

STATISTICA MEDICA

Lo studio clinico (lezione 1 – 11.04.2018)

Gli strumenti cardine per introdurre elementi scientifici e sistematizzare le conoscenze sono gli studi clinici, gli studi clinici randomizzati, le metanalisi e le rassegne della letteratura.

Cos'è un trial clinico: esperimento che di solito viene condotto per misurare efficacia di un trattamento tp, che ha il significato più ampio possibile, di un singolo farmaco, di un protocollo, di una combinazione etc.

Terminologia non coincidente; fasi di complessità e validità dello studio crescenti progressivamente.

Gruppo soggetti o pz su cui si applica trattamento. Quando due gruppi con trattamenti diversi da comparare come obiettivo: trial controllato. Se assegnazione pz fatta secondo procedura randomizzata si ha trial randomizzato e controllato.

Essendo esperimento deve seguire regole generali di validità degli studi, stesse regole per studi su uomo, animali, in economia etc.

Il primo aspetto da tenere in conto è la differenza tra studio osservazionale e sperimentale. Spesso soprattutto negli studi senza assistenza di uno statistico si trovano delle confusioni tra questi due studi.

Quando tratto dei pz e vedo come reagisco: studio sperimentale, perché il medico governa l'esposizione, decide quali pz si trattano in un certo modo e quali in un altro. Il fatto che ci sia osservazione è ovvio, ma studio sperimentale perché osservatore decide cosa viene fatto ai soggetti, secondo criteri di ragionevolezza scientifica e criteri etici.

Negli studi osservazionali il ricercatore non ha nessuna disponibilità di decidere chi è trattato in un modo e chi in un altro.

Principi da seguire per ragionare su validità disegno sperimentale:

- Bisogna avere qualcosa alle spalle: studio della letteratura. Parte essenziale anche dal punto di vista della valutazione dei progetti, che siano a scopo di finanziamento o altro. Aver chiaro lo stato delle conoscenze a quel momento. Dallo studio di fase 1: prima somministrazione ai soggetti umani; allo studio di fase 4: successivo all'entrata del farmaco nel mercato.
- Ridurre il rischio per i soggetti: attenzione perché si ha a che fare con i soggetti umani. Studio con un disegno di ricerca e disegno statistico adeguati e studio con dimensioni appropriate, non troppo piccolo (lavoro fatto per niente; non si riesce a mettere in evidenza il risultato che si spera di ottenere. Mettere in evidenza significa trovare un risultato statisticamente significativo. Spesso si cerca di fare degli studi piccoli per questioni economiche e di luogo. Con delle formule statistiche si trova il numero di soggetti che servono per lo studio) e non troppo grande. Quando si raggiunge la dimensione ottimale dello studio si parla di prendere una decisione su quale farmaco è meglio: se scelgo il farmaco nuovo, i pz trattati con il farmaco vecchio sono trattati in modo ottimale? Uno studio troppo grande porta al trattamento non ottimale di una parte dei pz. Si sa in partenza che uno dei due gruppi è trattato con il farmaco non ottimale, per questo serve il numero appropriato. Le case farmaceutiche, che hanno soldi e pz, tendono ad allargare gli studi, che si confondono con studi per la pubblicità del prodotto (invece che studi per la ricerca). Qualità. Sorveglianza degli effetti avversi. Non si replica uno studio quando i risultati ci sono già da un altro. Quando si contano gli studi per la registrazione di un farmaco non sono uno solo, ma più studi che vengono confrontati e soprattutto programmati prima.
- Quando si conduce uno studio si ha un progetto, delle ambizioni. La valutazione a priori del ricercatore deve essere messa da parte altrimenti rischia di influire sui risultati portando deriva.

Questa è applicazione delle aspettative del ricercatore che riflettono, non è ricerca scientifica. Bisogna mettere qualcosa in mezzo.

Una dose di farmaco e due pz, uno che sembra andare un po' meglio e un altro che sembra andare un po' peggio e che si sospetta non risponderà alla tp. Su chi investire? Questi due pz simbolizzano i due gruppi, a cui dare il farmaco nuovo e il farmaco vecchio/placebo. Quando si investe il farmaco sul primo pz, ci si aspetta che vinca il farmaco nuovo, anche perché nel braccio del farmaco nuovo sono stati messi i pz "migliori". Per evitare questa situazione erronea: randomizzazione, cioè assegnazione casuale (secondo regole precise che tengono conto calcolo di probabilità ed estrazione casuale; la situazione più semplice è la costruzione progressiva di due gruppi con assegnazioni casuali). Procedura in cui bisogna essere onesti e provare che non si è avuto modo di alterare la procedura (provare id aver utilizzato un protocollo che impediva di influire su esito della randomizzazione). La randomizzazione prevede delle sequenze (dal risultato di uno si fa la stima sull'assegnazione del successivo; non deve succedere) si deve essere in condizione di non conoscere l'assegnazione del singolo soggetto.

