

# Costruzione di indicatori statistici per l'epidemiologia

**XIX CORSO DI AGGIORNAMENTO AIRTUM PER  
OPERATORI DEI REGISTRI TUMORI**

**BARI 6-8 novembre 2019**

IRCCS Istituto Tumori

"Giovanni Paolo II"

# Definizione di indicatore

**Informazioni selezionate allo scopo di conoscere fenomeni di interesse, misurandone i cambiamenti e conseguentemente contribuendo ad orientare i processi decisionali dei diversi livelli istituzionali**

UN INDICATORE E' UNO STRUMENTO METODOLOGICO SEMPLICE COSTITUITO DA ESPRESSIONI QUANTITATIVE (CONTEGGI, PERCENTUALI, TASSI, PROPORZIONI) CHE FORNISCE UNA SINTETICA RAPPRESENTAZIONE DEL FENOMENO IN ESAME

# SCHEMA DI PROCESSO PER LA DEFINIZIONE DI UN INDICATORE

- INDIVIDUARE UN AMBITO PER LA VALUTAZIONE ED IL MONITORAGGIO
- DEFINIRE LA POPOLAZIONE OGGETTO DELLA MISURAZIONE
- DEFINIRE LA TIPOLOGIA DI EVENTO DA MISURARE
- INDIVIDUARE POSSIBILI FATTORI DI DISTORSIONE DELLA MISURAZIONE

# INDIVIDUARE UN AMBITO PER LA VALUTAZIONE ED IL MONITORAGGIO

[...]molteplici sfide per un Registro Tumori basate su una raccolta minima di dati: fornire rapidamente le giuste informazioni sulle varie epidemie di cancro attraverso i vari settori di ricerca sovrapposti, migliorando così:

➤ Salute Pubblica:

- ✓ contributo alla descrizione di frequenza e prognosi,
- ✓ ricerca eziologica,
- ✓ prevenzione,
- ✓ valutazione dello screening di massa;

➤ Salute del paziente:

- ✓ contributo alla descrizione della qualità dell'assistenza (processo e risultato),
- ✓ prognostica e qualità della vita (effetti collaterali, sopravvivenza).

# OBIETTIVI

Campo di interesse	Area di Ricerca	Misure epidemiologiche	Coorti, altre fonti di dati	Disegno di studio, sviluppo di scenari	Analisi approfondite, con obiettivi specifici
Sanità Pubblica	Frequenza (carico per il sistema)	Incidenza, prevalenza, sopravvivenza	UO patologia, cartelle cliniche, schede di dimissione, altre fonti, anagrafe	Sviluppo di scenari per prevenzione e allocazione di risorse	Quali risorse per gestire le neoplasie nei prossimi 10-15 anni? Quante evitabili?
	Etiologia e prevenzione	Analisi dei trend per età e sesso, relazione con stili di vita e rischi occupazionali	Linkage con banche dati legate ad esposizione	Analisi di coore, supporto al reclutamento di casi per studi di popolazione	Calcolo dei RR

