{sm,sm'}^{e,g}(f, a)$  è la probabilità che un individuo passi da stato di salute  $sm$  a stato di salute  $sm'$ , che dipende dall'esposizione al fumo e alla sedentarietà. Tutte le probabilità dipendono in linea di principio da età e genere.

Questa assunzione dice due cose fondamentali:

- non si considerano effetti di accoppiamento nelle transizioni, cioè ad esempio la transizione  $sm \rightarrow sm'$  dipende da  $f$ , ma non dallo stato di arrivo  $f'$ ;
- le transizioni negli stati di fumo ed attività fisica sono indipendenti dagli altri stati, mentre le transizioni di stati di salute dipendono dallo stato di fumo e di attività.

Ad esempio, consideriamo indipendenti le probabilità che un individuo smetta di fumare e si ammali in uno stesso anno, ma consideriamo il fatto che un individuo fumatore si ammala con maggior probabilità rispetto ad uno non fumatore, che è un elemento chiave ai nostri fini. Nelle prossime sezioni descriviamo le transizioni relative a fumo, attività, salute.

#### 3.1 Transizioni relative al fumo

Le transizioni relative al fumo seguono le seguenti regole:

## IL PERCORSO DEL LABORATORIO PREVENZIONE

- i non fumatori non possono cominciare a fumare;
- i fumatori ogni anno possono smettere di fumare, diventando quindi ex fumatori da 1 anno, con probabilità  $\alpha = 0.02$  ;
- gli ex fumatori da  $i$  anni possono ricominciare a fumare con probabilità

$$\phi_i = ABe^{-12i \cdot B}, \quad (3)$$

dove per i maschi utilizziamo i valori

$$A = 1.177, \quad B = 0.150, \quad (4)$$

e per le femmine

$$A = 1.197, \quad B = 0.113. \quad (5)$$

Si noti che le probabilità di transizione relative al fumo sono assunte indipendenti da età e genere. La relazione (3) è ricavata da [4].

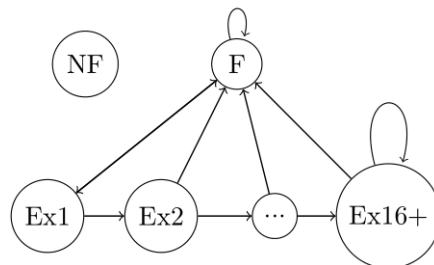


Figura 1: Transizioni relative al fumo.

## 3.2 Transizioni relative all'attività fisica

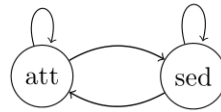


Figura 2: Transizioni relative all'attività fisica

Le transizioni ammissibili sono mostrate in Figura 1. Data l'assenza di dati affidabili sulle probabilità di transizione, imponiamo una matrice di transizione 2x2 che permetta di avere per ogni età e genere la prevalenza di individui sedentari ed attivi stazionaria nel tempo. Siano  $p_{att}^{e,g}$  e  $p_{sed}^{e,g}$  la frazione di attivi e sedentari tra gli individui di genere  $g$  ed età  $e$ , e sia  $A^{e,g}$  la matrice di transizione nello spazio delle sedentarietà. Per ogni età e genere deve quindi valere la seguente relazione:

$$\begin{cases} p_{att}^{e+1,g} = (1 - A_{att,sed}^{e,g}) \cdot p_{att}^{e,g} + A_{sed,att}^{e,g} \cdot p_{sed}^{e,g} \\ p_{sed}^{e+1,g} = (1 - A_{sed,att}^{e,g}) \cdot p_{sed}^{e,g} + A_{att,sed}^{e,g} \cdot p_{att}^{e,g} \end{cases} \quad (6)$$

dove le incognite sono  $A_{att,sed}^{e,g}$  e  $A_{sed,att}^{e,g}$ , coi vincoli

$$0 \leq A_{sed,att}^{e,g}, A_{att,sed}^{e,g} \leq 1. \quad (7)$$

Il sistema è sottodeterminato, dato che le due equazioni in (6) non sono indipendenti. Arbitrariamente, si sceglie la matrice di transizione che minimizzi la somma delle transizioni tra stati diversi, cioè che gli individui rimangano il più stabilmente possibile nei propri stati di attività fisica. In definitiva, la matrice di transizione si trova come soluzione del seguente programma lineare:

$$\begin{array}{ll} \min & A_{att,sed}^{e,g} + A_{sed,att}^{e,g} \\ \text{subject to} & (6), (7). \end{array} \quad (8)$$