Conduzione dello studio e valutazione dei risultati. Se il medico è coinvolto nello studio tende ad avere occhio di riguardo verso farmaco di maggiore interesse con sopravvalutazione dei risultati. Tanti risultati che sono risultato della valutazione del singolo clinico. Qualche margine di incertezza c'è (margini di soggettività del giudizio).

Una seconda barriera: chi fa la valutazione del risultato non deve sapere nulla del trattamento. Talvolta vale anche per il pz, per evitare riscontri soggettivi.

Evitare le analisi in interim: quando si fanno analisi dei dati si rischia di vedere qualcosa che è anche effetto del caso, che non si vede alla fine degli studi singolarmente. Quando si vedono queste cose si rischia di immaginare/focalizzarsi su qualcosa, pensando che lo studio sia finito. Le analisi ad interim sono ammesse ma programmate prima per valutare se lo studio sta procedendo, non per dire dove è arrivato. Ci sono degli aspetti più sofisticati relativi al fissare l'errore di primo tipo ripartendo il valore di p (0.05%) sul risultato finale delle analisi ad interim per il rischio di risultati fuori tempo.

- Controllare i fattori di confondimento: i precedenti erano tutti fattori interni, qua invece si parla di fattori esterni che rendono i due gruppi diversi. Per via della randomizzazione corretta è difficile che i due gruppi siano diversi, quindi la randomizzazione è il primo atto di prevenzione vs il confondimento. Negli studi epidemiologici ovviamente non si può fare, si prende quello che c'è. Gruppi uguali "on the long run": tendendo all'infinito, su numeri grandi. Sul piccolo e singolo studi qualche disomogeneità resta: prepararsi alla misurazione del confondimento. I fattori di confondimento si fanno studiando il problema su cui si lavora; generalmente sono lo stadio clinico, le patologie concomitanti, i FR (effetto sulla mortalità complessiva e sull'aggravamento) ma bisogna studiare l'argomento dello studio.
- Diffusione dei risultati: aspetto etico e di utilizzabilità delle informazioni che escono dallo studio. Studi che per effetto del caso danno risultato positivo. Degli studi negativi non se ne fa nulla, si buttano. Gli studi positivi si pubblicano. Alla fine se uno vuole fare la revisione, se non sa nulla dell'argomento, vedrà solo i risultati positivi e l'idea si sposta vs quello che dicono questi studi. Questo è il "publication bias". A qualsiasi modello ci si muove, i risultati di uno studio vanno pubblicati, cioè resi pubblici, accessibili in modo che in una ricerca sistematica anche quelli negativi compaiano. È difficile applicare questa cosa soprattutto chi studia per guadagnare con lo scopo di ottenere risultati positivi. Anche problema di disinteresse per studi che non hanno dato nulla. È un aspetto che va assolutamente trattato.

Non bisogna magnificare effetto del caso, bisogna avere un progetto prima. La strada principale va conosciuta prima, non si va a caso a seconda dei risultati. Anche per questo serve il ruolo dello

statistico; lo statistico va considerato al momento del disegno dello studio, quando si inizia a pensare al problema.

- Pubblicazione. Ci sono diverse filosofie degli studi: studi che dimostrano che A è meglio di B in astratto, nelle condizioni migliori possibili (trattamento fornito in modo ideale, in pz senza problemi etc. Sono gli studi "autorizzativi", per portare il dossier al FDA o EMA, AIFA e così via. Sono fatti da chi ha investito sulla produzione del farmaco); studi pragmatici (farmaco applicato nella realtà, su pz con comorbidità, poca compliance etc) che si chiedono se comunque c'è impatto complessivo del farmaco (spesso sono studi post-marketing. Il problema del prezzo è in secondo piano, si deve pensare se il farmaco serve o meno, nelle condizioni reali di vita. Molto di questo lavoro è fatto da gruppi di ricerca, che disegnano studi di valutazione dell'efficacia di combinazioni di farmaci). Problema dell'utilizzare un farmaco che deriva da ottimi studi che però sul pz "normale" non va tanto bene; applicazione del farmaco alla pratica clinica.

Immagine: ogni curva è curva di probabilità; distribuzione dei campioni tra cui farmaco nuovo e vecchio non c'è differenza. La seconda curva da popolazione in cui il farmaco nuovo serve a qualcosa.

Potenza dello studio: capacità di vedere il risultato come statisticamente significativo. Se si cade nella curva a dx si dichiara statisticamente significativo.

