

Sommario

| | |
|---|----|
| EPIDEMIOLOGIA..... | 1 |
| INSORGENZA DELLE INFEZIONI | 3 |
| POSTULATI koch ed Evans | 4 |
| ASSOCIAZIONE | 5 |
| CONFONDIMENTO | 6 |
| DETERMINANTI | 12 |
| POPOLAZIONE | 13 |
| PREVALENZA E INCIDENZA | 14 |
| MISURE EPIDEMICHE | 19 |
| CAMPIONAMENTO | 23 |
| VALUTAZIONE TEST DIAGNOSTICI | 26 |
| TEOREMA DI BAYES | 32 |
| VEROSIMIGLIANZA | 36 |
| MISURE DI ASSOCIAZIONE | 39 |
| STUDI EPIDEMIOLOGICI | 41 |
| STUDIO CASO-CONTROLLO | 42 |
| STUDIO DI COORTE | 45 |
| STUDI DI PREVALENZA | 47 |
| EPIDEMIOLOGIA MOLECOLARE | 51 |
| EPIDEMIOLOGIA DELLE MALATTIE INFETTIVE | 55 |
| BIOMARKER | 68 |

https://www.quadernodiepidemiologia.it/epi/screen/vp_test.htm

EPIDEMIOLOGIA MOLECOLARE

02/03

EPIDEMIOLOGIA

Statistica ed epidemiologia sono diverse. La statistica prende info da altre scienze.

Ci servono dati affidabili per avere una valutazione critica e scientifica

EPIDEMIOLOGIA: definizioni

1. È la scienza dei fenomeni generali legati alle malattie infettive. Una scienza induttiva che riguarda non solo la distribuzione delle malattie ma che costituisce una solida filosofia.
2. È un metodo di ragionamento sulla malattia che riguarda le interferenze biologiche derivanti dall'osservazione delle malattie nelle popolazioni
3. Studia la distribuzione della malattia e la frequenza dei determinanti (cause) della malattia

Serve per ricostruire la storia della malattia, la causa e i fattori di rischio (es ambienti, patogeni, situazioni di lavoro ecc)

L'epidemiologia veterinaria in particolare si occupa anche degli eventi correlati alla sanità, nella produzione.

Opera su soggetti che sono nella loro realtà, nel loro ambiente.

L'EPIDEMIOLOGIA è UNA COMPARAZIONE DI GRUPPI E PRODUCE DATI, VERIFICABILI E VALUTABILI, CHE POSSONO ESSERE CONFERMATI O MENO.

DIVERSE SONO LE CREDENZE CHE SONO BASATE SULLA COMUNE IMMAGINAZIONE, NON SONO VALUTABILI NE QUANTIFICABILI E NON POSSONO ESSERE CONFERMATI (es efficacia di un vaccino)

SCOPI:

- determinare l'origine della malattia ad eziologia conosciuta (individuazione dei fattori di rischio, ricerca della fonte di entrata e diffusione, individuazione dell'origine del focolaio, messa a punto di metodi diagnostici)
- determinare l'origine di una malattia ad eziologia sconosciuta o poco conosciuta (individuare fattori di rischio, valutazione malattie multifattoriali, sviluppo di programmi di controllo sulla base dei fattori di rischio individuati)
A volte si riscontrano delle malattie ma non si riesce ad individuare l'agente eziologico ne come agisce. Con l'epidemiologia posso trovare degli stratagemmi. Non elimino il patogeno ma metto in atto soluzioni che evitino che infetti. Ad es nei bovini da latte con le mastiti ambientali dovuti a batteri presenti nelle lettiere dove vivono. Posso anche non conoscere cosa c'è nella lattiera ma so che se pulisco o mungo correttamente non passano i batteri ed evito le mastiti. Non ho bisogno di sapere come si chiama il batterio, se giacomo o aldo, metto in atto soluzioni e basta.
- Acquisire informazioni sulla storia di una malattia.
Se agisco antagonizzando le azioni negative tra ospite ambiente e agente eziologico posso limitare o eliminare la malattia
Se conosco la storia naturale della malattia (interazione ospite-ambiente-agente eziologico) sarò in grado di mettere in atto strategie di controllo
- Sviluppare e monitorare programmi di controllo delle malattie
Si deve considerare il rapporto costo/beneficio
- valutare gli effetti economici delle malattie e i benefici dei piani di controllo
Sceglierò se contenere solamente e forme cliniche della diffusione o eradicare. Nel caso di ZONOSI si prevede l'eradicazione (brucellosi e tubercolosi). Non si fa vaccinazione xke vaccinando non elimino al 100%, quindi l'uomo può cmq infettarsi.
A seconda del rapporto costo beneficio deciderò se:
 - contenere le forme cliniche
 - contenere la diffusione
 - eradicare

7 PASSI DELL'INDAGINE EPIDEMIOLOGICA:

1. Definire la popolazione
 2. Definire le misure di esposizione e di salute
 3. Se la popolazione è grande prendere un campione su cui si fa l'indagine
 4. Stimare l'associazione tra esposizione (attività, comportamenti, luogo) e sanità
 5. Escludo associazioni casuali. Più è grande il numero di soggetti e di dati più aumenta la casualità
 6. Verificare se le evidenze sono biologicamente credibili e dimostrabili (es che il colore di capelli sia correlato all'allele valgo, non è scientifico)
 7. Valutare la possibilità che si possa estendere alla popolazione target. L'obiettivo è di comprendere la totalità
- **EPIDEMIOLOGIA DESCRITTIVA.** Quantifica la frequenza della malattie
es 20% hanno il raffreddore

- **EPIDEMIOLOGIA ANALITICA.** Analizza le osservazioni raccolte nella popolazione mediante diagnostica e statistica
- **EPIDEMIOLOGIA SPERIMENTALE.** Analisi in popolazione quando applico io il fattore di rischio (esperimenti)
- **EPIDEMIOLOGIA TEORICA.** Sviluppo di modelli matematici che simulano il pattern epidemiologico delle malattie
(es l'influenza aviaria che secondo i modelli teorici si prevedeva 100 milioni di morti. Si basavano sui dati della spagnola, proiettando quei dati su una popolazione più ampia il numero era più alto. Senza considerare che erano in guerra, non c'erano gli ospedali ne gli antibiotici ecc)
- **EPIDEMIOLOGIA CLINICA.**
- **EPIDEMIOLOGIA COMPUTAZIONALE**
- **EPIDEMIOLOGIA GENETICA**
- **EPIDEMIOLOGIA MOLECOLARE**
- **EPIDEMIOLOGIA AMBIENTALE**
- **EPIDEMIOLOGIA NUTRIZIONALE**

EPIDEMIOLOGIA CLASSICA VS MOLECOLARE

La grande differenza è come ottengo i dati.

INSORGENZA DELLE INFEZIONI

L'agente patogeno vuole entrare, replicare e uscire. Le vie d'entrata sono molteplici. Dalle stesse vie, o altre, ci può essere l'uscita del patogeno, e quindi la diffusione.

Fasi dell'insorgenza dell'infezione:

1. **Esposizione** al patogeno. se non c'è esposizione non c'è malattia
in questa fase possiamo distinguere la **colonizzazione** (patogeno presente ma con assenza di danno) o **infezione** (danno e segni di malattia)
2. **Entrata.** Deve superare delle difese immunitarie o barriere.
Vie di entrata: inalazione, ingestione, penetrazione (soluzioni di continuo, contatto), inoculazione, coito
3. **Diffusione.** Del patogeno nell'ospite. La maggior parte dei patogeni hanno predilezione per un organo bersaglio. Di solito si tratta di recettori, dove non ci sono recettori non diffonde.
4. **Moltiplicazione.** Condizionata da disponibilità di nutrienti e fattori fisici (es temperatura e PH)
5. **DANNO.** Dell'ospite da parte del patogeno o della reazione infiammatoria. A volte può non esserci il danno es i coli interemorragici. Se mangio o bevo questo patogeno, spt i bambini possono avere patologia che può causare morte. Xke i recettori sono nei reni e da sindrome uremica. Nell'adulto che non ha i recettori non c'è sindrome uremica ma colite. Il bovino che è il bersaglio si ammala ma non ha nessun danno quindi non si riesce ad individuare da sintomatologia se il bovino è malato oppure no. Si può fare solo su esame feci. Su un numero grande di capi non è semplice
Tipo di danno: Danni meccanici (ostruzione organi e vasi), alterazioni metaboliche (es tossine e risposta immunitaria eccessiva), morte cellulare (necrosi, infarto), risposta infiammatoria eccessiva
6. **ESCREZIONE.** Non avviene necessariamente subito dopo la malattia (LATENZA). Non è necessario che l'ospite sopravviva.

Tutti questi passaggi prevedono che le difese immunitarie vengano superate.

DEFINIZIONI:

COLONIZZAZIONE: presenza del patogeno senza danni. Può essere un vantaggio come no.

INFEZIONE: presenza di danni e sintomatologia

INCUBAZIONE: T che intercorre tra l'entrata del patogeno e la comparsa della malattia. A seconda del periodo di incubazione stabilito cambiano gli indici.

SERBATOIO: il serbatoio di un agente infettivo è l'ambiente nel quale l'agente patogeno normalmente vive, cresce e si moltiplica. Può essere umano, animale o ambientale. È la fonte diretta o indiretta da cui l'agente patogeno viene trasferito all'ospite. Conoscere il serbatoio è fondamentale per poter sviluppare ed applicare un programma di controllo efficace

FONTI: **Fonte di infezione:** soggetto od oggetto attraverso il quale viene acquisita l'infezione da parte dell'ospite sensibile. **Fonte di contaminazione:** soggetto od oggetto attraverso il quale viene contaminato l'ambiente da parte dell'ospite o da fonti esterne

MALATTIA INFETTIVA: malattia che si trasmette da un individuo infetto ad un altro individuo sensibile mediante un contatto diretto ed indiretto, in questo caso generale si definisce la malattia come trasmissibile. Una malattia infettiva si definisce come contagiosa quando è richiesto un contatto stretto fra soggetti malati e sani per una sua diffusione, questa rappresenta un caso particolare delle malattie infettive trasmissibili

MALATTIA CONTAGIOSA: quando il contatto deve essere STRETTO perché avvenga il contatto. È importante perché evitando il contatto evito la malattia
Le malattie infettive non sono tutte contagiose

La trasmissione di una malattia implica che vi sia

1. un serbatoio (es. mammella infetta)
2. una via di uscita attraverso il quale l'agente patogeno esce dal serbatoio (es. latte)
3. una via di trasmissione attraverso la quale l'agente patogeno lascia il serbatoio (es. guaina contaminata)
4. una via di entrata attraverso la quale l'agente patogeno entra nell'ospite sensibile (es. capezzolo)

PREVENZIONE PRIMARIA. Prevengo l'ingresso

PREVENZIONE SECONDARIA. Prevengo la diffusione

PREVENZIONE TERZIARIA. Ho perso le prime due

POSTULATI koch ed Evans

POSTULATI DI KOCH:

- L'agente eziologico deve essere presente in tutti i casi di malattia.
- Non è presente in altre malattie come agente opportunisto o non casuale.
- Se isolato e re inoculato dà malattia

Sono stati usati principalmente nel 900 ma non prevedono gli agenti opportunisti. Per questo c'è stato un dibattito.

PROBLEMA: Sono "deterministici". Non prevedono le malattie multifattoriali. Non prevedono gli agenti opportunisti. Richiedono un'infezione sperimentale talvolta di difficile realizzazione

| KOCH | EVANS |
|--|--|
| - Sono "deterministici" | - Sono "probabilistici" |
| - Non prevedono le malattie multifattoriali | - Richiedono una valutazione statistica dei dati |
| - Non prevedono gli agenti opportunisti | - Comprendono le malattie multifattoriali |
| - Richiedono un'infezione sperimentale talvolta di difficile realizzazione | - Prevedono nelle dimostrazioni l'utilizzo della prevenzione |

POSTULATI DI EVANS: sono **probabilistici**, ovvero non il "o tutto o nulla" come koch. Richiedono una valutazione statistica. Comprendono malattie multifattoriali. Prevedono l'utilizzo della prevenzione come "diagnosi" ovvero se prevengo e la malattia non c'è allora è malattia

1°- la proporzione di ammalati negli esposti deve essere più grande rispetto ai non esposti. non è necessario che tutti gli esposti siano ammalati. Anche tra i non esposti potrebbero esserci ammalati. La % di esposti ammalati deve essere più alta dei non esposti ammalati

Ammalati esposti > ammalati non esposti

- 2° - l'esposizione alla causa di malattia deve essere più frequente negli ammalati rispetto ai non ammalati, quando tutti gli altri fattori sono costanti.
- 3° - il numero di nuovi casi (INCIDENZA) deve essere maggiore negli esposti rispetto ai non
- 4° - la malattia deve seguire il periodo di incubazione della malattia e rispettare le sue sequenze temporali.
- 5°
- 6° - la risposta in seguito all'esposizione deve essere misurabile o maggiore in seguito ad esposizione, rispetto a quanto misurato prima dell'esposizione – MISURAZIONE DELL'EFFETTO
- 7° - la riproduzione sperimentale deve portare ad un aumento dei casi nei soggetti esposti rispetto ai non
- 8° - l'eliminazione della sospetta causa o della sua modificazione deve determinare la riduzione della frequenza di malattia – INTRODUZIONE DEL CONTROLLO
- 9° la modulazione della risposta dell'ospite deve significativamente ridurre la comparsa della malattia negli esposti – INTRODUZIONE DI PREVENZIONE
- 10° - TUTTO DEVE ESSERE BIOLOGICAMENTE CREDIBILE

ASSOCIAZIONE

Quando parliamo di associazioni dobbiamo parlare di qualcosa di misurabile. Si basa infatti sulla misura delle VARIABILI e ne studia le relazioni mediante la statistica.

Misuro la relazione tra le variabili con la statistica.

Una VARIABILE è ogni evento osservabile

- Variabile studiata – evento osservato
- Variabile risposta – malattia
- Variabile esplicativa – esposizione alla malattia

Faccio una previsione in base a quello che osservo sulla malattia

Dal punto di vista pratico l'associazione può essere statistica o non statistica (non ci facciamo nulla)

Un'**associazione non statistica** è quando osserviamo due cose che non hanno un'associazione tra loro (es diarrea e escherichia coli. Il fatto di trovare in molti casi di diarrea la presenza di coli non vuol dire che il coli sia causa di quella malattia, potrebbe, ma non è detto), ovvero la loro associazione è semplicemente dovuta al CASO

Un'**associazione statistica** è quando due variabili sono associate e la loro frequenza congiunta è superiore a quella attesa

La % per essere statisticamente significativa dipende da numero preso in esame. Più il numero è grande più la % può essere bassa per essere significativa. Meno soggetti ho più la % mi serve alta.

Se A è associato a B ed anche a C, allora B e C avranno una relazione. Di che tipo?

Es la mastite. Aumento cellule somatiche nel latte sono causa di minor produzione di latte? NO, ma la mastite ne è causa. Se uso un antinfiammatorio diminuisco le cellule somatiche e aumento la produzione di latte. Ma non ho risolto il problema, la mastite c'è ancora. Smetto gli antibiotici e riparto da 0.

Nello sviluppare un programma di prevenzione è necessario distinguere:

- **FATTORI DI RISCHIO.** Quando l'esposizione al fattore aumenta la frequenza di malattia
- **INDICATORI DI RISCHIO.** Qualcosa associato alla frequenza di malattia. La sua presenza è associata alla malattia ma non ne è la causa (es cellule somatiche nel latte in mastite)
- **ASSOCIAZIONI DIRETTE:** A causa B. (es trauma causa ferita)

- **ASSICIAZIOI INDIRETTE:** A causa B che causa C. (es leptospirosi causa emolisi che causa emoglobinuria)
- **ASSOCIAZIONE DIRETTA + INDIRETTA.** A causa B. B è veicolo per C. (es l'enterite da coli provoca diarrea, la diarrea sporca la mammella e lo sporco può dare mastiti da coli). Ma in questo caso bisogna capire come agire. Ad es usare antibiotici per il coli non ha senso sulla mastite xke il problema sono le tossine liberate dal coli e l'antibiotico non serve a nulla.
ES. le volpi attaccano la rabbia all'uomo + le volpi attaccano la rabbia al cane che l'attacca all'uomo

L'associazione e l'interazione tra cause dirette ed indirette produce due modelli

1. semplice, lineare
2. complesso, intrecciato
 - additivi: i due fattori x e y si sommano
 - moltiplicativi: i due fattori x e y si moltiplicano

CAUSA SUFFICIENTE: si definisce tale una causa che inevitabilmente produce un effetto. Generalmente comprende diversi fattori di rischio (semplici).

+ fattori = Causa complessa. -> In una causa sufficiente "complessa" tutti le cause semplici sono necessarie, ma una volta che una singola viene rimossa, la causa complessa non è più sufficiente

CAUSA NECESSARIA: se fattore di rischio (causa semplice) è una componente di ogni causa sufficiente complessa, si dice necessaria La sua eliminazione rende la causa complessa non sufficiente

CONFONDIMENTO

Gli studi epidemiologici consentono di indagare il ruolo di un particolare fattore nel determinare una malattia, Separandolo da quello svolto da altri fattori (variabili concomitanti o confondenti)

Tutti gli studi analitici devono cominciare con un'ipotesi correttamente formulata. Le ipotesi devono essere QUANTITATIVE e SPECIFICHE (devono considerare una relazione misurabile)

Esempio: i vitelli non allevati in gabbietta si ammalano !

Definire: 1. alternative alla gabbietta 2. tipo di malattia 3. altri fattori concomitanti

Perché UNO STUDIO EPIDEMIOLOGICO SIA VALIDO DEVE ESSERE CONDOTTO IN MODO DA POTER PERMETTERE UNA VALUTAZIONE STATISTICA DEI DATI

IPOTESI:

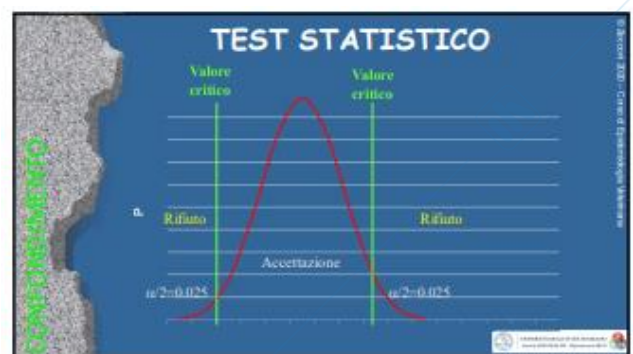
- **ipotesi NULLA:** si intende l'ipotesi preesistente all'osservazione dei dati campionari.
- **ipotesi ALTERNATIVA:** si intende l'ipotesi che si contrappone a quella nulla e che potrebbe essere considerata più verosimile.

TEST STATISTICO: Il test statistico è quello che ci farà dire se l'ipotesi nulla è **falsa**

Il rifiuto o l'accettazione dell'ipotesi nulla dipende dal campione osservato. Se l'informazione che si ricava dal campione contrasta in maniera evidente con l'ipotesi nulla, saremo costretti a rifiutarla.

Un TEST STATISTICO è una regola che porta ad accettare/rifiutare l'ipotesi H_0 (nessuna differenza tra "gruppi")

- **NEGAZIONE DELL'IPOTESI NULLA:** quando faccio uno studio non voglio aver ragione. Voglio dimostrare che non ho torto. Devo dimostrare che l'ipotesi nulla è falsa (che è il contrario di quello che ipotizziamo)
Es se voglio dimostrare che il coli dà enteriti devo verificare se il coli NON dà enteriti. Se non lo dimostro allora è vera la mia intuizione.



Normalmente quando dico che un'associazione è statisticamente significativa sto dicendo che c'è un 95% di probabilità che sia vero. Si può usare anche il 90% ma bisogna giustificare le condizioni

ERRORI:

- α = probabilità di commettere l'errore di I tipo è anche il livello di significatività del test. Definito a priori e convenzionalmente posto a 0.05. Permette di definire il valore critico in base al test statistico (in automatico)
- β = probabilità di commettere l'errore di II tipo. Dipende da α (e dal test statistico)
- $1-\beta$ = potenza del test e corrisponde alla probabilità di rifiutare l'ipotesi nulla quando è falsa

errore di I tipo: osservare una differenza quando in realtà non c'è ($\alpha=0.05$)

Errore di II tipo: NON osservare una differenza quando in realtà c'è ($\beta=0.20$)

ERRORE SECONDO TIPO (BETA) succede quando H_0 è falsa ma la considero vera. Di solito succede quando i campioni sono pochi. Bisogna avere un beta di almeno 80% e quindi con una formula calcolare quanti individui servono per avere beta 80%

gli errori (BIAS) sono generalmente prodotti da:

- Errori casuali
- Errori sistematici
- Distorsione (BIAS)
- Confondimento

DISTORSIONE E CONFONDIMENTO SONO DIVERSI

BIAS (distorsione)

La distorsione crea un'associazione che non è vera. Associa un fattore di rischio ad una malattia ma è un'associazione sbagliata

È un'erronea rappresentazione dell'effetto associato all'esposizione. Errore sistematico inserito nel disegno dello studio. Dovuto a:

- Erronea selezione soggetti dello studio
- Raccolta dei dati imprecisa
- Errori nella definizione di esposizione/malattia (errore diagnostico)

DISTORSIONE, quali sono le sue variabili:

- **SELEZIONE** – la popolazione non è sempre rappresentativa xke di solito si prendono i soggetti malati, bisogna includere i sani
- **RACCOLTA INFO:** il ricercatore influenza la risposta (es. è vero che con quello shampoo ti sei trovato bene? Più corretto sarebbe dire come giudichi la qualità dei tuoi capelli dopo lo shampoo?)
- **CLASSIFICAZIONE:** classificazione dei soggetti non corretta

CONFONDIMENTO

Classification bias

Situazione vera

| Esposizione | Malato | Sano | C/NC |
|-------------|--------|------|------|
| SI | 300 | 100 | 3 |
| NO | 200 | 400 | 0.5 |

OR=6

Test diagnostico con Se 60% e Sp 90%

| Esposizione | Malato | Sano | C/NC |
|-------------|--------|------|------|
| SI | 200 | 100 | 2 |
| NO | 300 | 400 | 0,75 |

OR=2.7

© Zecconi 2020 – Corso di Epidemiologia Veterinaria
UNIVERSITÀ DEGLI STUDI DI MILANO
Sezione ONE HEALTH - Dipartimento SBCO

OR= probabilità di ammalarsi

Nel test diagnostico ho avuto una distorsione (????)

COME EVITARLO:

- piu gruppi di controllo
- verificare classificazione dei soggetti
- usare procedure blind (l'operatore che raccoglie i dati non deve esserne influenzato)
- linee guida nella raccolta dati

Confondimento.

Il confondimento modifica il valore di un'associazione vera. Mi dà una risposta sbagliata ma non ne modifica l'effetto. Mi da un numero sbagliato pur non cambiando la sostanza, mi dice se il fatto è o meno associato

Il fattore confondente è un fattore di rischio che non considero (o xke non lo so o xke non lo conosco)

Deve essere associato al fattore di rischio, ma NON esserne conseguenza

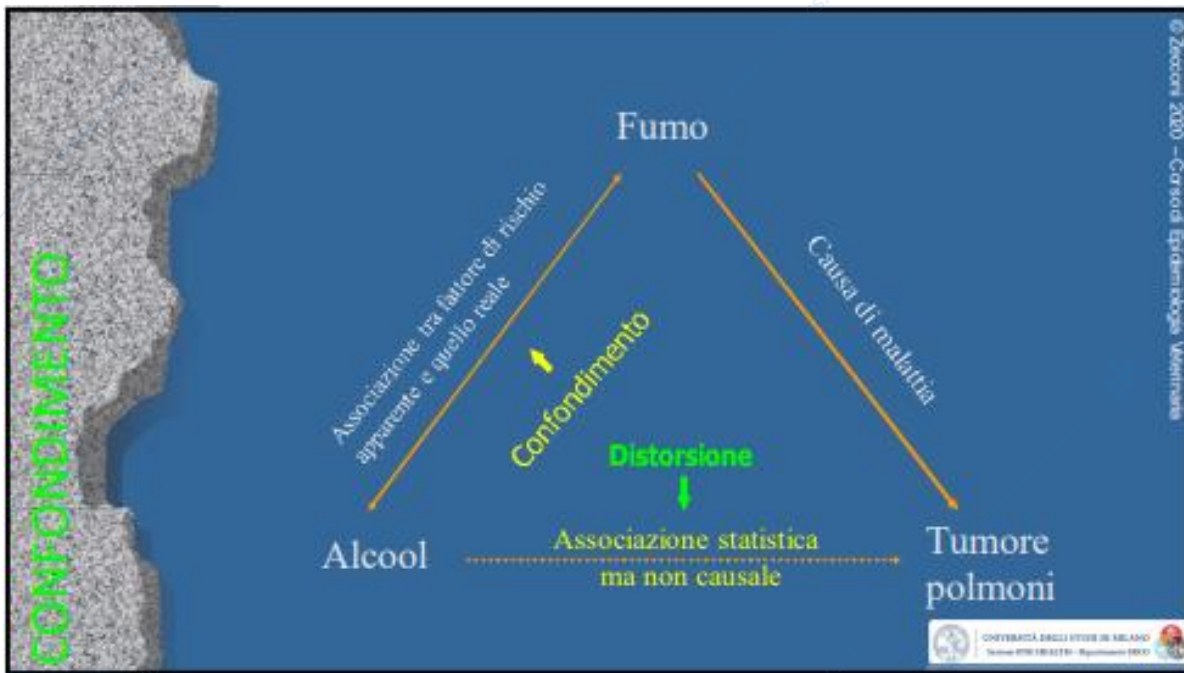
Il fattore confondente non è quindi un indicatore, perchè quest'ultimo è conseguente all'insorgenza della malattia
esempi di fattori confondenti:

- età, razza, sesso, stato fisiologico

Il fattore confondente può essere predittivo per la presenza di malattia, indipendentemente dall'esposizione che stiamo studiando

Ho un fattore di rischio apparente di malattia che xo non è causale.

Se io non lo considero uso un altro fattore di rischio, che è causa di malattia e mi va ad aggravare la situa



ESEMPIO

Il fumo causa il tumore ai polmoni. Posso dire che se uno usa un tot di sigarette ha un certo rischio di avere il tumore.

Se prendo le cartelle cliniche di questi soggetti potrei dire magari che anche l'alcol causa il tumore ai polmoni. Se faccio un'associazione statistica potrei dire che anche l'alcol causa il tumore.

C'è un effetto confondimento sul fumo. Chi fuma tanto, tra quelli che fumano tanto, magari sono persone che passano anche tanto tempo nei bar. Avendo una debilitazione dall'alcol e dal fumo è più probabile che abbiano il tumore.

Come faccio a risolvere il problema?

Stratifico i fumatori+ bevitori e fumatori meno bevitori. Così ottengo due rischi, del gruppo A e del gruppo B

Come controllare il confondimento?

- scelta casuale dei soggetti
- appaiamento (matching)
- stratificazione

ESEMPIO

Valutazione dell'associazione tra Stabulazione e mortalità neonatale

Allevamento con vitelli di 2 razze diverse (Frisona e Jersey) e due sistemi diversi di stabulazione (box e gabbia)

Devo capire dove mettere i vitelli, se in gabbia o in box

Quanti ne muoiono in gabbia?

Quanti in box??

Esempio

Mortalità gabbia

Mortalità box

$$= \frac{\text{morti .Gabbia}}{\text{nati .Gabbia}} = \frac{40}{205} = 19.51\%$$

$$= \frac{\text{morti .Box}}{\text{nati .Box}} = \frac{105}{165} = 63.63\%$$

© Zecconi 2020 - Corso di Epidemiologia Veterinaria

UNIVERSITÀ DEGLI STUDI DI MILANO
Sezione ONE HEALTH - Dipartimento SBCCO

Probabilmente la mortalità in box è più alta ma è statisticamente significativa?? Se guardo le percentuali potrei avere le stesse percentuali con numeri diversi, ad es se la popolazione fosse più piccola.

Faccio una valutazione fatta correttamente con 3 studi:

1. studio effetto stabulazione sul gruppo
2. studio la variabile predisposizione di razza
3. studio unificato tipo di stabulazione + predisposizione di razza

1. Studio: effetto stabulazione

| | | Malattia | | |
|-------------|---------------|------------|------------|------------|
| | | Morto | Vivo | Totale |
| Esposizione | Box | 105 | 60 | 165 |
| | Gabbia | 40 | 165 | 205 |
| | Totale | 145 | 225 | 370 |

| Tipo | Valore | Lim. Infer. | Super. |
|----------------------------|--------------|--------------|--------------|
| Rischio Esposti | 63.64% | 56.06 | 70.59 |
| Rischio non Esposti | 19.51% | 14.64 | 25.51 |
| Rischio Complessivo | 39.19% | 34.35 | 44.25 |
| Rapporto dei Rischi | 3.261 | 2.414 | 4.407 |
| Differenza dei Rischi | 44.12% | 35 | 53.25 |

Faccio una tabella dove metto morti e vivi e l'esposizione (una tabella 2x2)

Si calcola il rischio negli esposti, il rischio complesso e il rapporto dei rischi (rapporto tra esposti e nn esposti)

Il rischio di morire xke si è nel box è circa del 3%, che nella peggiore delle ipotesi va al 4%, nella migliore al 2.4%.