A differenza di quella del fumo, la matrice di transizione dell'attività fisica dipende da età e genere, dato che è ricavata dalle prevalenze, a loro volta dipendenti da età e genere. Si noti che il



nostro modello generalizza [1], nel quale gli effetti di un intervento di prevenzione sulla sedentarietà vengono valutati assumendo che non ci siano transizioni tra lo stato di attivo e sedentario al di fuori di quelle indotte dall'intervento. Il nostro modello invece considera la possibilità che gli individui modifichino le proprie abitudini anche spontaneamente.

### 3.3 Transizioni relative agli stati di salute

Facciamo la seguente assunzione.

**Assunzione 2** *Le malattie sono croniche, cioè un individuo con una patologia continua ad avere la patologia fino al decesso.*

Al fine di non aumentare la complessità del calcolo dei parametri, associamo ad ogni stato di malattia una malattia dominante che ne descrive l'evoluzione.

**Assunzione 3** *L'evoluzione di individuo malato di più patologie è descritto dalla patologia più grave tra le presenti. Le patologie sono ordinate come segue, dalla più grave alla meno grave in ordine di letalità: CP, STR, MC, BP, DIA.*

**Assunzione 4** *Un individuo non si ammala di due patologie nello stesso anno.*

La probabilità di ammalarsi di una patologia  $m$  dipende da età, genere, fumo e attività. Si assume inoltre che essere malati di una patologia  $n$  correli con l'insorgere di una qualsiasi patologia  $m$  (la correlazione può anche essere nulla). Definiamo la seguente notazione.

- $\beta_m^{e,g}$  indica la probabilità che un non fumatore, attivo, sano di età  $e$ , genere  $g$  si ammali della patologia  $m$ . Tali parametri non sono noti a priori, in questa sezione illustreremo il metodo usato per stimarli.

- $RR_{f,a,m}^{e,g}$  indica il rischio relativo (moltiplicativo) per un individuo di età  $e$ , genere  $g$ , stato di fumo  $f$ , attività  $a$ , relativamente alla malattia  $m$ . Vale la seguente relazione:

$$\beta_{f,a,m}^{e,g} = \beta_m^{e,g} \cdot RR_{f,a,m}^{e,g} \quad (9)$$

I rischi relativi legati all'attività fisica sono ricavati da [1] e i rischi relativi legati al fumo da [9]. Per quanto riguarda la costruzione del rischio globale, si ottiene il rischio additivamente a partire dai rischi separati per sedentarietà ( $RR_{a,m}^{e,g}$ ) e fumo ( $RR_{f,m}^{e,g}$ ),

$$RR_{f,a,m}^{e,g} = 1 + (RR_{a,m}^{e,g} - 1) + (RR_{f,m}^{e,g} - 1). \quad (10)$$

•  $\theta_{nm}$  indica la correlazione tra patologia  $m$  e  $n$ . Assumiamo che i rischi relativi e i  $\theta$  agiscano in modo moltiplicativo, e indipendente, cioè la probabilità che un individuo in un stato di salute  $sm$  la cui malattia dominante è  $n$ , in stato di fumo  $f$ , attività  $\alpha$ , si ammali della patologia  $m$  è

$$\beta_{f,\alpha,m}^{e,\theta} = \theta_{nm}$$

dove  $\beta_{f,\alpha,m}^{e,\theta}$  si calcola usando (9). Per semplicità, i  $\theta_{nm}$  si assume indipendente da genere ed età, e si assume che  $\theta_{sm} = 1$  per ogni malattia  $m$ , dove  $S$  indica lo stato di sano. Al momento, per mancanza di dati, tutti gli elementi  $\theta_{nm}$  sono uguali a 1, cioè non si considerano correlazioni nella probabilità di insorgenza delle patologie.

