

Università degli studi di Milano-Bicocca

Big Data in Public Health

Progetto Finale

ANALISI DEL TASSO DI INCIDENZA DEL TUMORE POLMONARE IN EUROPA

Autori:

Beatrice Fumagalli
Luca Gabellini
Matteo Provasi
Stefano Rola
Pierluigi Tagliabue

784549
777786
782922
790383
835211

b.fumagalli9@campus.unimib.it
l.gabellini1@campus.unimib.it
m.provasi2@campus.unimib.it
s.rola@campus.unimib.it
p.tagliabue7@campus.unimib.it



Introduzione

Il **carcinoma del polmone** o **tumore polmonare** è la neoplasia con il maggior tasso di incidenza e di mortalità nel mondo (1,35 milioni di nuovi casi all'anno e 1,18 milioni di morti), con la massima frequenza negli Stati Uniti d'America e in Europa¹. Infatti, esso rappresenta l'11% di tutte le nuove diagnosi di tumore nella popolazione generale (13% nei maschi, 6% nelle femmine). Colpisce prevalentemente le popolazioni ubicate nei paesi più industrializzati ed è invece meno comune nei paesi in via di sviluppo.

Esistono diversi fattori che portano allo sviluppo di questo tumore. Il **fumo di tabacco** rimane la causa di gran lunga più importante dell'epidemia di casi di carcinoma polmonare osservato in tutti i paesi dove questa abitudine si è diffusa. In particolare, il **fumo attivo** è responsabile di circa l'85% dei casi di tumore polmonare e una quota ulteriore, meno facilmente quantificabile, è attribuibile all'esposizione a fumo ambientale (**fumo passivo**). Altre cause in successione sono lo **smog** e l'**inquinamento atmosferico** prodotto dalla combustione di derivati del petrolio, quindi le lavorazioni che comportano l'uso di metalli particolari (nichel, cromo, ecc.) e di sostanze radioattive. Molte sostanze di origine lavorativa sono riconosciute essere carcinogeni polmonari, anche se sono meno rilevanti del tabacco, dal punto di vista epidemiologico. A rischio risultano, pertanto, i lavoratori del catrame, delle ferrovie e delle raffinerie, gli autisti di camion e autobus, i vigili urbani. Queste sostanze chimiche hanno la prerogativa di permanere a lungo nei polmoni per la loro stabilità e per la loro difficile eliminazione. Infine, ai fattori causali si affiancano i fattori di rischio, ossia le forme che facilitano l'insorgenza del tumore, come le **malattie croniche irritative** preesistenti dei polmoni, quali tubercolosi, bronchite cronica e bronchiectasia.

Si tratta di una delle forme tumorali a più alta letalità: la sopravvivenza a 5 anni dei pazienti con tumore del polmone è del 15% circa. Ciò è dovuto al fatto che la diagnosi di tumore del polmone è spesso tardiva. Solitamente vengono rilevate le metastasi (che generalmente danno sintomatologia), e poi si arriva alla diagnosi di tumore al polmone, ormai in uno stadio avanzato.



Figura 1 - tumore polmonare

¹Cancer Research UK, Commonly diagnosed cancers worldwide, <https://web.archive.org/web/20080110233828/http://info.cancerresearchuk.org/cancerstats/geographic/world/moncancers/>

Obiettivo

L'obiettivo del progetto è quello di calcolare l'incidenza del tumore ai polmoni (codice identificativo C33-34) in diversi paesi europei per analizzarne l'andamento nel tempo, mettendo in luce eventuali *trend* o andamenti di interesse. Si effettuerà un'analisi critica per valutare eventuali differenze significative dei tassi tra nazioni o anni di rilevazione distinti.

Dataset utilizzati – variabili

1. Europe.csv

Continent	Codice identificativo del continente. 5 = Europa
Registry	Codice identificativo della Regione del Paese
Name	Paese e, talvolta, anche provincia
YEAR1	Primo anno di disponibilità dei dati
YEAR2	Ultimo anno di disponibilità dei dati

2. Europe_Cases.csv

REGISTRY	Codice identificativo della Regione del paese
YEAR	Anno di riferimento dei casi
SEX	Variabile booleana indicante il genere (1 m ,2 f)
CANCER	Codice identificativo della tipologia di tumore
TOTAL	Totale della popolazione afflitta da tumore
N0_4, N5_9, N10_14, N15_19, N20_24, N25_29, N30_34, N35_39, N40_44, N45_49, N50_54, N55_59, N60_64, N65_69, N70_74, N75_79, N80_84, N85+, N_UNK	Numero di casi di tumore stratificati per 18 classi d'età