Un confondente è la RAZZA

Esempio

**Mortalità
Frisone**

$$= \frac{\text{morti .Frisone}}{\text{nati .Frisone}}$$

$$= \frac{80}{200} = 40.00\%$$

**Mortalità
Jersey**

$$= \frac{\text{morti .Jersey}}{\text{nati .Jersey}}$$

$$= \frac{65}{170} = 38.23\%$$

I dati dicono che la razza non influenzano la mortalità. NON BISOGNA FERMARSI QUI, sarà statisticamente vero???

2. Studio: effetto razza

| | Malattia | | |
|---------------------|------------|------------|------------|
| | Morto | Vivo | Totale |
| Esposizione Jersey | 65 | 105 | 165 |
| Esposizione Frisone | 80 | 120 | 205 |
| Totale | 145 | 225 | 370 |

| Tipo | Valore | Lim. Infer. | Super. |
|----------------------------|---------------|--------------|--------------|
| Rischio Esposti | 38.24% | 31.26 | 45.73 |
| Rischio non Esposti | 40.00% | 33.46 | 49.92 |
| Rischio Complessivo | 39.19% | 34.35 | 44.25 |
| Rapporto dei Rischi | 0.9559 | 0.740 | 1.234 |
| Differenza dei Rischi | -1.765% | -11.74 | 8.208 |

Risposta. Non c'è differenza

Quando i limiti di confidenza sono tra 1 e 2 il dato non è statisticamente significativo

CONCLUSIONE

La razza non risulta essere associata ad un aumento della mortalità (analisi 1)

La stabulazione risulta essere associata ad un aumento della mortalità (analisi 2)

MA NON È DEL TUTTO CORRETTO !!! devo tener presente che i dati potrebbero essere falsati dal fatto che molte più jersey sono finite nelle gabbiette piuttosto che nei box

Esempio

$$\text{Mortalità Frisone/gabbia} = \frac{\text{morti .Frisone .Gabbia}}{\text{nati .Frisone .Gabbia}} = \frac{15}{85} = 17.64\%$$

$$\text{Mortalità Frisone/box} = \frac{\text{morti .Frisone .Box}}{\text{nati .Frisone .Box}} = \frac{65}{115} = 56.52\%$$

$$\text{Mortalità Jersey/gabbia} = \frac{\text{morti .Jersey .Gabbia}}{\text{nati .Jersey .Gabbia}} = \frac{25}{120} = 20.83\%$$

$$\text{Mortalità Jersey/box} = \frac{\text{morti .Jersey .Box}}{\text{nati .Jersey .Box}} = \frac{40}{50} = 80.00\%$$

Devo matchare i dati.

È più veritiero. Notare cmq che la distribuzione degli animali non è omogenea

Valori grezzi

| Tipo | Valore | Lim. Infer. | Super. |
|------------------------------|---------------|---------------|---------------|
| Rischio Esposti | 50.75% | 45.41 | 56.06 |
| Rischio Non esposti | 29.63% | 25.39 | 34.26 |
| Rischio Complessivo | 39.19% | 35.74 | 42.75 |
| Rapporto dei Rischi | 1.713 | 1.426 | 2.058 |
| Differenza dei Rischi | 21.12% | 14.16 | 28.08 |
| Rapporto corretto | 1.598 | 1.315 | 1.942 |
| Differenza corretta | 23.21% | 16.48% | 29.94% |

Adesso tenendo conto della distribuzione degli animali il rischio è vero 1,6%, ma prima era di 3 %, quindi siamo scesi alla metà.

DETERMINANTI

Un determinante di malattia è qualsiasi caratteristica che influenza la salute di una popolazione.

La conoscenza dei determinanti è fondamentale per:

- Identificare i soggetti a rischio
- Prevenire la malattia
- Migliorare l'accuratezza della diagnosi

I determinanti possono essere:

- Primari & Secondari
- Endogeni ed esogeni
- Associati (ospite, agente, ambiente)

La malattia può anche derivare da interazione tra determinanti appartenenti a diverse categorie



POPOLAZIONE

La **frequenza** di un evento è misurata attraverso una proporzione:

- numeratore : **CASI**
- denominatore : **POPOLAZIONE A RISCHIO**

La **morbilità** è il numero di soggetti che **si ammalano**

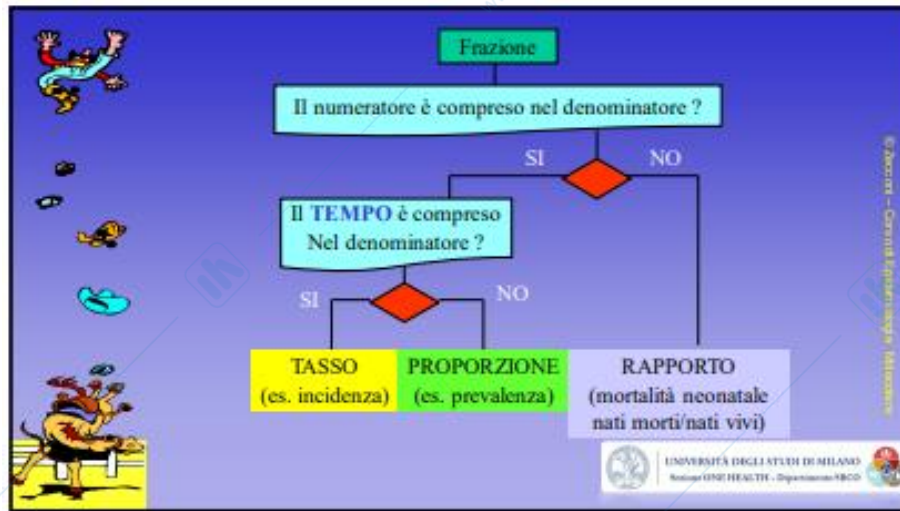
La **mortalità** è il numero di soggetti che **muoiono**

Morbilità e mortalità dipendono dal numero di soggetti ma anche dalla popolazione

La struttura della popolazione influenza la sua "misurazione" e la diffusione della malattia:

- **POPOLAZIONE CONTIGUA:** popolazione in cui vi è un notevole contatto tra gli individui della popolazione e con gli individui di altre popolazioni.
 - La diffusione delle malattie è molto facilitata (facili scambi e rimescolamenti)
 - Il conteggio degli individui è difficile: registri e anagrafe / cattura e rilascio
- **POPOLAZIONE SEPARATA:** individui segregati in unità semplici (ed allevamento, greggi..)
 - Il conteggio degli individui è relativamente semplice (registro allevamento)
 - può essere CHIUSA o APERTA. Se chiusa la diffusione delle malattie è ridotta

PREVALENZA E INCIDENZA



PREVALENZA numero di casi di malattia (o sieropositività o reperti microbiologici positivi) nella popolazione a rischio in un dato tempo

- Indica la probabilità di essere ammalato in un dato istante
- misura adimensionale
- può variare da **0 a 1**

ammalati / popolazione = % di ammalati al momento dell'indagine epidemiologica

- PUNTUALE : se riferita ad un dato istante di tempo
- DI PERIODO : se somma i casi avvenuti in un dato tempo (anno, mesi...)

PREVALENZA (%) = numero di soggetti ammalati in un particolare istante / popolazione a rischio in quel particolare istante

Esempio covid:

Al 9 marzo 2020: 53824 tamponi

- 9172 positivi
- 2936 isolamento domiciliare (asintomatici ?)
- 463 morti
- 724 guariti

«Prevalenza» infezione = tamponi + /tamponi totali -> $9172/53826 \approx 17\%$

In realtà i numeri sono molto diversi perché non sappiamo quanti si siano infettati per cui il calcolo della prevalenza dovrebbe essere:

Positivi per tampone o sierologia/popolazione italiana

Mortalità: morti / positivi -> $463 / 9172 \approx 5\%$

Guarigioni (tampone neg.) : guariti / positivi -> $724 / 9172 \approx 8\%$

La frequenza della mortalità è ampiamente sovrastimata . La frequenza delle guarigioni è ampiamente sottostimata.

MOTIVO: Perché non conosciamo la vera frequenza di infezione

INCIDENZA numero di **nuovi** casi di malattia (o sieropositività o reperti microbiologici positivi) nella popolazione a rischio in un dato periodo tempo

È necessario conoscere:

- numero di nuovi casi
- periodo di tempo durante il quale avvengono i nuovi casi (almeno 2 controlli)

Incidenza (%) = numero di NUOVI casi osservati in un dato periodo di tempo / popolazione a rischio in un dato periodo di tempo

Esempio influenza: virus influenzale colpisce (OMS):

- tra il 5 e il 15% della popolazione adulta
- tra il 20-30% dei bambini
- Tra i 3 e i 5 milioni di casi di influenza riportati annualmente evolvono in complicanze che causano il decesso in circa il 10% dei pazienti

INCIDENZA CUMULATIVA (o RISCHIO): proporzione di soggetti "non ammalati" all'inizio dell'indagine che si ammalano durante il periodo di indagine

- è la probabilità di ammalarsi in un dato periodo
- è una misura adimensionale
- può variare da 0 a 1

ci dà un'indicazione del rischio medio di contrarre una malattia in un particolare periodo di tempo ma può essere applicata solo a popolazioni statiche (chiuse) o a brevi intervalli di tempo

PROBLEMA: degli ammalati "rimossi" e della popolazione sensibile

Incidenza cumulativa (%) = numero di NUOVI casi osservati in un dato periodo di tempo / numero di soggetti sani all'inizio del periodo

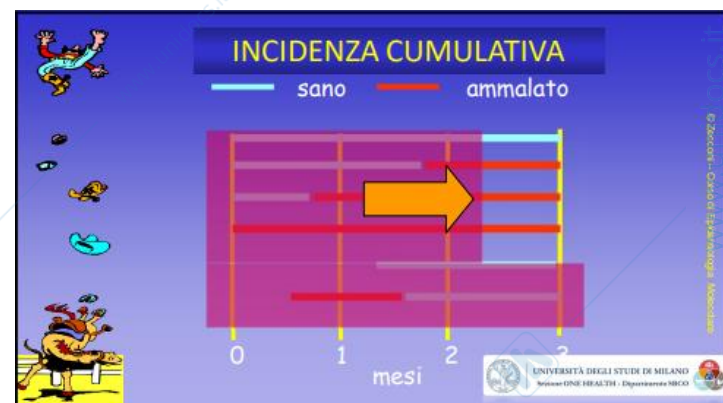
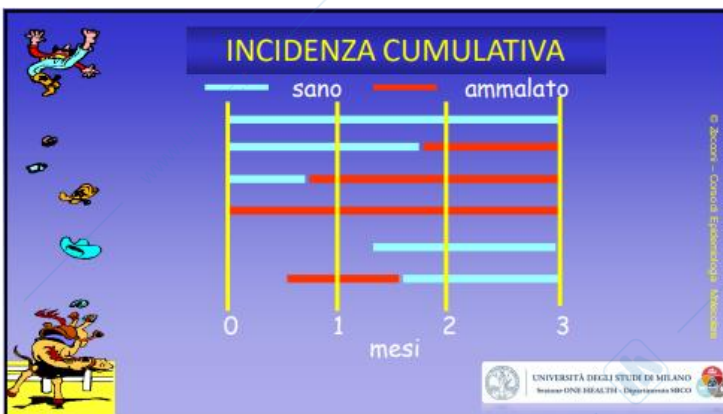
Esempio:

totale di 100 bovine, 100% sane
BVD (diarrea virale bovina) = 10
Tempo di indagine = 1 mese

Incidenza cumulativa 1 mese = $10/100 = 0,1$

totale di 100 bovine, 100% sane
BVD (diarrea virale bovina) = 20
Tempo di indagine = 2 mesi

Incidenza cumulativa 2 mesi = $20/100 = 0,2$



NOMINATORE: quanti nuovi casi? 3 (due sani che poi si ammalano, 1 sempre malato da tempo 0)

DENOMINATORE: quanti erano sani all'inizio? 4 (tutti perché assumo che il sempre malato era sano a T0)

Con questo tipo di incidenza non considero i nuovi ingressi, perché considera solo sistemi chiusi, quindi non prevede che ci sia l'ingresso di altri animali.

$IC = 3/4 = 0,75$ casi/3 mesi
ovvero 25 casi x mese x 100 animali

TASSO DI INCIDENZA

L'incidenza cumulativa è semplice da calcolare ma ci sono dei vincoli. La popolazione deve essere chiusa e/o i tempi brevi. In casi diversi si usa quindi al posto dell'incidenza cumulativa il tasso di incidenza che nella pratica misura la velocità con cui si sviluppano i NUOVI casi di malattia.

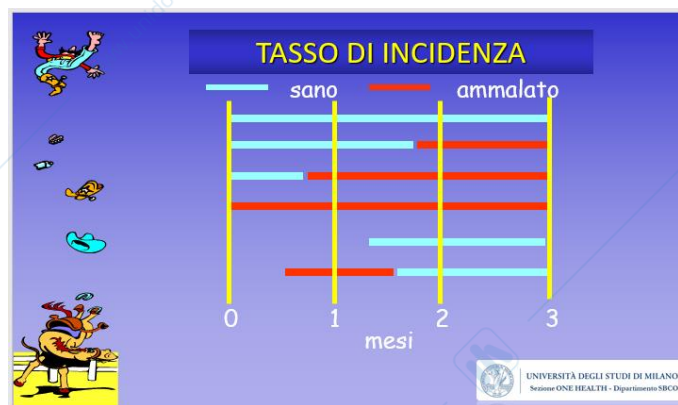
Il **TASSO DI INCIDENZA** misura la velocità con cui si sviluppano i nuovi casi di malattia

Tasso di incidenza (%) = numero di NUOVI casi osservati in un dato periodo di tempo / somma del tempo apportato da ciascun individuo

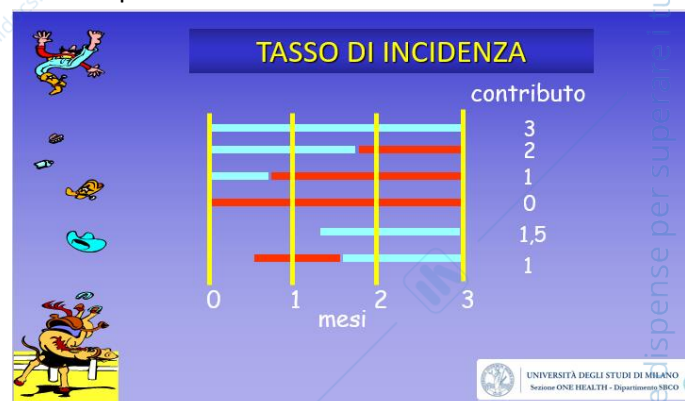
Ho un tempo 0 e un tempo x, verifico degli animali SANI (considero solo i sani) quanti di questi si sono ammalati alla fine del mio intervallo di tempo.

A differenza dell'incidenza cumulativa dove partivo dal presupposto (non verificato) che tutti gli animali fossero sani, qui lo verifico e scarto già i non sani.

Inoltre al denominatore ho la SOMMA del tempo, ovvero tutto il tempo in cui l'animale sano rimane sano e quindi suscettibile a contrarre malattia.



Esempio bovine di prima:



o dopo (perché non è una popolazione chiusa quindi possono entrare nuovi soggetti)

Vediamo cosa succede al tempo 0 e cosa succede al tempo 3 (momento in cui facciamo la valutazione finale)

Vediamo che di 4 soggetti iniziali 1 è sempre stato ammalato (non lo considero perché non è a rischio malattia, ce l'ha già).

NOMINATORE: C'è un caso in cui un animale è entrato malato ma poi guarisce. In questo caso considero il caso della malattia. In sostanza valuto se quella malattia poteva essere presente già prima dell'ingresso (quindi già malato, lo escludo) o se è una malattia che può aver contratto dopo (lo considero). Es. considero un animale al parto, il piccolo è malato. In sostanza se la malattia NON può essere trasmessa al parto allora non ce l'aveva da prima, è un nuovo caso.

Quindi ho 3 NUOVI casi di malattia (escludo: il sempre sano, il sempre malato, il malato che poi guarisce. Includo: 2 casi sani che si ammalano, il malato in ingresso che poi guarisce ma si è ammalato all'ingresso e non aveva la malattia pregressa) -> **3 al nominatore**

DENOMINATORE: Poi devo considerare QUANTO TEMPO RIMANGONO A RISCHIO gli animali. Quando sono malati non sono a rischio perché sono già malati. Sono a rischio quando sono sani. Nel caso del malato che guarisce devo considerare qual è la malattia in questione. Se è una malattia per la quale può aver fatto anticorpi allora non lo considero a rischio, altrimenti lo considero a rischio anche dopo che è guarito. (es. mastiti, anche dopo che guarisce è cmq suscettibile a altre mastiti)

Il sempre sano rimane a rischio per 3 mesi (vale 3). Il sano che poi si ammala rimane a rischio per 2 mesi (vale 2). Il sano che si ammala rimane a rischio per 1 mese (vale 1). Il sempre malato lo escludo (vale 0). Il nuovo ingresso sano rimane a rischio per 1 mese e mezzo (vale 1,5). Il malato che guarisce rimane a rischio 1 mese (vale 1). PS: questi sono valori che conosco perché so quando sono entrati gli animali e li ho monitorati costantemente. Se non so quando esattamente si sono ammalati perché non posso fare sempre monitoraggi stimo il tempo di rischio la metà del totale (su 3 mesi, 1,5)

FACCIO LA SOMMA di tutti i tempi: $3+2+1+0+1,5+1 \rightarrow$ **8,5 al denominatore**

IR (tasso incidenza) = $3/8,5 = 0,35$ casi / mese animale

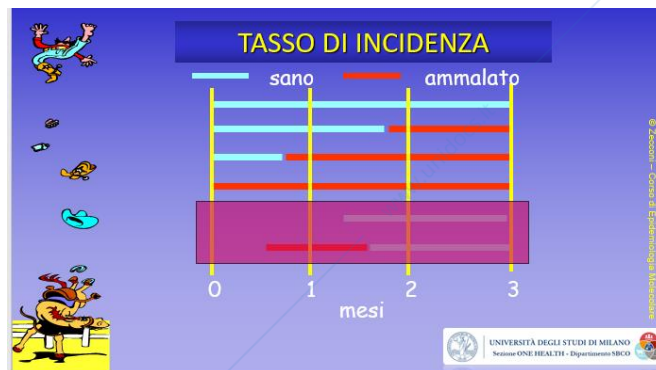
Quindi la velocità di diffusione della malattie è di 35 casi x mese x 100 animali.

Matematicamente dovrei dire che con malattie di questo tipo tutti i soggetti sarebbero ammalati, per cui il tempo che ho per intervenire è molto breve (es. influenza aviaria)

TIPS:

- Se un soggetto ammalato guarisce, ritorna ad essere calcolato come "a rischio" se NON immune
- se non so quando un soggetto si è ammalato contribuisce al denominatore per il 50% dell'intervallo di osservazione

CONFRONTO IC VS IR



PER FARE IL CONFRONTO TRA I DUE NON CONSIDERO I NUOVI INGRESSI (perche ic non li prevede)

IC (incidenza cumulativa) = $\frac{3}{4} = 0,75$ / 3 mesi – ovvero – 25 casi/ 100 animali – mese

IR (tasso incidenza) = $2 \text{ casi}/3 + 1,5 + 1,5 = 2/6$ – ovvero – 0,33 casi / animale - mese. Cioè 33 casi/100 animali – mese

L'incidenza cumulativa sottostima di molto qual è la velocità di diffusione della malattia

La diffusione di una malattia dipende da:

- dimensione della popolazione (misure precise su popolazioni molto grande è molto impegnativo)
- periodo di osservazione (se una malattia è diffusiva, piu tempo aspetto piu il tasso di incidenza non è veritiero)
- forza della morbilità (*misura teorica del rischio di contrarre la malattia in un istante di tempo (hazard rate) stimato dal tasso di incidenza*) <- la capacità della malattia di diffondersi

ESEMPI PREVALENZA E INCIDENZA

Esempio : 100 bovine in lattazione

Problema : infezioni mammarie

Analisi : latte di quarto

Descrizione risultati:

- prevalenza bovine infette
- prevalenza quarti infetti
- prevalenza agenti microbici
- proporzione agenti microbici

HO 100 BOVINE E PRELEVO I QUARTI (LATTE, 4 PER BOVINA)

| Descrizione risultati: | | |
|------------------------|--------|--------|
| | bovine | quarti |
| batt.neg. | 67 | 294 |
| batt.pos. | 33 | 106 |

PREVALENZA = n ammalati / tot soggetti a rischio

- Prevalenza bovine infette = $33/100 = 33\%$

- Prevalenza quarti infetti = $106/400 = 26,5\%$


- PREVALENZA DEGLI AGENTI PATOGENI (DOPO BATTERIOLOGI)
 - P. STAPHIOCCOCUS aureus NELLE BOVINE= $12/100$, cioè 12%
 - P. STRA DYSGAL nelle bovine = $4/100 = 4\%$
E così via
 - P. STAPH AUREUS NEI QUARTI = $46/400 = 11,5\%$
 - P. STRA DYSGAL NEI QUARTI = $6/400 = 1,5\%$
E così via

Descrizione risultati:
• prevalenza agenti microbici

| | bovine | quarti |
|--------------|--------|--------|
| Staph. aur. | 12 | 46 |
| Str. dysgal. | 4 | 6 |
| Staph. crom. | 13 | 46 |
| E. coli | 3 | 5 |
| Altri | 1 | 3 |

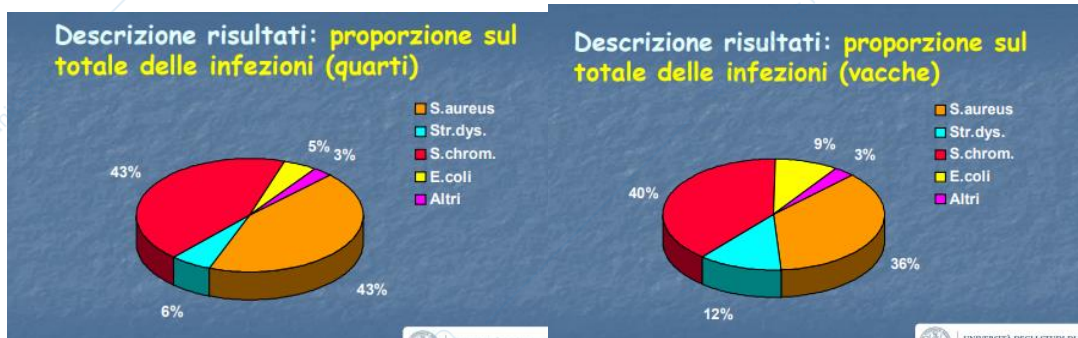
Descrizione risultati:
• prevalenza agenti microbici

| | bovine | quarti |
|--------------|--------|--------|
| Staph. aur. | 12% | 11.5% |
| Str. dysgal. | 4% | 1.5% |
| Staph. crom. | 13% | 11.5% |
| E. coli | 3% | 1.2% |
| Altri | 1% | 0.7% |



- Proporzionamento degli agenti microbici:

| Proporzionamento agenti microbici su bovine (solo distribuzione sulle tot positive: divido per 33) | Proporzionamento agenti microbici su quarti (solo distribuzione su tot quarti positivi: divido 106) |
|--|--|
| staph aureus = $12/33 = 36\%$ stra dysgal = $4/33 = 12\%$ staph crom = $13/33 = 40\%$ e coli = $3/33 = 9\%$ altri $1/33 = 3\%$ | staph aureus = $46/106 = 36\%$ stra dysgal = $6/106 = 6\%$ staph crom = $46/106 = 43\%$ e coli = $5/106 = 5\%$ altri $3/106 = 3\%$ |



INCIDENZA:

stesso allevamento un mese dopo rifaccio di nuovo le stesse analisi.

Ho cmq sempre in tot 100 animali ma di solito in allevamento succede che avrò animali in asciutta (non produce latte) e quindi non sono più a rischio per mastiti, altri animali verranno eliminati, altri saranno stati messi per sostituire i vecchi

Esempio : 100 bovine in lattazione
Problema : infezioni mammarie
Analisi : latte di quarto

Controllo 1 mese dopo
 • 10 bovine in asciutta
 • 10 animali eliminati
 • 20 manze partorito 3 gg prima

Descrizione risultati:
 • prevalenza agenti microbici

| | bovine | quarti |
|-------------|--------|--------|
| Staph.aur. | 12 | 46 |
| Str.dysgal. | 4 | 6 |
| Staph.crom. | 13 | 46 |
| E.coli | 3 | 5 |
| Altri | 1 | 3 |

La distribuzione degli agenti patogeni sono gli stessi

Calcolo dell'incidenza:

| | bovine | gg prod. | contrib. |
|------------|--------|----------|----------|
| Infette | 15 | 30 | 0 |
| Sane | 57 | 30 | 30 |
| Nuove inf. | 8 | 30 | 15 |
| Manze sane | 10 | 3 | 3 |
| Manze inf. | 10 | 3 | 1,5 |

CALCOLO INCIDENZA:

Tasso di incidenza (%) = numero di NUOVI casi osservati in un dato periodo di tempo / somma del tempo apportato da ciascun individuo

N soggetti sani per il periodo a rischio (30gg) ← colonna gg prod

15 erano infetti la prima volta e sono infetti anche adesso. Quanto contribuiscono al numeratore? 0. Quanto al denominatore? 0

I nuovi infetti sono 8. Quanto contribuiscono al numeratore? 8. Quanto al denominatore? 15 gg (la metà del tempo)

10 manze entrano sane e rimangono sane. Quanto al numeratore? 0. Al denominatore? 3gg

10 manze partoriscono e al momento del controllo sono infette. Al numeratore? 10. Al denominatore? 1,5 (sempre la metà del tempo)

Ora posso calcolare il tasso di incidenza

Calcolo dell'incidenza:

$$RI = \frac{10+8}{(57 \times 30) + (8 \times 15) + (10 \times 3) + (10 \times 1,5)} = \frac{18}{1875} = 0,0096 \text{ casi/giorno}$$

29 casi/100 animali x mese

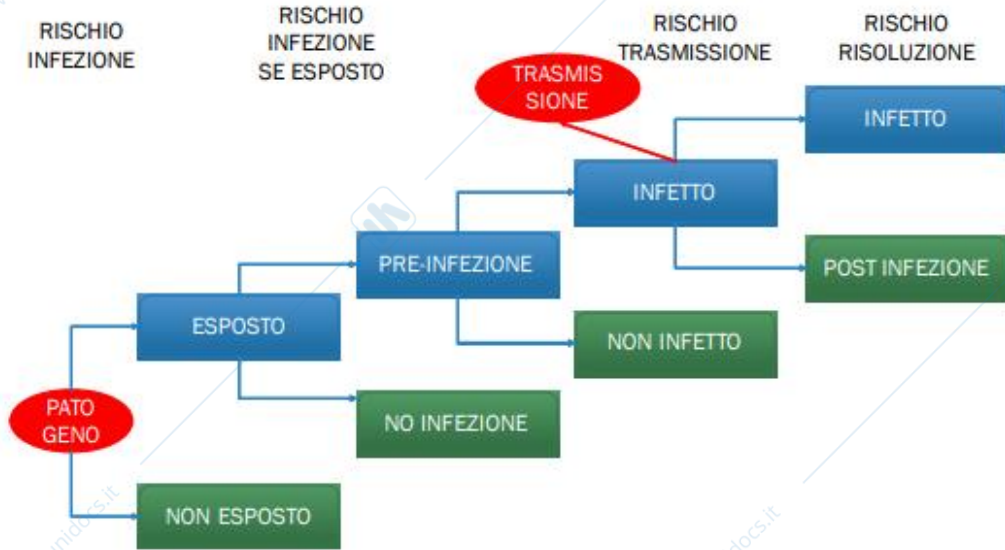
(il 57 è sbagliato, è un 67)

La prevalenza è più semplice ma da meno informazioni e tiene conto solo di quel momento. Come se fosse una foto. Se ho una malattia endemica e faccio la prevalenza non avrei una situazione chiara sotto mano.

MISURE EPIDEMICHE

SLIDE 35

LE MALATTIE INFETTIVE DA UN PUNTO DI VISTA DELLA TRASMISSIONE



MISURE DELLO STATO DI INFEZIONE



RISCHIO INFEZIONE = Rischio di esposizione * Rischio infezione se esposto

RISCHIO MALATTIA = Rischio di esposizione * Rischio di infezione se esposto * Rischio di malattia se infetto

RISCHIO MORTALITA' = Rischio di malattia se infetto * Rischio mortalità se infetto

RISCHIO CURA = rischio cura tra infetti non clinici + rischio di malattia * rischio di cura tra clinici

FATTORI CHE INFLUENZANO LA MISURA DELLA FREQUENZA DI UNA MALATTIA INFETTIVA



Il test diagnostico accurato solo dopo un certo tempo dopo l'infezione (problema risposta immunitaria/patogenesi)

La diagnosi non è sufficientemente accurata (problema test diagnostico)

Segni clinici comuni a diverse malattie (problema clinico)

Episodi ripetuti da patogeni geneticamente diversi (es. sierotipi) danno stessi segni clinici o positività al test

Immunità di breve durata favorisce episodi ripetuti in intervalli temporali brevi

Non tutti i soggetti sono disponibili per i test diagnostici

NEL MOMENTO IN CUI ABBIAMO LA PATOLOGIA ABBIAMO BISOGNO DEGLI INDICATORI EPIDEMIOLOGICI CHE CI PERMETTONO DI CAPIRE LE CARATTERISTICHE DELLA MALATTIA IN TERMINE DI SOGG COLPITI VELOCITÀ DI DIFFUSIONE ECC.