•  $I_m^{e,\theta}$  indica il numero di incidenze annuali nella fascia di età  $e$ , genere  $\theta$  per la patologia  $m$ . Questi dati sono riferiti al 2019 e sono ricavati da GBD.

•  $mal(sm)$  indica la malattia dominante associata allo stato  $sm$ , cioè la patologia più importante tra quelle che caratterizzano lo stato  $sm$ .

Il parametro  $\beta_m^{e,\theta}$  viene ricavato imponendo che il numero atteso di incidenze per ogni età, genere e patologia sia quello dato dal GBD. A tal fine, scriviamo una relazione per il valore atteso delle incidenze ( $E[X]$  indica il valore atteso della variabile stocastica  $X$ ). In particolare, il numero atteso di incidenze è uguale al numero di individui in ogni stato, tra quelli che non comprendono già la patologia  $m$  ( $P_{f,\alpha,sm}^{e,\theta}$ ), moltiplicato per la probabilità di ammalarsi, dove per semplicità di notazione chiamiamo  $n := mal(sm)$ . Di conseguenza:

$$E[I_m^{e,\theta}] = \sum_{sm:m \notin sm} \square \sum_f \square \sum_\alpha \square P_{f,\alpha,sm}^{e,\theta} \beta_m^{e,\theta} RR_{f,\alpha,m}^{e,\theta} \theta_{nm} \quad (11)$$

Noti i rischi relativi  $RR$ , le correlazioni  $\theta$ , le prevalenze di fumo e malattia nello stato iniziale, e le incidenze annue per la patologia in esame, l'unica incognita dell'equazione è  $\beta_m^{e,\theta}$ , cioè

$$\beta_m^{e,\theta} = \frac{I_m^{e,\theta}}{\sum_{sm:m \notin sm} \square \sum_f \square \sum_\alpha \square P_{f,\alpha,sm}^{e,\theta} RR_{f,\alpha,m}^{e,\theta} \theta_{nm}} \quad (12)$$

A questo punto, tramite i rischi relativi, ricaviamo  $\beta_{f,\alpha,m}^{e,\theta}$  per ogni stato di esposizione al rischio  $(f, \alpha)$ .

Per quanto riguarda i parametri di mortalità, dividiamo le patologie in due categorie: malattie fulminanti (ictus e infarto) e no (cancro ai polmoni, broncopneumopatia, diabete). Le patologie fulminanti sono caratterizzate dal fatto che una frazione degli individui colpite dalla patologia muore immediatamente a seguito dell'evento.

**Assunzione 5** *Assumiamo per semplicità che un individuo affetto da una certa patologia abbia una probabilità di decesso uguale in ogni anno ad esclusione del primo, indipendentemente da quando ha sviluppato la patologia.*

Utilizziamo la seguente notazione:

- $v_{f,a,m}^{e,g}$  indica la probabilità che un individuo di età  $e$  e genere  $g$ , colpito da patologia  $m$ , con esposizione ai rischi  $f, a$  muoia a causa della patologia nell'anno di insorgenza della patologia stessa.
- $\delta_{f,a,m}^{e,g}$  indica la probabilità che un individuo in stato di fumo  $f$ , attività  $a$  e malattia  $m$  muoia in un qualsiasi anno (diverso dall'anno della diagnosi) per la patologia  $m$ ;
- In analogia a quanto fatto con i  $\beta$ , in caso di non fumatori attivi si omettono i pedici, cioè si definisce  $\delta_m^{e,g}$  indica la probabilità che un individuo di età  $e$ , genere  $g$ , malato della patologia  $m$ , non fumatore ed attivo muoia a causa della patologia stessa. Inoltre, assumiamo che la mortalità di un esposto ai fattori di rischi si ottenga a partire da quella di un non esposto tramite un fattore moltiplicativo  $\omega_{f,a,m}$ , cioè

$$\delta_{f,a,m}^{e,g} = \delta_m^{e,g} \cdot \omega_{f,a,m}$$

con  $\omega_{NF,att,m} = 1$ . Per semplicità (e mancanza di dati) si assume che i vari  $\omega$  non dipendano da genere ed età. Allo stato attuale gli  $\omega$  sono tutti uguali ad 1, cioè stiamo lavorando sotto l'assunzione che le patologie non influenzino il decorso delle patologie traccianti, ma solo la loro insorgenza.