3. Europe_Pops.csv

REGISTRY	Codice identificativo della Regione del paese
YEAR	Anno di riferimento della popolazione
SEX	Variabile booleana indicante il genere (1 m ,2 f)
TOTAL	Popolazione totale
P0_4, P5_9, P10_14, P15_19, P20_24, P25_29, P30_34, P35_39, P40_44, P45_49, P50_54, P55_59, P60_64, P65_69, P70_74, P75_79, P80_84, P85+	Popolazione stratificata per 18 classi d'età

4. std_world_pop18classes.csv

Age group	Gruppi di età
World	Popolazione standard mondiale

Dati

I dataset sono stati scaricati dal sito ufficiale della IARC², un'agenzia mondiale che colleziona dati relativi ai casi di cancro e che coordina e conduce ricerche in questo ambito. I dati scaricati sono relativi alla collezione più generale in quanto l'obiettivo è quello di analizzare una tipologia di tumore senza andare nel dettaglio delle diverse classificazioni o stadi di evoluzione della malattia.

Valutate le diverse possibilità si è scelto di analizzare il tumore al polmone nell'area dell'Europa per i seguenti motivi:

- Come già citato il tumore al polmone è una delle patologie più comuni e quindi oltre ad avere un importante interesse sociale permette di avere un discreto numero di casi per condurre le analisi.
- Per questo specifico cancro i dati sono disponibili, per alcuni stati, già a partire dagli anni '70. Ciò permette di coprire al meglio una delle componenti dell'obiettivo, ovvero valutare l'andamento storico dell'incidenza della patologia.
- L'Europa, dopo gli Stati Uniti d'America, è il continente con l'incidenza maggiore per il tumore al polmone.

I dati sono presentati ad un livello di stratificazione differenziato a seconda del paese di riferimento: per l'Italia sono principalmente a livello di provincia, ma esistono anche eccezioni in cui l'aggregazione è fatta ad un livello intermedio fra le province e le regioni. Nel Regno Unito al contrario i dati sono raccolti a livello di regioni.

I file relativi ai casi e alla popolazione presentano un codice identificativo che specifica unicamente la "regione" di un paese: questa variabile è stata utilizzata per fare un *merge* fra i due *dataset* in modo da averne uno unico contenente sia il numero di casi che la popolazione. È presente sia il totale dei casi sia il totale della popolazione oltre alla stratificazione per 18 classi d'età, partendo dalla classe 0-4 anni fino ad arrivare a quella di 85 o più.

I file scaricati non specificano se la popolazione rappresentata è espressa in persone o persone-anno. Controllando manualmente si è trovato che i dati sono relativi al numero di persone, in particolare per quanto riguarda l'Italia i dati relativi alla popolazione per province sono stati raccolti dal sito dell'ISTAT³. Per poter effettuare appropriatamente i calcoli dei valori di incidenza si trattano questi valori come persone-anno, supponendo che ogni persona sia osservata per un singolo anno.

Oltre ai dati di popolazione e ai casi non si hanno ulteriori variabili che possono risultare confondenti od offrire ulteriori informazioni per le analisi.

Da conoscenze pregresse si è al corrente che l'età è un fattore molto rilevante per il tumore ai polmoni (l'età più comune in cui si sviluppa la patologia è intorno ai 70 anni⁴). Si è reso necessario quindi effettuare una standardizzazione, sia per eliminare l'effetto della distribuzione della popolazione fra i diversi paesi europei (effetto limitato in quanto si trattano paesi allo stesso livello di industrializzazione) sia per rendere possibile un confronto con i dati passati di una stessa nazione.

La mancanza di dati riferiti a tutte le "regioni" porta ad inevitabili approssimazioni nelle analisi. Infatti, fatta eccezione per Regno Unito, per tutti gli altri paesi sono disponibili osservazioni solo per alcune province.

Dovendo raggruppare i dati a livello di nazione, l'analisi dei risultati porterà ad intendere paese anche quei casi nei quali si è considerata solo una piccola parte della popolazione. Questa assunzione è valida solamente nel caso in cui ci sia una distribuzione uniforme dell'incidenza dei tumori e dell'aspettativa di vita in tutte le regioni, anche di quelle di cui non si hanno a disposizione dati. La condizione in Italia non è strettamente verificata in alcuni

² IARC, <http://ci5.iarc.fr/Ci5plus/Pages/download.aspx>

³ ISTAT, http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1

⁴ National Cancer Institute, Cancer Statistics, <https://seer.cancer.gov/statfacts/html/lungb.html>

Metodi di analisi e risultati

$$TI_i = 100000 \cdot \frac{\text{casi di tumore nella classe d'età } i}{\text{anni/persona nella classe d'età } i}$$
$$TI_{tot} = \sum_{i=1}^n TI_i \cdot \frac{pop_i}{\sum_{i=1}^n pop_i}$$