1. EMERGE IL RISCHIO
2. PERCEZIONE DEL RISCHIO
3. INTERVENTI DI CONTENIMENTO: BARRIERE, VACCINI, QUALITÀ ARIA ECC
4. EFFETTI DEL CONTENIMENTO

QUALI SONO I FTTORI CHE CI MISURANO LA VALUTAZIONE:

- **TEST DIAGNOSTICI** → ALL INIZIO DELL'INFEZIONE NON CI SI PUÒ ASPETTARE CHE SIANO ACCURATI DA SUBITO. SONO ACCURATI SOLO DOPO UN CERTO PERIODO DI TEMPO (ES TEST SNAP E TEST MOLECOLARI SU PCR)
- **SEGNI CLINICI** → POTREBBERO ESSERE COMUNI AD ALTRE MALATTIE QUINDI SE FACCIO UNA DIAGNOSI CLINICA POSSO AVERE SOLO UN SOSPETTO. DEVO ASSOCIARLO AD ALTRE INDAGINI DIAGNOSTICHE
- **VARIANTI** → VARIANTI DELLO STESSO PATOGENO POSSONO DARE SINTOMATOLOGIE DIVERSE. DI SOLITO I SIEROTIPI CHE ARRIVANO DOPO IL PICCO DI EPIDEMIA SONO MENO GRAVI XKE IL VIRUS CERCA DI SOPRAVVIVERE NON UCCIDENDO LA POPOLAZIONE MA INFETTANDOLA E FACENDOSI REPLICARE. CON IL VACCINO SPINGO ALLA COMPARSA DI VARIANTI COME STRATEGIA ADATTATIVA DEL PATOGENO. PER LO STESSO MOTIVO LE VARIANTI SONO MENO AGGRESSIVE.
- **IMMUNITÀ** → FAVORISCE NUOVE INFEZIONI. L'IMMUNITÀ DI BREVE DURATA FAVORISCE EPISODI BREVI E RIPETUTI
- **SOGGETTIVITÀ** → NON TUTTI I SOGGETTI SONO DISPONIBILI A TEST DIAGNOSTICI, IL CHE VUOL DIRE UNA SOTTOSTIMA DEGLI EPISODI. (es NEL CASO DEL COVID SE ESCO A FARE IL TEST SONO CONTAGIOSO QUINDI MOLTI PER SENSO DI RESPONSABILITÀ NON LO FANNO). O ANCHE UNA SOVRASTIMA XKE SE IO POSITIVO FACCIO IL TEST 5 VOLTE CONTO COME 5 POSITIVI (PENSA AD ES A CHI FA TAMPONI X LAVORO TIPO CALCIATORI ECC)

DAL PUNTO DI VISTA PRATICO QUALI SONO I **PROBLEMI**???

- **IGNORARE LA CICLICITÀ SETTIMANALE:**

I DATI NON RIGUARDANO MAI IL GIORNO STESSO MA IL GIORNO OPRIMA DURANTE IL FINESETTIMANA I DATI CAMBIANO DRASTICAMENTE (XKE GLI UFFICI SONO CHIUSI E DATI SI ACCUMULANO)

LA FREQUANZA DEI CONTAGI È DETERMINATA DAL CONTEGGIO DEI TEST. SE IO NE FACCIAMO 5 CONTO COME 5 PERSONE DIVERSE

→ BISOGNEREBBE FARE MEDIE MOBILI A 7 GIORNI. ACCUMULO IL DATO DA LUNEDÌ A DOMENICA E IL LUNEDÌ LI FACCIAMO USCIRE. COMINCERÒ A CONTARE DA MARTEDÌ A LUNEDÌ E IL MARTEDÌ USCIRANNO E COSÌ VIA

- NON DIVIDERE PER DENOMINATORE: È FONDAMENTALE. DOBBIAMO SAPERE QUANTI INDIVIDUI A RISCHIO CI SONO A DENOMINATORE.

DEVO CONSIDERARE LA POPOLAZIONE CHE PRENDO IN ESAME. ES COVID NELLE REGIONI CONSIDERO I MALATI MA IN RELAZIONE ALLA POPOLAZIONE REGIONALE. SE LA LOMBARDIA HA UNA GRANDE POPOLAZIONE LA PERCENTUALE SARÀ DIVERSA DA QUELLA DI UNA REGIONE CON UNA POPOLAZIONE MINORE
SERVE ANCHE PER CAPIRE SE C'È UN FATTORE CHE INFLUISCE LA DIFFUSIONE DELLA MALATTIA O SE C'È QUALCHE PECCA NEL CONTROLLO

- CONFONDERE INCIDENZA E PREVALENZA: TUTTI I CASI CHE ARRIVANO – CASI CHE MUONO/GUARISCONO SOLITAMENTE L'INCIDENZA SI NORMALIZZA (DIVIDO LE FREQUENZE CHE OSSERVO PER LA MEDIA), CIOÈ LE AMMISSIONI SONO PARAGONABILI ALLE DIMISSIONI

- QUANDO ATTRIBUIRE LA MORTE? DEVO CONTARE LA MORTE VEDENDO LA FREQUENZA DEI DECESSI E DEI POSITIVI. SE HO IL PICCOLO DI POSITIVI A MARZO E DI MORTE AD APRILE DEVO ESSERE PREPARATO FACENDO UNA STIMA SULLE MORTI

SI CONTA LA MEDIA DEL TEMPO TRA DIAGNOSI E DECESSO E SI STIMA LA LETALITÀ DIVIDENDO IL NUMERO DEI DECESSI PER IL NUMERO DI POSITIVI X GIORNI PRIMA (MEDIA TEMPO TRA DIAGNOSI E MORTE)

- TEST DIAGNOSTICI: NON HA SENSO PARLARE DI TASSO DI POSITIVITÀ PERCHÉ IL MOTIVO PER L'ESECUZIONE DEL TEST CAMBIA. ES I TEST SOTTO GREEN PASS SONO AUMENTATI MOLTO, O PER MOTIVI DI LAVORO, O PER ACCEDERE AD EVENTI. SE NE FACCIAMO 10 CONTO COME 10. SENZA CONTARE L'USO DI TAMPONI PIÙ O MENO EFFICACI (FAI DATE, RAPIDI, PCR).

SI DOVREBBERO FARE TAMPONI A CAMPIONE CASUALE

TASSO DI INCIDENZA → RISCHIO DI INFEZIONE TRA I SOGGETTI ESPOSTI SENSIBILI. DA NON CONFONDERE CON IL RISCHIO DI MALATTIA CHE LA MALATTIA PREVEDE SEGNI CLINICI. NON TUTTI GLI INFETTI DIVENTANO CLINICI. IL RISCHIO DI MALATTIA PIÙ ESSERE DIVERSO AL TASSO DI INFEZIONE

INFETTIVITÀ → L'infettività (trasmissione dell'infezione) è generalmente al livello più alto subito prima della comparsa dei sintomi della malattia e svanisce in genere prima della scomparsa dei sintomi (non vale per tutte le malattie). Non è generalmente possibile calcolare l'infettività se non dopo che si sono osservati casi clinici (o confermati da un punto di vista diagnostico). Se presenti casi confermati si può calcolare l'INTERVALLO SERIALE (lag)

INTERVALLO SERIALE → TEMPO CHE INTERCORRE TRA LA COMPARSA DI MALATTIA E L' INFEZIONE DI NUOVI SOGGETTI. PRIMA DI ESSERE MALATO SEI GIÀ STATO INFETTATO E QUINDI HAI GIÀ ELIMINATO E CONTAGIATO ALTRI SOGGETTI

BASIC REPRODUCTIVE NUMBER (R0) → NUMERO MEDIO DEI CASI SECONDARI IN UNA POPOLAZIONE SENSIBILE DOPO L'INTRODUZIONE DI UN SINGOLO CASO (PAZIENTE 0). DAL PAZIENTE 0 AI NUOVI CASI SI CALCOLA R0. MA CHI È IL PAZIENTE 0?? È QUINDI DIFFICILE DA STABILIRE.

non è una misura della gravità di una malattia infettiva o della rapidità di diffusione di un agente patogeno in una popolazione

R0 SONO I PIÙ FAMOSI MA SONO LEGATI PIÙ ALLA PARTE TEORICA CHE PRATICA

RISCHIO DI INFEZIONE SECONDARIA (R1) → RISCHIO DI INFEZIONE TRA CONTATTI SENSIBILI E OSPITE INFETTO (COMPRENDE INFETTABILITÀ E TRASMISSIBILITÀ)

EFFECTIVE REPRODUCTIVE NUMBER (R) → numero medio di trasmissione di infezione per ogni singolo caso.

A differenza di R_0 che prevede una completa sensibilità della popolazione, in questo caso si assume che vi sia un'immunità che potrà cambiare nel tempo (e così anche R)

- $R < 1$ EPIDEMIA IN RISOLUZIONE
- $R > 1$ EPIDEMIA SI STA DIFFONDENDO FINO A CHE $R < 1$

COSA FACCIAMO? MOLTIPLICHIAMO R_0 CALCOLATO IPOTETICAMENTE CON IL NUMERO DI POPOLAZIONE.

Se assumiamo una popolazione omogenea per i fattori di rischio legati alla diffusione dell'infezione il numero di nuovi casi è proporzionale alla probabilità casuale di contratto tra soggetti infetti e soggetti sensibili

$R = R_0 \times x$ dove x sono soggetti sensibili → SE SIAMO ALL'EQUILIBRIO SARÀ UN VALORE = 1

SE ABBIAMO 60 OVINI SENSIBILI E ARRIVANO UNA NUOVA INFEZIONE CON PICCO A 4 SETTIMANE. LA POPOLAZIONE SENSIBILE COMINCIA A DIMINUIRE (SI AMMALANO E SVILUPPANO IMMUNITÀ)

INDICE R_t → è il rapporto tra una stima degli eventi al tempo t e degli eventi al tempo $t-s$ dove s sono i giorni che si stima siano necessari affinché un caso possa riprodurre un altro.

ABBIAMO BISOGNO DEL T CHE CI METTE UN SOGGETTO AD INFETTARE I SOGGETTI SENSIBILI. SI BASA SUI SEGNI CLINICI. Ai fini dei calcoli $L'R_t$ utilizza la data dell'inizio sintomi

INDICE R_{Dt} → Come R_t a livello di concetto, diverso il metodo di calcolo. Si è PROPOSTO DI USARE R_{Dt} CHE SI BASA NON SULLA CLINICA MA SULLA DIAGNOSTICA, OVVERO CALCOLO I SOGGETTI POSITIVI A TEST PIUTTOSTO CHE QUELLI CLINICI. Ai fini del calcolo dell'intervallo seriale $L'R_{Dt}$ utilizza la data di registrazione della positività $RTD = 2$ VUOL DIRE CHE OGNI SOGGETTO NE INFETTA DUE QUINDI HA UN ANDAMENTO ESPONENZIALE. Più AUMENTA IL TEMPO Più IL VALORE AUMENTA

CAMPIONAMENTO

SCELTA DEI CAMPIONI:

Devo prendere dei campioni per non sfalzare i risultati finali.

Ci sono delle regole da seguire.



Ho una popolazione target (una popolazione grande). Partendo da questa identifico una popolazione studiata (così da restringere il campo) e all'interno di questa tiro fuori il mio campione.

Es ho un milione di bovine frisone. Prendo le bovine lombarde come popolazione studiata e da loro prendo il campione. Quello che ottengo lo posso trasferire alla popolazione di origine

OBIETTIVI:

- Estrarre un campione rappresentativo
- Precisione ed accuratezza. Una volta che identifico il campione devo fare valutazioni che siano sufficientemente accurate
- Minimizzare il n di campioni da prendere, per non sprecare risorse

Quello che voglio ottenere è centrare il mio obiettivo

- ✓ Devo avere assenza di confondimento
- ✓ Devo essere preciso
- ✓ Quando le cose non vanno nel verso giusto ho bassa accuratezza e distorsione (BIAS)

SCELTA DEL CAMPIONE:

- **Casuale semplice** → moneta bilanciata, tombola (finisce a 90 numeri, devo avere massimo 90 soggetti), numeri casuali, computer (excel con funzione random) ← scelta a caso dei soggetti
- **Sistematico** → selezionare ad intervalli precisi (ogni x soggetti) ← es ogni 3 bovine ne scelgo una

- **Casuale stratificato** → Gli animali vengono suddivisi in base alla caratteristica di stratificazione. Si selezionano in modo casuale (semplice) i soggetti da ciascun strato
- **Cluster** → quando ho una popolazione molto grande si mettono in aggregati. Ad es box di cavalli o suini. In questo caso il soggetto non è più l'animale singolo ma il cluster a cui appartiene. Si seleziona casualmente il cluster invece del singolo
- **Cluster multistrato** → come il precedente ma oltre a selezionare casualmente il cluster si selezionano casualmente dei soggetti all'interno del cluster. Per popolazioni veramente grandi ad es tutta l'Italia. Seleziono tutte le asl (1 cluster) da qui indico i paesi sotto la singola asl (2 cluster) e seleziono casualmente il cluster di riferimento (1 paese)
- **Multistadio**

CALCOLO DELLE DIMENSIONI:

- Per calcolare la dimensione del campione riguardante VARIABILI QUANTITATIVE bisogna:
 - stabilire la variabile e la deviazione standard (quanto può essere variabile la misura). Non mi interessa sapere quanto sono alte x persone. Mi interessa sapere qual è la variabile tra loro, in questo caso l'altezza media (valore che si prende di solito da dati già studiati e pubblicati)
 - qual è il livello di errore accettabile (non può essere grande quanto la deviazione standard. Cmq deve essere un valore piccolo.)

Campionamento calcolo delle dimensioni

♦ esempio

! num.camp. = $4S^2/L^2$

S = deviazione standard

L = livello di errore

$4 \cong 1,96^2$

Il n di campioni è il n da prendere CASUALMENTE

Campionamento calcolo delle dimensioni

♦ esempio

! num.camp. = $4S^2/L^2$
SD=20; L=1

animali = $4*400/1$

animali = 1600

Cosa succede se cambio solo SD, solo L, o entrambi?

Campionamento calcolo delle dimensioni

♦ esempio

! num.camp. = $4S^2/L^2$
SD=40; L=5

animali = $4*1600/25$

animali = 256

Campionamento calcolo delle dimensioni

♦ esempio

! num.camp. = $4S^2/L^2$
SD=40; L=1

animali = $4*1600/1$

animali = 6400

Campionamento calcolo delle dimensioni

♦ esempio

! num.camp. = $4S^2/L^2$
SD=20; L=1

animali = $4*400/1$

animali = 1600

ESEMPIO

Campionamento calcolo delle dimensioni
 Pop: 2500 vitelloni
 Quanti campioni per stimare il peso medio?
 Peso medio vero: 191.9 ± 98.6 kg

Formula per n. campioni= $4S^2/L^2$
 Errore accettato (L)=10 kg

$$N = \frac{4 \times 98^2}{10^2} = \frac{38416}{100} \approx 385$$

Campionamento calcolo delle dimensioni
 Media vera: 191.9 ± 98.6

- Se peso 385 animali a caso → 197.6 ± 100.2
- Se peso 50 animali a caso → 205.4 ± 116.9
- Se peso 100 animali a caso → 227.9 ± 99.7
- Se peso 200 animali a caso → 193.3 ± 108.3
- Se peso 600 animali a caso → 199.7 ± 94.9

Anche prendendo un campione più ampio del dovuto (di 385) non è detto che sia meglio.

VARIABILE DICOTOMICA= PREVALENZA

- VARIABILI QUALITATIVE → per calcolare le dimensioni del campione (var. dicotomica) bisogna:
 - stimare la prevalenza
 - stabilire il livello di errore sopportabile
 - stabilire limite confidenza

Campionamento calcolo delle dimensioni
 ♦ esempio leucosi

! num.camp.= $4 \cdot P \cdot Q / L^2$

- P= prevalenza
- Q=1-P
- L=livello di errore
- $4 \approx 1,96^2$

SLIDE 43

CAMPIONAMENTO PER MALATTIA

In programmi di monitoraggio è necessario valutare se la malattia nella popolazione (allevamento) è presente e se è assente (perché non è presente)

Campionamento per malattia

| | | | |
|---------------------|-------|--|----|
| Confidence | 0,95 | Direct calculation | |
| Population size (N) | 5,00% | Percentage of diseased animals in the population | |
| | 50 | | 35 |
| Confidence | 0,95 | Direct calculation | |
| Population size (N) | 5,00% | Percentage of diseased animals in the population | |
| | 100 | | 44 |
| Confidence | 0,95 | Direct calculation | |
| Population size (N) | 5,00% | Percentage of diseased animals in the population | |
| | 250 | | 52 |
| Confidence | 0,95 | Direct calculation | |
| Population size (N) | 5,00% | Percentage of diseased animals in the population | |
| | 1000 | | 57 |

Numero di campioni da prendere

La domanda è: Quanto soggetti devo prendere per giungere ad una delle due conclusioni con una sicurezza del 95%-99% ?

Circoscrivo dei soggetti che possono essere più a rischio

Anche se aumento di molto il n di soggetti (1000 ad es) non mi cambia molto, lavoro solo di più.

Se ne trovo 1 o ne trovo 5 di positivi non mi cambia nulla. Dirò che la malattia nella popolazione c'è. Quanta ce ne sia non lo so, dovrei usare un campionamento per variabile di controllo

Più sono in una condizione sanitaria svantaggiosa meno lavoro devo fare

Campionamento per malattia

- Si applica un formula

$$[1 - (1 - a)^{\frac{1}{D}}] \times (N - \frac{D-1}{2})$$
- Dove
 - a = livello di confidenza (es. 95%)
 - D = numero di soggetti presunti ammalati
 - N = popolazione totale

Il principio è che:

- c'è il livello di confidenza
- il n di soggetti sospetti ammalati (in base alla prevalenza presunta di malattia e il n di soggetti della popolazione)
- la popolazione totale

Delimito la popolazione con maggior rischio all'interno della popolazione di studio

VALUTAZIONE TEST DIAGNOSTICI

SLIDE 51

Approfondisci su https://www.quadernodiepidemiologia.it/epi/screen/def_tst.htm

Vi è una crescente necessità di test diagnostici validati a livello internazionale

- Per la pratica veterinaria per aggiornare i metodi diagnostici tradizionali
- Per i programmi di sorveglianza per identificare il numero di campioni necessario
- Per l'epidemiologia per stimare in modo più accurato i rischi relativi
- Per la gestione del rischio, per ottenere stime più accurate

Che cos'è un test diagnostico ?

- Ogni sistema che riduce l'incertezza sullo stato (di salute) di un soggetto
- Infetto/non infetto
- Protetto/non protetto
- Gravido/non gravido

Cosa prevede l'effettuazione di un test?

- Che vi sia un protocollo operativo standard

L'OIE ha stabilito che:

- Deve essere definito un protocollo preciso per effettuare un test

- Prima di procedere alla valutazione dell'accuratezza del test stesso
- Devono essere inoltre previste delle procedure per mantenere e migliorare l'accuratezza del test stesso

Le performances di un test diagnostico sono descritte da

- Precisione: concordanza (valori simili) tra misure ripetute sullo stesso campione
- Accuratezza: concordanza (vicinanza) tra il valore misurato ed il valore vero
Un sinonimo di accuratezza è validità

Il test diagnostico **ideale** da sempre la risposta corretta:

- positivo nel soggetto ammalato
- negativo nel soggetto sano

NON ESISTE UNA POSSIBILITÀ CHE SIA SENSIBILE AL 100%
CI SONO SISTEMI CHE PERMETTONO DI IDENTIFICARNE LA SENSIBILITÀ

I TEST (IDEALI) DEVONO ESSERE:

- VELOCI → NON È DETTO CHE UN TEST VELOCE SIA MIGLIORE IN TERMINI DI ACCURATEZZA
- SICURO → ACCURATO
- SEMPLICE → X EVITARE DI COMPIERE ERRORI
- NON INVASIVO → MENO È INVASIVO MEGLIO È
- AFFIDABILE → DEVE ESSERE POSSIBILE UTILIZZARLO ANCHE IN CONDIZIONI NON OTTIMALI (ES CONDIZIONI DI CAMPO)
- ECONOMICO → SPT IN VETERINARIA IL COSTO DIPENDE DAL BENEFICIO

REGOLE PER VALUTARE O SVILUPPARE UN NUOVO TEST DIAGNOSTICO

1. CHIEDERSI: È NECESSARIO AVERE UN NUOVO TEST?? SE HO GIÀ UN TEST ACCURATO CON I CRITERI GIUSTI ME NE SERVE UNO NUOVO? PROBABILMENTE NO
2. CAPIRE COME SONO STATI INCLUSI I SOGGETTI UTILIZZATI PER PROVARE CHE IL TEST FUNZIONA SE AD ES IL TEST È STATO DEFINITO SOLO SU SOGGETTI MALATI ANCHE IL PEGGIOR TEST SARÀ ACCURATO (TRUFFA COMMERCIALE). QUINDI DEVO SAPERE I DATI DA DOVE SONO STATI PRESI.
3. DEFINIRE UN GOLD STANDARD (CONFRONTANDOLO CON UN TEST DI RIFERIMENTO) RAGIONEVOLE, CREDIBILE E UTILIZZABILE. IL GOLD STANDARD LO STABILISCE LA LEGGE. X STABILIRE CHE UN SOGGETTO ABBAIA UNA MALATTIA LO STABILISCE UN TEST A LIVELLO STATISTICO, DAL QUALE POI VENGONO CONFRONTATI TUTTI GLI ALTRI
4. ASSICURARSI CHE IL TEST E IL GOLD STANDARD SIANO APPLICABILI AI SOGGETTI IN MODO EGUALE E STANDARD
5. definire una dimensione del campione adeguata ad avere stime di sensibilità e specificità con limiti di confidenza almeno 95%
6. assicurarsi di avere abbastanza soggetti per raggiungere le dimensioni richieste
7. alla fine dichiarare : •sensibilità •specificità •valori predittivi •prevalenza

IN CHE STUDI UTILIZZO TEST DIAGNOSTICI:

- Stima della prevalenza → stimare la frequenza di un "tratto" nella popolazione.
- Studio sui fattori di rischio → Stimare associazione tra fattori di rischio e "tratti" di interesse (sanitari o meno)
- Quantificazione dei rischi → stimare il rischio associato al movimento (commercio) di animali o di prodotti biologici

TIPO DI DATI:

- Continui (es. valori ELISA OD)
- Ordinali (es. OD corrispondente diluizione)
- Dicotomici (es. OD x positivo OD x negativo)

PER VALUTARE UN TEST DIAGNOSTICO È NECESSARIO FARE UNO STUDIO OSSERVAZIONALE DOVE HO DUE VARIABILI: RISULTATO DEL TEST (variabile predittiva) E MALATTIA (variabile risposta). PER LA MALATTIA VIENE DEFINITO UN GOLD STANDARD.

IL **GOLD STANDARD** IDEALE È POSITIVO PER GLI AMMALATI E NEGATIVO NEI SANI

| DEFINIZIONI | | malattia | |
|-------------|---------|----------|------|
| | | pres. | ass. |
| test | | | |
| pos. | vero + | falso + | |
| neg. | falso - | vero - | |

| DEFINIZIONI | | malattia | | |
|-------------|-----|----------|------|-----|
| | | pres. | ass. | tot |
| test | | | | |
| pos. | a | b | a+b | |
| neg. | c | d | c+d | |
| Tot. | a+c | b+d | | |

LA **SENSIBILITÀ** DEL TEST È LA PROPORZIONE DI SOGGETTI AMMALATI CHE SONO POSITIVI AL TEST DIAGNOSTICO (PIÙ È SENSIBILE PIÙ È VICINO AL GOLD STANDARD)

$$\text{SENSIBILITÀ} = a/a+c$$

LA **SPECIFICITÀ** DI UN TEST È LA PROPORZIONE DI SOGGETTI SANI CHE SONO NEGATIVI AL TEST DISGNOSTICO

$$\text{SPECIFICITÀ} = d/b+d$$

IL **VALORE PREDITTIVO POSITIVO** È LA PROBABILITÀ CHE UN SOGGETTO POSITIVO AL TEST SIA AMMALATO

$$\text{VPP} = a/a+b$$

IL **VALORE PREDITTIVO NEGATIVO** È LA PROBABILITÀ CHE UN SOGGETTO NEGATIVO AL TEST SIA SANO

$$\text{VPN} = d/c+d$$

UN BUON TEST HA:

ELEVATA SENSIBILITÀ → pochi falsi negativi

BUONA SPECIFICITÀ → pochi falsi positivi

FATTORI CHE INFLUENZANO SENSIBILITÀ E SPECIFICITÀ:

I valori di sensibilità e specificità variano in genere da uno studio all'altro

- Differenti popolazioni di riferimento
- Differenti strategie di campionamento
- Errori sistematici e casuali
- Altre variabili "tecniche": Abilità tecnica
 - Scelta del gold standard
 - Valore di cut-off
 - Gestione dei dati dubbi

La valutazione di sensibilità e specificità per strato permette delle stime più precise rispetto ad una valutazione complessiva (cruda). Devono essere considerati in generale I fattori biologici che influenzano la sensibilità e la specificità di un test diagnostico

FATTORI BIOLOGICI CHE INFLUENZANO LA SENSIBILITÀ:

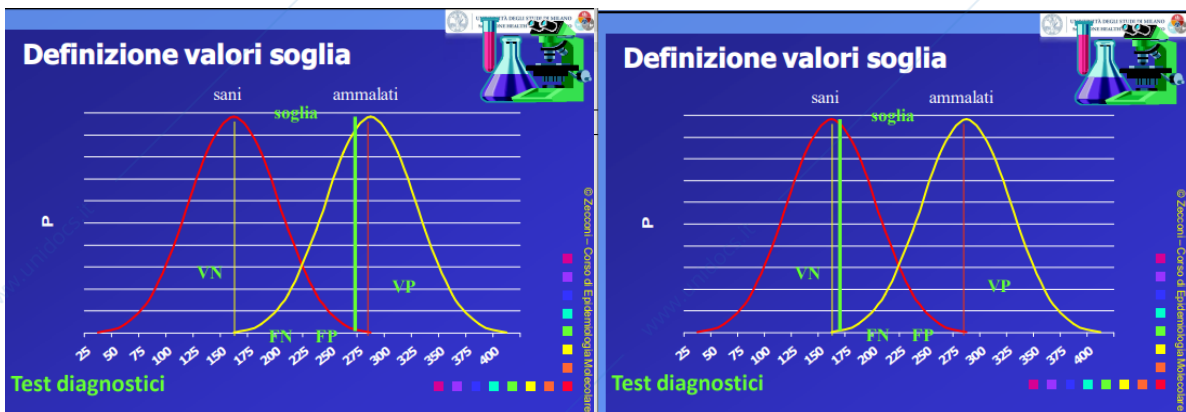
- Stadio dell'infezione (curva epidemica)
- Reazioni crociate
- Falsi positivi da reazioni aspecifiche (es. positività antibiotici forme cliniche)
- Selezione popolazione

- Programmi vaccinali
- Anticorpi colostrali
- Programmi di eradicazione

Quando il risultato del test è una variabile continua è necessario definire un valore soglia (cutoff) ovvero il valore di equilibrio tra sensibilità e specificità

Se devono essere evitati i falsi positivi (es. positivi macellati) → il valore soglia deve essere posto in modo da massimizzare la specificità

Se devono essere evitati i falsi negativi (es. zoonosi) → il valore soglia deve essere posto in modo da massimizzare la sensibilità



Un sistema per definire il valore soglia appropriato è l'utilizzo della curva Receiver Operator Characteristics (ROC)
 Si deve determinare: la sensibilità e la specificità per diversi valori soglia quindi si confrontano graficamente sensibilità vs 1-specificità (tasso falsi +)

NON LO CAPISCO CI SONO SOLO GRAFICI

Copiato da https://www.quadernodiepidemiologia.it/epi/screen/vp_test.htm


Anche per la comprensione del valore predittivo di in test conviene, come già avvenuto a proposito della sensibilità e della specificità, tenere presente la ormai ben nota (v. Unità precedente) Tabella di contingenza a due entrate, in cui le colonne indicano lo stato reale dell'animale (M+, malato; M-, sano) e le righe l'esito del test (T+, positivo, T-, negativo):

| | | | |
|-----------|-----------|-----------|-----|
| | M+ | M- | |
| T+ | a | b | a+b |
| T- | c | d | c+d |
| | a+c | b+d | |

totale test-positivi (a+b)
 totale test-negativi (c+d)
 totale malati (a+c)
 totale sani (b+d)

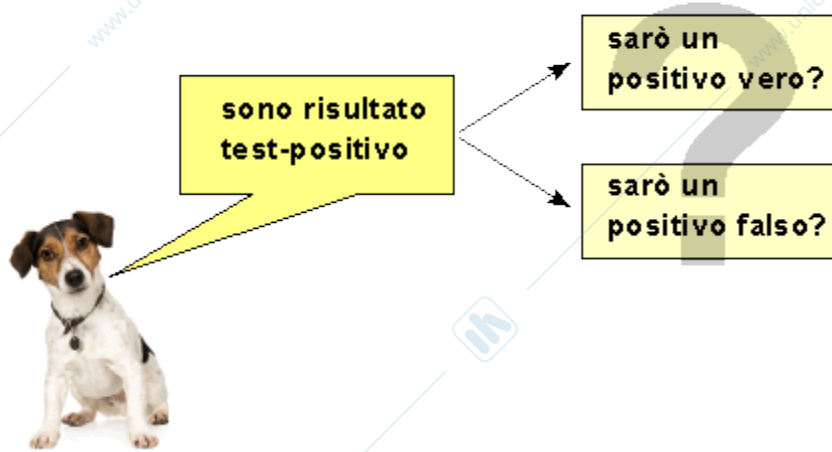
È importante tenere ben separati i concetti di sensibilità (Se) e specificità (Sp) dal valore predittivo. La Se e la Sp sono caratteri propri del test e legati al suo intimo funzionamento. In genere essi sono

dichiarati dal produttore e sono (o dovrebbero essere) noti prima dell'applicazione del test sulla popolazione: Se e Sp sono probabilità *pre-test*.