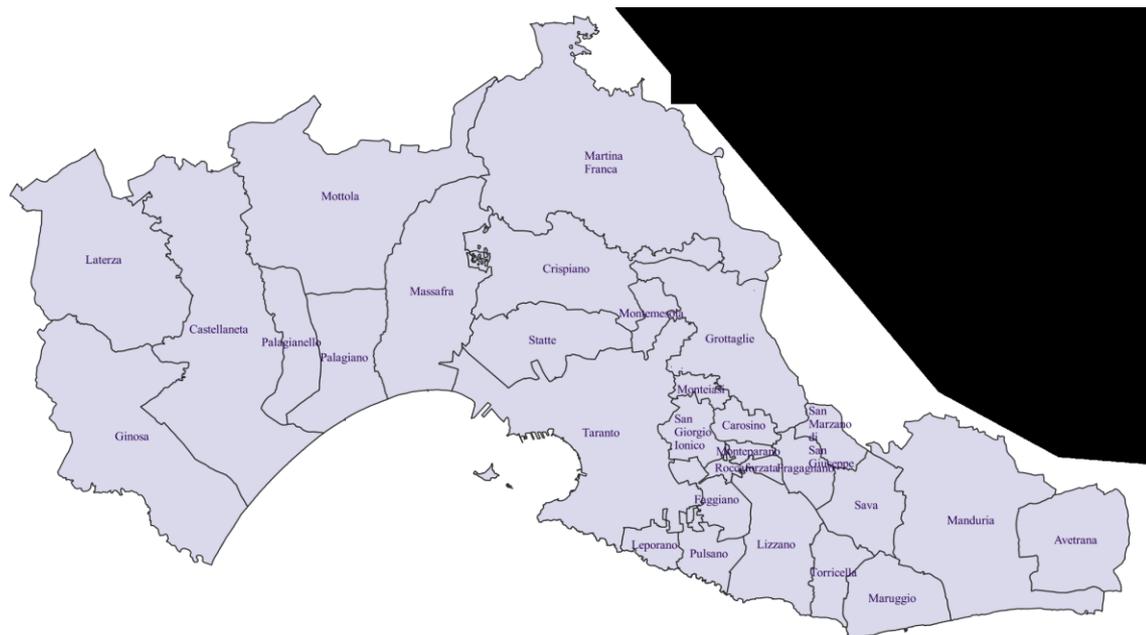
# DEFINIRE LA POPOLAZIONE OGGETTO DELLA MISURAZIONE

Si tratta in genere dei residenti dell'area geografica da cui si rilevano i dati.

Si tratta più estensivamente della regione di cui il registro è campione (quando non copre l'intera popolazione)

## AREA IN STUDIO

L'area oggetto di studio è l'intera provincia tarantina, essa si estende su una superficie di 2.467,35 km<sup>2</sup>, ha una popolazione di 586.061 abitanti (ISTAT- 1° Gennaio 2016) e si distribuisce amministrativamente su 29 comuni (Figura1).



# DEFINIRE LA TIPOLOGIA DI EVENTO DA MISURARE

E' indispensabile definire correttamente il caso, individuandolo nella banca dati secondo le codifiche o altri criteri, che siano ben dettagliati

- relativi al numero di nuovi casi di tumore maligno (casi incidenti) diagnosticati nella popolazione residente nella provincia di Taranto negli anni 2006-2011 (Fonte: Registro Tumori Taranto). Per la definizione di caso incidente è stata utilizzata la Classificazione internazionale delle malattie per l'Oncologia, terza revisione (ICD-O-3) secondo i raggruppamenti riportati in Tabella 3.

Descrizione	Morfologie	Sedi ICD-O-3	Comportamento
Tutte le neoplasie / Tutte le neoplasie con età 0-14		C00-C43;C45-C96	3
		C67	1-2
	8720-8790;9590-9989;9140	C44	3

# TASSO DI INCIDENZA

$$I = \frac{\text{nuovi casi}}{\text{popolazione a rischio}} t$$

$$I = \frac{\text{nuovi casi}}{\sum p * y}$$

## ANNI-PERSONA DI OSSERVAZIONE

### Numero di anni-persona a rischio.

E' il numero totale di anni/persona durante i quali la popolazione in studio è stata seguita, ovvero è stata esposta al rischio di sviluppare il secondo tumore. Si calcola sommando il tempo in cui ogni singolo paziente è stato osservato. Per esempio, se si segue 1 paziente per 7 anni e un altro per 5, si ottiene un totale di 12 anni-persona a rischio, lo stesso valore che si osserverebbe seguendo un solo paziente per 12 anni, o tre pazienti per 4 anni e così via.

Pz1: tempo di follow-up 7 anni

Pz2: tempo di follow-up 5 anni

Totale tempo di follow-up=7+5=12

$$\sum p * y = 1 * 7 + 1 * 5 = 12$$

# TASSO DI INCIDENZA

## Incidence

Crude IRs of rare cancers were calculated as the number of new cancers occurring in 2000-2010 divided by the population at risk (male and female also for gender-specific rare cancers) over the same period, expressed as person-years.

$$I = \frac{\text{nuovi casi}}{\sum p*y}$$

*Epidemiol Prev* 2016; 40 (1) Suppl 2:1-120

## RESULTS

In total, 339,403 tumours were included in the incidence analysis. The annual incidence rate (IR) of all 198 rare cancers in the period 2000-2010 was 147 per 100,000 per year, corresponding to about 89,000 new diagnoses in Italy each year, accounting for 25% of all cancer.

# TASSO DI PREVALENZA

$$Prev. = \frac{\text{casi presenti}}{\text{Totale della popolazione}}$$

## RARE HAEMATOLOGICAL DISEASES

17 464	RARE LYMPHOID DISEASES
3 572	ACUTE MYELOID LEUKAEMIA AND RELATED PRECURSOR NEOPLASMS
3 610	MYELOPROLIFERATIVE NEOPLASMS
2 371	MYELODYSPLASTIC SYNDROME AND MYELODYSPLASTIC/ MYELOPROLIFERATIVE DISEASES
68	HISTIOCYTIC AND DENDRITIC CELL NEOPLASMS