Se è vera ipotesi di lavoro (curva azzurra) si dichiara il risultato statisticamente significativo solo per la parte verde della curva.

Curva più stretta: aumentando la dimensione del campione diminuisce errore standard. Effetto è che area verde cresce in proporzione.

Questo è il significato di dare la dimensione adeguata. In base alla dimensione e all'ampiezza delle curve devo dare dimensione per avere area verde sufficientemente grande. Questo è tutto ipotizzato dal ricercatore.

Se si fa crescere ancora lo studio le curve si assottigliano e si distacca area di sovrapposizione con aumento della potenza. Punto in cui non c'è aumento della potenza ma si ha spreco e si ricade in discorso per cui spreco costa e non è etico.

Randomizzazione: pz che hanno passato criteri di selezione, verificati (selezioni o screening, cioè "passare al setaccio"). Scelta casuale e assegnazione al primo o al secondo trattamento. Disegni a più bracci, fattoriali, con randomizzazione successiva per successivi trattamenti e così via.

Obiettivo principale è eliminare distorsioni sistematiche, scelta del ricercatore nell'assegnazione etc.

Dal punto di vista statistico ipotesi nulla si materializza nell'avere due gruppi coincidenti. Ipotesi nulla se vera corrisponde a dire che gruppo del trattamento A è uguale a gruppo del trattamento B.

I test statistici dicono quale è probabilità di ottenere quel risultato sperimentale se è vera ipotesi nulla. Se probabilità alta vince ipotesi nulla e se è bassa probabilità vinco ipotesi di lavoro.

Controllo del confondimento: problema meno grave dopo la randomizzazione, ma che comunque va considerato concettualmente.

Immaginiamo 3 trattamenti e pz classificati in pz molto gravi, gravi, intermedi e lievi. Probabilità di sopravvivenza è molto diversa. Suddivisione random dei pz per avere più omogeneità. Se non faccio randomizzazione il farmaco nel gruppo dove pesa di più la quota dei gravi andrà peggio.

Analisi separata strato per strato: si scopre che non ci sono differenze (ricorda che c'è sempre incertezza).

La tabella 1 dello studio dice quanti pz si sono considerati e come sono stati suddivisi. Serve per capire se ci sono presupposti per problema in cui randomizzazione non abbia funzionato. Se non ci sono problemi le analisi del trial sono molto semplici. Concettualmente se due gruppi sono simili il risultato che interessa viene trovato semplicemente, perché non si ha il problema del confondimento.

Il risultato totale è una media pesata delle sopravvivenze per ogni strato dove il peso è la numerosità di ogni strato per ogni colonna. Nei quadrati della griglia dovrei avere sempre 50.

- Randomizzazione
- Appaiamento individuale (matching, studi cross-over): il primo soggetto è maschio e finisce in un gruppo, mi aspetto di avere un altro soggetto maschio con le stesse caratteristiche (ad es non fuma, fa attività fisica adeguata etc). Appaiamento viene in natura generalmente, altrimenti grande complessità pratica. Gli studi appaiati di solito vengono fatti quando c'è già appaiamento naturale (coppie di gemelli, misure ripetute nel tempo nello stesso soggetto etc).
- Bilanciamento dei gruppi a confronto (stratificazione)
- Limitazione ad un solo strato: situazione estrema. Gli studi esplorativi (delle case farmaceutiche) in fondo sono studi stratificati con criteri molto stretti in previsione di adeguatezza e proseguimento del trattamento.

Terapia del confondimento:

- Analisi stratificate
- Analisi multivariate

Se si ha il sospetto che il farmaco funzioni meglio tra i fumatori ma ci si è dimenticati di chiedere se i soggetti fumano (non era previsto, non è venuto in mente etc; si mettono insieme pezzi di informazioni). Nell'ipotesi migliore si può tornare indietro e chiedere ma se si parla ad es di campioni la laboratorio non si può tornare indietro. Per questo bisogna ragionare molto all'inizio dello studio. Capire cosa potrebbe servire; il momento chiave è la partenza dello studio, è meglio avere qualche informazione in più che qualche informazione in meno fondamentale.

Interazione: concetto spesso usato a sproposito. Immaginiamo due trattamenti che su qualche scala portano diversi livelli di un parametro. Poi si scopre che il parametro cambia ad esempio con l'età. Se si mettono insieme: due trattamenti mantengono la stessa distanza nell'effetto su tutte le età (funzionano allo stesso modo) oppure sui giovani non c'è differenza ma sugli anziani ce ne è molta. Non è un problema piccolo. Per analisi dell'interazione è necessario avere studi designati apposta. La base del disegno è il "disegno fattoriale": nello stesso studio sugli stessi pz verifico effetto dei due diversi fattori. Distribuzione dei fattori e dei trattamenti. Valutazione dell'effetto congiunto dei due fattori.