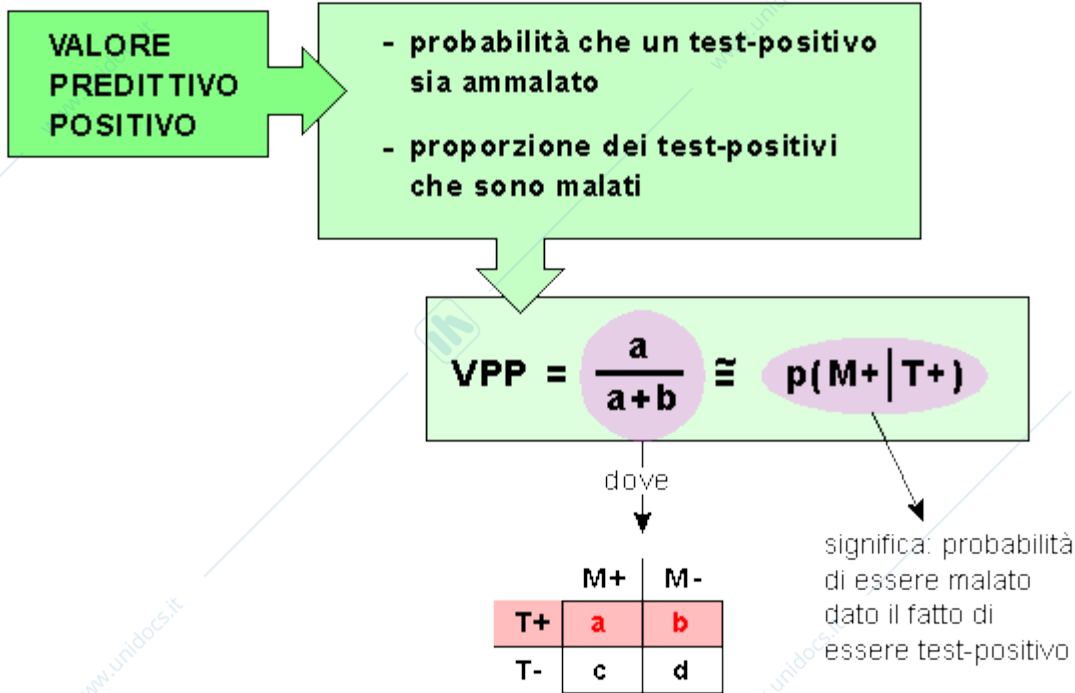
Tuttavia, una volta eseguito il test, Se e Sp perdono importanza. Infatti, dopo uno  *screening* di popolazione, o nell'attività diagnostica ambulatoriale, oppure - più in generale - quando si deve interpretare il risultato di un test, diventano importanti due probabilità *post-test*. Infatti, è auspicabile avere un'idea della quota di soggetti realmente ammalati (positivi veri, cella a) sul totale degli animali risultati positivi al test (celle a+b). Analogamente, è bene conoscere la quota di soggetti realmente sani, (negativi veri, cella d) sul totale dei negativi al test (celle c+d).

VALORE PREDITTIVO POSITIVO:

Non tutti gli animali positivi a un test sono realmente ammalati. Quindi, in presenza di un animale positivo a un test, occorre rispondere alla seguente domanda:

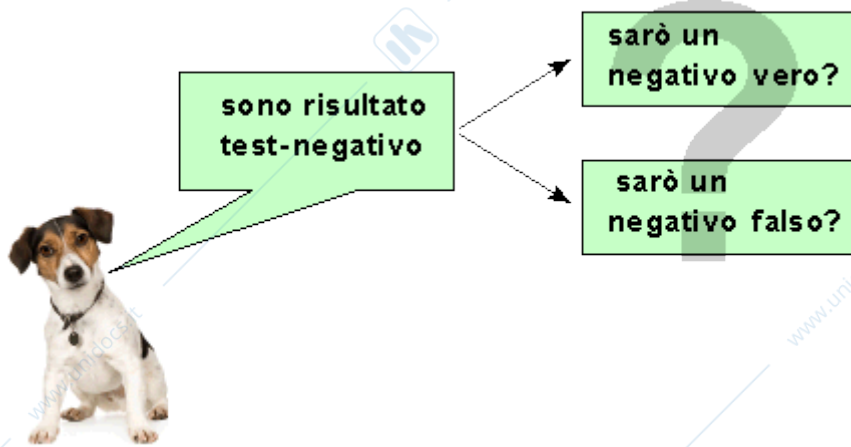


La risposta a questa domanda può essere data solo *in termini di probabilità*, calcolando il valore predittivo positivo (VPP) che indica la probabilità che un animale test-positivo sia davvero ammalato. Esso rappresenta la proporzione di animali test-positivi che sono ammalati. A partire dalla tabella di contingenza, il VPP si calcola così: $a/(a+b)$.

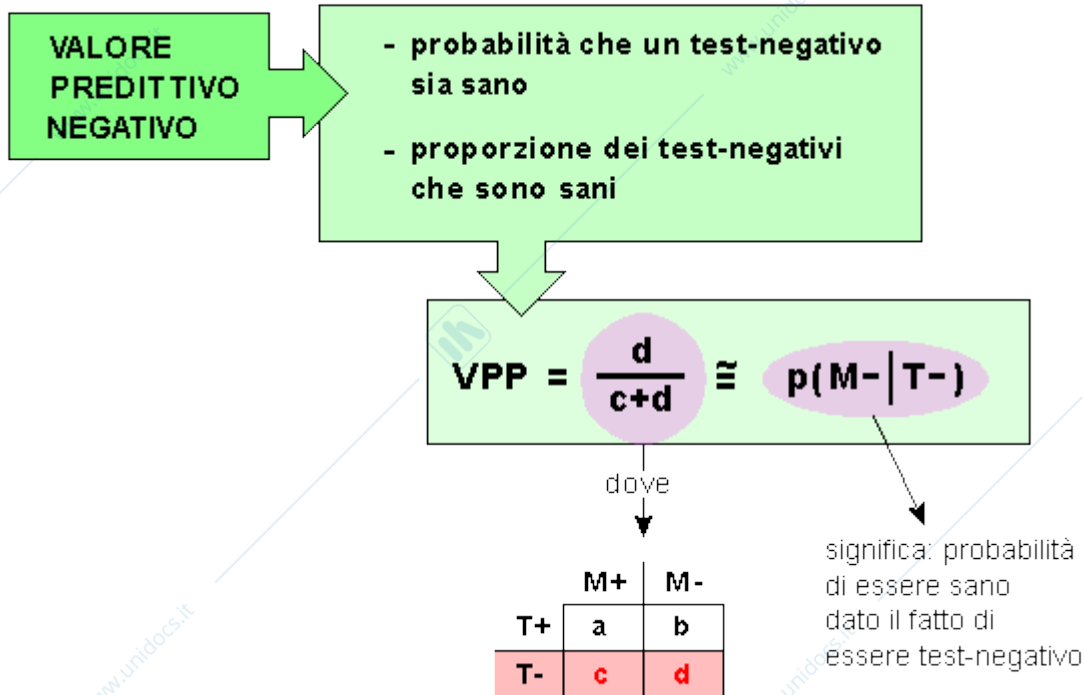


VALORE PREDITTIVO NEGATIVO:

Non tutti gli animali negativi a un test sono realmente sani. Ecco quindi che, in presenza di un animale negativo a un test, occorre rispondere alla seguente domanda:



Anche in questo caso, la risposta può essere data solo *in termini di probabilità*, calcolando il valore predittivo negativo (VPN) che indica la probabilità che un animale test-negativo sia davvero sano. Esso rappresenta la proporzione di animali test-negativi che sono sani. A partire dalla tabella di contingenza, il VPN si calcola così: $c/(c+d)$.



TEOREMA DI BAYES

FIN'ORA ABBIAMO PARLATO DI VALORI PREDITTIVI, ASSUMENDO DI CONOSCERE NON SOLO IL RISULTATO DEL TEST MA ANCHE LO STATO REALE (SANO O MALATO) DELL'ANIMALE. QUESTA ASSUNZIONE PUO ESSERE FATTA SOLO A SCOPO DIDATTICO, PERCHE NELLA REALTÀ NON È PROPRIO COSÌ.

BAYES CI PERMETTE DI CALCOLARE IL VPP E IL VPN CONOSCENDO **SPECIFICITÀ** E **SENSIBILITÀ** DI UN TEST, MA SOPRATTUTTO LA **PREVALENZA**

DI SOLITO QUANDO SVILUPPO UN TEST LO COSTRUISCO IO E USO UNA PREVALENZA IDEALE, CHE DECIDO IO PER COME MI È COMODO, MA NEL MOMENTO IN CUI TORNO ALLA CONDIZIONE REALE LA PREVALENZA POTREBBE ESSERE MOLTO DIVERSA. QUINDI DISTINGUO UNA PREVALENZA APPARENTE DA UNA REALE.

BAYES MI PERMETTE DI APPLICARE ALLA REALTÀ IL TEST SVILUPPATO CON DATI IDEALI, SENZA DOVERMI CALCOLARE NUOVAMENTE SENSIBILITÀ E SPECIFICITÀ (se voglio posso calcolarle ma perdo tempo)

Sensibilità e specificità sono calcolate utilizzando una popolazione con una prevalenza definita, motivo per cui i valori predittivi sono prefissati Nella realtà la prevalenza delle malattie possono essere anche molto diverse, e quindi possono essere molto diversi i valori predittivi

IL TEOREMA DI BAYES CI CONSENTE DI OTTENERE, TRAMITE UNA FORMULA, LA PREVALENZA REALE A PARTIRE DA QUELLA APPARENTE, CONOSCENDO SP E SE DI UN TEST.

VIENE IMPIEGATO PER CALCOLARE LA PROBABILITÀ DI UNA CAUSA CHE HA PROVOCATO L'EVENTO PERIFICATO. NEL CASO DELLE MALATTIE CALCOLA LA PROBABILITÀ CHE UN ANIMALE SOFFRA DELLA MALATTIA PER CUI HA ESEGUITO IL TEST DISGNOSTICO

VPP= PROBABILITÀ CHE UN SOGGETTO POSITIVO AL TEST SIA REALMENTE AMMALATO

VPN = PROBABILITÀ CHE UN SOGGETTO NEGATIVO AL TEST SIA REALMENTE SANO

il VPN (ed anche il VPP) sono correlati alla sensibilità e alla specificità del test, ma dipendono anche dalla prevalenza della malattia nella popolazione.

- Il VPP diminuisce con il diminuire della prevalenza della malattia → la diminuzione della prevalenza comporta l'incremento degli animali sani; ciò, a sua volta, fa sí che aumenti il numero di esiti positivi falsi.
- il VPN aumenta con il diminuire della prevalenza → aumentano le possibilità che ci siano test falsi negativi

Per questo motivo, anche conoscendo la sensibilità e la specificità del test, non è possibile rispondere direttamente alla domanda dell'allevatore, a meno di non conoscere (o stimare) la prevalenza. In tal caso, si può applicare il teorema di Bayes, usando la seguente formula:

Il teorema di Bayes

$$VP+ = \frac{\text{sens} * \text{prev}}{\text{sens} * \text{prev} + [(1-\text{spec}) * (1-\text{prev})]}$$

Test diagnostici

Ovvero:

Il teorema di Bayes

$$VP+ = \frac{\text{prob (vero +)}}{\text{prob (vero+) + prob (falso+)}}$$

Test diagnostici

Il teorema di Bayes

$$VP- = \frac{\text{spec} * (1-\text{prev})}{\text{spec} * (1-\text{prev}) + [(1-\text{sens}) * \text{prev}]}$$

Test diagnostici

Il teorema di Bayes

$$VP- = \frac{\text{prob (vero -)}}{\text{prob (vero-) + prob (falso-)}}$$

Test diagnostici

ESEMPIO. Un bovino appartenente a un gruppo in cui si stima che la prevalenza della brucellosi sia pari a 0.20 viene sottoposto con esito negativo al test di agglutinazione rapida al Rosa bengala ($Se=0.620$ e $Sp=0.995$). Ci si domanda qual è: la probabilità quell'animale sia ammalato. Applicando la formula soprastante ottieni:

$$1.VPN = \frac{(0.2 * 0.380)}{[(0.2 * 0.380) + (0.8 * 0.995)]} = 0.0872 \text{ (8.72\%)}$$

$$1 - VPN = \frac{P * (1 - Se)}{P * (1 - Se) + [(1 - P) * Sp]}$$

probabilità che un animale test-negativo sia ammalato

dove: P = prevalenza
 Se = sensibilità del test
 Sp = specificità del test

La probabilità di un evento è compresa tra 0 (non si verifica) e 1 (si verifica) per questo 1-vpn è la probabilità della variabile predittiva negativa

Esempio valutazione test diagnostico

| | SN PRESENTE | | ASSENTE | | tot | % |
|----------|-------------|--------|---------|--------|-----|--------|
| | n | % | n | % | | |
| ELISA | | | | | | |
| POSITIVO | 131 | 32,83% | 52 | 13,03% | 183 | 45,86% |
| NEGATIVO | 8 | 2,01% | 208 | 52,13% | 216 | 54,14% |
| tot | 139 | 34,84% | 260 | 65,16% | 399 | 100% |

Diagnosi IBR confronto SN vs Elisa
Test diagnostici

| | |
|----------------------------|--------|
| sensibilità | 94,24% |
| specificità | 80,00% |
| accuratezza | 84,96% |
| prevalenza | 34,84% |
| valore predittivo positivo | 71,58% |
| valore predittivo negativo | 96,30% |
| likelihood positivo | 4,71 |
| likelihood negativo | 0,07 |

Diagnosi IBR confronto SN vs Elisa
Test diagnostici

DAL PUNTO DI VISTA PRATICO IL RISULTATO SARÀ CHE ALCUNI TEST SONO PIU INDICATI PER RILEVARE I SANI, IN ALTRI PER RILEVARE I MALATI

Esempi Bayes

MedCalc: Bayesian Analysis Model

Enter PREVALENCE, SENSITIVITY, and SPECIFICITY:

PREV: 0.2 SENS: 0.98 SPEC: 0.9

Calculate Reset

Or enter TP, FN, TN, and FP:

| N = 1000 | Disease | No Disease |
|---------------|---------|------------|
| Positive Test | TP: 196 | FP: 80 |
| Negative Test | FN: 4 | TN: 720 |

PREV (ipotizzata) 20%
SENS 95%
SPEC 90%

VPP=70% → probabilità del 70% ammalato, 30% falsi positivi, non malati

VPN=99% → probabilità 99% che i negativi siano veramente sani, 1% falsi negativi, malati.

Esempi Bayes

MedCalc: Bayesian Analysis Model

Enter PREVALENCE, SENSITIVITY, and SPECIFICITY:

PREV: 0.4 SENS: 0.9 SPEC: 0.98

Calculate Reset

Or enter TP, FN, TN, and FP:

| N = 1000 | Disease | No Disease |
|---------------|---------|------------|
| Positive Test | TP: 360 | FP: 12 |
| Negative Test | FN: 40 | TN: 588 |

Positive Predictive Value = 0.9677
Negative Predictive Value = 0.9383
Positive Likelihood Ratio = 45
Negative Likelihood Ratio = 0.102

Pre-test Probability = 0.4 (prevalence)
Post-test Probability = 0.9677

QUESTO TEST SARÀ PIU INDICATO PER VEDERE QUANTI SANI CI SONO, PIUTTOSTO CHE QUANTI MALATI

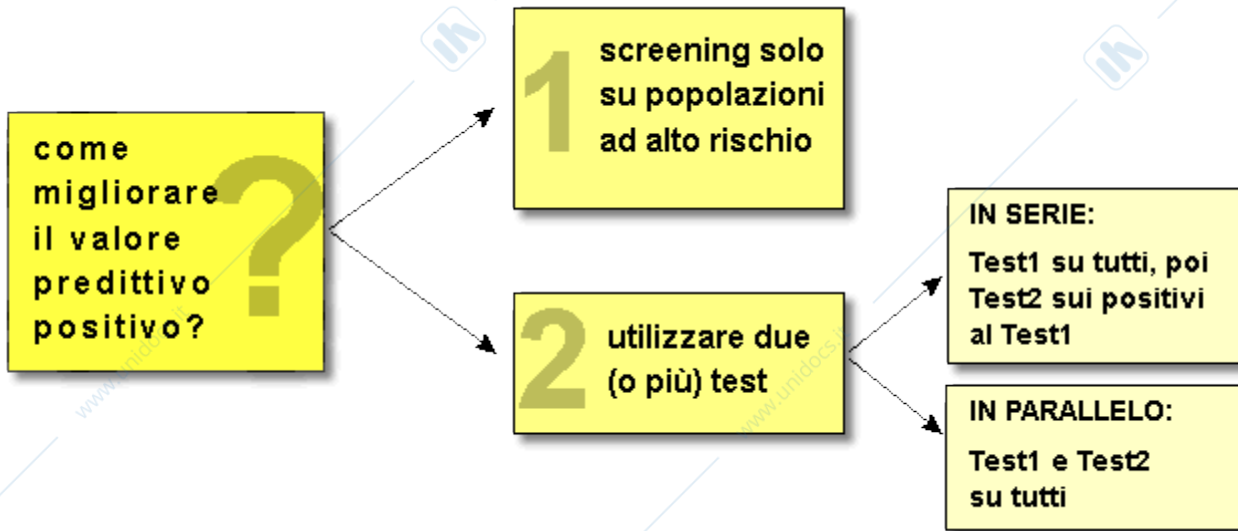
PREV 40%
SENS 90%
SPEC 98%

VPP= 96%
VPN=90% → rispetto all'esempio di prima ho un

problema sui falsi negativi, ho perso 9%

PER AVERE UN VPP ELEVATO, COS' DA NON AVERE FALSI POSITIVI DEVO CONSIDERARE CHE A SECONDA DELLA PREVALENZA DOVRÒ PREFERIRE UN TEST PIU SENSIBILE, O UN TEST PIU SPECIFICO.

ANCHE CAMBIANDO LA PREVALENZA, UN TEST SARÀ MEGLIO DI UN ALTRO



Es:

- TEST1: Se=0.937, Sp 0.987
- TEST2: Se=0.981, Sp=0.990; questo test è più costoso e più sensibile del TEST1

Sottoponi al TEST1 tutti gli 8000 soggetti della popolazione, e poi applichi il TEST2 solo su quelli risultati positivi al TEST1.

Per fare l'esempio (nella realtà non lo potrei sapere) poteri soprannaturali che ci fanno conoscere lo stato reale (malato/sano) di ciascuno di questi 8000 animali. In particolare, sappiamo che 111 di essi sono ammalati e 7889 sono sani.

| TEST 1 | malati | sani | totale |
|--------|--------|------|--------|
| test + | 104 | 100 | 204 |
| test - | 7 | 7789 | 7796 |
| totale | 111 | 7889 | 8000 |

questi animali vengono sottoposti al TEST2

| TEST 2 | malati | sani | totale |
|--------|--------|------|--------|
| test + | 102 | 1 | 103 |
| test - | 2 | 99 | 101 |
| totale | 104 | 100 | 204 |

- 103 positivi (di cui 102 positivi veri e 1 positivo falso)
- 7897 negativi (7796 negativi al TEST1 + 101 negativi al TEST2). Fra questi 7987 negativi, 7888 sono negativi veri e 7+2 negativi falsi.

Ora puoi calcolare Se, Sp e VPP complessive, ossia quelle ottenute utilizzando i due test in serie:

SENSIBILITA' = $\frac{\text{positivi veri TEST2}}{\text{ammalati}} = \frac{102}{111} = 0.919$

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 104 | 100 | 204 |
| test - | 7 | 7789 | 7796 |
| totale | 111 | 7889 | 8000 |

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 102 | 1 | 103 |
| test - | 2 | 99 | 101 |
| totale | 104 | 100 | 204 |

SPECIFICITA' = $\frac{\text{neg. veri TEST1} + \text{neg. veri TEST2}}{\text{sani}} = \frac{7789 + 99}{7889} = 0.999$

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 104 | 100 | 204 |
| test - | 7 | 7789 | 7796 |
| totale | 111 | 7889 | 8000 |

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 102 | 1 | 103 |
| test - | 2 | 99 | 101 |
| totale | 104 | 100 | 204 |

VALORE PREDITTIVO POSITIVO = $\frac{\text{positivi veri TEST2}}{\text{totale positivi TEST2}} = \frac{102}{103} = 0.990$

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 104 | 100 | 204 |
| test - | 7 | 7789 | 7796 |
| totale | 111 | 7889 | 8000 |

| | malati | sani | totale |
|--------|--------|------|--------|
| test + | 102 | 1 | 103 |
| test - | 2 | 99 | 101 |
| totale | 104 | 100 | 204 |

VEROSIMIGLIANZA

(Rapporto di verosimiglianza = RV)

IL RAPPORTO DI VEROSIMIGLIANZA (O LIKELIHOOD RATIO) È UN SISTEMA CHE CI PERMETTE DI VALUTARE IL TEST DIAGNOSTICO IN CLASSI, E NON SOLO IN POSITIVO O IN NEGATIVO.

MOLTI TEST CHE SI UTILIZZANO SONO DIVISI IN CLASSI O VARIABILI CONTINUE (difficili da usare).

IL RV VA AD INTEGRARE NEL SUO VALORE LE INFORMAZIONI CHE CI SONO NELLA TABELLA 2X2

VEROSIMIGLIANZA = RV (rapporto di verosimiglianza) = LR (likelihood ratio)

LR += VEROSIMIGLIANZA POSITIVA

| | | malattia | | |
|------|-----|----------|------|-----|
| test | | pres. | ass. | tot |
| pos. | a | b | a+b | |
| neg. | c | d | c+d | |
| Tot. | a+c | b+d | | |

$LR+ = [a/(a+c)]/[b/(b+d)]$

COME ABBIAMO VISTO PRIMA

$a/a+c \rightarrow$ sensibilità

$b/b+d \rightarrow$ complemento della specificità, ovvero 1-specificità

Quindi la verosimiglianza + sarà il rapporto tra sensibilità e specificità, ovvero tra la probabilità che un soggetto con la malattia abbia test + (vero +) e la probabilità che un soggetto sano abbia test + (falso +)

$LR+ =$ Veri positivi/ falsi positivi

SE CONOSCO UN TEST DI CUI CONOSCO QUESTI DUE VALORI POSSO CALCOLARE LA VEROSIMIGLIANZA

$c/a+c \rightarrow$ complemento della sensibilità, ovvero 1-sensibilità

$d/b+d \rightarrow$ SPECIFICITÀ

| | | malattia | | |
|------|-----|----------|------|-----|
| test | | pres. | ass. | tot |
| pos. | a | b | a+b | |
| neg. | c | d | c+d | |
| Tot. | a+c | b+d | | |

$LR- = [c/(a+c)]/[d/(b+d)]$

quindi la verosimiglianza – sarà il rapporto tra la probabilità che un soggetto ammalato abbia test – (falso -), e la probabilità che un soggetto sano abbia test – (vero -)

LR -= FALSO POSITIVO / VERO NEGATIVO

VEROSIMIGLIANZA POSITIVA= PROBABILITÀ CHE UN SOGGETTO POSITIVO AL TEST SIA AMMALATO → STESSA COSA DEL VPP

VEROSIMIGLIANZA NEGATIVA= PROBABILITÀ CHE UN SOGGETTO NEGATIVO AL TEST SIA AMMALATO **NON è LA STESSA COSA DEL VPN!!!**

Il LR+ ideale ha valore tendente a infinito

Esempio

| | Malattia | Sanità | Totale |
|--------|----------|--------|--------|
| Test + | 731 | 270 | 1001 |
| Test - | 78 | 1500 | 1578 |
| Totale | 809 | 1770 | 2579 |

Sensibilità 731/809= 90%
 Specificità 1500/1770=85%
 Val.Pred.Pos. 731/1001=73%
 Val.Pred.Neg. 1500/1578=95%
 Prevalenza 809/2579=31%

IL LR- ideale ha valore = 0

Esempio:

731 veri +

78 falsi -

Sensibilità (Veri + / tot malati) 90%

Specificità (Veri - / tot sani) 85%

Da qui calcolo

VPP (veri + / tot +)73%

VPN (veri -/tot -) 95%

E possono conoscere

Prevalenza reale (Tot sogg malati / tot soggetti) 31%

CALCOLO LR +

(cambio le percentuali in decimali)

Esempio

Sensibilità 731/809= 90%
 Specificità 1500/1770=85%
 Val.Pred.Pos. 731/1001=73%
 Val.Pred.Neg. 1500/1578=95%
 Prevalenza 809/2579=31%

Probabilità di un test positivo in presenza di malattia

$$LR+ = \frac{Se}{1 - Sp}$$

$$LR+ = \frac{0,90}{1 - 0,85} = 6$$

Test diagnostici

CALCOLO LR -

Esempio

Sensibilità 731/809= 90%
 Specificità 1500/1770=85%
 Val.Pred.Pos. 731/1001=73%
 Val.Pred.Neg. 1500/1578=95%
 Prevalenza 809/2579=31%

Probabilità di un test negativo in presenza di malattia

$$LR- = \frac{1 - Se}{Sp}$$

$$LR- = \frac{1 - 0,90}{0,85} = 0,11$$

Più LA VEROSIMIGLIANZA POSITIVA è ALTA MEGLIO è. TRA DUE TEST SCEGLIERO QUELLO CON LR + PIU ALTO

NEL CASO DELLA VEROSIMIGLIANZA – VOGLIO UN NUMERO PIU BASSO POSSIBILE

SE IL VALORE DI LR + E LR - SI AVVICINA A 1 IL TEST è INUTILE

COME SI INTERPRETANO I RISULTATI DELLA VEROSIMIGLIANZA POSITIVA??

Rapporto di verosimiglianza positivo

| LR | Interpretazione | Utilità |
|------|--------------------------------|---------------------|
| >10 | Aumento della P di malattia | Conclusivo |
| 5-10 | Moderato aumento P di malattia | Moderatamente utile |
| 2-5 | Piccolo aumento P di malattia | A volte utile |
| 1-2 | Minimo aumento P di malattia | Raramente utile |
| 1 | Nessun aumento P di malattia | Inconclusivo |

NELLA PRATICA SE DEVO UTILIZZARE UN TEST DI QUESTO TIPO DEVO STARE NELLE PRIME DUE CLASSI, SE USO LE ALTRE SONO DISPERATO PERCHE NON HO UN RISULTATO UTILE.

SE VOGLIO UN TEST CHE MI INDIVIDUI SOGGETTI AMMALATI DEVO AVERE UN RISULTATO TRA 5 E 10

COME SI INTERPRETANO I RISULTATI DELLA VEROSIMIGLIANZA POSITIVA?? (ricerca dei soggetti ammalati)

Rapporto di verosimiglianza negativo

| LR | Interpretazione | Utilità |
|---------|----------------------------------|---------------------|
| <0,1 | Riduzione della P di malattia | Conclusivo |
| 0,1-0,2 | Moderata riduzione P di malattia | Moderatamente utile |
| 0,2-0,5 | Piccola riduzione P di malattia | A volte utile |
| 0,5-1 | Minima riduzione P di malattia | Raramente utile |
| 1 | Nessuna riduzione P di malattia | Inconclusivo |

SE VOGLIO UN TEST CHE MI INDIVIDUI SOGGETTI AMMALATI DEVO AVERE UN RISULTATO <0.2

COME ABBIAMO VISTO CON BAYES ANCHE QUI DOBBIAMO RICONDURRE IL TEST ALLA REALTÀ, QUINDI DOBBIAMO CONSIDERARE LA PREVALENZA.

SE PRIMA PARLAVAMO DI CONDIZIONI TEORICHE ADESSO PARLIAMO DI REALTÀ.

COMBINIAMO QUINDI LA VEROSIMIGLIANZA CON LA **PROBABILITÀ A PRIORI (ODDS)**, OVVERO QUELLA PRE TEST CALCOLIAMO L'ODDS PRE TEST E LA METTIAMO IN RELAZIONE CON LA LR +, E POI CON LA LR -. DOPODICHE NE CALCOLIAMO LA PROBABILITÀ POST TEST. COSÌ FACENDO POSSIAMO INDIVIDUARE LA PROBABILITÀ IN UNA SITUAZIONE DI REALTÀ

Esempio

Sensibilità 731/809= 90%
 Specificità 1500/1770=85%
 Val.Pred.Pos. 731/1001=73%
 Val.Pred.Neg. 1500/1578=95%
 Prevalenza 809/2579=31%

Probabilità pre-test
(Prevalenza)

$$Pt.odds = \frac{Pr}{1 - Pr}$$

$$Pt.odds = \frac{0,31}{1 - 0,31} = 0,45$$

Esempio

Sensibilità $731/809=90\%$
 Specificità $1500/1770=85\%$
Val.Pred.Pos. $731/1001=73\%$
 Val.Pred.Neg. $1500/1578=95\%$
 Prevalenza $809/2579=31\%$

Odds post test = Pre-test.odds x LR+
 Odds post test = $0,45 \times 6 = 2,7$

$$Pr.post = \frac{Odds.post}{1 + Odds.post} = \frac{2.7}{1 + 2.7} = 73\%$$

Il soggetto ha il 73% di probabilità di essere ammalato, dato il test positivo

Esempio

Sensibilità $731/809=90\%$
 Specificità $1500/1770=85\%$
 Val.Pred.Pos. $731/1001=73\%$
 Val.Pred.Neg. $1500/1578=95\%$
 Prevalenza $809/2579=31\%$

Odds post test = Pre-test.odds x LR-
 Odds post test = $0,45 \times 0,11 = 0,05$

$$Pr.post = \frac{Odds.post}{1 + Odds.post} = \frac{0,05}{1 + 0,05} = 4,7\%$$

Il soggetto ha il 4,7% di probabilità di essere ammalato, dato il test negativo

TIRANDO LE SOMME, A COSA MI SERVE TUTTO QUESTO?