**Osservazione** *Per le patologie fulminanti,  $v_m^{e,g}$  è un parametro indipendente che si ricava dalla letteratura, e ci dice la frazione di individui che in seguito a un evento muoiono in un arco di tempo breve. Assumiamo che tale parametro non dipenda da fumo e sedentarietà per mancanza di dati. Per le patologie non fulminanti assumiamo invece  $v_{f,a,m}^{e,g} = \frac{\delta_{f,a,m}^{e,g}}{2}$ . L'assunzione sottostante è che in media gli individui si ammalino a metà anno, e che quindi possano morire per la patologia solo nei sei mesi successivi. Per questo la probabilità di morire nell'anno della diagnosi si considera dimezzata rispetto ad ogni altro anno.*

Le morti attese per la patologia  $m$  sono:

$$\begin{aligned}
 E[M_m^{e,g}] &= \sum_{sm:m=mai(sm)} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} \delta_{f,a,m}^{e,g} \\
 &+ \sum_{sm:m \notin sm} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} \beta_{f,a,m}^{e,g} \theta_{nm} v_{f,a,m}^{e,g}
 \end{aligned}
 \tag{13}$$

dove nel primo termine  $mai(sm)$  è la malattia dominante relativa allo stato  $sm$ , e  $n$  nel secondo termine denota la malattia dominante relativa allo stato  $sm$ . I due termini vanno interpretati nel seguente modo:

- il primo termine sono le morti per causa  $m$  degli individui in stato di salute  $sm$  tale che  $m$  è la malattia dominante di  $sm$  ;
- il secondo sono le morti di individui in stati di salute  $sm$  che non contengono la patologia  $m$ , che nell’arco dello stesso anno si ammalano di  $m$  e muoiono di  $m$  ;

Tutti questi termini sono stratificati in base all’esposizione ai fattori di rischio. Note le morti  $M_m$  (da GBD), i vari parametri  $\omega$ ,  $\theta$  e  $v$  (per le patologie fulminanti in cui queste rappresentano un parametro indipendente) e i  $\beta$  (calcolati in precedenza) si ricavano i  $\delta$ . In particolare, per le malattie non fulminanti, in cui  $v_{f,a,m}^{e,g} = \frac{\delta_{f,a,m}^{e,g}}{2}$ , vale la relazione

$$\delta_m^{e,g} = \frac{M_m^{e,g}}{D_m^{e,g}},$$

dove

$$\begin{aligned}
 D_m^{e,g} &= \sum_{sm:m=mai(sm)} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} \omega_{f,a,m} \\
 &+ \sum_{sm:m \notin sm} \square \sum_f \square \sum_a \square \frac{P_{f,a,sm}^{e,g} \beta_{f,a,m}^{e,g} \theta_{nm} \omega_{f,a,m}}{2}.
 \end{aligned}$$

Per le patologie fulminanti, in cui i parametri  $v$  e  $\theta$  sono indipendenti,

$$\delta_m^{e,g} = \frac{M_m^{e,g} - \sum_{sm:m \notin sm} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} \beta_{f,a,m}^{e,g} \theta_{nm} v_m^{e,g}}{\sum_{sm:m=mai(sm)} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} \omega_{f,a,m}}.
 \tag{14}$$

**Osservazione** Si noti che così facendo si assume implicitamente che un individuo che ha una patologia  $a$  e poi si ammala di  $b$ , con  $a$  più grave di  $b$ , può morire di  $b$  nell'anno dell'insorgenza di  $b$ , ma negli seguenti viene descritta come un individuo malato solo di  $a$ , e quindi non può morire di  $b$ . Questa approssimazione è una debolezza del modello ma è necessaria per evitare che la calibrazione del modello diventi troppo complessa. Al contempo, il modello tiene conto della mortalità istantanea per infarto e ictus anche in individui con altre patologie.