Quando ipotesi di interazioni, queste si possono far entrare in modo pianificato nel disegno dello studio.

La sequenza delle sperimentazioni è una sequenza logica che parte dalla sperimentazione nelle condizioni di non rischio e passa al rischio maggiore. Condizione di minor rischio è quella in laboratorio (su colture cellulari, su animali, simulazioni etc), che fa parte dello **studio pre-clinico**, che serve per anticipare caratteristiche del farmaco e della sua interazione con organismo. Anche provate sottoponendo il farmaco ad animali di laboratorio, sani o malati. Per lo studio di farmaci oncologici si tende a impiantare cellule tumorali sugli animali, in modo da farli ammalare.

Evoluzione a costo più basso è costruire aggregati di cellule in laboratorio, che danno migliore idea dell'efficacia del farmaco.

Non si parte da zero.

Dopo il pacchetto pre-clinico si passa agli **studi di fase 1**: pochi posti in Italia. Sono le prime occasioni di somministrazione del farmaco. In altri paesi si fanno molto di più. Come reagisce organismo umano. Pochi soggetti, scaglionati in modo sequenziale, in modo da interrompere lo studio quando necessario. Obiettivo: stabilire quando interrompere e quali sono gli effetti avversi. C'è un gruppo solo (no studi randomizzati). Disegno statistico: da dose iniziale definita sulla base dei risultati della sperimentazione pre-clinica; se nessuno dei pz ha dei problemi si passa ad aumento della dose (algoritmo di incremento della dose è predefinito; all'inizio aumenta del 30% e poi a scaglioni che si riducono secondo sequenza

predefinita) e si espongono altri tre pz e così via). A un certo punto uno o più soggetti hanno effetto avverso. Si passa alla dose precedente. Se è più di uno ci si ferma e questo è il limite. Possibile che nello studio di fase non ci siano soggetti sani ma malati, comunque obiettivo è valutazione effetti avversi, che sono classificati, anche in base alla gravità (generalmente classificazione su scala di 4 livelli).

Studi di fase 1b: applicazione di un farmaco già studiato e già definito sulla base degli aspetti regolatori (pronto per passare in commercio) in condizioni diverse. Ad esempio farmaco per la leucemia provato per il linfoma. Soprattutto in onco i farmaci hanno effetti su diverse condizioni patologiche. Soggetti aati della patologia a cui si interessa portar eil farmaco ma il disegno dello studio è sostanzialmente lo stesso, di ottimizzare la dose (che può essere diversa) in relazione agli effetti avversi.

Studi di fase 2: obiettivo è quello di avere una valutazione preliminare dell'efficacia del trattamento. Sulla base di uno studio piccolo ho un suggerimento sul fatto che il farmaco funzioni. Piccolo perché gli studi costano. 9 farmaci su 10 si fermano alla fase 2, perché non hanno dimostrato indicazioni di capacità tp sufficienti. Incremento dell'investimento più contenuto possibile.

C'è una tempistica: uno studio piccolo condotto in tempo breve è meglio di studio più grande che richiede aa.

Alcune volte si aggiunge aspetto degli end point surrogati: end point è effetto che vado a misurare. Tra i tanti effetti scelgo quello su cui si concentra attenzione, su cui anche si misura la dimensione statistica dello studio. Di solito effetto di maggiore interesse è la sopravvivenza. Ci sono situazioni in cui i tempi di sopravvivenza sono molto lunghi e non si può aspettare tutto quel tempo. End point surrogati sono qualcosa di intermedio nella storia clinica della malattia, che può essere qualcosa id realmente intermedio (es dimensioni della massa) o qualcosa che è parallelo (risultato di laboratorio, marcatore tumorale etc). Obiettivi: valutazione preliminare di efficacia e valutazione ulteriore degli effetti avversi.

Informazioni preziose per la valutazione della fattibilità degli studi successivi. Si cerca di utilizzare i risultati degli studi di fase 2 non sono per decidere se proseguire con la fase 3, ma c'è sempre di più la tendenza, approvata qualche aa fa, di considerare con certi limiti gli studi di fase 2 anche per autorizzazioni provvisorie all'uso del farmaco in contesto clinico allargato, non sono sperimentale. Si parla di farmaci biologici (immunotp, Ab monoclonali etc) che sono molto costosi; in condizioni particolari (almeno due studi fase 2 abbastanza rossi con risultati coerenti) si danno autorizzazioni provvisorie mentre va avanti al programmazione. Situazione discutibile perché comunque sono studi in cui si sposta l'asticella dell'errore nella curva gaussiana e si porta la soglia per il rifiuto a un livello più alto (probabilità più alta di dare come rilevanti gli effetti del farmaco che invece magari non ci sono). Si fa quando si conducono degli studi randomizzati, che sono studi 3 in miniatura; quando serve un gruppo di controllo per valutare efficacia del farmaco. Se si vede che la massa neoplastica si riduce il farmaco è efficace (non serve gruppo di controllo perché la massa da sola non si riduce); se invece studio farmaco contro la cefalea serve studio di comparazione (si fa studio piccolo di fase 2 se non ci sono fondi).