## PREVALENCE

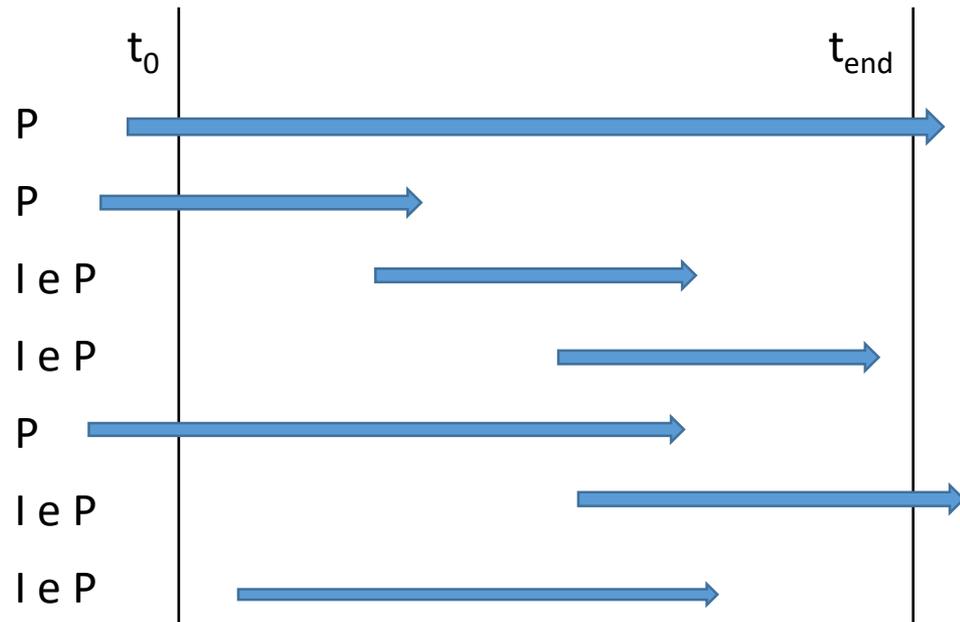
**225 872**

ESTIMATED PREVALENT CASES  
ITALY, 2010

## INCIDENCE

**27 084**

ESTIMATED NEW CASES  
ITALY, 2015



# TASSO DI PREVALENZA

$$Prev. = \frac{\text{casi presenti}}{\text{Totale della popolazione}}$$

$$Prev(x) = Prev_L^{obs}(x) + Prev_L^{unobs}(x)$$

Epidemiol Prev 38 (6) Suppl 1:1-144

**RARE HAEMATOLOGICAL DISEASES.** Observed prevalence (proportion per 100,00 and 95% confidence interval - 95% CI) by duration ( $\leq 2$ , 2-5,  $\leq 15$  years) prior to prevalence date (1<sup>st</sup> January 2007), and complete prevalence. Estimated prevalent cases in 2010 in Italy.

PROPORTION	95% CI	ESTIMATED PREVALENT CASES 2010
277.80	271.10-284.50	225 872

# LA VARIABILITA' DELLE MISURE

Si tratta in genere dei residenti dell'area geografica da cui si rilevano i dati.

Si tratta più estensivamente della regione di cui il registro è campione (quando non copre l'intera popolazione)



L'insieme dei dati dei RT in Italia è utile per trarre conclusioni sulla distribuzione dei tumori nell'intera nazione



**STIME**

Incidenza: 147 per 100,000 per year,

Sorge il problema dell'incertezza dei risultati dovuta all'effetto del caso

**Calcolo degli intervalli di confidenza**

## Incidenza dei tumori

In mancanza di dati osservati per l'intero territorio nazionale, il numero di nuovi casi di tumore diagnosticati ogni anno in Italia viene quantificato mediante stime (vedi [www.tumori.net](http://www.tumori.net)).

Nel 2008 si stima vi siano stati oltre 250.000 nuovi casi di tumore nei soggetti di età compresa tra 0 e 84 anni: 132.141 tra gli uomini e 122.052 tra le donne.

E&P, anno 33 (1-2) gennaio-aprile 2009 supplemento 2

# INTERVALLI DI CONFIDENZA

## Calcolo degli intervalli di confidenza

Con un certo grado di probabilità (livello di confidenza, solitamente 95%) il vero valore sarà all'interno di due limiti: l'indicatore può variare nella popolazione tra quei due limiti con la probabilità del 95%

### ***Variabilità (gli intervalli di confidenza)***

Solitamente quando si presenta un indicatore (nel nostro caso SIR ed EAR) lo si correda anche di un'informazione sul grado di fiducia che abbiamo nella stima prodotta;

I tumori in Italia, Rapporto AIRTUM 2013, TUMORI MULTIPLI, *Epidemiol Prev* 37 (4-5) Suppl 1:1-152

The **CONFIDENCE INTERVAL (CI)** is the measure of the uncertainty of the figures we present. The larger the interval, the more uncertain the figures. When you read the value of an indicator we suggest you also look at the CI in which the value is included.



E.g., incidence rate  
1.84, CI:1.78-1.89.  
This CI is narrow, showing  
reliability of the provided  
estimation.