⁵ ANSA, In Trentino si vive 3 anni più che in Campania, gap Nord-Sud, 2018, http://www.ansa.it/canale_salutebenessere/notizie/sanita/2018/02/19/in-trentino-si-vive-3-anni-piu-che-in-campania-gap-nord-sud_00eb6200-6bfe-4934-be23-da94c265942d.html

Per meglio evidenziare graficamente l'andamento dei tassi si è deciso di selezionare un sottoinsieme di paesi: le prime tre nazioni (Norvegia, Svizzera e Irlanda) per indice di sviluppo umano (HDI), i cinque principali paesi europei (Francia, Germania, Italia, Spagna, Regno Unito) e gli ultimi tre in Europa, fra quelli disponibili, sempre per HDI (Croazia, Bulgaria e Bielorussia).

L'**HDI** è l'indice comparativo dello sviluppo dei vari paesi calcolato tenendo conto delle diverse aspettative di vita, livello di istruzione e reddito nazionale lordo pro-capite. Viene considerata una misura di riferimento per stimare il benessere di un paese. Pertanto, è stato costruito un campione di nazioni il più possibile rappresentativo secondo il sopracitato indice. I dati sono stati scaricati dal sito dell'HNDP⁶.

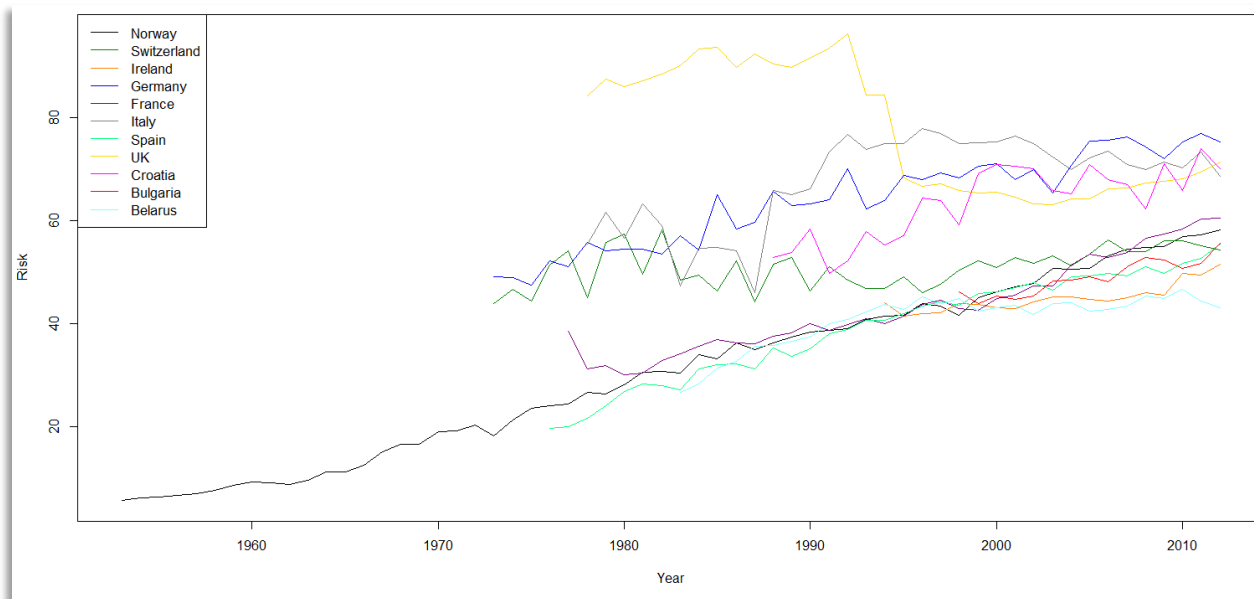


Figura 3 - Non-Standardized Incidence Rate (per 100000) degli 11 paesi selezionati

Insieme ai tassi grezzi sono stati calcolati anche quelli standardizzati per età, in modo da eliminare l'eventuale differenza strutturale tra le popolazioni d'interesse. In particolare, è stata effettuata una **standardizzazione diretta** prendendo come riferimento una popolazione standard di 100 mila individui ripartiti nelle stesse 18 classi d'età dei dati analizzati.