Gli studi di fase 2 non ci sono informazioni sufficienti per decidere se il farmaco va bene; sono informazioni per decidere se proseguire con la ricerca.

Studi di fase 3: studi chiave, comparativi. Si ha un trattamento corrente e un trattamento nuovo che ha passato tutte le fasi precedenti; in particolare ha passato la fase 2 e vengono investite ulteriori risorse. Confronto del farmaco vecchio con quello nuovo e porta a decidere quale sia meglio.

Efficacie ed effetti avversi sono i due aspetti. Analisi in modalità indipendenti, come viene fatto di solito: obiettivo principale di efficacie, si costruisce uno studio per misurare efficacia. Parallelamente si raccolgono informazioni sugli eventi avversi. Recentemente cominciano a comparire studi che dal disegno integrano le due dimensioni di efficacia ed effetti avversi in unica valutazione e portano a costruire conclusione definitiva per lo studio che pesa entrambi gli aspetti. Normalmente succede che il farmaco A migliora la sopravvivenza

ma aumenta gli EC, quindi che fare? Questi altri disegni statistici invece pesano insieme i due aspetti di vantaggio e svantaggio con valutazione complessiva.

Disegno generale di uno studio: risultato osservato individuale e poi sintesi di gruppo che viene poi confrontata. Test dell'ipotesi nulla assenza di differenza. Molti studi valutano che non ci sia assenza di differenza (studi di equivalenza). Ci sono poi gli studi di superiorità, che valutano se un farmaco è meglio del vecchio: tutte le tappe sono i processi del protocollo. Da una ricognizione di quello che si conosce (effetti positivi, negativi, costi etc), problema su come affrontare eventuali problemi etici che possono essere suscitati, disegnare e dimensionare lo studio, dare i criteri di accettazione, criteri per decidere se il pz ha avuto successo o meno, sintesi dei risultati e analisi dei dati.

Studio clinico (lezione 2 – 02.05.2018)

Disegno dello studio di fase 1. Tabella standard che si può modificare negli studi; logica della escalation della dose: primi salti e poi incremento si riduce. Ogni step sono almeno 3 pz in più salvo step di conferma; almeno 24 soggetti al punto 8.

Algoritmo che descrive come procedere.

Per gli studi di fase 2 modalità dal punto di vista statistico per disegni a 1 e 2 stadi.

Per il disegno a 1 stadio: calcolo IC di una proporzione (formula). Proporzione di successi nel gruppo, calcolata in IC della proporzione; si vede se inclusione del valore che si ritiene il valore normale per il trattamento standard relativo allo studio. Se non contiene c'è prospettiva di condurre studio che porta a evidenza di risultato superiore dello standard se contiene meglio lasciar perdere (farmaco non promettente).

Disegni a 2 stadi: separazione dei problemi. (1) Il nuovo farmaco dà qualche risultato superiore a minimo accettabile, che fisso come proporzione di successi (pi greco 0)? (2) Se superamento soglia minima che viene verificata con pochi soggetti si valuta se superamento soglia del risultato standard del trattamento (pi greco 1). Lettura dei risultati su tabelle.

Studi di fase 3: valutazione della soglia del singolo soggetto e valutazione di gruppo.

I protocolli delle case farmaceutiche hanno termini precisi; quando protocolli fai da te meno chiari.

- **Obiettivo generale** dello studio.
- **Outcome**: qualcosa che corrisponde di solito a un aspetto di tipo sanitario; qualcosa relativo alla salute dei pz. Noi intendiamo o speriamo di poter modificare con il nuovo trattamento. Non è per sé misurabile, ha livello generale concreto, più ristretto come vantaggio rispetto all'obiettivo. È qualcosa che non si può misurare direttamente.
- Emergono delle quantità misurabili dall'outcome: qualcosa che misuro che serve per decidere se pz ha raggiunto il suo risultato e mettere insieme i risultati di tutti i pz. **End point**.

Esempio: outcome variazione AP. End point è il risultato numerico della PA allo sfigno. Obiettivo generale è la riduzione della mortalità o degli eventi CC conseguenti alla ipn.