A CAPIRE SE UN TEST MI PUO ESSERE UTILE O NO.

MI FACCIO DARE SE E SP DEL TEST, CONOSCO LA PREVALENZA DELLA MALATTIA, FACCIO I CONTI PER CAPIRE QUANT'È LA VEROSIMIGLIANZA E LA PROBABILITÀ POST TEST. COSÌ FACENDO HO ANCHE UNO STRUMENTO PER PARAGONARE DUE TEST DIVERSI E SCEGLIERE IL MIGLIORE

MISURE DI ASSOCIAZIONE

DEVO POTER MISURARE UN FATTORE DI RISCHIO PER POTER DIRE CHE È LA CAUSA DI QUALCOSA

Tabella 2x2

| | Malattia | Sanità | Totale |
|----------|----------|--------|--------|
| Esposto | a | b | a+b |
| Non espo | c | d | c+d |
| Totale | a+c | b+d | n |

È LA STESSA TABELLA DEI TEST DIAGNOSTICI MA ABBIAMO DI DIVERSO A SX ESPOSTO O NON ESPOSTO (CHE POI SONO CMQ MISURATI CON UN TEST CHE MI DICE SE SONO POSITIVI, E QUINDI ESPOSTI, O NO)

3 MISURE DI ASSOCIAZIONE:

- **RISCHIO RELATIVO:** è IL RAPPORTO TRA INCIDENZE (INCIDENZA ESPOSTI/INCIDENZA NON ESPOSTI)

Misure di associazione

Rischio relativo (rapporto incidenze)

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Incidenza esposti
 Incidenza non esposti

| | Malattia | Sanità | Totale |
|----------|----------|--------|--------|
| Esposto | a | b | a+b |
| Non espo | c | d | c+d |
| Totale | a+c | b+d | n |

QUI NON È CONSIDERATO IL TEMPO, MA CMQ QUANDO CALCOLO L'INCIDENZA C'È IL TEMPO

$$RR = \text{INCIDENZA ESPOSTI} / \text{INCIDENZA NON ESPOSTI}$$

INCIDENZA = numero di NUOVI casi osservati in un dato periodo di tempo / somma del tempo apportato da ciascun individuo

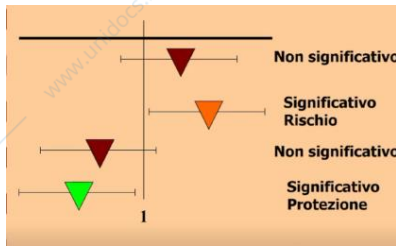
RR > 1 = esposizione associata alla malattia

RR = 1 = nessuna associazione

RR < 1 = esposizione protegge dalla malattia (es vaccini)

A patto che il test yquadro sia statisticamente significativo

LIMITI DI CONFIDENZA: devo considerare anche i limiti di confidenza, che vanno da x a y. Se un limite di confidenza comprende x e y tra 1 allora lo considero cmq non significativo, perché considero significativi SOLO quelli >1



← questi sono i limiti di confidenza, il primo, anche se è >1 per la maggior parte, lo considero cmq non significativo, perché la x è <1, stessa cosa per il terzo, la y è >1 quindi non significativo. QUESTI DUE CASI NON HANNO VALIDITÀ SCIENTIFICA

L'ASSOCIAZIONE NON MISURA LA FREQUENZA DELLA MALATTIA, MI DICE SOLO SE QUEL RISCHIO PUO ESSERE CAUSA O MENO DELLA MALATTIA

- **RAPPORTO PREVALENZE**: STIMA DEL RISCHIO RELATIVO (a differenza del RR che è il GOLD STANDARD, queste altre due sono delle stime)

Rapporto prevalenze (stima RR)

$$RP = \frac{a/(a+b)}{c/(c+d)}$$

Prevalenza esposti (a/(a+b))
Prevalenza non esposti (c/(c+d))

| | Malattia | Sanità | Totale |
|----------|----------|--------|--------|
| Esposto | a | b | a+b |
| Non espo | c | d | c+d |
| Totale | a+c | b+d | n |

ESSENDO UN RAPPORTO DI PREVALENZE QUINDI NON CONSIDERO IL TEMPO

$$RP = \text{PREVALENZA ESPOSTI} / \text{PREVALENZA NON ESPOSTI}$$

PREVALENZA (%) = numero di soggetti ammalati in un particolare istante / popolazione a rischio in quel particolare istante

PER ESSERE VALIDO DEVE ESSERE STATISTICAMENTE SIGNIFICATIVO E I LIMITI DI CONFIDENZA NON DEVONO COMPRENDERE 1

- **ODDS RATIO**: STIMA DEL RISCHIO RELATIVO SI CALCOLA FACENDO IL RAPPORTO DELLE DIAGONALI

Odds ratio (stima RR)

$$OR = \frac{a \times d}{b \times c}$$

| | Malattia | Sanità | Totale |
|----------|----------|--------|--------|
| Esposto | a | b | a+b |
| Non espo | c | d | c+d |
| Totale | a+c | b+d | n |

Odds ratio (stima RR)

Odds negli esposti: $(\text{casi} / \text{non_casi})_{\text{esposti}}$
Odds nei non esposti: $(\text{casi} / \text{non_casi})_{\text{non_esposti}}$

$$OR = \frac{(\text{casi} / \text{non_casi})_{\text{esposti}}}{(\text{casi} / \text{non_casi})_{\text{non_esposti}}}$$

$$OR = \frac{8 / 39}{10 / 134} = 2,74$$

$$OR = \text{ODDS ESPOSTI} / \text{ODDS NON ESPOSTI}$$

$$\text{ODDS} = (\text{CASI ESPOSTI} / \text{NON CASI ESPOSTI}) / (\text{CASI NON ESPOSTI} / \text{NON CASI NON ESPOSTI})$$

ESEMPIO:

Rapporto BCS e chetosi

| | Chetosi | No chet | Totale | Mesi/anim |
|--------|---------|---------|--------|-----------|
| BCS>4 | 60 | 41 | 101 | 284 |
| BCS<4 | 157 | 359 | 516 | 1750 |
| Totale | 217 | 400 | 617 | 2034 |

SONO TUTTE PREVALENZE (numero di soggetti ammalati in un particolare istante / popolazione a rischio in quel particolare istante)

Incidenza:

• **RP : 59% / 30% = 1.95**
 • **RR : 0,21 / 0,09 = 2.33**
 • **OR : (359*60)/(157*41)=3.35**
 • **RA=0.59-0.30=0.29**
 • **FA = (0.59-0.30)/0.59=0.49→49%**

Pr Chetosi in BCS + = 60/101 = 59%
 Pr Chetosi in BCS - = 157/516 = 30%
 Tinc Chetosi in BCS + = 60/284 = 0,21 casi/mese-animale
 Tinc Chetosi in BCS - = 157/1759 = 0,09 casi/mese animale
 Tinc Chetosi totale = 217/2034 = 0,11 casi mese animale

PREVALENZE

numero di soggetti ammalati in un particolare istante / popolazione a rischio in quel particolare istante

• **Chetosi : 217/617 = 0,352 = 35%**
 • **BCS>4: 101/617 = 0,163 = 16 %**
 • **Pr Chetosi in BCS + = 60/101 = 59%**
 • **Pr Chetosi in BCS - = 157/516 = 30%**

| | Chetosi | No chet | Totale | Mesi/anim |
|--------|---------|---------|--------|-----------|
| BCS>4 | 60 | 41 | 101 | 284 |
| BCS<4 | 157 | 359 | 516 | 1750 |
| Totale | 217 | 400 | 617 | 2034 |

INCIDENZE

numero di NUOVI casi osservati in un dato periodo di tempo / somma del tempo apportato da ciascun individuo

• **Tinc Chetosi in BCS + = 60/284 = 0,21 casi/mese-animale**
 • **Tinc Chetosi in BCS - = 157/1750 = 0,09 casi/mese animale**
 • **Tinc Chetosi totale = 217/2034 = 0,11 casi mese animale**

| | Chetosi | No chet | Totale | Mesi/anim |
|--------|---------|---------|--------|-----------|
| BCS>4 | 60 | 41 | 101 | 284 |
| BCS<4 | 157 | 359 | 516 | 1750 |
| Totale | 217 | 400 | 617 | 2034 |

IN QUESTO CASO QUALE MISURA SCELGO?

DIPENDE DAL TIPO DI TEST. AD ESEMPIO PER USARE OR MI SERVE UNO STUDIO CASO CONTROLLO

IN QUESTO CASO SCELGO RR PERCHE LE ALTRE DUE SONO STIME

VOUL DIRE CHE UN ANIMALE GRASSO HA RISCHIO DOPPIO DI AVERE LA CHETOSI

STUDI EPIDEMIOLOGICI

GLI STUDI SI APPLICANO TUTTE LE VOLTE CHE VOGLIAMO ASSOCIARE UN FATTORE DI RISCHIO A UNA MALATTIA (ad es se un ceppo di coli mi causa enterite, o sapere se c'è un fattore di patogenicità comune a batteri di diversa specie o della stessa specie)

DISTINZIONE:

1. STUDI OSSERVAZIONALI
 - a. DESCRITTIVI – mi danno le misure di frequenza della malattia (prevalenza incidenza ecc)
 - b. ANALITICI - con gruppi di controllo (individui non esposti)
 - i. TRASVERSALI o di prevalenza (non allungati nel tempo)
 - ii. LONGITUDINALI PROSPETTICI (di incidenza, follow up, di coorte) – andamento della malattia nel tempo
 - iii. LONGITUDINALI RETROSPETTIVI (caso- controllo) – misuro quello che è già successo (non mi riferisco al passato, ma a quando è successo l'evento. Se misuro la malattia oggi io mi sono già infettato. Se invece faccio lo studio oggi e aspetto che il soggetto si ammala sto facendo uno studio prospettico. La variabile è QUANDO succede la malattia)
2. STUDI SPERIMENTALI

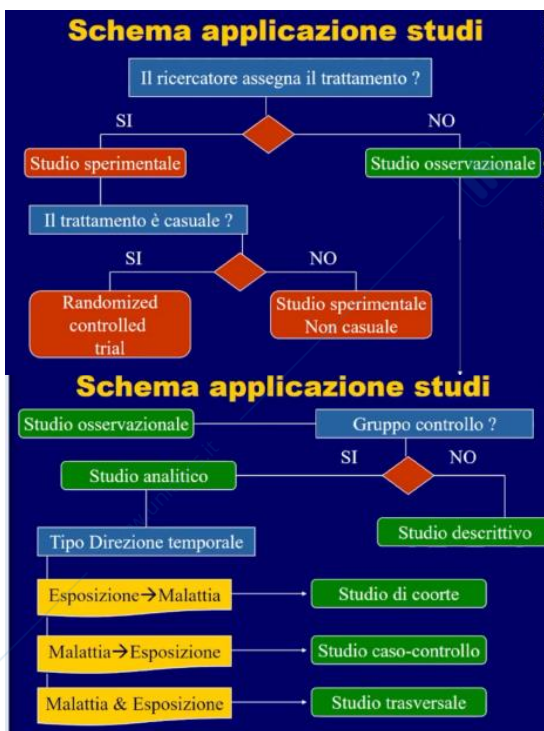
- a. NON CONTROLLATI – non c'è un controllo, non sono dei veri e propri studi scientifici perché non si fa un controllo. Sono chiamati quasi esperimenti
- b. CONTROLLATI
 - i. RANDOMIZZATI – casuali nell'assegnazione del trattamento
 - ii. CROSS OVER – lo stesso soggetto viene trattato per il trattamento o meno. Si riduce la variabile biologica. Ovviamente lo posso fare solo quando il trattamento non ha conseguenze una volta terminato

Schema studi

| | Ezio | Diag | Prog | Tera |
|---------|------|------|------|------|
| Trasv | + | +++ | - | - |
| Coorte | +++ | ++ | +++ | + |
| Caso-C. | ++ | + | ++ | - |
| Rand.C. | - | - | - | +++ |

← Quando è più efficace uno studio rispetto ad un altro

UN ALTRO ASPETTO DA CONSIDERARE È QUELLO DEI TEMPI.



IL RICERCATORE ASSEGNA DEI TRATTAMENTI I (quindi non è osservazionale ma sperimentale xkè è lo scienziato a assegnare un trattamento)

IL TRATTAMENTO È ASSEGNATO IN MODO CASUALE O NO? SE NON È ASSEGNATO IN MODO CASUALE NON HA VALENZA SCIENTIFICA PERCHÉ NON POSSO FARE CONFRONTI. SE L'ASSEGNAZIONE È CASUALE È UNO STUDIO RANDOMIZZATO CONTROLLATO

SE INVECE NON ASSEGNO TRATTAMENTO MA OSSERVO È UNO STUDIO OSSERVAZIONALE. C'È UN GRUPPO CONTROLLO? NO → STUDIO DESCRITTIVO (prevalenza/ incidenza) / SI → STUDIO ANALITICO

LO STUDIO ANALITICO IN FUNZIONE DELL'ESPOSIZIONE TEMPORALE SARÀ:

- SE POSSO DIMOSTRARE CHE L'ESPOSIZIONE È PRECEDENTE LA MALATTIA – **STUDIO DI CORTE**
- SE VEDO LA MALATTIA E POI VADO A VERIFICARE SE C'È STATA ESPOSIZIONE – **STUDIO CASO CONTROLLO**
- SE VEDO CONTEMPORANEAMENTE ESPOSIZIONE E MALATTIA – **STUDIO TRASVERSALE**

STUDIO CASO-CONTROLLO

LA MALATTIA È GIÀ AVVENUTA, L'ESPOSIZIONE È GIÀ AVVENUTA. MISURO QUALCOSA CHE È GIÀ AVVENUTO. SCELGO I CASI DALLA POPOLAZIONE DEI CASI (in realtà nella pratica poi mi devo accontentare di tutti i casi su cui posso mettere le mani). IN BASE AI CASI SCEGLIERÒ I CONTROLLI (quelli senza malattia, o quella cosa di interesse). VALUTO SE C'È UNA DIVERSA ESPOSIZIONE VALUTO I MALATI, VALUTO I SANI E VEDO SE C'È STATA ESPOSIZIONE O MENO. FACCIO UN'ASSOCIAZIONE CON L'ODDS RATIO (ovvero una valutazione di rischio)

1. DEFINIRE L'OBIETTIVO = associazione tra 1 o più fattori di rischio e la malattia
2. DEFINIRE CRITERI DI SCELTA DEI CASI E DEI CONTROLLI. Devo scegliere una soglia di controllo. Una volta che ho scelto i casi ho indirettamente scelto anche i controlli
3. DEFINIRE IL METODO DI CAMPIONAMENTO – nella realtà veramente il caso controllo si fa quando ci sono pochi casi quindi succede che prendo tutto e mi accontento
4. DEFINIRE IL METODO DI MISURA DELLE VARIABILI – Devo decidere dall'inizio come misurare i dati
5. VERIFICARE LA CORRETTEZZA DELLE IPOTESI

VANTAGGI → poco sforzo, buoni risultati. Posso lavorare anche con numeri piccoli (con il rischio che i dati non siano statisticamente significativi).

Posso avere una valutazione di diversi fattori di rischio senza una spesa. La malattia è una ma i fattori di rischio possono essere tutti quelli che posso rilevare. È importante però che la malattia sia SOLO 1

DIFETTI → non posso calcolare prevalenza e incidenza perché nel 99% dei casi il campionamento non è casuale. Ad es anche se ho 20 casi e 100 controlli non posso dire che ha il 20% di prevalenza perché i 20 casi me li sono presi io, e anche i 100 controlli.

NON POSSO VALUTARE PIU MALATTIE CONTEMPORANEAMENTE

Un altro difetto è che valuto retrospettivamente quindi ho il problema di poter affermare con certezza che l'esposizione sia antecedente alla malattia

METODI DI SCELTA:

- CAMPIONARE CASI E CONTROLLO CON LO STESSO METODO
- CONTRONTARE CASI E CONTROLLI DIRETTAMENTE. FACCIO **MATCHING**, OVVERO IL MIO CONTROLLO DEVE ESSERE IL PIU VICINO AI SOGGETTI (es. in un allevamento di suini prenderò tutti quelli nella stessa fase di crescita o produttiva)
- USO UNO O PIU GRUPPI DI CONTROLLO. Aumentare in n di controlli
- CERCARE CONTROLLI IN ALTRE POPOLAZIONI. ES i registri tumori in ambito umano si possono usare come controllo in ambito veterinario, quelli umani sono più vasti. Ovviamente è limitato a casi particolari come l'oncologia

METODI DI VALUTAZIONE:

- USARE DATI PRECEDENTI AL RILEVAMENTO DELLA MALATTIA
- VALUTAZIONE BLIND – CHI FA LE ANALISI NON DEVE SAPERE SE IL SOGGETTO È UN CASO O UN CONTROLLO, POTREBBE ESSERE INFLUENZATO DALLO SPINGERE LA VALUTAZIONE VERSO UN RISULTATO O UN ALTRO (es nel caso delle bande elisa potrebbe inconsciamente spingerlo a valutarli positivamente o negativamente)

ALLA FINE AVRO DELLE MISURE:

- ASSOLUTE – PROPORZIONE DI ESPOSTI NEI CASI E NEI CONTROLLO
- ASSOCIATIVE – RAPPORTO TRA CASI E NON CASI DEGLI ESPOSTI E CASI E NON CASI DEI NON ESPOSTI (ODDS RATIO).

Studio CASO-CONTROLLO

Odds ratio (stima RR)

$$OR = \frac{(casi / non_casi)_{esposti}}{(casi / non_casi)_{non_esposti}}$$

Odds negli esposti

Odds nei non esposti

APPLICAZIONI PRATICHE:

SCOPO: C'è UNA RELAZIONE TRA CRIPTORCHIDISMO E NEOPLASIE DEL CANE?

Vado a verificare se i soggetti con la malattia siano stati esposti ad un fattore di rischio.

1. TIPO DI STUDIO: Sono soggetti che vede il vet in clinica, quindi la cosa più facile è fare uno studio caso-controllo.
2. DEFINIZIONE DEL CASO: il veterinario deve visitare il cane, necessariamente maschio, e definire se ha o meno la malattia
3. DEFINIZIONE DEL CONTROLLO: sarà un soggetto che non ha tumore testicolare, ma che cmq sia esposto al rischio. Inoltre il tumore ci mette un po' a svilupparsi, per cui non prenderò in esame il cucciolo, ma piuttosto cani relativamente anziani e nella stessa fascia d'età. Considero anche la variabilità di razza, ideale sarebbe avere la stessa razza o almeno dimensioni simili (taglie grandi, taglie medie, taglie piccole, taglie toy).
4. DEFINIZIONE DELL'ESPOSIZIONE: sceglierò dei soggetti la cui cartella clinica dice che il criptorchidismo era stato diagnosticato tempo prima e non curato
5. FONTE DEI DATI: animali portati in visita

DATI: 64 cani

- Tumore e criptorchidismo: 30
- Tumore senza criptorchidismo: 10
- No tumore con criptorchidismo: 8
- No tumore no criptorchidismo: 16

I controlli dovrebbero essere numericamente lo stesso n dei casi. Nella realtà invece non è quasi mai così perché non si riesce ad arrivare al numero

MI COSTRUISCO LA TABELLA 2X2 e tengo un liv confidenza 95%

Relazione tra criptorchidismo e neoplasie nel cane

| Input of DATA: | | | | Level of Confidence | |
|-----------------------|---------|----|-------|---|--|
| Observed frequencies: | Exposed | | Total | <input type="radio"/> 90 % <input checked="" type="radio"/> 95 % <input type="radio"/> 97.5 % <input type="radio"/> 99 % <input type="radio"/> 99.5 % | |
| | Yes | No | | | |
| Case | 30 | 10 | 40 | | |
| Control | 8 | 16 | 24 | | |
| Total | 38 | 26 | 64 | | |

| RESULTS: | | | | Confidence Limits of Odds Ratio | |
|-----------------------|---------|-------|------------|---------------------------------|--|
| Expected frequencies: | Exposed | | Lower Lim. | Upper Lim. | |
| | Yes | No | | | |
| Case | 23,75 | 16,25 | 1,977 | 18,210 | |
| Control | 14,25 | 9,75 | 2,043 | 17,620 | |

Odds Ratio (OR): 6,000
 Attributable propor. among exposed: 0,833
 Attributable proportion: 0,625

Calculate Close

Limits are valid.

UNIVERSITÀ DEGLI STUDI DI MILANO

LA MIA ODDS RATIO è 6. VUOL DIRE CHE HO IL 6 VOLTE LA POSSIBILITÀ DI SVILUPPARE UN TUMORE SE CRIPTORCHIDE

È SIGNIFICATIVO? Vado a vedere **Xquadro** (si legge chi quadro), cioè il limite di confidenza, che è tra 2 e 17.

E' SIGNIFICATIVO PERCHE NON COMPRENDE L'1

Quindi c'è un'associazione statistica tra criptorchidismo e neoplasie. In termine pratici da medico ho un'indicazione per dire al proprietario di castrare il cane

SE VOGLIO ANDARE AVANTI CON LO STUDIO POSSO FAR SEGUIRE UNO STUDIO DI COORTE PROSPETTIVO PER CONFERMARE L'ASSOCIAZIONE

STUDIO DI COORTE

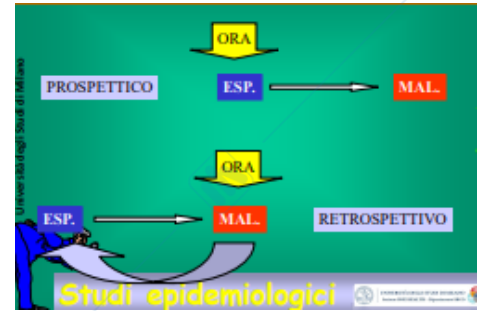
Gli studi di coorte prevedono di seguire un gruppo di soggetti per un dato tempo

Scopi:

1. descrizione dell'incidenza di un dato evento
2. analisi dell'associazione tra esposizione e malattia

varianti:

- **studi prospettivi:** lo studio viene pianificato e quindi realizzato nel futuro
- **studi retrospettivi:** lo studio viene pianificato e quindi realizzato con dati del passato



STUDI PROSPETTIVI:

SI PUO' CALCOLARE L'INCIDENZA DELLA MALATTIA IN RELAZIONE ALLA PRESENZA O MENO DELL'ESPOSIZIONE

PRO:

- unico studio in cui la relazione temporale causa-effetto può essere dimostrata
- potendo misurare l'esposizione prima della malattia, si ottengono misure più accurate e precise
- Se l'esposizione (fattore di rischio) è la stessa, si possono misurare più malattie

CONTRO:

- gli studi di coorte sono costosi ed inefficienti (tempi lunghi e molte analisi)
- impossibili per malattie rare
- relativamente sensibile agli effetti del confondimento

STUDI RETROSPETTIVI:

il ricercatore sceglie una coorte di soggetti privi della malattia oggetto di indagine (NEL PASSATO)

PRO:

- minore bias

CONTRO:

- mancanza di controllo sulle modalità e tecniche di esecuzione delle analisi

Misure che si possono fare in uno studio di coorte:

- ASSOLUTE
 - o Tasso di incidenza della malattia negli esposti e nei non esposti
- ASSOCIAZIONE



- Rischio relativo (rapporto tra tassi)
- Rischio attribuibile (differenza tra tassi)

APPLICAZIONI PRATICHE:

SCOPO: C'è UNA RELAZIONE TRA CRIPTORCHIDISMO E NEOPLASIE DEL CANE? (segue l'esempio di prima)
SELEZIONO SOGGETTI:

- CHE NON ABBIANO TUMORE (da cartelle cliniche o introspettivo), quindi cani di 4-5 anni senza tumore. Non solo, devono essere animali che devono essere stati visitati anni prima a cui è stato trovato criptorchidismo
- NON CRIPTORCHIDI SANI, CON LE STESSA CARATTERISTICHE DI Età E RAZZA
- NON CRIPTORCHIDI CON TUMORE CON LE STESSA CARATTERISTICHE DI Età E RAZZA
- CON TUMORE CRIPTORCHIDI (caso)

DATI:

30 cani con neoplasia

- 24 criptorchidi
- 6 non criptorchidi

FATTORE TEMPO (servirà per il tasso di incidenza):

- 500 mesi a rischio (criptorchidi)
- 600 mesi a rischio (non criptorchidi) ← è un numero più grande anche se i casi sono solo 6, ma ha senso perché sono sempre rimasti sani quindi hanno contribuito più tempo



Incidenza esposti = 24/500

Incidenza non esposti = 6/600

RR (RISCHIO RELATIVO)= 4,8 ← VUOL DIRE CHE UN CRIPTORCHIDE HA 5 VOLTE DI PIU LA POSSIBILITA DI SVILUPPARE UN TUMORE

E' STATISTICAMENTE SIGNIFICATIVO? Vado a Xquadro che è tra 2.1 e 10 quindi si, è statisticamente significativo

STUDI DI PREVALENZA

Negli studi crosssectional (trasversali) il ricercatore misura tutto in una singola occasione

Hanno due scopi principali:

1. descrizione della prevalenza di un dato evento
2. analisi dell'associazione tra esposizione e malattia

Gli studi crosssectional (trasversali) sono simili agli studi di coorte ma non c'è coorte e follow-up

Molto utili per descrivere le diverse variabili (fattori di rischio e malattia) e le loro distribuzioni

Possono essere utilizzati per studiare le associazioni tra fattori e malattia

COME DISEGNARE UN STUDIO:

1. definire l'obiettivo dello studio
2. definire criteri di scelta della popolazione e dei campioni:

DEVO CONSIDERARE DUE TIPI DI PREVALENZA: LA PREVALENZA DELLA MALATTIA (che la conosco da studi scientifici e letteratura) E LA PREVALENZA DELL'ESPOSIZIONE.

SE PRENDO LA PREVALENZA DELLA MALATTIA OTTENGO UN CERTO NUMERO X CHE NON È DETTO CHE SIA SUFFICIENTE PERCHÉ C'È DA CONSIDERARE LA PREVALENZA DELL'ESPOSIZIONE

QUINDI QUANDO DEVO SCEGLIERE QUANTI CAMPIONI PRENDERE CALCOLERO LA PREVALENZA ATTESA DELLA MALATTIA, LA PREVALENZA ATTESA DELL'ESPOSIZIONE, OTTERRÒ DUE NUMERI X E Y E SCEGLIERÒ SEMPLICEMENTE IL NUMERO PIÙ GRANDE TRA I DUE. COSÌ AVRÒ UN NUMERO DI CAMPIONI RAPPRESENTATIVO SICURAMENTE PER ENTRAMBI.