Descriviamo adesso come sono stati trovati i parametri di mortalità per cause diverse dalle traccianti. Indichiamo queste morti con l'indice  $oc$  (other causes). Non definiamo una classe di malattia per  $oc$ , si suppone cioè che si muoia per altre cause arrivando direttamente dagli altri stati. Si assume che anche le altre cause siano correlate ai fattori di rischio, e si definisce quindi un rischio relativo per tutte le morti escluse le traccianti, ricavati a partire da [9]. La mortalità per altre cause è ottenuta tramite un ragionamento analogo ai precedenti. Innanzitutto si ottengono le morti per altre cause dal GBD, sottraendo alle morti totale le morti per le 4 traccianti, i.e.,

$$M_{oc}^{e,g} = M_{tot}^{e,g} - \sum_m M_m^{e,g}. \tag{15}$$

Siano:

- $\gamma^{e,g}$  la mortalità per altre cause di un non fumatore attivo sano di una età  $e$  e genere  $g$  ;
- $\chi_m$  un coefficiente moltiplicativo che correla la mortalità per altre cause alla presenza di malattie  $m$ , assunto indipendente da genere e età.

La mortalità per altre cause di un individuo con patologia dominante  $m$  ed esposizione al fumo  $f$  è:

$$\gamma_{f,a,m}^{e,g} = \gamma^{e,g} \cdot RR_{f,a,oc}^{e,g} \cdot \chi_m. \tag{16}$$

Le morti per altre cause sono quindi

$$E[M_{oc}^{e,g}] = \sum_{sm} \sum_f \sum_a P_{f,a,sm}^{e,g} \gamma_{f,a,m}^{e,g}. \tag{17}$$

dove nuovamente  $n$  identifica la malattia dominante dello stato  $sm$ . Noti i fattori di rischio  $RR_{f,a,oc}^{e,g}$  e i fattori di correlazione  $X$ , utilizzando (16) e le morti per altre cause da (15), otteniamo  $\gamma^{e,g}$  e di conseguenza i  $Y_{f,a,m}^{e,g}$  per ogni età e genere. In particolare,

$$\gamma^{e,g} = \frac{M_{oc}^{e,g}}{\sum_{sm} \square \sum_f \square \sum_a \square P_{f,a,sm}^{e,g} RR_{f,a,oc}^{e,g} X_n} \quad (18)$$

### 3.4 Output del modello

Tramite questo modello si può simulare l'andamento della popolazione con un orizzonte temporale arbitrario. In particolare, le grandezze di interesse sono le seguenti:

- Gli YLL (years life lost) sono calcolati a partire dalle aspettative di vita Istat. Sia  $v_g(e)$  l'aspettativa di vita di un individuo di età  $e$  e genere  $g$ . Per un individuo di genere  $g$  che muore ad età  $e$ , gli anni di vita persi sono

$$v_g(e) - e.$$

- Gli YLD sono calcolati utilizzando i pesi del GBD. Ad ogni individuo di età  $e$  e genere  $g$  con patologia  $m$  si associa un peso  $w_m^{e,g}$ . Tali pesi possono essere quindi ricavati a partire dalla formula

$$YLD_{m1}^{e,g} = w_{m1}^{e,g} \cdot P_{m1}^{e,g} \quad (19)$$

Ricavando dal GBD prevalenze e YLD e invertendo la formula si ricavano i pesi. Si noti che, nonostante il modello non ammetta stati con più di una patologia, nel conteggio degli YLD si possono considerare gli YLD di più patologie in contemporanea.

- I DALYs sono la somma di YLL e YLD.

### 3.5 Intervento di prevenzione

Il guadagno relativo ad un intervento di prevenzione si misura in DALYs. In particolare, vogliamo stimare la differenza di DALYs prodotti tra uno scenario baseline e uno scenario di prevenzione. Gli interventi di prevenzione sono descritti da un'efficacia teorica e una fascia target della popolazione, generalmente identificata dall'età. Si suddividono gli interventi in due categorie:

- interventi che agiscono sull'inizializzazione della popolazione, diminuendo di una certa frazione la prevalenza dei fattori di rischio nella popolazione target (un esempio è il consueing);
- interventi che agiscono sia sull'inizializzazione che sull'input di nuovi individui che entra nella coorte ad ogni anno (ad esempio l'aumento del prezzo delle sigarette, che inibisce i giovanissimi dall'iniziare a fumare)

## Bibliografia

[1] Nana K Anokye, Joanne Lord, and Julia Fox-Rushby. Is brief advice in primary care a cost-effective way to promote physical activity? *British journal of sports medicine*, 48(3):202–206, 2014.