Outcome ed end point sono divisi: outcome principale ed end point principale (di maggiore importanza; spesso non si trovano negli studi vere gerarchie di importanza, ma sulla base della possibilità di valutazione nei tempi e nelle risorse). Su questo si calcola la potenza statistica dello studio. Obiettivo primario non sempre misurabile.

Si decide come obiettivo primario qualcos'altro: end point surrogati.

Chiara la definizione di obiettivo primario e corrispondente quantità e quello su cui si disegna lo studio.

Obiettivi II sono tutti gli altri, insieme a end point ed outcome II.

Esempio: riduzione lesione neoplastica è outcome, che viene misurato con una rx confrontata con altra prima e dopo trattamento (la variazione che corrisponde alla pratica clinica, con criteri standardizzati di valutazione delle dimensioni; è end point).

Quando si parla di sintomi traduzione della sensazione in qualcosa che diventa quantitativo: scale visuali analogiche (SVA); disegno di una riga, in cui il pz dice dove si colloca il sintomo su una scala ad es da 1 a 100. Traduzione di dato soggettivo in dato misurabile oggettivo.

Quando si parla di sopravvivenza si parla di scala temporale.

Spesso no misurazioni dirette, ma si usano degli indicatori surrogati che spesso sono prodotti dal laboratorio. Anche rx è un indicatore surrogato perché ci dà solo un aspetto della lesione neoplastica che può anche avere localizzazioni diverse.

Il problema è non sono necessariamente correlati in modo così stretto con obiettivo e risultato e quantità che ci interessa misurare. Riduzione marker non necessariamente predittore cambiamento di sopravvivenza.

Poi valutazione della risposta a livello individuale (studi in cieco, in doppio cieco etc).

Esempio: studio di equivalenza tra farmaci. Secondo il giudizio del clinico il farm nuovo è anche molto meglio del precedente. Revisione indipendente revisiona lo studio. Obiettivo raggiunto ma giudizio del clinico fortemente ottimista rispetto al giudizio di valutatore esterno.

Trial di superiorità: dimostrare che un trattamento (E: trattamento sperimentale) è superiore a un altro (C: trattamento controllo). Se superiorità avremo differenza a vantaggio di un trattamento; delta va a favore di un trattamento e IC di questa differenza tra le proporzioni non comprende il valore nullo, cioè quello che corrisponde alla assenza di effetto del farmaco, che corrisponde alla ipotesi nulla di assenza di differenze tra farmaci; delta di 0. Differenza che va in favore di una alternativa; IC della differenza non comprende lo 0 quindi differenza ss in modo corrispondente all'ampiezza del IC (es con $p < 5\%$ se IC al 95%).

Esempio: testati gruppi di FANS con placebo.

Trial di equivalenza: altra situazione in cui obiettivo/ipotesi di lavoro è che 2 trattamenti siano equivalenti tra loro. Delta vicino allo 0 e IC in fascia di interesse; la fascia di interesse non è definita dallo statistico ma dal clinico. Per la variabilità casuale mai studi in cui perfettamente identici i risultati con differenza esattamente sullo 0; ci sarà qualche escursione mai tale per cui limite IC esce dalla fascia di rispetto. Fascia di rispetto (margine di equivalenza) è quello che accetta il clinico, che ritiene tollerabile come differenza di trattamento tra due farmaci per poterli usare come equivalenti. Deve essere abbastanza piccolo.

Trial di non inferiorità: spesso obiettivo è cercare situazione in cui il farmaco nuovo è equivalente o superiore rispetto al farmaco vecchio.

Il limite superiore dice che il vecchio non può superare il nuovo per un certo limite (il farm nuovo può essere un po' peggio del vecchio ma non più di un tot; valore guardato rispetto non alla differenza ma rispetto al IC). Se il farm nuovo è meglio non si pongono limiti e c'è fascia di rispetto che guardo solo in relazione a una delle possibilità. Come dire non guardo IC per quanto riguarda la coda migliore. Dal punto di vista statistico vantaggio perché IC come se fosse unilaterale al 97,5%.

Nei trial di equivalenza si guardano entrambe le code, di superiorità e non inferiorità mentre nei trial di non inferiorità si guarda solo la non inferiorità.

Quando usare trial di non inferiorità: qualcosa per cui non ci si aspetta superiorità rispetto al trattamento standard (farm o placebo; attenzione alla scelta del farm se tanti farmaci diversi, farm nuovo rispetto al farm più forte che dà migliori risultati tra i vecchi. Non deve essere rispetto al peggiore dei vecchi). Altra situazione è quando ci si aspetta che tra farm che sono simili vi siano delle differenze relative al costo, agli EC, alla via di somministrazione, all'accettazione da parte dei pz.