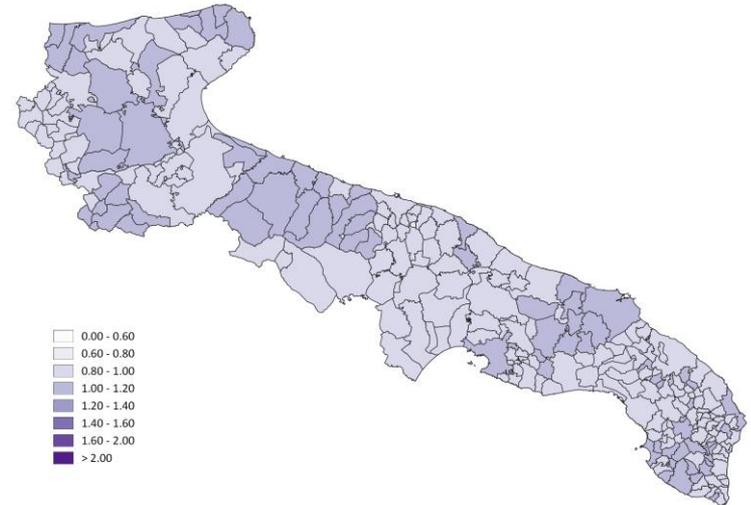
# INDIVIDUARE POSSIBILI FATTORI DI DISTORSIONE DELLA MISURAZIONE

La misurazione grezza (crude incidence),  
SE USATA PER COMPARAZIONI O PER DECISIONI SULLE  
POPOLAZIONI può portare a false conclusioni (bias).

Si devono individuare

le fonti di bias

le modalità di correzione del bias



**ATTENZIONE!** I SIR sono calcolati su una popolazione specifica, quindi forniscono un'informazione relativa a quella data popolazione (che avrà una specifica distribuzione per età e genere eccetera); per questo, quando si confrontano SIR di popolazioni differenti (come uomini *vs.* donne, giovani *vs.* anziani) bisogna ricordare che questo indicatore è influenzato dalla distribuzione per età o per sede di primo tumore, che è diversa, per esempio, nei due sessi.

# Misure comparative

## **DISEGNO DELLO STUDIO E METODI STATISTICI**

Nel presente studio è stato utilizzato il disegno di coorte retrospettivo, con l'obiettivo di valutare se il numero di secondi tumori diagnosticati nei pazienti oncologici fosse **superiore o inferiore rispetto a** quanto avviene nella popolazione generale.

## **Rapporto standardizzato di incidenza (osservati/attesi)**

E' stato calcolato dividendo il numero osservato di secondi tumori per il numero di casi che sarebbero stati diagnosticati tra i pazienti della coorte se questi avessero sperimentato gli stessi livelli di incidenza della popolazione generale. Il rapporto osservati/attesi è solitamente definito "rapporto standardizzato di incidenza" (**Standardized Incidence Ratio, SIR**) ed è frequentemente interpretato come un rischio relativo, poiché confronta il tasso di incidenza dell'evento di interesse nella popolazione studiata con quello nella popolazione di riferimento.

# Il rapporto standardizzato di incidenza

## SIR

SIR (*Standardized Incidence Ratio*) è il rapporto standardizzato di incidenza, dato dal rapporto tra:

**Casi osservati (misura reale)**

**Casi attesi (misura teorica)**

dove

**Casi osservati** = è il numero di secondi tumori registrati nella popolazione in studio (misura reale);

**Casi attesi** = è il numero di secondi tumori che avremmo osservato nella popolazione in studio se questa avesse sperimentato i livelli di incidenza della popolazione generale di pari sesso ed età (è una misura di confronto fittizia, teorica).

I tumori in Italia, Rapporto AIRTUM 2013, TUMORI MULTIPLI, *Epidemiol Prev* 37 (4-5) Suppl 1:1-152

FIRST CANCER SITE	MEN					WOMEN				
	OBSERVED	EXPECTED	PY RISK	SIR	EAR	OBSERVED	EXPECTED	PY RISK	SIR	EAR
Oral cavity	1 151	596	34 591	1.93*	16.03*	286	192	17 303	1.48*	5.40*
Pharynx	937	439	28 319	2.13*	17.55*	157	78	8 572	1.99*	9.13*

# METODI DI STANDARDIZZAZIONE

Sono note due modalità di standardizzazione dei tassi:

Il metodo diretto

Il metodo indiretto

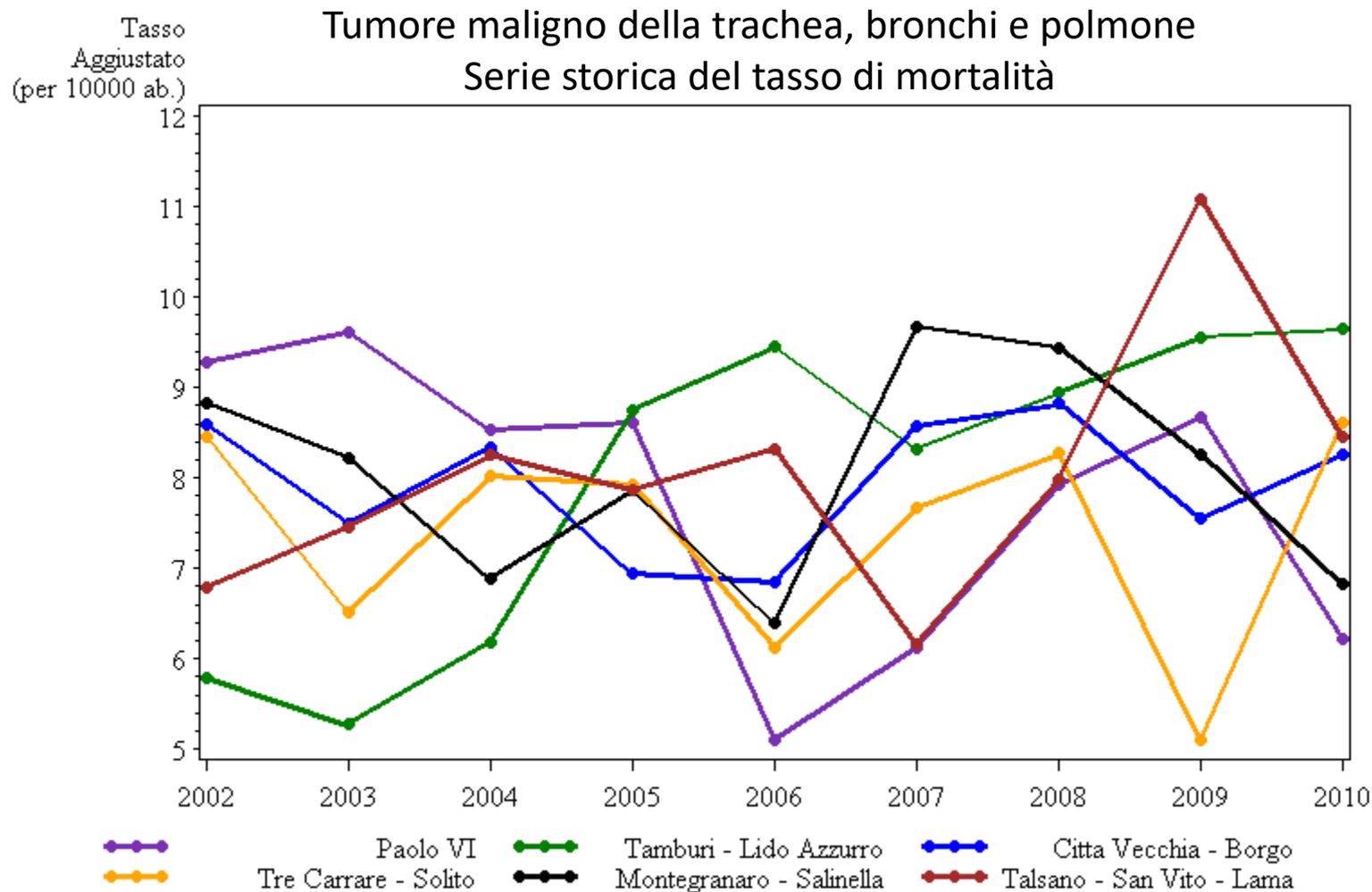
In entrambi i casi si determina un numero di casi «attesi»

Tumore maligno della trachea, bronchi e polmone, tassi di mortalità

Area	Osser.	Pop.	Attesi	Tasso grezzo ogni 10.000 ab.	Tasso Adj. ogni 10.000 ab.
Paolo VI	9	18625	11,59045	4,832215	6,224588
Tamburi	17	18103	14,11801	9,390709	9,652594
Città Vecchia	41	44314	39,78359	9,252155	8,261301
Tre Carrare Solito	40	41677	37,21405	9,59762	8,616314
Montegranaro- Salinella	34	46568	39,93128	7,301151	6,825496
Talsano-S.Vito- Lama	32	46526	30,36273	6,877875	8,448463

# METODI DI STANDARDIZZAZIONE

I tassi sono standardizzati usando come popolazione di riferimento la regione Puglia



# Metodo di standardizzazione diretto

Il metodo diretto.

Si supponga che le fonti di bias siano l'età e il sesso.