[2] Adam DM Briggs, Linda J Cobiac, Jane Wolstenholme, and Peter Scarborough. Primetime ce: a multistate life table model for estimating the cost-effectiveness of interventions affecting diet and physical activity. *BMC health services research*, 19(1):1–19, 2019.

[3] Rob Carter, Marj Moodie, Alison Markwick, Anne Magnus, Theo Vos, Boyd Swinburn, and Michele M Haby. Assessing cost-effectiveness in obesity (ace-obesity): an overview of the ace approach, economic methods and cost results. *BMC public health*, 9(1):1–11, 2009.

[4] Rudolf T Hoogenveen, Pieter HM van Baal, Hendriek C Boshuizen, and Talitha L Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost effectiveness and resource allocation*, 6(1):1, 2008.

[5] Susan F Hurley and Jane P Matthews. The quit benefits model: a markov model for assessing the health benefits and health care cost savings of quitting smoking. *Cost effectiveness and resource allocation*, 5(1):1–20, 2007.

[6] Peter R Killeen. Markov model of smoking cessation. *Proceedings of the National Academy of Sciences*, 108(Supplement 3):15549–15556, 2011.

[7] David Levy, Silvano Gallus, Kenneth Blackman, Giulia Carreras, Carlo La Vecchia, and Giuseppe Gorini. Italy simsmoke: the effect of tobacco control policies on smoking prevalence and smoking attributable deaths in italy. *BMC Public Health*, 12(1):1–13, 2012.

[8] Stefan K Lhachimi, Wilma J Nusselder, Henriette A Smit, Pieter Van Baal, Paolo Baili, Kathleen Bennett, Esteve Fernández, Margarete C Kulik, Tim Lobstein, Joceline Pomerleau, et al. Dynamo-hia—a dynamic modeling tool for generic health impact assessments. *PloS one*, 7(5):e33317, 2012.

[9] Michael J Thun, Dena G Myers, Cathy Day-Lally, Mohan M Namboodiri, Eugenia E Calle, W Dana Flanders, Stacy L Adams, Clark W Heath, et al. Age and the exposure-response relationships between cigarette smoking and premature death in cancer prevention study ii. *Changes in cigarette-related disease risks and their implications for prevention and control*, 383:413, 1997.

## Appendice A: distribuzione ex fumatori

La probabilità che un individuo sia ex fumatori da  $i + 1$  anni è uguale alla probabilità che fosse ex fumatore da  $i$  anni all'anno precedente, moltiplicato per la probabilità che non abbia ricominciato a fumare all'anno  $i$ -esimo. Assumendo che il numero di individui che ha smesso di fumare negli anni passati sia costante, vale la seguente formula per la prevalenza di ex fumatori:

$$P_{ex(i+1)}^{e,g} = P_{ex_i}^{e,g} \cdot (1 - ABg^{-12i \cdot B}). \quad (20)$$



Iterando questa formula,

$$P_{ex_i}^{e,g} = P_{ex_1}^{e,g} \prod_{j=1}^{i-1} (1 - AB e^{-12j \cdot B}). \quad (21)$$

Di conseguenza, chiamando  $P_{ex}^{e,g}$  la totalità degli ex fumatori di una certa età e genere,

$$\begin{aligned} P_{ex}^{e,g} &= \sum_{i=1}^{i_{max}} P_{ex_i}^{e,g} \\ &= \sum_{i=1}^{i_{max}} P_{ex_1}^{e,g} \prod_{j=1}^{i-1} (1 - AB e^{-12j \cdot B}), \end{aligned} \quad (22)$$

dove  $i_{max} = e - 18$  (si assume che nessuno smetta di fumare prima dei 18 anni). Invertendo la formula si ottiene  $P_{ex_1}^{e,g}$  tramite

$$P_{ex_1}^{e,g} = \frac{P_{ex}^{e,g}}{\sum_{i=1}^{e-18} \prod_{j=1}^{i-1} (1 - AB e^{-12j \cdot B})}. \quad (23)$$

Inserendo questa in (21), si ottengono tutte le prevalenze di ex fumatori.