Non considerati aspetti statistici per il calcolo della potenza dello studio (calcolo diverso tra non inferiorità e superiorità). Cambiando uno dei termini della formula dell'equazione fondamentale: applicazione della formula agli studi di superiorità. Per gli studi di equivalenza la formula va leggermente modificata, per considerare ampiezza IC più dell'errore di primo tipo.

Analisi dei dati

Differenze tra clinico e revisore esterno: CAP vs 5FU. Revisore esterno non a conoscenza di quale trattamento ha ricevuto il pz. Confronto anche con le lastre.

Non facendo così ci sarebbe bias di riconoscimento (differenza di attenzione del malato rispetto al caso negli studi caso-controllo).

Attenzione: quando non applicazione delle metodiche di valutazione e solo risultato del clinico, lo statistico non viene incontro; dall'analisi statistica dei dati non si arriva a dato corretto perché dati sbagliati, valutati in modo sbagliato perché diversa sensibilità connessa al tipo di trattamento.

Importante la valutazione indipendente.

Come si analizzano i risultati di un trial: se condotto correttamente con randomizzazione analisi è facile, senza confondimento (che è il problema principale degli studi caso-controllo). A seconda del tipo di variabile da analizzare ci sono diverse metodiche di analisi.

Assenza del confondimento; applicazione corretta della randomizzazione, verificata attraverso risultati (ci sono degli esempi, delle tabelle, in cui si descrive baseline dei gruppi; no risultati diversi dal punto di vista ss se non risultati diversi per il valore di p. Qualche differenza da valutare in termini di frequenza di risultati ss confrontando con proporzione risultati ss attesi su base di ipotesi di effettiva randomizzazione (non più di 5 su 100) e non concentrati su variabili critiche o di interesse). Questo fa parte di standard di metodo di presentazione dei risultati.

Le metodologie sono quelle standard. Utile saper utilizzare metodi multivariati; se dubbio che il caso abbia lasciato qualche sbilanciamento tra gruppi, utile usare più variabili insieme per dare spiegazioni. Le analisi non sono fatte tutte allo stesso livello gerarchico; prima analisi corrispondenti all'end point primario (se significativo è risultato principale dello studio). Se ripetizione dello stesso esperimento (es lanciare moneta n volte) per n volte ci si aspetta certa frequenza di risultati ss; quando ripetizione analisi nello stesso studio con stessi dati tenere conto di questo e modificare valore dell'errore di I tipo per tenere conto della molteplicità delle analisi effettuate.

Se errore I è p, dividere per il numero di test effettuati (che si intendono effettuare nella pianificazione complessiva delle analisi). Tanto più sono numerose le analisi tanto più basso è il valore di p accettabile per ottenere risultato ss. Se stesso livello gerarchico non si trova nulla. Si procede stabilendo livelli gerarchici: prima il livello principale poi tutte le altre che si leggono allo stesso modo ma dichiarate come meno importanti.

Esempio: se 10 anni con 10 analisi, aumento di 10 volte probabilità di ottenere risultato ss.

Dividere p (0,05) se non si vuole per effetto del caso probabilità per effetto del caso, risultato ss oltre certa soglia. Fissazione del valore di p a livello basso (0,001 ad es) e valore di p per analisi vere a 0,049) in modo che somma sia 0,05, uguale a quella accettabile. Succede che uno mai ss ma in complesso dà sensazione di come va lo studio anche considerando il valore di p. Io 0,049 è quasi uguale a p iniziale e consente di

concludere lo studio esattamente come pensavo e fa fare analisi ad interim. Nel protocollo analisi ad interim devono essere dichiarate prima (vedere se studio va come voluto).

Le analisi multivariate consentono di tenere conto di più variabili insieme senza limite delle analisi ripetute. No correzione per la ripetizione delle analisi. Evitare il problema delle analisi ripetute con modifica necessaria del valore di p.

Metodi non parametrici si usano poco nei trial perché con randomizzazione siamo sicuri che distribuzione gaussiana valida.

Attrition bias: abbiamo due gruppi di soggetti randomizzati per due AD. Se EC del farmaco assegnato passaggio ad un altro trattamento, ma se solo due farmaci disponibili si passa da un braccio all'altro dello studio di trattamento, ma come analizzare questi soggetti? Per il braccio iniziale di assegnazione o per il secondo farmaco a cui il soggetto è stato assegnato dopo gli EC?