**Si stabilisce un riferimento.**

**Nel caso del metodo diretto il riferimento è la distribuzione della popolazione rispetto alle classi di età e al sesso.**

Si determina il numero di casi «attesi» moltiplicando il tasso osservato per la numerosità della popolazione di riferimento in ogni classe di età e sesso.

Si sommano i casi attesi e si dividono per il totale della popolazione di riferimento.

	Pop A	Dec A	Pop B	Dec B	Pop Rif N	Tax A t	Tax B	Att A	Att B
30-39	72450	14	5552	3	6000	19,32	5,403	1,16	3,24
40-44	78575	48	4570	5	6000	61,09	10,94	3,66	6,56
45-49	74600	113	3378	10	6000	151,47	29,603	9,09	17,76
50-54	65425	220	2792	16	5000	336,26	57,31	16,81	28,65
55-59	62025	347	2228	23	4000	559,45	103,23	22,38	41,29
60-64	52450	539	1660	26	4000	1027,64	156,63	41,11	62,65
Tot	405525	1281	20180	83	31000			94,21	160,17
	315,886		411,298			303,904	516,66		

$$\text{Tasso st.} = \frac{\sum_i N_i t_i}{\sum_i N_i}$$

# Metodo di standardizzazione indiretto

Il metodo indiretto.

Si supponga che le fonti di bias siano l'età e il sesso.

Si stabilisce un riferimento.

Nel caso del metodo indiretto il riferimento è il tasso di incidenza per classe di età e sesso.

Si determina il numero di casi «attesi» moltiplicando il tasso di riferimento per la numerosità della popolazione in esame in ogni classe di età e sesso.

Si calcola il rapporto tra casi osservati e casi attesi :

**RAPPORTO STANDARDIZZATO**

Il tasso si ottiene moltiplicando il rapporto standardizzato per il tasso globale del riferimento.

	Pop A n	Dec A	Pop B	Dec B	Pop Rif T	Att A	Att B
30-39	72450	14	5552	3	24,3	17,6	1,35
40-44	78575	48	4570	5	69,8	54,84	3,19
45-49	74600	113	3378	10	182,5	136,14	6,16
50-54	65425	220	2792	16	350,7	229,44	9,79
55-59	62025	347	2228	23	378,5	234,76	8,43
60-64	52450	539	1660	26	1050,2	550,83	17,43
Tot	405525	1281	20180	83		1223,64	46,36

$$\text{RAPPORTO STANDARDIZZATO} = \frac{\sum_i n_i}{\sum_i T_i n_i} = \frac{O_i}{E_i} = \frac{1281}{1223,64} = 1,05$$

# Interpretazione del Rapporto Standardizzato

Rapporto Standardizzato  $\rightarrow$  RS



OSSERVATI > ATTESI

La popolazione in studio presenta **maggior rischio** di presenza di casi RISPETTO ALLA POPOLAZIONE USATA COME RIFERIMENTO



OSSERVATI = ATTESI



OSSERVATI < ATTESI

La popolazione in studio presenta **minor rischio** di presenza di casi RISPETTO ALLA POPOLAZIONE USATA COME RIFERIMENTO

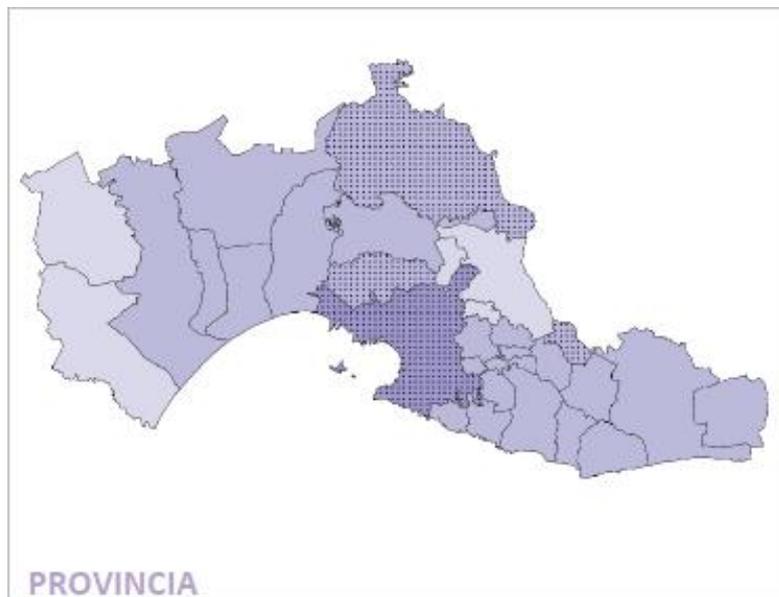
# Confronto tra aree

1 – applicazione di un modello di regressione logistica per la stima dei casi attesi, con metodo indiretto e riferimento i tassi della Regione Puglia