3. definire il fenomeno da studiare
4. definire il metodo di misura delle variabili

CALCOLO LA PREVALENZA TOTALE E POI VALUTO LA PREVALENZA DELLA MALATTIA NEGLI ESPOSTI FATTO QUESTO CALCOLO IL RAPPORTO TRA PREVALENZE CHE IN SOLDONI È UNA BUONA APPROSSIMAZIONE DEL RISCHIO RELATIVO

PRO:

- SONO DEI TEST RAPIDI E POCO COSTOSI. HO UN BUDGET FACILMENTE DEFINIBILE PERCHÉ HO UN TOT DI CAMPIONI DEFINITI, QUINDI POSSO STIMARE I COSTI.
- POSSO CALCOLARE LA PREVALENZA, CHE NON SARÀ L'INCIDENZA MA È CMQ UNA BUONA MISURA

CONTRO:

- DIFFICILE STABILIRE UNA RELAZIONE CAUSALE PERCHÉ NON HO UNA SEQUENZA TEMPORALE
- NON POSSO USARLO PER LE MALATTIE RARE
- SOGGETTI A BIAS PREVALENZA/INCIDENZA. Se ho delle malattie acute che poi scompaiono mi influiscono sull'incidenza ma non le rilevo con la prevalenza (che non prevede la variabile tempo). L'associazione dipende dalla durata della malattia, se ad es dura due anni ho il tempo di vederla, se dura poco no. Ma cmq non è un grossissimo problema se lo conosco

Misure che si possono fare in studio di prevalenza:

- ASSOLUTE
 - o Prevalenza della malattia negli esposti e nei non esposti
 - o Prevalenza della malattie e dell'esposizione
- ASSOCIAZIONE
 - o Rapporto di prevalenza (Stima del rischio relativo)

studi di PREVALENZA

$$RP = \frac{\text{prevalenza_esposti}}{\text{prevalenza_non_esposti}}$$

RISPETTO AGLI STUDI DI CORTE, PUR ESSENDO MENO PRECISO NEL VALUTARE L'ASSOCIAZIONE, ESSENDO SCELTI IN MODO CASUALE I CAMPIONI SONO PIÙ RAPPRESENTATIVI → **PIÙ RAPPRESENTATIVO E VELOCE**
OVVIAMENTE I NUMERI DEVONO ESSERE PIÙ GRANDI

APPLICAZIONI PRATICHE:

SCOPO: associazione tra stabulazione dei vitelli e comparsa della mortalità neonatale (in questo caso da patologia enterica)

Allevamento di bovine da latte con mortalità neonatale:

- Due diverse razze allevate (Frisona e Jersey)
- Due diversi modi di stabulare di vitelli (gabbia o box)
- Altri possibili fattori di rischio (colostratura alimentazione, vaccinazioni, stagionalità parti simili) ← devo prima verificare che i fattori di rischio siano comuni per tutti i soggetti. In questo caso tutti i vitelli dovranno aver preso il colostro, tutti vaccinati, tutti parti nello stesso periodo)

INFORMAZIONI DA RACCOGLIERE:

- dati oggettivi sulla frequenza di malattia
- dati oggettivi sull'esposizione ai diversi fattori di rischio

Analisi associazione tra fattori di rischio e malattia (mortalità)

SOGGETTI:

- Numero di vitelli nati: 370
 - da Frisona: 200
 - da Jersey: 170
- Numero di vitelli in gabbia : 205
 - Frisona / gabbia: 85
 - Jersey / gabbia 120
- Numero di vitelli in box : 165
 - Frisona / box: 115
 - Jersey / box 50

MORTALITÀ:

- Morti totali 145
- Morti gabbia 40
- Morti box 105
- Morti da Frisona: 80
- Morti da Jersey: 65
- Morti Frisona / gabbia: 15
- Morti Frisona / box: 65
- Morti Jersey / gabbia 25
- Morti Jersey / box 40

Da questi dati posso sospettare che la gabbia sia più protettiva del box

| Misure epidemiologiche per razza | | |
|-------------------------------------|--|-------------------------------|
| Mortalità totale (Prevalenza) | $= \frac{\text{morti}}{\text{nati}}$ | $= \frac{145}{370} = 39.19\%$ |
| Mortalità Frisona (Prevalenza) | $= \frac{\text{morti.Frisona}}{\text{nati.Frisona}}$ | $= \frac{80}{200} = 40.00\%$ |
| Mortalità Jersey (Prevalenza) | $= \frac{\text{morti.Jersey}}{\text{nati.Jersey}}$ | $= \frac{65}{170} = 38.23\%$ |

FACCIO TABELLA 2X2

FACCIO LA PREVALENZA DELLA MORTALITÀ.

39% DI MORTALITÀ TOT È MOLTO ALTA

Avere la stessa mortalità in due razze diverse non è un dato atteso. La frisona è una razza più robusta, la jersey è più debole, quindi perché la prevalenza è maggiore nelle frisona? La verità è che l'allevatore ha messo più frisona in box. Quindi non bisogna farsi ingannare dai numeri perché potrebbero essere falsati come in questo caso

Misure di associazione

Tabella 2x2

| | Morti | Vivi | Totale |
|---------|-------|------|--------|
| Jersey | a | b | a+b |
| Frisone | c | d | c+d |
| Totale | a+c | b+d | n |

Misure di associazione

Tabella 2x2

| | Morti | Vivi | Totale |
|---------|-------|------|--------|
| Jersey | 65 | 105 | 170 |
| Frisone | 80 | 120 | 200 |
| Totale | 145 | 225 | 370 |

Effetto razza

Input of DATA:
 Observed frequencies: Exposed Level of Confidence 95 %

| | Jersey | Frisone | Total |
|----------|--------|---------|-------|
| Diseased | 65 | 80 | 145 |
| Healthy | 105 | 120 | 225 |
| Total | 170 | 200 | 370 |

RESULTS:
 Expected frequencies: Exposed Yes No

| | Yes | No |
|----------|--------|--------|
| Diseased | 66,62 | 78,38 |
| Healthy | 103,38 | 121,62 |

 Attributable Risk: -0,018
 Attributable propor. among exposed: 0,044
 Attributable proportion: 0,020
Confidence Limits
 Logarithmic approx. Lower Lim. Upper Lim. Lower Lim. Upper Lim.
 Odds Ratio (OR): 0,929 0,611 1,412 0,610 1,413
 Prevalence Ratio (PR): 0,956 0,740 1,234 0,740 1,234
 Bias: 0,027
 Limits are valid.

Perche uso il rapporto di prevalenza e non odds ratio? PERCHE STO FACENDO UNO STUDIO DI PREVALENZA e non un caso- controllo

RP = 0,95 → SE FOSSE SIGNIFICATIVO SAREBBE UN FATTORE DI PROTEZIONE PERCHE <1

E' SIGNIFICATIVO? Xquadro è tra 0,7 e 1,2 quindi passa per 1. NON è SIGNIFICATIVO PERCHE PASSA PER 1

Visto che non è significativo possiamo concludere che LA RAZZA non è un fattore di rischio per la mortalità sei vitelli. (non sto considerando la stabulazione)

PRENDO UN'ALTRA STRADA: LA STABULAZIONE HA UN EFFETTO SULLA MORTALITÀ?

Misure epidemiologiche per stabulazione

| | |
|------------------|---|
| Mortalità totale | $= \frac{\text{morti}}{\text{nati}} = \frac{145}{370} = 39.19\%$ |
| Mortalità Gabbie | $= \frac{\text{morti.Gabbie}}{\text{nati.Gabbie}} = \frac{40}{205} = 20.00\%$ |
| Mortalità Box | $= \frac{\text{morti.Box}}{\text{nati.Box}} = \frac{105}{165} = 63.63\%$ |

Effettivamente dai numeri sembrerebbe che le morti in box sono maggiori rispetto a quelle in gabbia

Misure di associazione

Tabella 2x2

| | Morti | Vivi | Totale |
|--------|-------|------|--------|
| Box | 105 | 60 | 165 |
| Gabbia | 40 | 165 | 205 |
| Totale | 145 | 225 | 370 |

Effetto stabulazione

Input of DATA:
Observed frequencies:

| | Box | Gabbia | Total |
|----------|-----|--------|-------|
| Diseased | 105 | 40 | 145 |
| Healthy | 60 | 165 | 225 |
| Total | 165 | 205 | 370 |

Level of Confidence: 95 %

RESULTS:

| Expected frequencies: | Yes | No |
|-----------------------|--------|--------|
| Diseased | 64.66 | 80.34 |
| Healthy | 100.34 | 124.66 |

Odds Ratio (OR): 7.219
 Prevalence Ratio (PR): 3.261
 Bias: 3.957

Rapporto prevalenze (stima RR)

$$RP = \frac{a/(a+b)}{c/(c+d)}$$

Prevalenza esposti: $a/(a+b)$
 Prevalenza non esposti: $c/(c+d)$

| | Malattia | Sanità | Totale |
|-------------|----------|--------|--------|
| Esposto | a | b | a+b |
| Non esposto | c | d | c+d |
| Totale | a+c | b+d | n |

RP = (105/165) / (40/205) = 3.2 → Il soggetto in box ha 3 volte più probabilità di morire

E' STATISTICAMENTE SIGNIFICATIVO? SI PERCHE Xquadro è tra 2 e 4

POSSO AFFERMARE CHE LA STABULAZIONE è UN FATTORE DI RISCHIO E CHE LA STABULAZIONE IN BOX è PIU ASSOCIATA RISPETTO A QUELLA IN GABBIA

PASSO SUCCESSIVO: UNISCO I DUE FATTORI RI RISCHIO RAZZA+ STABULAZIONE

Misure epidemiologiche per razza e stabulazione

Mortalità Frisone/gabbia = $\frac{\text{morti .Frisone .Gabbia}}{\text{nati .Frisone .Gabbia}} = \frac{15}{85} = 17.64\%$

Mortalità Frisone/box = $\frac{\text{morti .Frisone .Box}}{\text{nati .Frisone .Box}} = \frac{65}{115} = 56.52\%$

Mortalità Jersey/gabbia = $\frac{\text{morti .Jersey .Gabbia}}{\text{nati .Jersey .Gabbia}} = \frac{25}{120} = 20.83\%$

Mortalità Jersey/box = $\frac{\text{morti .Jersey .Box}}{\text{nati .Jersey .Box}} = \frac{40}{50} = 80.00\%$

LA STABULAZIONE è IL RISCHIO PRINCIPALE (il rischio principale è sempre quello dove posso intervenire, sulla razza non posso fare niente, quella è)

LA RAZZA è IL FATTORE CONFONDENTE

Studio: analisi statistica effetto razza e stabulazione

Input of DATA:

| Stratum | Diseased | | Healthy | | Total |
|---------|----------|----------|---------|----------|-------|
| | Exp. | Non Exp. | Exp. | Non Exp. | |
| 1 | 65 | 15 | 50 | 70 | 200 |
| 2 | 40 | 25 | 10 | 95 | 170 |
| 3 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 |
| Total | 105 | 40 | 160 | 165 | 370 |

RESULTS:

| PR | Logarithmic approx. | | I ² Approximation | | X |
|-------------------|---------------------|------------|------------------------------|------------|-------|
| | Lower Lim. | Upper Lim. | Lower Lim. | Upper Lim. | |
| Crude | 3.261 | 2.414 | 4.407 | 2.941 | 3.616 |
| Pooled (Direct) | 3.589 | 2.655 | 4.830 | | |
| Pooled (Mantel-H) | 3.496 | 2.565 | 4.765 | 2.652 | 4.609 |
| | 3.840 | 2.638 | 5.589 | 2.664 | 5.535 |

Chi-square test for homogeneity of PR's across strata using MH-PR:
 GBD = 15.774 with 1 degree of freedom;
 p-value: < 0.0001 Level of Confidence (N) > 99.9

CALCOLO PRIMA I RP PER LE FRISONE IN GABBIA, POI RP DELLE JERSEY IN GABBIA (il numero due con la freccia mi indica la colonna 2 delle jersey)

LE FRISONE HANNO RP 3.2

LE JERSEY HANNO RP 3.8

SONO ENTRAMBI SIGNIFICATIVI PER xquadro MAGGIORE DI 1, E MI DICE CHE LE JERSEY SONO PIU A RISCHIO SE STABULATE IN GABBIA

RISULTATI:

RP razza= 0.95 – Non significativo (razza non associata a mortalità)

RP stabulazione = 3 – P <0.05 (stabulazione associata a mortalità)

RP per razza e stabulazione (stratificato) = 3,5

- Frisona = 3.2
- Jersey = 3.8 → rischio maggiore per le jersey

COSA FARÀ L'ALLEVATORE? Metterà tutte le jersey in box. Una volta finire le jersey se ci sono box liberi potrà metterci le frisone. Consiglio: comprare piu box

EPIDEMIOLOGIA MOLECOLARE

1. *Una scienza che si occupa dell'eziologia, della distribuzione e del controllo delle malattie nelle famiglie e delle cause ereditarie di malattia nelle popolazioni.*
2. *Una scienza che si occupa dei contributi dei fattori di rischio genetici e ambientali identificati a livello MOLECOLARE e biochimico allo studio dell'eziologia, distribuzione e controllo delle malattie delle famiglie e popolazioni*
3. *Non è solo un termine per descrivere nuove tecniche, ma rappresenta un'opportunità per usare strumenti con un potere risolutivo maggiore e quindi di poter sviluppare teorie sulla patogenesi della malattie che comprendano le complesse interazioni che avvengono durante il loro sviluppo*

Permette di condurre ricerche sull'eziologia della malattia, di conoscere i determinanti della malattia, di sviluppare quindi nuovi approcci per la prevenzione.

- Si basa sulla genetica delle popolazione
- Utilizza tecniche statistiche per valutare agli aspetti genetici delle malattie croniche
- Enfasi sui fattori di rischio ambientali limitata o assente
- Molti studi sono rivolti ad identificare i geni di sensibilità alle malattie:
 - o SNP e mappe aplotipi
 - o Analisi di linkage

DAL PUNTO DI VISTA PRATICO NORMALMENTE CON UN APPROCCIO TRADIZIONALE È DIFFICILE CLASSIFICARE UNA POPOLAZIONE TARGET, COSA CHE SI PUÒ INVECE FARE CON L'EPIDEMIOLOGIA MOLECOLARE

Per test molecolari viene richiesto di valutare:

- la standardizzazione
- la validazione analitica
- la validazione clinica

Utilizza i classici studi epidemiologici

- Coorte
- Caso-controllo
- Prevalenza

LA PRIMA APPLICAZIONE DELL'EPIDEMIOLOGIA MOLECOLARE GENETICA È STATA QUELLA DELL'**HUMANE GENOME PROJECT**, CON IL QUALE SI È POTUTO DESCRIVERE IL GENOMA UMANO

Esempio pratico: si è visto come ci sia un gene associato alla malattia clinica di paratubercolosi

OBIETTIVI:

cosa faccio a livello pratico? Posso:

- Condurre studi analitici e descrittivi per valutare le interazioni gene/ambiente nell'eziologia delle malattie
- Identificare tassi di morbilità specifici per i diversi fattori di rischio al fine di intervenire in caso di focolaio e prevenire le malattie

La stima di specifici fattori di rischio permette di intervenire anche a livello individuale per identificare individui a rischio e quindi intervenire su:

- Stile di vita, dieta (uomo)
- Caratteristiche allevamento (animali)
- Esposizione ambiente
- Approcci mirati sono più efficienti nella prevenzione delle malattie

EPIDEMIOLOGIA MOLECOLARE DESCRITTIVA:

- Definisce la distribuzione delle malattie
 - per individuo, luogo e tempo
 - definisce le conseguenze per la popolazione
- Incidenza
- Prevalenza
- Tassi di morbilità/mortalità
- Misura gli effetti precocemente nello sviluppo della malattie (non devo aspettare della manifestazione di malattie ma posso rilevare con i biomarker)
- Riduce gli errori nella classificazione delle malattie (diagnosi precisa)
- Valuta correttamente i marker di esposizione/sensibilità

POSSO FARE UNA VALUTAZIONE PRECOCE PERCHE NON DEVO ASPETTARE LA MANIFESTAZIONE DELLA MALATTIA
INTERVENGO PRIMA= PIU SUCCESSO

EPIDEMIOLOGIA MOLECOLARE ANALITICA:

- valuta le associazioni tra fattori di rischio e malattie tenendo conto di
 - Caratteristiche ospite
 - Esposizione ambientali
- Rischio relativo
- Odds ratio
- Rapporto prevalenze
- Utilizza i biomarker per migliorare la valutazione di esposizione, malattia e presenza dei potenziali fattori confondenti

Posso valutare anche gli effetti ambientali. Posso a livello cellulare vedere se sono stato esposto ad una sostanza, senza la manifestazione clinica.

Posso valutare anche la sensibilità genetica.

- ➔ **BIOMARKER DA ESPOSIZIONE:** posso vedere a livello cellulare se sono stato esposto ad una sostanza che mi da alterazione, senza vederne necessariamente aspettare la manifestazione clinica (vale per i chimici ma anche per infettive ecc)
- ➔ **Marker genetici:** polimorfismo proteico e polimorfismo DNA

Posso combinare le informazioni che ottengo correlando la presenza o meno di modificazioni genetiche con l'esposizione + o - e quindi assegnare un livello di rischio piu o meno alto, di conseguenza quindi adottare strategie diverse.

EPIDEMIOLOGIA MOLECOLARE ONCOLOGICA

Un approccio con il quale metodi di laboratorio avanzati sono usati assieme all'EPIDEMIOLOGIA MOLECOLARE analitica per identificare a livello biochimico o molecolare agenti specifici esogeni e/o fattori dell'ospite che giocano un ruolo nello sviluppo dei tumori.

L'epidemiologia molecolare è il risultato della fusione delle tecniche molecolari con l'epidemiologia. Le tecniche molecolari sono parte integranti del disegno dello studio epidemiologico.

Le metodiche molecolari sono più discriminanti, sensibili e specifiche, e riducono i bias dovuti all'assegnazione erronea di un soggetto ad un gruppo.

LEZ 27 APRILE

SLIDE APPLICAZIONI MOLECOLARI IN MEDICINA VETERINARIA

L'obiettivo dell'epidemiologia molecolare non è solo quello di classificare i microrganismi, ma soprattutto di identificare:

- Gli agenti responsabili delle patologie
- Le loro fonti e reservoirs
- Le relazioni biologiche e patogenetiche
- Le vie di trasmissione
- I geni responsabili della resistenza ai farmaci
- I geni importanti ai fini dello sviluppo di vaccini

COSA POSSO FARE?

- Tipizzazione comparativa
- Identificazione delle fonti d'infezione (es se so che in un paese c'è una malattia posso bloccare gli ingressi da quel paese)
- Dinamica della trasmissione dell'infezione
- Studi sulla persistenza, reintroduzione e reinfezione
- Specificità dell'ospite e nicchie ecologiche
- Sviluppo di vaccini

TIPIZZAZIONE COMPARATIVA:

Utilizzato per lo studio di microrganismi in un contesto. → Valutazione se ceppi provenienti da matrici biologiche diverse (latte, cute ecc) sono simili o meno.

Posso così definire quelli che sono pericolosi in quel contesto e quelli che no.

In base a quello che risulta posso mettere dei blocchi. Ad esempio per gli staphilococchi meticillino resistenti che sono trasmessi dai suini, in alcuni paesi se eri un allevatore ti facevano all'ingresso in ospedale test per meticillino resistenza e potevi entrare solo se negativo (stesso principio di tempone covid)

IDENTIFICAZIONE DELLE FONTI D'INFEZIONE:

applicazione comune nell'ambito di focolai di malattie infettive e tossinfezione alimentare (trasmissione animale-uomo, animale-animale, uomo-animale, trasmissione ambiente-alimento)

Le fonti d'infezione possono essere a livello geografico o di allevamento

Per alcuni patogeni ci sono proprio delle linee guida universali in modo da poter identificare i ceppi e verificare se il mio ceppo è solo qui o anche da altre parti geografiche

STUDI SULLA PERSISTENZA REINTRODUZIONE E REINFEZIONE

Per distinguere tra nuova infezione e reinfezione (Se curo un animale e poi si reinfetta devo sapere se è il ceppo di prima e quindi non è mai guarito o se è un altro ceppo)

Identifico se ci sono problemi nell'efficacia terapeutica, se ci sono cronicità nell'infezione o se è un problema di biosicurezza

SPECIFICITÀ DELL'OSPITE E NICCHIE ECOLOGICHE

La specificità dell'ospite e le eventuali modificazioni di tale specificità sono importanti sia nell'ambito delle zoonosi che dello sviluppo di piani vaccinali.

Una volta che eliminiamo un patogeno liberiamo una nicchia. Es la salmonella nel tempo ne abbiamo eliminate diverse specie (non la salmonella in toto). Questo perché le salmonelle eliminate hanno liberato una nicchia che è stata rimpiazzata da un nuovo ceppo. La speranza è che i nuovi siano meno aggressivi dei precedenti. Di solito cmq i nuovi sono più deboli in quanto emergenti.

SVILUPPO DI VACCINI:

associato molto al fatto che spt in veterinaria si stanno limitando l'utilizzo di alcuni farmaci come ad es gli antibiotici e quindi in sostituzione il vaccino è un'ottima profilassi

Lo sviluppo di vaccini marker permette di distinguere animali malati da quelli vaccinati. Es per la brucellosi non ho un vaccino marker quindi il protocollo prevede che vaccino l'animale per tot anni fino al livello di avere immunità di gregge e sospendo i vaccini. Faccio test per qualche anno e se trovo dei positivi saranno delle nuove infezioni. Se avessi un vaccino marker non dovrei farlo. Vaccino con i marker e a parte le prime generazioni che si sovrapporranno malati e sani, avrò poi modo di sapere se ho nuovi malati.

Fare un vaccino marker è difficile. Un vaccino marker in zootecnica è l'herpesvirus.

SVILUPPO TEST DIAGNOSTICI:

sviluppo di test aggiornati per sorveglianza sanitari considerando anche i costi (un proprietario di un cane spende per un test, un allevatore se costa troppo e deve farlo su tanti capi non lo farà)

RIDURRE USO DI ANTIBIOTICI:

l'antibiotico resistenza è una delle maggiori cause di sviluppo di nuove malattie

HOT TOPICS IN MED VET:

1. Test diagnostici aggiornati
2. Sviluppo di vaccini (spt marker)
3. Riduzione trattamenti antibiotici

SORVEGLIANZA E CONTROLLO: sanità pubblica

L'OIE (organizzazione mondiale della sanità animale) a livello mondiale sorveglia l'andamento delle malattie nella sanità animale.

C'è bisogno di nuove tecniche diagnostiche e più efficienti, lavorando su metodiche genomiche e proteomiche.

Alcune malattie per cui si fa sorveglianza:

- Foot-and-mouth disease FMD
- **Swine vesicular disease SVD (ita)**
- Vesicular stomatitis VS
- Classical swine fever CSF
- African swine fever ASF
- **Bluetongue BT (ita)**
- African horse sickness AHS
- Newcastle disease ND
- Highly pathogenic avian influenza HPAI

Visto che i vaccini fanno parte della profilassi, non potendo usare antibiotici, c'è bisogno di svilupparne di nuovi. I vaccini proteggono le forme cliniche, e cmq non al 100%. I virus cambiano anche nel tempo e dobbiamo capire come cambiano per svilupparli. Ad es il vaccino dell'influenza ogni anno cambia., o anche devo considerare i sierotipi, ad es l'influenza equina ha due varianti, l'americana e l'europea perché il virus è mutato, e non posso usare il vaccino per il virus americano su quello europeo e viceversa.

Caso del virus BVDV (diarrea virale bovina) che da nelle bovine aborti (inizialmente si pensava fosse una problematica GI perché associata a diarrea ma in realtà il problema è riproduttivo)

negli anni 80 si comincia a vaccinare e sembrava andare bene. A un certo punto è comparsa una sintomatologia emorragica nei vaccinati. Il virus era probabilmente mutato ad un ceppo emorragico (tipo 2). Il BVD tipo 2 è

comparso anche in Italia e si è scoperto che era all'interno del vaccino tipo 1. Cos'è successo? L'azienda produttrice non si era accorta che all'interno del vaccino si era sviluppato un ceppo nuovo che è stato trasmesso proprio dalla vaccinazione.

SICUREZZA ALIMENTARE:

Diventata una priorità dopo il caso della BSE (mucca pazza)

È importante individuare l'origine del focolaio della malattia in modo di andare all'origine e sapere quale punto della produzione è responsabile, ovvero la tracciabilità.

Si fanno indagini di laboratorio per tipizzare i patogeni isolati in corso di focolaio e si usano:

- Aspetti fenotipici → biotipo, sierotipo, tipizzazione, antibiogramma, ma non danno info sufficienti
- Metodi molecolari → stanno soppiantando i fenotipici

Solo i laboratori accreditati possono fare indagini di laboratorio valide a livello di sicurezza alimentare.

Con le metodiche molecolari io posso caratterizzare i geni del patogeno e i diversi sierotipi.

IDENTIFICAZIONE E FONTI DI RESERVOIRS

NON è SUFFICIENTE vedere il fingerprint (su base del genoma) di un patogeno ma bisogna considerare anche i fattori di patogenicità

Es staphilococcus aureus ha due ceppi A e B, uno cutaneo, uno della mammella, che si è visto essere specifici per quel distretto. Quindi oltre alla comparazione del fingerprint si è associata la presenza o meno di cluster relativi ai fattori di patogenicità per proteina A, fattore legante collagene e fattore legante fibrinogeno. Facendo così ottengo una nuova classificazione.

Fatti i cluster però non basta, devo cmq metterli in relazione alla malattia (tabella 2x2) altrimenti non sto facendo epidemiologia.

SVILUPPO DEL VACCINI

Esempio di mastite bovina: si è cercato da 60 anni di sviluppare vaccini ma al momento ce ne sono solo due.

Il problema è che la risposta della mammella è specifica. Se sviluppo il vaccino mi aspetto che ci sia un aumento degli anticorpi e devo fare un'iniezione ma non a livello di mammella, a livello generico. Il problema è che alla mammella gli anticorpi non arrivano, o meglio ne arrivano in maniera non sufficiente. Il problema è che il batterio è localizzato in mammella che non ha una difesa umorale (ovvero con immunoglobuline). Per cui i vaccini non hanno tanto senso. Allora si è pensato di agire sulle adesine (lo staphilococcus ha bisogno di aderire per avere effetto, no adesine, no adesione del batterio).

Dal punto di vista sperimentale c'era una buona risposta. È stato poi fatta la stessa cosa applicata a diversi ceppi di staphilococcus, ma la risposta era diversa. Il vaccino sarebbe funzionato solo per il ceppo per cui era stato prodotto, quando in natura ci sono infezioni da diversi ceppi (es i ceppi che sono in Lombardia non ci sono in Campania).

EPIDEMIOLOGIA DELLE MALATTIE INFETTIVE

VIDEO LEZ 4 MAGGIO da min 20

SLIDE EPIDEMIOLOGIA MOLECOLARE DELLE MALATTIE INFETTIVE

La distribuzione di una malattia infettiva è determinata o "programmata" dal materiale genetico del patogeno che evolve in risposta al sistema difensivo dell'ospite e in funzione dell'ambiente in cui risiede.

Questo sistema è in evoluzione, tutti gli agenti patogeni mutano a velocità differenti in funzione di una risposta anticorpale (naturale o artificiale). -> situazione dinamica

Questo vuol dire che una situazione epidemiologica attuale può non avere valore tra qualche anno (es influenza stagionale, ogni anno si cambia il vaccino)

Con i metodi molecolari posso identificare:

- Agenti eziologici
- Reservoirs
- Vie di trasmissione

- Geni di patogenicità
- Antigeni per vaccini
- Geni per resistenza antibiotici

DIFFERENZE::

- Tassonomia → semplice classificazione. Mettere una specie in un gruppo piuttosto che in un altro
- Filogenesi → oltre a classificare studio anche l'evoluzione nel tempo

Sono diversi dall'epidemiologia!

- Epidemiologia molecolare → studia le **interazioni** tra ospite, ambiente ed agente patogeno (devo avere almeno una tabella 2x2 esposizione-malattia)

Con l'epidemiologia molecolare posso determinare la dinamica della trasmissione della malattia, identificarne i rischi, distinguere i ceppi patogeni dai non patogeni, indagare i focolai di malattia e individuare i fattori di patogenicità.

Le tecniche di biologia molecolare mi devono permettere di poter analizzare tanti dati (efficienza, utilizzo di sistemi automatici o possibilità di utilizzare campioni in pool, ovvero usare un mix di campioni piuttosto che i campioni in singolo. Questo lo faccio di solito quando mi aspetto una certa prevalenza, se l'analisi risulta negativa vuol dire che tutti sono negativi, se risulta positiva devo farne altre escludendo di volta in volta i campioni), essere semplici, appropriate e di un costo adeguato (costo sia in termini di materia che del tempo necessario)

Le metodiche molecolari devono essere in grado di dare un risultato non ambiguo, di darmi sempre lo stesso risultato quando lo ripeto, essere capace di discriminare ceppi differenti, essere facile da interpretare e da usare

CLASSIFICAZIONE:

1. METODO CLADISTICO O TASSONOMICO

Metodo detto anche di sistematica evolutiva che si basa sulla presenza di caratteristiche comuni tra organismi (taxa).

Vogliamo conoscere gli ANTECEDENTI, ovvero sapere come si è evoluto l'organismo → si usa la tassonomia

2. METODO FENETICO O NUMERICO

Metodo che non assume un modello evolutivo. Gli organismi vengono ordinati in base alle loro similarità o differenze (distanze) espresse in modo numerico. → si usa per l'epidemiologia

Gruppi di organismi con le stesse caratteristiche si chiamano OTU OPERATIONAL TAXONOMIC UNIT

METODI FENETICI:

- SIMILARITÀ → sistema che ci permette di vedere la distanza genetica tra un OTU e l'altro. Se la distanza è grande gli OTU sono diversi, se sono vicini potrebbero essere considerati uguali
- CLUSTER ANALYSIS

CALCOLO SIMILARITÀ:

HO DUE OTU, A e B, quante bande (dopo elettroforesi) hanno in comune?


a=bande presenti in entrambi OTU (5)

b=bande presenti solo in OTU B (1)

c=bande presenti solo in OTU A (2)

d=bande assenti in entrambi OTU (2)

qui ho solo due OTU ma in realtà ce ne sarebbero centinaia o più da analizzare quindi di solito ci pensa un software



Similarità

| Band position | OTU A | OTU B |
|---------------|-------|-------|
| 1 | — | — |
| 2 | — | — |
| 3 | — | — |
| 4 | — | — |
| 5 | — | — |
| 6 | — | — |
| 7 | — | — |
| 8 | — | — |
| 9 | — | — |
| 10 | — | — |

applico quindi delle formule per calcolare la similarità. Posso usare alcuni metodi:

| | |
|--|--|
| <p>Semplice $S = \frac{a+d}{a+b+c+d} = \frac{7}{10} = 0,7$</p> <p>Sneath & Sokal $S_{ss} = \frac{2(a+d)}{2(a+d)+c+b} = \frac{14}{17} = 0,82$</p> | <p>Jaccard $S_j = \frac{a}{a+b+c} = \frac{5}{8} = 0,62$</p> <p>Dice $S_D = \frac{2a}{2a+c+b} = \frac{10}{13} = 0,78$</p> |
|--|--|

Ma qual è la formula più adatta? Anche perché i risultati sono diversi. Non c'è una formula più adatta delle altre, potrebbero andare tutte bene, ma devo cmq essere consistente, perché non posso usare una formula per un alvoro e una per l'altra, le valutazioni non sarebbero consistenti e i risultati non comparabili. Quindi si cerca di usare quella che usano tutti, che di solito è **Dice**

Ora mi devo costruire l'albero di classificazione.