Caricato il file con le popolazioni standard si sono moltiplicate queste, ripartite nelle diverse classi d'età, per i valori TI_i calcolati precedentemente. Sommando i valori per le singole classi d'età si trova l'**Age Standardized Rate (ASR)**. Un tasso standardizzato per età è una misura sommaria; è il tasso che una popolazione avrebbe se avesse una struttura per età standard. La standardizzazione è necessaria quando si confrontano diverse popolazioni che differiscono rispetto alla struttura per età, perché l'età ha una forte influenza sul rischio di tumore. Viene calcolato come:

$$ASR = \sum_{i=1}^n \frac{\text{casi di tumore nella classe di età } i}{\text{popolazione nella classe di età } i} \cdot \frac{\text{popolazione std nella classe di età } i}{\text{popolazione std totale}}$$

⁶ UNDP, Human Development Report, <http://hdr.undp.org/en/data>

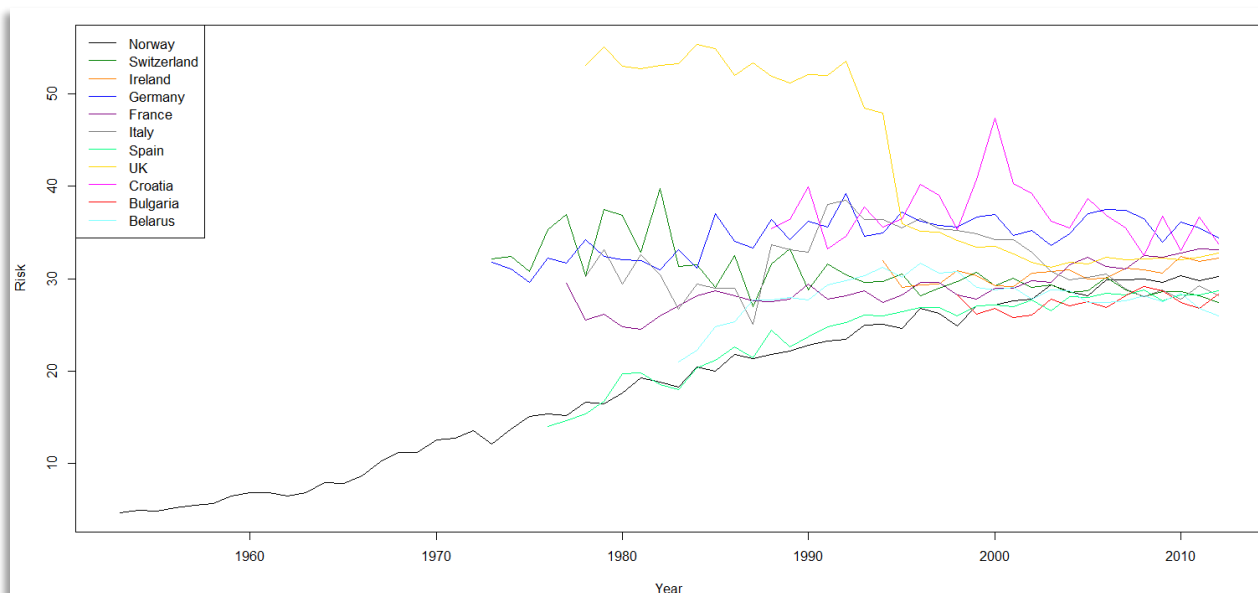


Figura 4 – Standardized Incidence Rate (per 100.000) degli 11 paesi selezionati

I risultati ottenuti con la standardizzazione sono affini a quelli raggiunti con i tassi crudi, in quanto tutti i paesi europei sono industrializzati e le rispettive piramidi delle età sono generalmente molto simili tra loro. In Fig. 5 è rappresentata la piramide dell'età di tre paesi: Norvegia (primo per HDI), Italia (uno dei 5 più grandi paesi europei) e Bielorussia (ultimo per HDI).