2- valori attesi e valori osservati sono utilizzati in un modello Bayesiano per la stima del rischio relativo che tenga conto di:

confondimento legato allo stato socio economico

distribuzione geografica dei casi e struttura dell'area geografica



$$O_i \sim \text{Poisson}(\theta_i E_i)$$

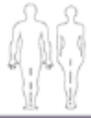
(3)

$$\log(\theta_i) = \alpha + \beta ID_i + h_i + b_i$$

dove  $O_i$  è il numero di casi/decessi/ricoveri osservati nell'area  $i$ -esima,  $\theta_i$  è il rischio relativo nell'area  $i$ -esima,  $E_i$  è il numero di casi attesi nell'area  $i$ -esima (riferimento: la Puglia per la mortalità e l'ospedalizzazione e il Pool Sud per l'incidenza),  $ID_i$  è l'indice di deprivazione relativo alla stessa area,  $h_i$  è la componente non strutturata degli effetti spaziali per la quale si assume una distribuzione a priori Normale  $(0, \tau)$ ,  $b_i$  è la componente strutturata degli effetti spaziali attribuita a priori secondo il modello Conditional AutoRegressive (CAR).

La stima puntuale del RR è stata corredata da un intervallo di credibilità al 95%.

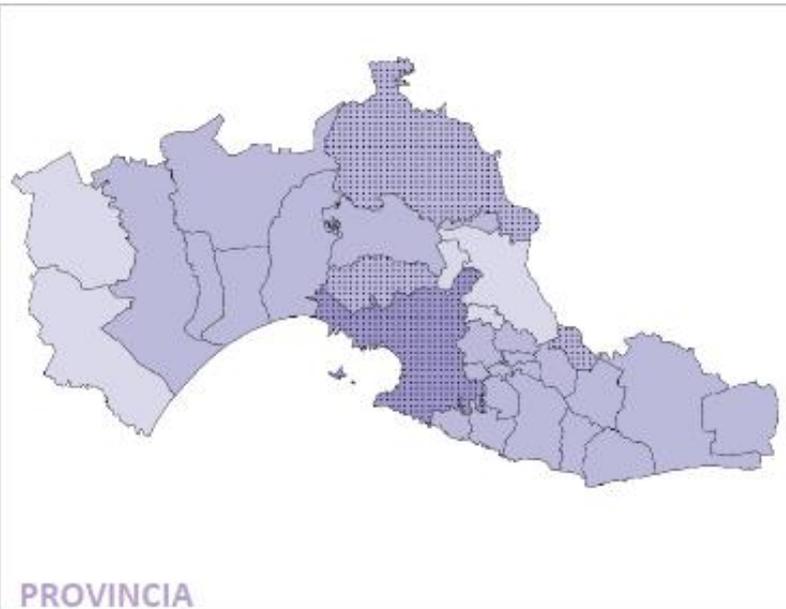
# Confronto tra aree - Interpretazione Rischi e relativi intervalli



TUTTE LE NEOPLASIE

SEDI ICD-O-3: C00-C96 /3; C67 /1-2

DISTRETTO	COMUNE	MASCHI+FEMMINE			MASCHI			FEMMINE		
		RR	I.C.95%		RR	I.C.95%		RR	I.C.95%	
QUARTIERI di Taranto	Isola-Borgo	1.18	1.12	1.24	1.17	1.09	1.26	1.20	1.10	1.29
	Italia Montegranaro	1.11	1.05	1.17	1.06	0.98	1.14	1.18	1.09	1.27
	Paolo VI	1.16	1.08	1.24	1.13	1.03	1.23	1.21	1.08	1.32
	Salinella	1.12	1.04	1.20	1.08	0.98	1.19	1.16	1.05	1.27
	San Vito-Lama-Carelli	1.04	0.95	1.14	0.96	0.85	1.08	1.15	1.01	1.30
	Solito Corvisea	1.11	1.04	1.18	1.03	0.94	1.11	1.22	1.12	1.33
	Talsano	1.08	1.01	1.15	1.01	0.92	1.10	1.17	1.07	1.28
	<b>Tamburi - Lido Azzurro</b>	<b>1.24</b>	<b>1.16</b>	<b>1.34</b>	1.17	1.07	1.28	1.30	1.19	1.45
	Tre carrare - Battisti	1.15	1.09	1.21	1.08	1.00	1.16	1.25	1.16	1.34



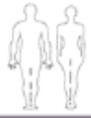
Il Rischio nell'area evidenziata è **MAGGIORE** rispetto a quello provinciale.

Il valore 1 è fuori dall'intervallo, il limite inferiore è 1.16.

Applicando l'interpretazione classica degli IC:  
In 95 casi su 100 otterremo sempre un rischio superiore a quello del riferimento.