Esempio UPGMA

unweighted pair group method with arithmetic mean

| | A | B | C | D | E |
|---|---|---|---|---|---|
| B | 2 | | | | |
| C | 4 | 4 | | | |
| D | 6 | 6 | 6 | | |
| E | 6 | 6 | 6 | 6 | |
| F | 8 | 8 | 8 | 8 | 8 |

Es. Dice d_{AB}

Calcolo nuove distanze:

$d_{(AB)C} = (d_{AC} + d_{BC}) / 2 = 4$

$d_{(AB)D} = (d_{AD} + d_{BD}) / 2 = 6$

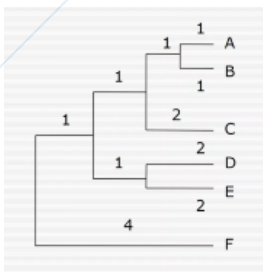
$d_{(AB)E} = (d_{AE} + d_{BE}) / 2 = 6$

$d_{(AB)F} = (d_{AF} + d_{BF}) / 2 = 8$

Con Dice mi calcolo le distanze di ogni OTU dall'altro. Distanza da A rispetto a B, distanza da A a C, distanza da A a D eccc..

Prima calcolo le distanze tra i singoli e le segno, poi calcolo le distanze dalla coppia (AB rispetto a C poi rispetto a B ecc. cmq lo fa un computer) ← quelli nel riquadro sotto

Se capita che le distanze coincidono vuol dire che i due gruppi appartengono allo stesso cluster.



Come si misurano nel grafico? I rami si misurano in orizzontale, in larghezza. Es la distanza tra A e F è 1+1+1+1+4, quindi in larghezza, non in altezza

Se io voglio dire che i miei OTU (organismi) sono diversi tra di loro devo dimostrare che sono distanti e quanto sono distanti tra di loro. Utilizzo i dati della tabella verticale che ho trovato prima e li porto in orizzontale con il grafico che vedo qui a sx. Metto sul grafico sia la distanza tra i singoli che la distanza tra le coppie.

Cosa uso per misurare distanza tra organismi? Dice, jakar ecc

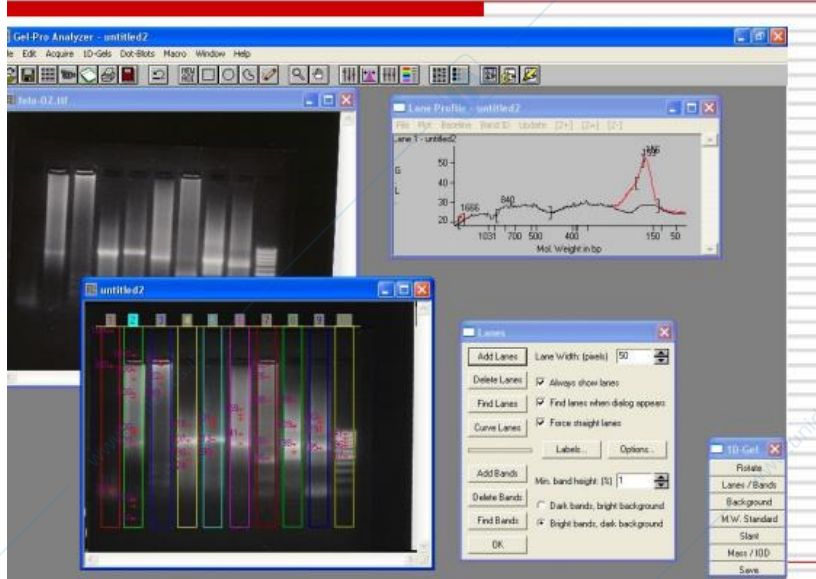
Che tipo di sistema uso per fare la valutazione complessiva della classificazione?

Devo darmi una regola di soglia perché non so quanto sono effettivamente grandi le distanze.

Posso anche dire "la soglia per definire che un organismo è diverso dall'altro è ad esempio 3". Quindi conto e posso dire ad es che da A ad E sono simili tra di loro e sono dissimili dagli altri.

Non è detto che perché sono su rami diversi allora sono diversi, geneticamente sono diversi sì, ma quanto diversi? Dipende da quello che sto studiando, magari non sono uguali ma sono simili, mentre rispetto ad altro sono completamente dissimili. Il valore soglia serve a questo, e lo scelgo in base a un ragionamento logico, ovvero con una percentuale (ad es, sopra il 20%, cioè 1, sono simili, sotto sono diversi). Per questo devo darmi una soglia.

Esempio pratico – Step 2



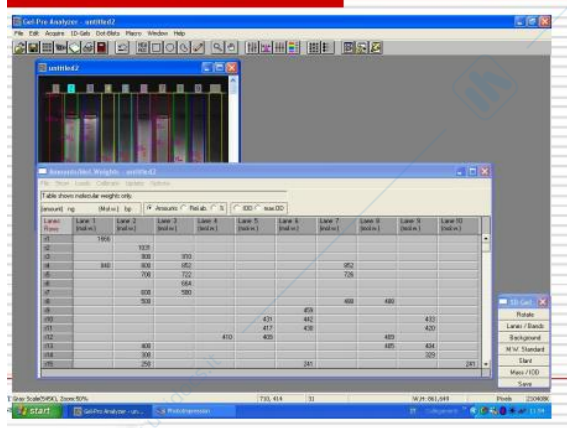
Le bande del gel di elettroforesi non sono chiare:

- Sono storte
- Il peso molecolare non è ben definito perché non si capisce la distanza tra le bande
- Se lo analizzo visivamente individuerò bande diverse rispetto ad un'altra persona

Per questo si deve usare un software che:

- Raddrizza le bande
- Riconosce i picchi delle bande, dove ci sono i numeri in viola (o meglio, mi da un numero ma io devo stabilire dei criteri in cui dire "sì è vero è una banda" o non è vero. Es: deve essere superiore a 40 di densità ottica, lo metto tra i filtri)

Esempio pratico – Step 2



Ottingo un foglio con dei dati che corrispondono ai pesi molecolari dei vari picchi delle bande

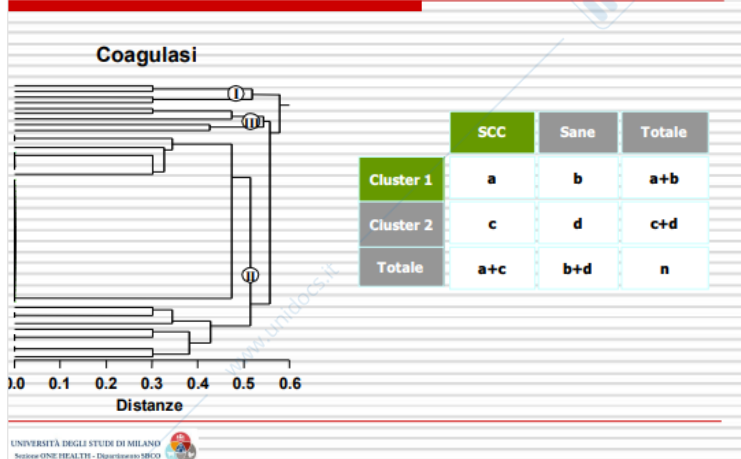
In questo caso ho un gene della coagulasi tagliato con enzima di restrizione. Quindi mi aspetto o una banda unica, se non ha tagliato niente, o più bande (fino a 4 ne può tagliare).

Se il gene è 1000 paglia di basi e viene tagliato, il totale della somma delle bande deve darmi 1000, se è di meno c'è qualcosa che manca. Non bisogna dare per scontato che perché l'ha fatto il pc allora è giusto. ←devo verificare io!

A questo punto ho le bande, in numeri, e posso usare le formule

viste prima.

Esempio pratico – Step 5



Esempio 2

Visto che sto parlando di epidemiologia devo associare i dati ad un evento: faccio tabella 2x2

- Animale sano - animale malato ← definiti a seconda della risposta cellulare
- Esposto – non esposto

Dove risposta cellulare:

- Animale sano: almeno 1 migliaio di neutrofili nel latte almeno
- Animale malato: almeno 10 miliardi di neutrofili nel latte

Faccio l'albero come detto prima con soglia 20% (cioè 1) e identifico 3 cluster principali 1,2,3.

Il CLUSTER è l'esposizione (fattore di patogenicità + fattore di rischio)

Voglio sapere la capacità di questi cluster di indurre una risposta infiammatoria

PROBLEMA: ho 2 cluster ma la tabella è 2x2, quale uso? Quale identifico come esposto e quale come non esposto? Devo scegliere il cluster di riferimento. Ma come faccio a decidere quale è migliore dell'altro? Il migliore, ovvero il riferimento, sarà quello di non esposizione, in questo caso quello che da risposta infiammatoria minore (risposta cellulare minore)

Potrei scegliere anche quello più alto come riferimento, ma se scelgo quello con n di cellule più alte come riferimento, ovvero esposizione, automaticamente tutti gli altri saranno i non esposti, ma avrò un numero molto maggiore di non esposti. Non è possibile perché non posso dire che se uno non ha una risposta infiammatoria allora non è stato esposto. Per questo scelgo come riferimento il dato che non dà il problema, ovvero poche cellule.

Scegliendo come riferimento il non esposto, avrò una situazione più bilanciata



Esempio pratico – Step 6

| GENE | ODDS RATIO x SCC |
|------|------------------------|
| cna | 1,3 |
| pvl | 1,5 |
| coa | Cluster I: 1,5 |
| | Cluster II: 1 |
| | Cluster III: referenza |
| efb | Cluster I: 0,5 |
| | Cluster II: 1 |
| | Cluster III: referenza |
| spa | Cluster I: 10 |
| | Cluster II: 19 |
| | Cluster III: referenza |

Come lo leggo??

Studio sulla collagenasi dove ho trovato più geni, i primi due con una sola forma, gli altri polimorfi

Cna e pvl non è un problema, non c'è polimorfismo quindi presenza/assenza = presenza del gene = esposto, assenza = non esposto

Lo leggo: in un ceppo di staphiloccocus a. (del mio studio) la presenza del gene cna da 1,3 volte il rischio di sviluppare la malattia rispetto a un ceppo che non ce l'ha, idem per pvl dove ho 1,5

volte un aumento della risposta infiammatoria rispetto a un ceppo che non ce l'ha.

Per i polimorfi (coa, efb, spa) devo scegliere un cluster di riferimento.

Il numero di ripetizioni di Spa (proteina A) mi aumento di 10 o di 19 volte l'aumento della risposta infiammatoria. In questo caso avere troppo poco spa o troppo spa diventa cmq patogenico

Così ho associato la presenza del gene (da risposta cellulare) a un rischio di presenza di malattia

Problema di un approccio di questo tipo? Dal punto di vista della patogenicità è utile o no?

In verità è più corretto dire che il gene che contiene questa combinazione di geni mi da questa risposta. Se li prendessi singolarmente darebbero la stessa risposta?

Attribuisco un effetto a un gene che è un effetto dovuto da quella combinazione di geni. -> CONFONDIMENTO, variabile da combinazione di dati

Analizzare gene per gene non ha senso xke nessun microrganismo avrà solo 1 gene ma una combinazione di essi

Combino l'associazione dei geni in un disegno matematico più grande. Immagina quando ho una combinazione di 10000 geni. Cmq lo fa un computer quindi ok

ESEMPIO 2 = VALUTAZIONE DELLA COMBINAZIONE DI ADESINE E TOSSINE IN PATOLOGIE DELL'UOMO

Sono stati identificati e analizzati una serie di fattori di patogenicità: i diversi tipi di adesine e i diversi tipi di tossine
Le tossine sono anche dei superantigeni e possono influenzare la risposta immunitaria

Si dichiara cosa si fa, ovvero che primer vengono usati e le sequenze, i cicli di pcr (t°, min, ecc), i kit usati, gli enzimi usati, da dove vengono prelevati i campioni ecc tutto quello che serve per replicare lo studio

ADESINE:



Esempio 2

| Virulence determinant | No. of isolates | | No. of genes ^a | | | |
|-----------------------|------------------------|------------------------|---------------------------|---------------------------|----------------------------------|---------------|
| | Carriage (n = 178) (%) | Invasive (n = 155) (%) | Unstratified analysis | | Stratified analysis | |
| | | | Odds ratio (99% CI) | P | Odds ratio ^b (99% CI) | P |
| Adhesins | | | | | | |
| <i>fimbA</i> | 154 (87) | 152 (98) | 7.9 (1.7–77) | 0.0001^c | 12.0 (1.6–89) | 0.0001 |
| <i>clfA</i> | 174 (98) | 155 (100) | | 0.13 | | 0.13 |
| <i>clfB</i> | 178 (100) | 155 (100) | | 1 | | |
| <i>cna</i> | 57 (32) | 81 (52) | 2.3 (1.3–4.3) | 0.0002 | 2.0 (1.1–3.7) | 0.003 |
| <i>spa</i> | 160 (90) | 145 (94) | 1.6 (0.54–5.4) | 0.24 | 1.7 (0.55–5.5) | 0.20 |
| <i>sdrC</i> | 178 (100) | 156 (100) | | 1 | | |
| <i>sdrD</i> | 73 (41) | 74 (48) | 1.3 (0.71–2.3) | 0.27 | 1.9 (0.8–4.4) | 0.05 |
| <i>sdrE</i> | 72 (40) | 87 (56) | 1.9 (1.0–3.4) | 0.006 | 3.1 (1.4–6.7) | 0.0001 |
| <i>bbp</i> | 76 (43) | 59 (38) | 0.82 (0.45–1.5) | 0.43 | 0.74 (0.37–1.5) | 0.27 |
| <i>ebpS</i> | 110 (62) | 105 (68) | 1.3 (0.70–2.4) | 0.30 | 1.5 (0.65–3.4) | 0.20 |
| <i>map/leap</i> | 165 (93) | 149 (96) | 2.0 (0.50–9.4) | 0.24 | 1.7 (0.42–6.5) | 0.34 |

Come lo leggo?

Tutti i ceppi con le risposte sulla loro presenza/assenza.

Su una colonna ci sono tutti i tipi di adesine, sull'altra i soggetti che hanno il batterio (staphilococcus nel caso dello studio) a livello dell'organo target (narici), sull'altra quelli che hanno la forma invasiva.

Nel caso di ClfB 178 soggetti hanno il batterio solo a livello di narici, 155 soggetti hanno un'infezione diffusa.

Si fa una tabella 2x2 dove carrier è sano, invasive è ammalato.

Si valuta presenza/assenza di ogni gene

Quando calcolo l'ODDS RATIO nel caso di ClfB dove tutti i soggetti del gruppo carriage e tutti i soggetti del gruppo invasive hanno il gene (100) avrò una odds ratio di 1, quindi non significativa, inutilizzabile. Idem nel caso di sdrC.

Vuol dire che tutti i soggetti hanno quel gene

Nel caso di *fimbA*, del *ClfA* e di *sdrE* ho invece dei risultati significativi, all'interno dei limiti di confidenza.

Vuol dire che questi 3 geni sono associati all'invasività del batterio

Successivamente hanno fatto un'ANALISI STRATIFICATA, in funzione dei complessi clonali (combinazione dei geni)

Vado a vedere se un'associazione specifica si genera sia più aggressiva. In termini numerici rimangono gli stessi geni significativi ma si alzano/abbassano i valori.

Anche da un'analisi stratificata questi 3 geni sono associati all'invasività

TOSSINE:

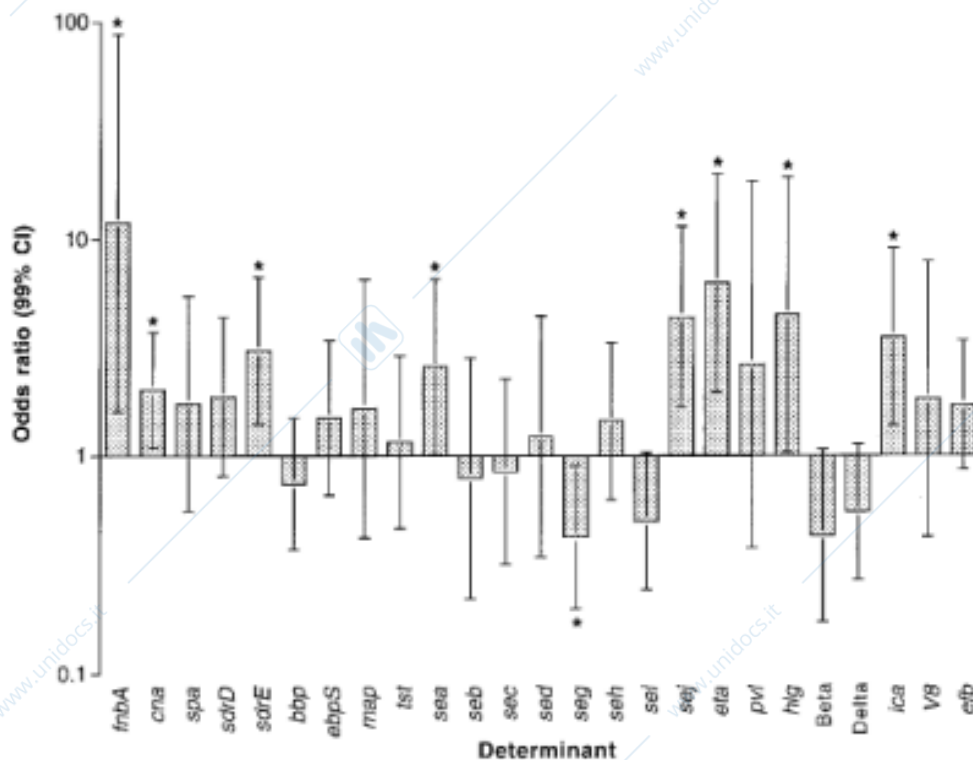
| Virulence determinant | No. of isolates | | No. of genes ^a | | | |
|-----------------------|------------------------|------------------------|---------------------------|------------------|----------------------------------|-------------------|
| | | | Unstratified analysis | | Stratified analysis | |
| | Carriage (n = 178) (%) | Invasive (n = 155) (%) | Odds ratio (99% CI) | P | Odds ratio ^b (99% CI) | P |
| Toxins | | | | | | |
| <i>tst</i> | 44 (25) | 46 (30) | 1.3 (0.68–2.6) | 0.27 | 1.2 (0.46–2.9) | 0.68 |
| <i>sea</i> | 30 (17) | 48 (32) | 2.3 (1.1–4.7) | 0.003 | 2.6 (1.0–6.6) | 0.007 |
| <i>seb</i> | 13 (7) | 14 (9) | 1.3 (0.41–3.9) | 0.72 | 0.79 (0.22–2.8) | 0.33 |
| <i>sec</i> | 20 (11) | 15 (10) | 0.84 (0.30–2.2) | 0.72 | 0.85 (0.32–2.3) | 0.66 |
| <i>sed</i> | 9 (5) | 8 (5) | 1.0 (0.24–4.2) | 1 | 1.2 (0.34–4.4) | 0.68 |
| <i>see</i> | 0 | 0 | | 1 | | |
| <i>seg</i> | 113 (64) | 85 (55) | 0.69 (0.38–1.3) | 0.12 | 0.42 (0.2–0.9) | 0.003 |
| <i>seh</i> | 18 (10) | 24 (15) | 1.6 (0.65–4.1) | 0.19 | 1.4 (0.63–3.3) | 0.25 |
| <i>sei</i> | 106 (60) | 81 (52) | 0.73 (0.40–1.3) | 0.18 | 0.50 (0.24–1.0) | 0.01 |
| <i>sej</i> | 12 (7) | 38 (25) | 4.5 (1.8–13) | <0.001 | 4.4 (1.7–11) | <0.0001 |
| <i>eta</i> | 11 (6) | 34 (22) | 4.2 (1.6–13) | <0.001 | 6.3 (2.0–20) | <0.0001 |
| <i>etb</i> | 6 (3) | 0 (0) | | 0.03 | | 0.06 |
| <i>pvl</i> | 3 (2) | 6 (4) | 2.3 (0.31–27) | 0.31 | 2.7 (0.38–19) | 0.18 |
| <i>hlg</i> | 157 (89) | 150 (97) | 3.6 (0.97–19) | 0.01 | 4.5 (1.0–19) | 0.004 |
| Alpha-toxin | 178 (100) | 153 (99) | | 0.21 | | 0.15 |
| Beta-toxin | 160 (89) | 122 (79) | 0.45 (0.18–1.0) | 0.02 | 0.43 (0.17–1.1) | 0.01 |
| Delta-toxin | 144 (80) | 106 (69) | 0.54 (0.27–1.1) | 0.02 | 0.55 (0.27–1.1) | 0.03 |

Nel caso delle tossine viene fatto lo stesso lavoro e si calcola l'Odds ratio per singolo gene, poi si fa un'analisi stratificata considerando il confondimento. Nella stratificazione la tossina SEG che prima non era significativa lo diventa.

Tutto sommato cmq i complessi clonali non influenzano in questo caso, la patogenicità dei geni

Ho preso 30-40 geni di patogenicità ma solo una decina sono significativi. Posso usarli per un vaccino? Posso usarli con funzione predittiva?

Ma la cosa più importante è che a volte la combinazione di geni (che singolarmente possono essere patogeni) possono dare un'effetto PROTETTIVO (quando l'odds ratio è negativa)



Sono andati poi a vedere quali sono le combinazioni più frequenti di geni, che possono poi essere utilizzati come cluster e confrontarli

TABLE 4. The 12 most common gene combinations

| Combination | Complement | | | | | | | n ^a | No. of isolates | Overall % of total | Cumulative % of total |
|-------------|-------------|------------|-------------|------------|------------|------------|------------|----------------|-----------------|--------------------|-----------------------|
| | <i>fnbA</i> | <i>ena</i> | <i>sdhE</i> | <i>sej</i> | <i>eta</i> | <i>hly</i> | <i>ica</i> | | | | |
| A | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 4 | 64 | 19.6 | 19.6 |
| B | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 4 | 64 | 19.6 | 39.1 |
| C | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 39 | 11.9 | 51.1 |
| D | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 5 | 35 | 10.7 | 61.8 |
| E | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 3 | 10 | 3.1 | 64.8 |
| F | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 6 | 9 | 2.8 | 67.6 |
| G | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 6 | 8 | 2.4 | 70.0 |
| H | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 8 | 2.4 | 72.5 |
| I | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 | 7 | 2.1 | 74.6 |
| J | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 7 | 2.1 | 76.8 |
| K | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 5 | 6 | 1.8 | 78.6 |
| L | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 5 | 6 | 1.8 | 80.4 |

^a Number of virulence-associated determinants within the gene combination (out of seven).

VIDEO 11 MAGGIO

SCOPO: TROVARE I CEPPI ISOLATI DA CHE SPECIE PROVENGONO

LA TECNICA CHE SCELGO LA SCELGO IN BASE SE PENSO POSSA FUNZIONARE O MENO PER QUEI CAMPIONI

ESEMPIO 4

RUOLO DELLO S. AUREUS NELLA MASTITE

DEFINISCO I FATTORI DI VIRULENZA (in base a lavori pregressi, prendendo info da li). In questo caso prendo le aree ripetute della proteina A. c'è un'area all'interno della proteina A, il gene X, di circa 26 paia di basi, si purifica, si vede il peso molecolare, si divide per 26 e si vede quante sono le ripetizioni (?)

Esempio 4

Table 2

Frequency of isolates belonging to the different clusters defined by of protein A-X gene repetitions region and mean somatic cell counts (SCC) observed in each cluster

| Repetitions (N) | Isolates (N) | Cluster | Isolates (N) | Frequency (%) | SCC (\pm std. dev.) |
|-----------------|--------------|---------|--------------|---------------|------------------------|
| 2 | 11 | I | 15 | 30.0 | 5.41 \pm 1.21 |
| 3 | 4 | | | | |
| 4 | 4 | | | | |
| 6 | 4 | II | 8 | 16.0 | 4.45 \pm 1.12 |
| 8 | 3 | | | | |
| 9 | 5 | III | 27 | 54.0 | 5.55 \pm 1.06 |
| 10 | 16 | | | | |
| 11 | 3 | | | | |

li classifico in cluster e identifico con una logica di distanza del 20% tra un cluster e l'altro.

I cluster sono caratterizzati da un contenuto cellulare diverso e quindi una risposta immunitaria diversa.

A questo punto assumo che la risposta immunitaria sia simile per tutti gli ospiti (ovviamente non è una cosa vera ma devo assumerlo perche non posso sapere gli individui come rispondono singolarmente)

Il cluster II è quello con la risposta cellula piu bassa quindi deduco che è quello con risposta immunitaria

minore e diventa il mio cluster di riferimento

Faccio adesso la tabella 2x2 con il cluster II di riferimento. Mi serve capire chi è esposto/non esposto, chi è malato/non malato.

I non esposti sono quelli che hanno bassa immunità, gli esposti sono gli altri due cluster

Come faccio a definire che un soggetto è ammalato? Definisco una soglia (200000 cellule dagli studi) sopra cui ho mastite. Quelli nella tabella 5.41 e 5.55 dei cluster I e III sono in logaritmo e sono >200.000 cellule, quindi loro sono ammalati

Chi è esposto e chi no?

Faccio 2 tabelle. Da una parte gli animali infetti del cluster II, da un'altra la risposta cellulare >200.000 degli animali ammalati (o del cluster I o del Cluster II perché prendo i due estremi)

Sono due tabelle. Cluster II vs Cluster III e Cluster II vs Cluster I

Sono poi andati a considerare i tipi di tossine

Table 3

Frequency of isolates harbouring the different genes considered and mean somatic cell counts (SCC) observed in each cluster

| Gene | Positive | | Negative | |
|------------|----------|--------------------------------------|----------|--------------------------------------|
| | N (%) | Log ₁₀ SCC ± std. dev. | N (%) | Log ₁₀ SCC ± std. dev. |
| <i>can</i> | 15 (30) | 5.42 ± 1.05 | 35 (70) | 5.28 ± 1.03 |
| <i>pvl</i> | 28 (56) | 5.32 ± 1.10 | 22 (44) | 5.34 ± 1.24 |
| <i>sea</i> | 9 (18) | 5.62 ± 1.14 | 41 (82) | 5.26 ± 1.16 |
| <i>seb</i> | 0 (0) | – | 50 (100) | 5.32 ± 1.16 |
| <i>sec</i> | 22 (44) | 5.11 ± 1.11 | 28 (56) | 5.49 ± 1.18 |
| <i>sed</i> | 31 (62) | 5.41 ± 1.03 | 19 (38) | 5.18 ± 1.35 |
| <i>see</i> | 1 (2) | 3.69 | 49 (98) | 5.36 ± 1.14 |
| <i>seg</i> | 35 (70) | 5.33 ± 1.19 | 15 (30) | 5.31 ± 1.11 |
| <i>she</i> | 0 (0) | – | 50 (100) | 5.32 ± 1.16 |
| <i>sei</i> | 50 (100) | 5.32 ± 1.16 | 0 (0) | – |
| <i>sej</i> | 7 (14) | 5.74 ± 1.11 | 43 (86) | 5.26 ± 1.16 |

Sono tutte le tossine che possono essere superantigeni

Ci sono frequenze diverse e risposte cellulari diverse.

La pvl da risultati simili mentre sec è più alta nei negativi rispetto ai positivi (si porta dietro probabilmente delle caratteristiche genetiche o altro che influenza la risposta cellulare)

Confronto con tabella 2x2 la presenza assenza dei fattori di patogenicità e la risposta cellulare.

Ad es cna presente/assente il X² (chi quadro) è non significativo, quindi possiamo dire che anche se ODDS ratio è > 1 (vorrebbe dire che la presenza del gene aumenta la frequenza di mastiti subcliniche)

Tabella:

presenza del gene = esposto

Assenza del gene = non esposto

L'ODDS ratio è > 1 e in teoria posso dire che la presenza del gene cna aumenta il rischio di mastite subclinica. Questo solo in teoria xke per poter fare quest'associazione l'odds ratio deve essere significativa al test X². Non lo è xke X² è 0.26 quindi devo dire che la mia assunzione è sbagliata.