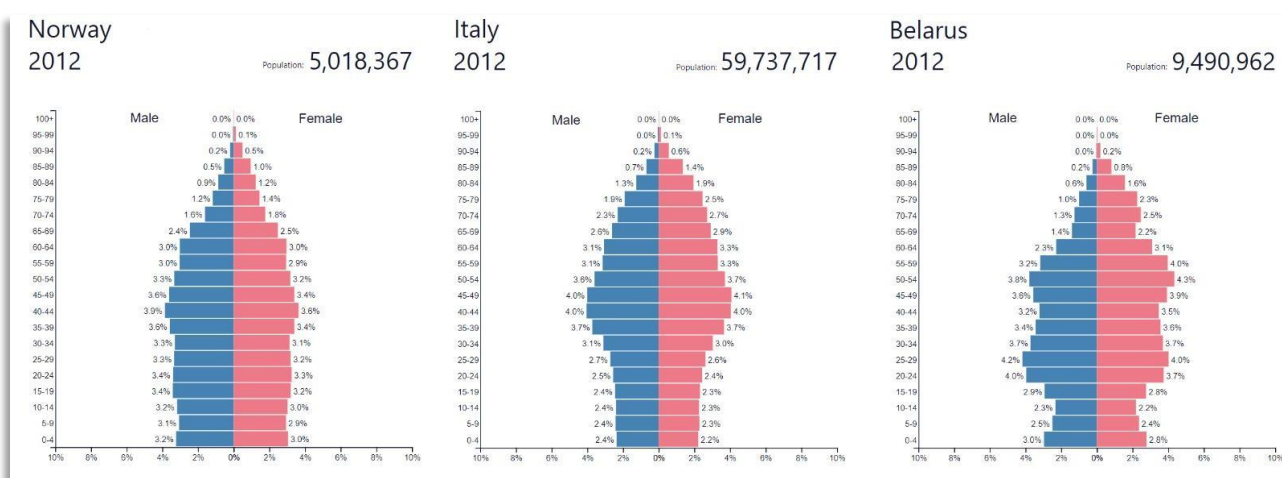


Figura 5 - Population Pyramid

Prendendo in considerazione la Fig. 4, salta all'occhio il brusco calo del tasso relativo al Regno Unito: negli anni '80 e '90 lo UK presentava un'incidenza molto superiore rispetto agli altri paesi, mentre negli anni recenti la misura si attesta su posizioni medie. Il crollo dell'incidenza in questo caso è molto probabilmente dovuto all'introduzione di politiche antifumo già dagli anni '80, con largo anticipo rispetto a molti altri paesi europei⁷. Paesi come Francia, Spagna e Norvegia presentano un costante aumento del tasso negli anni, mentre altri presentano un trend molto altalenante (Croazia, Svizzera e Germania).

⁷ Centre for Public Impact, Smoking ban in England and Wales, <https://www.centreforpublicimpact.org/case-study/smoking-ban-england-and-wales/>

Una misura utile per fare confronti e quantificare numericamente la variazione di un tasso in diverse unità temporali è l' **Incidence Rate Ratio**⁸, ovvero un rapporto tra tassi di incidenza:

$$IRR = \frac{\text{tasso di incidenza al tempo } (t + 1)}{\text{tasso di incidenza al tempo } t}$$

Lo **Standardized Incidence Ratio** è una misura simile alla precedente ma riferita al rapporto tra tassi standardizzati.

Per gli 11 paesi precedentemente considerati sono stati selezionati come anni di riferimento i due estremi in comune: come minimo il 1998 (l'Italia non ha dati precedenti) e come anno finale il 2012 (ultimo anno disponibile per tutti i paesi).