Questo fa concludere che il rischio è **STATISTICAMENTE SIGNIFICATIVO**

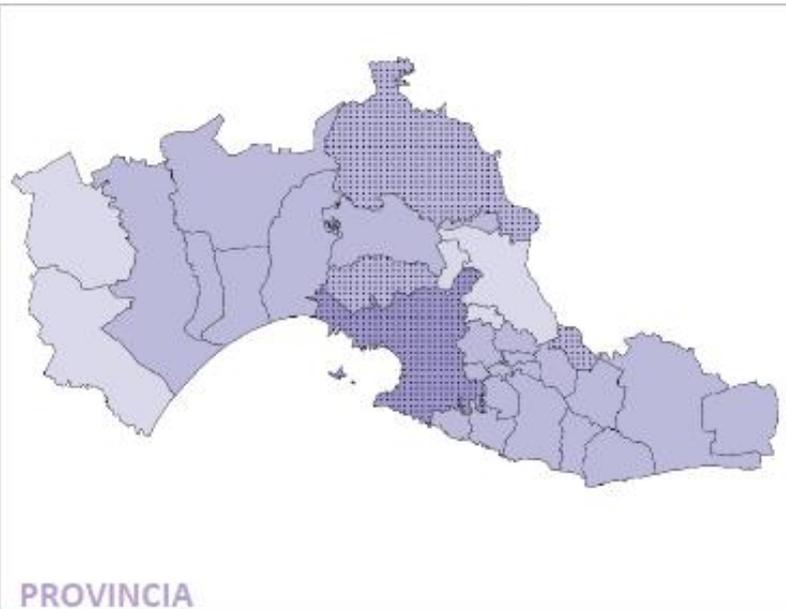
# Confronto tra aree - Interpretazione Rischi e relativi intervalli



TUTTE LE NEOPLASIE

SEDI ICD-O-3: C00-C96 /3; C67 /1-2

DISTRETTO	COMUNE	MASCHI+FEMMINE			MASCHI			FEMMINE		
		RR	I.C.95%		RR	I.C.95%		RR	I.C.95%	
QUARTIERI di Taranto	Isola-Borgo	1.18	1.12	1.24	1.17	1.09	1.26	1.20	1.10	1.29
	Italia Montegranaro	1.11	1.05	1.17	1.06	0.98	1.14	1.18	1.09	1.27
	Paolo VI	1.16	1.08	1.24	1.13	1.03	1.23	1.21	1.08	1.32
	Salinella	1.12	1.04	1.20	1.08	0.98	1.19	1.16	1.05	1.27
	San Vito-Lama-Carelli	1.04	0.95	1.14	0.96	0.85	1.08	1.15	1.01	1.30
	Solito Corvisea	1.11	1.04	1.18	1.03	0.94	1.11	1.22	1.12	1.33
	Talsano	1.08	1.01	1.15	1.01	0.92	1.10	1.17	1.07	1.28
	Tamburi - Lido Azzurro	1.24	1.16	1.34	1.17	1.07	1.28	1.30	1.19	1.45
	Tre carrare - Battisti	1.15	1.09	1.21	1.08	1.00	1.16	1.25	1.16	1.34



Il Rischio nell'area evidenziata è **MAGGIORE** rispetto a quello provinciale.

Il valore 1 è compreso nell'intervallo, il limite inferiore è 0,95.

Applicando l'interpretazione classica degli IC:

In 95 casi su 100 non otterremo sempre un rischio superiore a quello del riferimento.

Questo fa concludere che il rischio è

**NON STATISTICAMENTE SIGNIFICATIVO**

# Citazione finale

«Ho compreso che l'epidemiologia è una branca indispensabile della medicina. Gli epidemiologi e i biostatistici dovrebbero, sempre di più, divulgare in modo comprensibile informazioni corrette e validate, necessarie a incidere su certi stereotipi culturali, spesso basati su informazioni poco accurate, inoltre, hanno il compito importantissimo di indurre a rivedere modalità di follow-up che alla luce delle evidenze epidemiologiche potrebbero risultare sovradimensionate, soprattutto in oncologia.

I dati epidemiologici corretti sono strumenti indispensabili nella scelta dei sistemi di prevenzione e nell'indirizzare le scelte cliniche tramite un processo di traslazione dall'alto verso il basso, cioè dalle conoscenze oneste alle prescrizioni; così come inversamente, dal basso verso l'alto, è necessario analizzare i bisogni dei pazienti per pervenire alle norme.

I cittadini e i pazienti dovrebbero ricercare e prestare attenzione alle informazioni accurate e riscontrabili, ai dati scientifici ben documentati, alle conclusioni verificabili.

Un'informazione di qualità si impone come traguardo indispensabile che necessita di uno sforzo congiunto di operatori sanitari e pazienti.»