Anche i limiti di confidenza me lo lasciano intuire xkè comprendono 1. È più importante il limite di confidenza del X² xke su numeri grandi X² può diventare significativo pur avendo limite di confidenza compreso 1

Per la proteina A ho invece i 3 cluster di prima

Se faccio una tabella 2x2 e ho 3 cluster, uno fa da riferimento e gli altri li confronto

Table 5

Frequency of subclinical mastitis observed for the different clusters and genes considered and odds ratio for the outcome of subclinical mastitis evaluated by logistic regression analysis

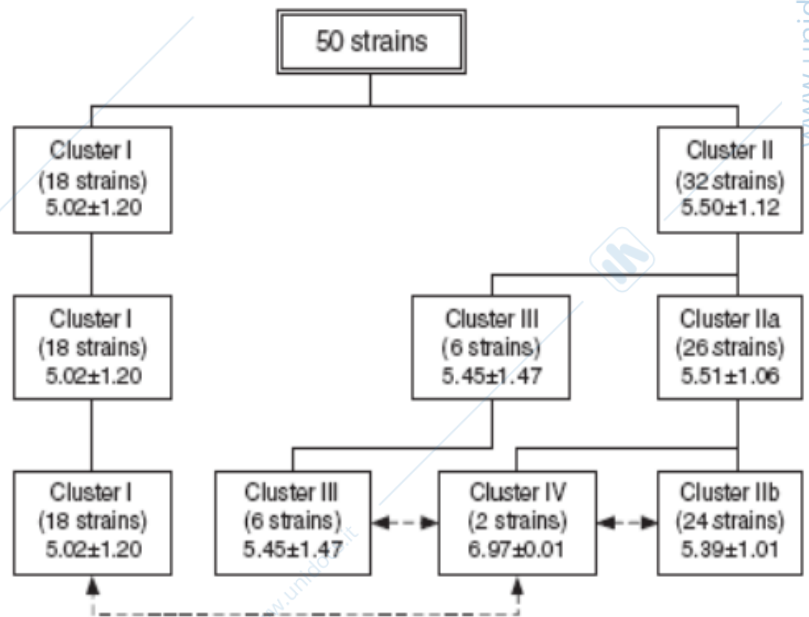
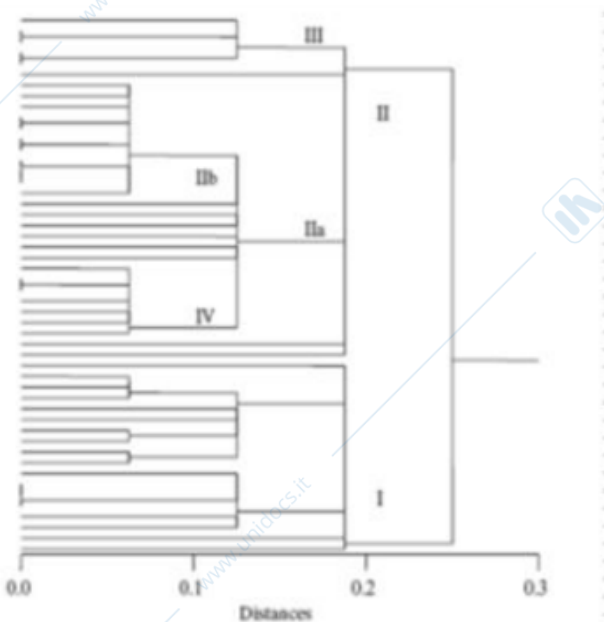
| Gene | Cluster | Frequency of subclinical IMI N (%) | χ^2 (P=) | Odds ratio (conf. lim. 95%) |
|------|----------|------------------------------------|---------------|-----------------------------|
| ona | Positive | 9 (32.14) | 0.26 (0.61) | 1.705 (0.224-12.974) |
| | Negative | 19 (67.86) | | |
| pvl | Positive | 17 (60.71) | 0.01 (0.95) | 1.051 (0.191-5.788) |
| | Negative | 11 (39.29) | | |
| efb | I | 8 (28.57) | 1.04 (0.59) | 2.361 (0.049-113.2) |
| | II | 17 (60.71) | | |
| | III | 3 (10.71) | | |
| apu | I | 11 (39.28) | 4.83 (0.0) | 48.548 (1.511-999.9) |
| | II | 1 (3.57) | | |
| | III | 16 (57.14) | | |
| ara | Positive | 5 (17.86) | 1.81 (0.18) | 0.128 (0.006-2.547) |
| | Negative | 23 (82.14) | | |
| arc | Positive | 10 (35.71) | 0.21 (0.64) | 0.624 (0.087-4.499) |
| | Negative | 18 (64.29) | | |
| aed | Positive | 17 (60.71) | 0.09 (0.76) | 0.768 (0.138-4.289) |
| | Negative | 11 (39.29) | | |
| are | Positive | - | 0.01 (0.98) | 0.001 (0.001-999.9) |
| | Negative | 28 (100) | | |
| arg | Positive | 19 (67.86) | 0.35 (0.55) | 1.641 (0.316-8.520) |
| | Negative | 9 (32.14) | | |
| sej | Positive | 5 (17.86) | 3.71 (0.05) | 24.142 (0.947-615.4) |
| | Negative | 23 (82.14) | | |

Quelli significativi sono la prot a cluster II 8 (con meno di 3 ripetizioni se vado a vedere la tabella sopra) aumenta di 48 volte il rischio di avere mastite ed è statisticamente significativo

Il cluster III che ha piu ripetizioni invece non è significativo. Il limite di confidenza è 0.826-301.9, quindi comprende 1, ovvero non è significativo

Per quanto riguarda sej potremmo dire (visto che il limite di confidenza è 0,947, cioè praticamente 1) che la presenza dell'enterotossina j aumento al 24% la frequenza di mastiti. Lo posso assumere ma devo sapere che dal punto di vista statistico non è significativo

Poi si puo fare un albero di cluster sulla base della similarità genetica. Faccio una classificazione, e poi una volta divisi i cluster associo i dati del contenuto cellulare nel latte



Vedo che il Cluster IV anche se sono solo 2 isolati ha una risposta vicino ai 10.000.000 cellule/ml mentre il Cluster I con 18 isolati ha una risposta di 100.000

Ma qual è il cluster più preoccupante dal punto di vista epidemiologico? In realtà il I.

Quando un patogeno non dà una risposta grave non viene rilevato e trattato quindi in realtà dal punto di vista epidemiologico è più "grave" la forma che non dà una risposta eclatante

Es: covid le forme cliniche gravi contro i ceppi meno virulenti. Si può agire facilmente sui ceppi virulenti mentre quelli che non lo sono fanno continuare a circolare il virus e dare manifestazioni cm nei soggetti più recettivi

Anche l'allevatore si preoccupa della mastite clinica, mentre quelle subcliniche tende ad ignorarle e il batterio continua a girare

METODI DI CALCOLO

SLIDE METODI CALCOLO

Con il metodo Linneo (TASSONOMICO) la classificazione veniva fatta da un gruppo di esperti (1759) quindi discutibili e basate su osservazioni umane soggettive

Successivamente nel 1963 viene proposto il metodo fenetico, la similarità o distanza in base alle loro similarità derivante da tratti genetici o fenotipici

METODO FENETICO: Gli organismi devono essere raggruppati e classificati in base alle loro similarità (distanza) derivanti da precise regole si misura sul maggior numero di "tratti" (genetici o in mancanza fenotipici)

Preve che devo cercare di analizzare caratteristiche di un gran n di soggetti e a quella caratteristica genetica o fenotipica do lo stesso valore (es se hanno coda più lunga o coda più corta vale allo stesso modo)

Vedo quanto i taxa (organismi) sono più o meno distanti tra di loro.

Anche se c'è una grande componente fenotipica è per lo meno un'osservazione oggettiva, rispetto a quella precedente che era soggettiva

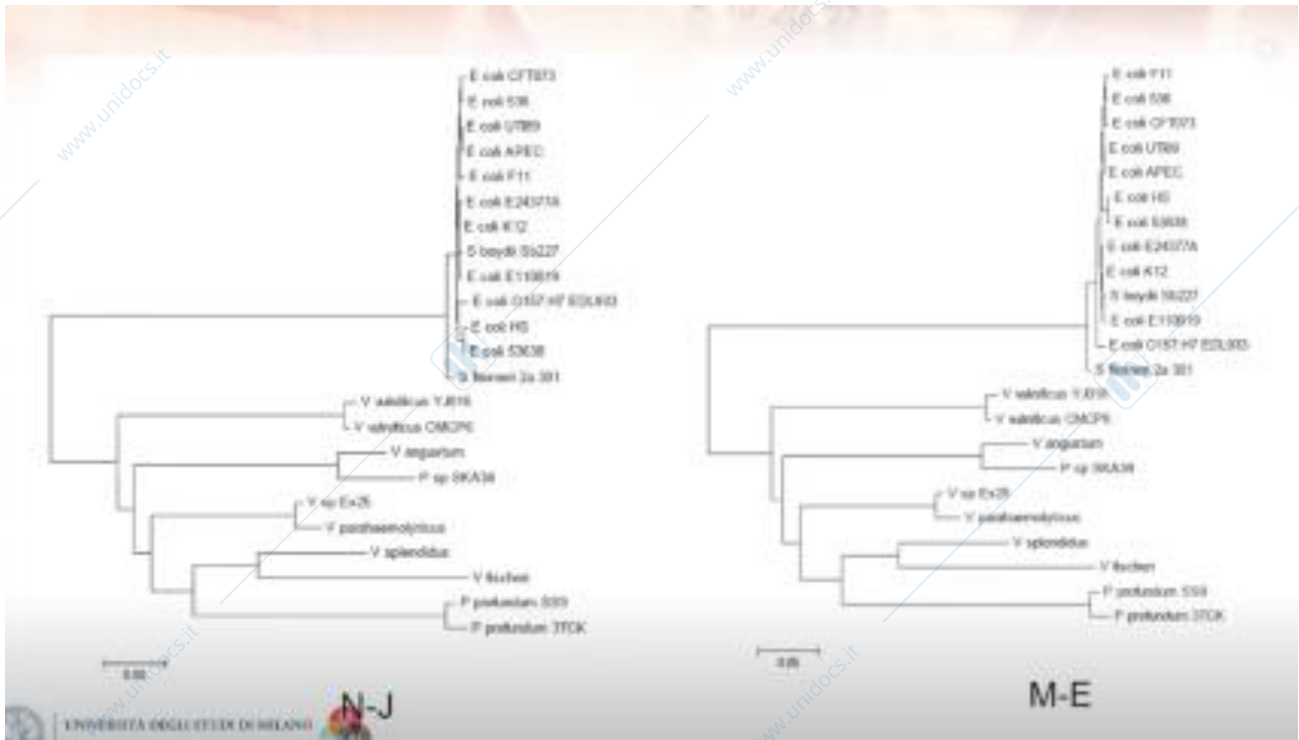
Negli anni 80 viene proposto il **METODO CLADISTICO**: tiene conto dei tratti sinapomorfici (comuni tra discendenti) Non gli interessa tanto vedere la distanza tra i gruppi, ma gli importa sapere che ci sia un ramo, il che indica una discendenza e quindi un'accomunamento tra i gruppi. I cladisti sono più interessati alla presenza di rami che alla loro lunghezza

Quando si analizzano grandi gruppi di organismi, i due metodi danno risultati sovrapponibili

Non esiste un metodo di calcolo unico ma piuttosto dei criteri per scartare quale metodo non è indicato per quel caso

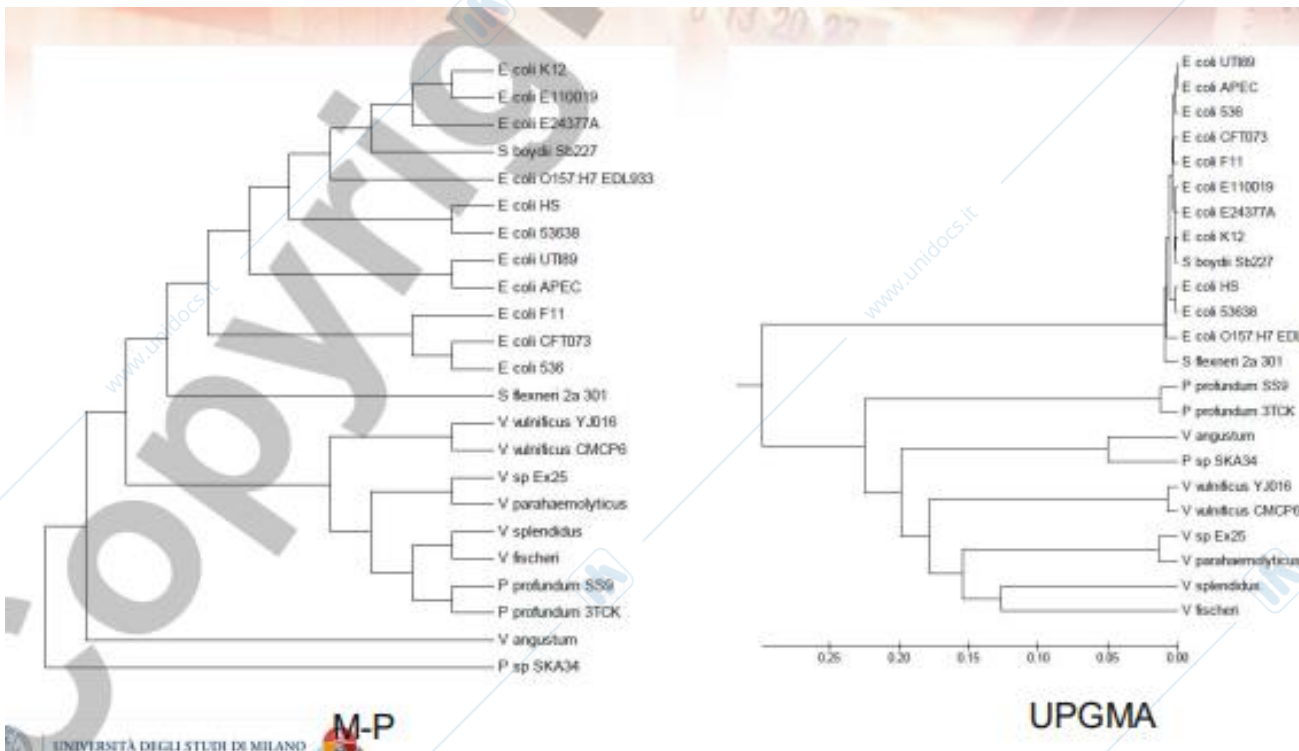
In ogni caso i metodi utilizzati per le valutazioni filogenetiche non sono gli stessi di quelli per classificazioni epidemiologiche

A seconda del metodo statistico che uso per costruire l'albero a partire dalla stessa matrice posso ottenere alberi diversi



Sono due alberi costruiti con metodi statistici diversi a partire da stessa matrice. Sono praticamente uguale ma cambiano i ceppi in alto.

Se sto facendo uno studio epidemiologico diventa un problema considerare tutti i ceppi come diversi. Per questo per quei ceppi considero le distanze e valutando quanto sono distanti tra di loro posso dire se sono uguali o meno. Se sono poco distanti posso clusterizzarli e contarli come uno, a quel punto ho meno dati da dover analizzare



A seconda del disegno che uso ottengo risultati diversi. Il piu usato è il metodo UPGMA

In un albero FILOGENETICO avrò dei NODI (esterni alla fine dell'albero, interni all'origine dell'albero) e dei RAMI (rappresentano la quantita di modificazione genetiche).

Il numero dei rami indica la FREQUENZA delle modificazioni mentre la lunghezza indica la QUANTITÀ delle modificazioni. Alla fine dell'albero ci sono i TAXA (microrganismi finali), all'origine gli ANTENATI.

L'UPGMA quantifica la distanza piu piccola tra i taxa (con metodo DICE) e fa la media, procedendo poi al taxa successivo.

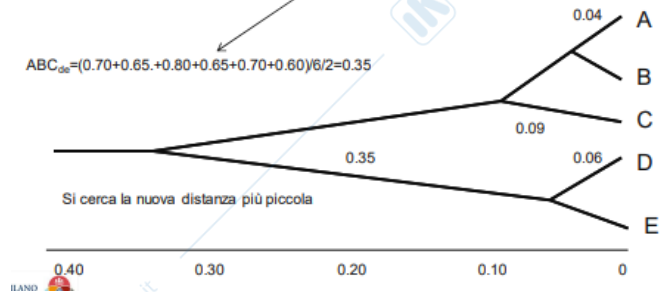
| | A | B | C | D | E |
|---|---|------|------|------|------|
| A | - | 0.08 | 0.19 | 0.70 | 0.65 |
| B | | - | 0.17 | 0.75 | 0.70 |
| C | | | - | 0.80 | 0.60 |
| D | | | | - | 0.12 |
| E | | | | | - |

Calcolo Es. DICE

UPGMA:

- L'albero è additivo (somma distanza da taxa)
- Ultramerico (i taxa sono tutti equidistanti dalla radice). Assunzione quest'ultima spesso non vera

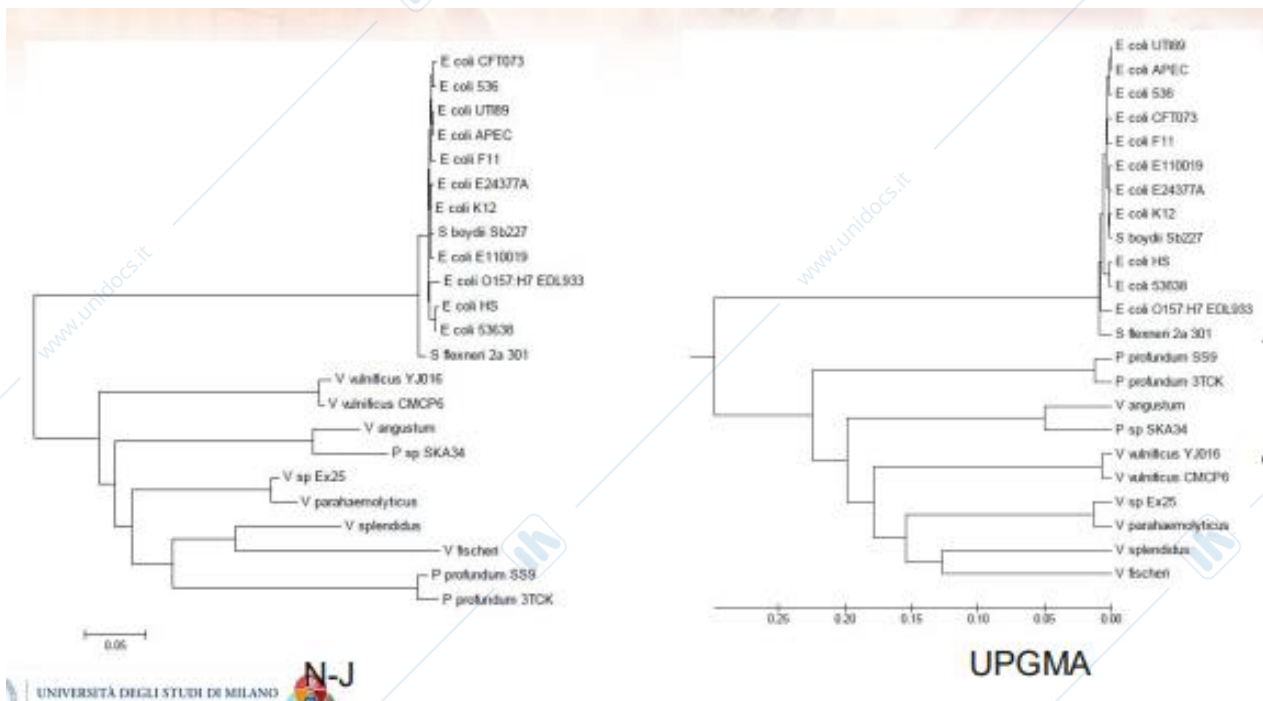
Utile per studi di epidemiologia molecolare ma SCONSIGLIATO per studi filogenetici



Il metodo NEIGHBOR JOINING ha principi simili ma cambia la matrice. Non si assume che i taxa siano equidistanti dalla radice

NEIGHBOR JOINING:

- Non costruisce in partenza cluster ma calcola una matrice di distanze tra taxa
- A partire da quelli più vicini ricalcola le distanze
- Processo iterativo che si conclude quando si ha una matrice finale
- Non assume che i taxa siano egualmente distanti dalla radice



Gli alberi sono molto simili ma i dati sono diversi. Le distanze di alcuni sono molto diverse (es v vulnificus)

Altri metodi utilizzati (definiti come character-based) sono

MAXIMUM PARSIMONY

- Assunzione che l'albero più probabile è quello che richiede il minor numero di modificazioni genetiche per spiegare l'allineamento

MINIMUM EVOLUTION

- Vengono calcolate tutte le possibili distanze tra i taxa e quindi calcolata la distanza minore che verrà a costituire il ramo. Simile a UPGMA, ma senza assunzione di eguali distanze da radice

Come faccio a verificare se i risultati sono giusti?

In realtà qualsiasi albero posso fare sarà sbagliato, xke possono essercene miliardi possibili.

Posso però cercare l'albero che abbia la maggior probabilità di essere corretto.

Per verificare questo si usa il **metodo bootstrap**:

- Replico random i possibili alberi (200 repliche consigliate)
- Valuto la % di persistenza di un ramo e comparo con un altro degli alberi creati

CLUSTERING ANALYSIS: individuo la presenza di gruppi (CLUSTER) in un insieme di dati al fine di classificare:

- La frequenza di comparsa di un dato
- La presenza/assenza di qualcosa (un gene, una caratteristica, una risposta ecc)
- Variabili continue

Sapendo che tipo di dato ho posso usare il sistema di misura delle distanze

Due tipi di clustering

- Sovrapposto (non considerato)
- Esclusivo
 - o Gerarchico (un cluster può essere contenuto in un altro cluster) <- IL PIU USATO
 - o Partizionato (un cluster NON può essere contenuto in un altro)
 - o Additivo (versione grafica – lunghezza=distanza tra cluster) viene fatto solo dal punto di vista grafico

Come misurare?

- Linkage: Average = UPGMA
- Distance: scelgo la tecnica più appropriata tra:
 - o Chi-quadro: per frequenze
 - o Euclidea: variabili quantitative (mancati non considerati)
 - o Gamma: scale ordinali o ranghi (es scale di gravità di una malattia in campo clinico)
 - o MW: quantitative
 - o Pearson: quantitative (mancanti non considerati)
 - o Percento: categoriche e nominali

Criteri per la misurazione: non ce ne sono definiti, un limite generalmente accettato è 20%

→ <20% sono uguali, >20% sono diversi

Sempre meglio verificare i risultati di un albero filogenetico con i dati molecolari rilevati

Quando possibile fare sia un albero basato sulle distanze ed uno basato sulle caratteristiche

Utilizzare un controllo (extra-taxa)

Usare il bootstrapping per confermare i dati

BIOMARKER

POSSO USARE I BIOMARKER PER PREVENIRE/DIAGNOSTICARE UNA MALATTIA

Posso fare un albero di classificazione e vedere che c'è un'associazione tra un cluster e la malattia e sapendo i fattori di patogenicità che ho messo nei cluster posso vedere se uno di questi può essere usato come biomarker

Es Aflatossine nel latte

Si è vista una correlazione tra il carcinoma epatocellulare (causato dal virus dell'epatite B) e l'esposizione prolungata all'aflatossina B1, nei paesi poveri

Voglio cercare di capire dei marker per verificare la presenza di aflatossina nell'alimentazione delle persone per vedere se diventa un fattore di rischio nello sviluppo del carcinoma

Il fattore scatenante è il virus dell'epatite B (400 milioni di persone portatrici croniche), ma solo il 10-25% sviluppa tumore

È ipotizzato che l'aflatossina possa essere un fattore favorente nelle persone affette da epatite B.

Domanda: se consumo tanta aflatossina dalla dieta, che possibilità di sviluppare il tumore ho? (considerando che l'agente del tumore è un virus)

IL TUTTO È FINALIZZATO A STUDIARE UN PROGRAMMA DI PREVENZIONE

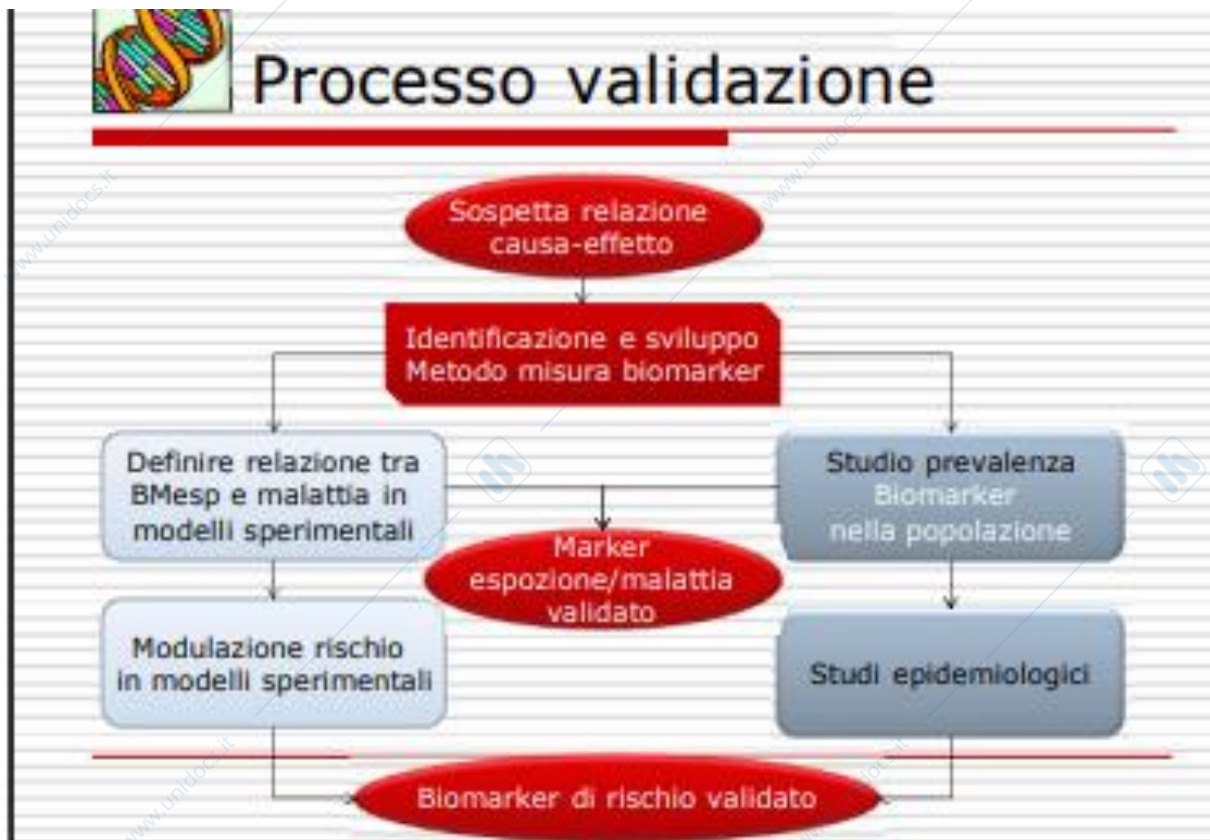
Cerco di definire:

- biomarker di esposizione (presenza di aflatossina in termine di quantità)
- biomarker di effetto (risposta dell'ospite una volta esposto al fattore di rischio aflatossina)
- biomarker di sensibilità (predisposizione innata della persona di sviluppare la malattia)

I biomarker devono poter essere misurati, e questo viene fatto per via molecolare

Una volta che ho il biomarker lo devo VALIDARE, devo verificare se il biomarker mi dà le risposte che mi aspetto.

Inizialmente devo valutare in condizioni sperimentali le performance del biomarker (colonna SX), poi lo applicherò alla popolazione (colonna DX)



È difficile quantificare quanta aflatossina viene ingerita, xke non posso sapere che tipo di alimenti sono stati assunti, in che quantità e quant'era la percentuale di contaminante (in zootecnia posso anche farlo xke so cosa e quanto do da mangiare in allevamento, ma in umana a livello globale non posso)

Devo usare un'altra strategia.

La correlazione virus-tumore è abbastanza chiara e misurabile.

Per l'aflatossina posso misurare le modificazioni del DNA dall'aflatossina, o le proteine che vengono indotte dall'assunzione dell'aflatossina

È un pò piu semplice ed economico misurare le proteine, oltre il fatto che hanno lunga emivita

Ora che ho definito cosa voglio analizzare (le proteine indotte dall'aflatossina) posso farlo a livello sperimentale.

Come:

- **metodo indiretto 1:** se riesco a bloccare (ridurre) effetto AFB1 con mezzi chimici dovrei trovare meno biomarker e quindi posso predire rischio tumore. Studi sperimentali con diversi dosaggi di AFB1 e inibitori. C'è un'effettiva riduzione del biomarker ma può essere utile al fine diagnostico e non preventivo, xke il tumore già c'è
- **metodo indiretto 2:** agente chemiopreventivo → significativa riduzione ALB-AFB → riduzione significativa HCC
ottengo un effetto con beneficio individuale ma non c'è correlazione tra il biomarker e la frequenza del tumore quindi non mi da nessun valore preventivo
- **metodo diretto:** valuto un metabolita dell'aflatossina come markatore, il AFM1, escreto con l'urina. In questo caso ho una relazione diretta tra il biomarker (il metabolita AFM1) e l'aflatossina. Piu AFM1 ho, piu AFB ho. Posso valitarne sia la quantità giornaliera che la quantità ad 1 settimana
- **valutazione del gene p53** (tumor suppressor gene) mutato in seguito all'esposizione dell'aflatossina

quest'ultimi due sono i biomarker che posso usare in uno studio

Studio caso controllo. L'OR=odds ratio

| Coorte | Casi | Controllo | Biomarker | OR |
|--------------------|---------|-----------|---------------------------|---|
| 18224 M | 50 | 267 | AFB urine | 3.4 AFB 7.3 AgHBV 59.4 AFB+Ag |
| 12040 M 13758 F | 56 | 220 | AFB metaboliti urinari | 1.7 AFB 22.8 AgHBV 111.9 AFB+HBV |
| 4691 M 1796 F | 33 | 123 | ALB-AFB | 5.5 AFB 129 AFB+Ag |
| 12024 M 13594 F | 79 Ag + | 149 Ag+ | ALB-AFB | 2 (1.1-3.7) |

Lo leggo:

utilizzando AFB nelle urine ho 3,4 volte la possibilità che si sviluppi un tumore se trovo l'aflatossina, 7,3 volte se oltre all'aflatossina rilevo anche la presenza di anticorpi per il virus, (?????)

poi mi divide maschi da femmine e ci sono diversi numeri di casi, e successivamente soggetti già positivi per il virus (in questo caso si riduce il peso che ha l'aflatossina)

come prevenzione posso dire che se anche sei positivo per il virus, agendo sulla dieta e l'introduzione di aflatossine si può ridurre delle metà la possibilità di sviluppare il tumore