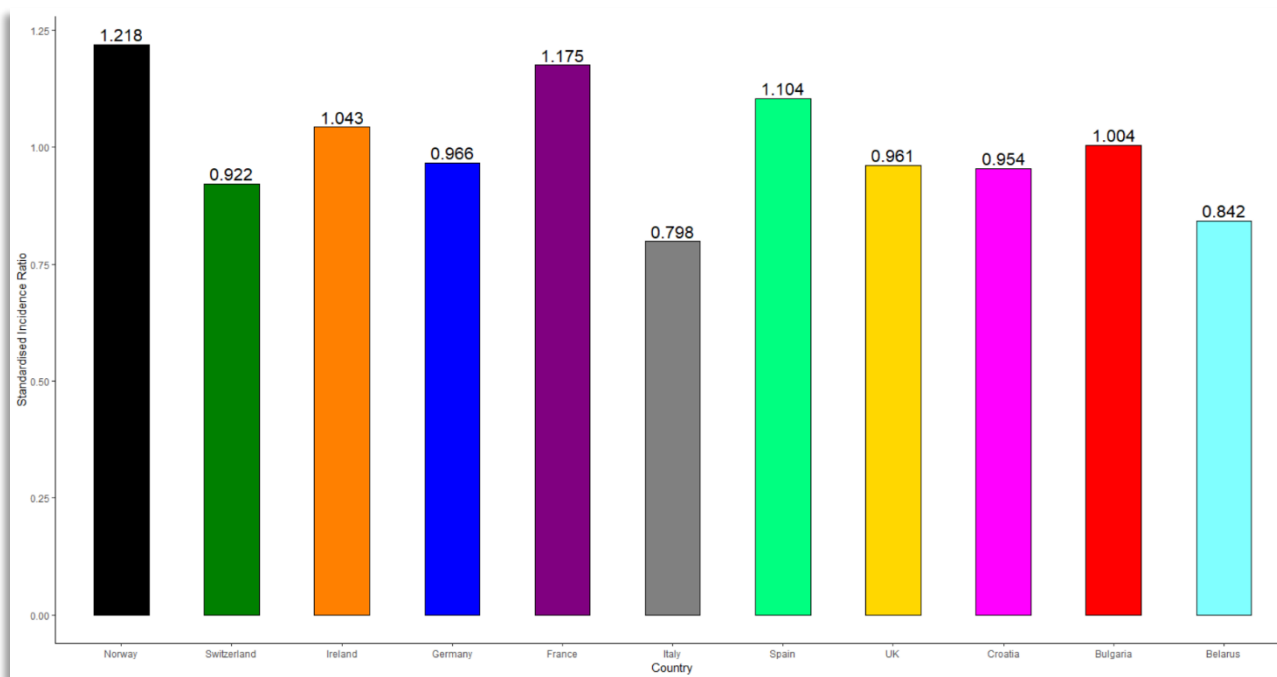


Figura 6 - Bar Chart SIR degli 11 paesi selezionati

La Fig. 6 conferma quanto detto prima sui tassi: la Norvegia in particolare mostra l'aumento maggiore con la Francia; l'Italia presenta il calo più rilevante mentre per paesi come Bulgaria e Germania il tasso calcolato nel 1998 e quello del 2012 risultano pressoché identici.

⁸ CTSPedia, <https://www.ctspedia.org/do/view/CTSpedia/RateRatio>

Country	SIR
Norway	1.2183473
Switzerland	0.9218843
Ireland	1.0426676
Germany	0.9656590
France	1.1752097
Italy	0.7984559
Spain	1.1038260
UK	0.9606694
Croatia	0.9541490
Bulgaria	1.0035469
Belarus	0.8415875

Figura 7 -SIR degli 11 paesi selezionati

Country	ASR
Norway	0.0003025639
Switzerland	0.0002739987
Ireland	0.0003219131
Germany	0.0003440840
France	0.0003313139
Italy	0.0002815684
Spain	0.0002866442
UK	0.0003275521
Croatia	0.0003373858
Bulgaria	0.0002832768
Belarus	0.0002592285

Figura 8 - ASR del 2012 degli 11 paesi selezionati

L'aggiustamento per età è stato poi condotto utilizzando un ulteriore metodo, ossia il **modello di regressione di Poisson**.

Fittando un modello di questo tipo, con un'unica variabile esplicativa contenente i valori per il confronto fra tassi nell'anno 2012 e 1998 si ottiene un **Incidence Rate Ratio (IRR)** di $\exp(0.3378) = 1.40$ per la Norvegia. Nel caso della Norvegia, il rischio di incidenza di tumore al polmone è significativamente maggiore (per $\alpha = 0.05$) nell'anno 2012 rispetto al 1998.

Utilizzando la formula $(\exp(\beta) - 1) \cdot 100$ troviamo la variazione percentuale del tasso, aumentato del 40% in 14 anni.

```
[1] "----- Poisson summary for Norway -----"

Call:
glm(formula = cases ~ year2012, family = "poisson", data = temp_df[!is.na(temp_df),
], offset = log(pop))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-19.44  -16.57  -12.25   14.36   28.74

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.78508     0.02329  -334.22 <0.0000000000000002 ***
year2012      0.33783     0.02974   11.36 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 11006  on 35  degrees of freedom
Residual deviance: 10874  on 34  degrees of freedom
(7560 observations deleted due to missingness)
AIC: 11047

Number of Fisher Scoring iterations: 6

              Incidence rate ratio      2.5 %      97.5 %
(Intercept)    0.0004158937  0.0003973331  0.0004353212
year2012        1.4018955154  1.3225210267  1.4860338674
```

Figura 9 – IRR (Poisson) tra il 2012 e il 1998 per la Norvegia

Costruendo un nuovo modello di Poisson per valutare i contrasti tra le diverse fasce di età, prendendo come classe di riferimento l'age group 70-74 della Norvegia, si sono ottenuti i risultati raffigurati in Fig. 10.

Per le tutte le fasce di età (escluse le prime tre) il valore della statistica test è negativo: il tasso di incidenza stimato è significativamente inferiore rispetto a quello della classe di riferimento.

```
Call:
glm(formula = cases ~ age_group, family = "poisson", data = temp_df[temp_df$year ==
  2012, ])

Deviance Residuals:
 [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    6.22258    0.04454 139.696 < 0.0000000000000002 ***
age_group0_4   -29.52516  69653.80045    0.000    0.999662
age_group10_14 -29.52516  69653.80072    0.000    0.999662
age_group15_19  -5.52943    0.70851  -7.804  0.00000000000000598 ***
age_group20_24  -6.22258    1.00099  -6.216  0.00000000050865062 ***
age_group25_29  -6.22258    1.00099  -6.216  0.00000000050865062 ***
age_group30_34  -4.43082    0.41067 -10.789 < 0.0000000000000002 ***
age_group35_39  -4.14313    0.35635 -11.627 < 0.0000000000000002 ***
age_group40_44  -3.04452    0.20893 -14.572 < 0.0000000000000002 ***
age_group45_49  -2.17953    0.13974 -15.597 < 0.0000000000000002 ***
age_group5_9    -29.52516  69653.80071    0.000    0.999662
age_group50_54  -1.47764    0.10334 -14.298 < 0.0000000000000002 ***
age_group55_59  -0.85194    0.08146 -10.459 < 0.0000000000000002 ***
age_group60_64  -0.24369    0.06720  -3.626    0.000287 ***
age_group65_69   0.09639    0.06153   1.567    0.117212
age_group75_79  -0.17520    0.06594  -2.657    0.007884 **
age_group80_84  -0.29565    0.06820  -4.335  0.00001455908811376 ***
age_group85     -0.72132    0.07788  -9.262 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 4545.67907369416844  on 17  degrees of freedom
Residual deviance:  0.00000000045488  on  0  degrees of freedom
AIC: 121.94

Number of Fisher Scoring iterations: 21
```

Figura 10 - Modello di regressione di Poisson per la Norvegia stratificato per classi di età nel 2012

In questo caso, dato che le stime dei coefficienti sono negative, conviene ribaltare l'esponenziale per rendere più facile l'interpretazione. Per fare un esempio, la classe 70-74 ha un tasso di incidenza circa $\exp(-6.22258) - 1 \sim 500$ volte più grande rispetto alla fascia 20-24.

Si ricorda che le stime a cui è corrisposto uno standard error elevato (Z-value molto piccolo) sono statisticamente non significative e quindi poco attendibili.

Viene infine proposto un modello logistico utile per quantificare l'evoluzione del tasso di incidenza negli anni, riferendoci sempre al caso della Norvegia:

Coefficients:					Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)
as.factor(year)1998	1.76977	0.07479	23.662	< 0.0000000000000002 ***	age_group0_4	-22.21924	252.46916	-0.088	0.929871
as.factor(year)1999	1.85056	0.07450	24.841	< 0.0000000000000002 ***	age_group10_14	-7.87453	0.31634	-24.893	< 0.0000000000000002 ***
as.factor(year)2000	1.87693	0.07439	25.231	< 0.0000000000000002 ***	age_group15_19	-6.99746	0.20430	-34.251	< 0.0000000000000002 ***
as.factor(year)2001	1.89803	0.07430	25.544	< 0.0000000000000002 ***	age_group20_24	-6.51242	0.16035	-40.614	< 0.0000000000000002 ***
as.factor(year)2002	1.90999	0.07424	25.726	< 0.0000000000000002 ***	age_group25_29	-6.01823	0.12432	-48.409	< 0.0000000000000002 ***
as.factor(year)2003	1.96644	0.07405	26.556	< 0.0000000000000002 ***	age_group30_34	-5.32051	0.08745	-60.844	< 0.0000000000000002 ***
as.factor(year)2004	1.95827	0.07404	26.449	< 0.0000000000000002 ***	age_group35_39	-4.16214	0.04969	-83.761	< 0.0000000000000002 ***
as.factor(year)2005	1.95499	0.07401	26.415	< 0.0000000000000002 ***	age_group40_44	-3.20462	0.03206	-99.942	< 0.0000000000000002 ***
as.factor(year)2006	1.99528	0.07386	27.015	< 0.0000000000000002 ***	age_group45_49	-2.27398	0.02172	-104.712	< 0.0000000000000002 ***
as.factor(year)2007	2.01586	0.07376	27.329	< 0.0000000000000002 ***	age_group5_9	-9.47516	0.70716	-13.399	< 0.0000000000000002 ***
as.factor(year)2008	2.01732	0.07372	27.366	< 0.0000000000000002 ***	age_group50_54	-1.52973	0.01656	-92.357	< 0.0000000000000002 ***
as.factor(year)2009	2.01691	0.07367	27.377	< 0.0000000000000002 ***	age_group55_59	-0.93718	0.01387	-67.550	< 0.0000000000000002 ***
as.factor(year)2010	2.04463	0.07355	27.798	< 0.0000000000000002 ***	age_group60_64	-0.48088	0.01248	-38.518	< 0.0000000000000002 ***
as.factor(year)2011	2.04270	0.07351	27.788	< 0.0000000000000002 ***	age_group65_69	-0.17423	0.01194	-14.593	< 0.0000000000000002 ***
as.factor(year)2012	2.05393	0.07344	27.969	< 0.0000000000000002 ***	age_group70_74	0.01724	0.01252	1.377	0.168583
					age_group80_84	-0.16668	0.01477	-11.282	< 0.0000000000000002 ***
					age_group85	-0.63523	0.01937	-32.801	< 0.0000000000000002 ***

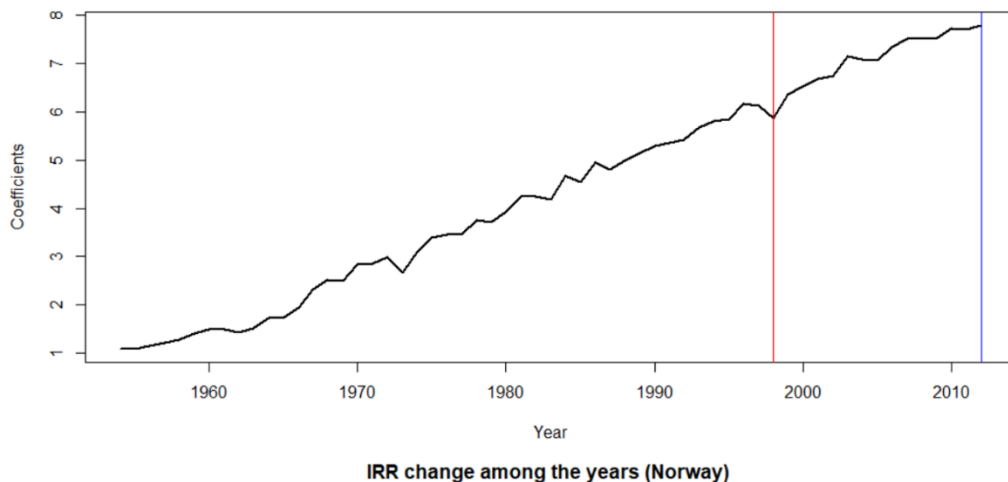


Figura 11 – IRR (Poisson) tra il 1953 e 2012 e per classi d'età in Norvegia

Limiti dello studio

- Il primo e l'ultimo anno di disponibilità dei dati di interesse non sono omogenei per tutti i paesi. Ciò ha portato a selezionare come anni di riferimento il primo e l'ultimo comuni ai paesi presi in considerazione, perdendo così una parte aggiuntiva di dati a disposizione per altre nazioni.
- I dati relativi alla Slovenia sono stati eliminati in quanto nel *dataset* non erano presenti le informazioni relative alla popolazione con età superiore agli 85 anni.
- Si è persa parte della specificità dell'analisi a causa della mancanza nel *dataset* di molte Regioni e Province, disponibili solamente per alcuni paesi.

Conclusioni

Analizzando i risultati ottenuti e sopra descritti, si può notare che il paese che presenta il più alto rapporto tra tassi calcolati tra gli anni 2012 e 1998 è la Norvegia, con un SIR pari a 1.218. Da evidenziare che la Norvegia è il primo paese in assoluto per indice di sviluppo umano (HDI), che, come precedentemente spiegato, rappresenta lo strumento standard per misurare il benessere di un paese. A seguire si trovano i principali paesi europei, rispettivamente Francia, Spagna, Irlanda e Germania. Anche quest'ultimi sono paesi fortemente industrializzati. I paesi con un SIR minore sono gli stessi che sono in coda alla lista dei paesi per indice di sviluppo, ossia Bulgaria, Croazia e Bielorussia. Da questa analisi emerge quindi una correlazione positiva tra HDI e SIR, al contrario di quanto ci si possa aspettare.

Un caso straordinario è rappresentato dall'Italia che, nonostante sia tra i principali paesi europei e abbia un medio grado di indice di sviluppo, presenta il rapporto di incidenza standardizzato minore tra tutto il campione analizzato.

Prendendo in considerazione il Tasso standardizzato per età (ASR) dell'anno 2012, ossia l'anno massimo disponibile per gli 11 paesi selezionati ai fini dell'analisi, si può notare che la Norvegia, al primo posto per SIR, scala alla quinta posizione, preceduta in ordine, da Germania, Croazia, Francia e Irlanda. Ciò significa che la popolazione norvegese, nonostante abbia subito un forte incremento del tasso di incidenza negli anni, presenti comunque un'insorgenza della patologia minore rispetto ai paesi sopra citati. Agli ultimi posti si trovano Italia, Svizzera e, per ultima, Bielorussia.

Sviluppi futuri

Sarebbe stato interessante poter valutare l'andamento del tasso di incidenza del tumore polmonare in associazione alle principali cause del medesimo, ossia il fumo di tabacco, lo smog e l'inquinamento atmosferico, informazioni non disponibili nel dataset considerato. Inoltre sarebbe utile ai fini dell'analisi reperire ulteriori dati circa i paesi europei, come ad esempio un maggior numero di province rispetto quelle disponibili ed estendere il campione per condurre un'analisi su scala globale anziché europea.