Il clinico dice di analizzare effetto del farmaco quindi analisi sul farmaco ricevuto e chi ha cambiato gruppo analisi nel gruppo trattato. Va bene se voglio vedere EC, per cui devo essere più sensibile possibile, senza mescolare le carte tra i gruppi. Obiettivo dello studio è decidere se un farm funziona diversamente dall'altro; se clinico vuole alterare i risultati deve sapere chi sta ricevendo cosa; poniamo di voler far vincere un farmaco. Prendo pz che vanno male che andrebbero male comunque e passo sull'altro gruppo. Prendo p che vanno bene che andrebbero bene comunque e passo sul primo gruppo. Se c'è differenza l'ho aumentata ma questo è barare.

Succede che alla fine analisi principali vengono condotte sulla base del trattamento inizialmente assegnato senza badare a quello che è successo dopo; chi cambia gruppo resta per le analisi nel gruppo originale. Se c'è differenza tra i due farmaci le persone passate da gruppo all'altro portano differenza più ridotta ma mai a differenza aumentata; così evito risultati arbitrariamente diversi tra farmaci. Logica applicata agli studi di superiorità.

Negli studi di equivalenza tutto quello che posso mescolare può dare risultati imprevedibili; se due trattamenti con poca differenza no alterazione risultati ma se differenza difficile previsione del risultato. Lasciare soggetti nel gruppo di assegnazione iniziale.

Differenza assoluta: effe soggetto trattato meno effetto gruppo standard. Rischia attribuibile: effetto proporzionato al valore gruppo di base. Possiamo anche (contesto studi di prevenzione) contare quanti soggetti devo trattare per prevenzione evento: number needed to treat (inverso della variazione assoluta del rischio).

Immaginiamo gruppo di soggetti standard, gruppo sperimentale, numero di eventi nei due gruppi. La frequenza degli eventi ha due % diverse; la riduzione assoluta del rischio ha due % diverse.

Misure che si utilizzano per fare revisioni sistematiche.

Articolo: *Brentuximab Vedotin with Chemotherapy for stage III or IV Hodgkin's Lymphoma*.

Farm nuovo con bleomicina aumenta la tossicità della stessa. È uno studio su protocolli di farmaci. Obiettivo è diminuzione di farmaci nel protocollo.

Trattamento standard: ABVD (doxorubicina, vinblastina, bleomicina e dacarbazina).

Trattamento sperimentale: A-AVD (brentuximab, doxorubina, vinblastina e dacarbazina).

End point: progression-free survival. End point II: overall survival.

Metanalisi (lezione 3 – 03.05.2018)

Le rassegne sistematiche e metanalisi sono continuazione uno dell'altro. Prima rassegna sistematica poi stimatori comuni tra i diversi studi (metanalisi).

Scala di importanza nel parere dove il valore più basso opinione degli esperti e valore più alto clinical trial randomizzati. Mettere insieme il prodotto di diversi studi sullo stesso argomento e confrontabili e fare una sintesi qualitativa (revisione) e sintesi numerica per la metanalisi.

Risultato metanalisi dal punto di vista grafico: una riga per studio con info relative allo studio (identificazione dello studio, come nome, rivista etc; informazione sul risultato dello studio, il limite delle metanalisi e delle revisioni è considerare un risultato per volta, quindi decidiamo noi il risultato del singolo studio da confrontare con altri risultati corrispondenti degli altri studi). Molte debolezze e vantaggi.

Punto debole: scelta del valore non priva di arbitrarietà e se diversi studi hanno problemi analoghi che portano risultato più o meno favorevole o contrario, facendo analisi congiunta, se non ce ne accorgiamo, rafforziamo il giudizio basato su difetti di singoli studi.

Attenzione alla lettura degli metanalisi. Dato che qualcuno ha già fatto una sintesi e selezione probabilmente ha già fatto il lavoro.

Risultato in termini numerici e grafici. Valore puntuale dello stimatore è il valore centrale. Quadrato proporzionale alle dimensioni dello studio o alla importanza dello studio nella metanalisi. Asse lineare o asse logaritmica. IC.

Forrest plot è questo schema di risultati con i quadratini.

Definizione di metanalisi: procedura statistica che mira a sintetizzare risultati di studi indipendente producendo stime di effetto complessive che sono medie pesate del prodotto di singoli studi. Obiettivo dare stima più precisa, riducendone la variabilità e anche, essendo la media di stime puntuali, che valga la regola del campionamento ripetuto.

Assunzioni. Scegliere misura del risultato che sia la stessa. Dati di solito sono disponibili perché le metanalisi cercano di usare i dati pubblicati. Assunzione è che variazione tra i diversi studi sia solo effetto del caso, che entrambi stimino la stessa cosa. Quando tra gli studi ci sono state delle scelte diverse può condurre a necessità di non condurre unica analisi ma analisi razzionate.

Pubblicazione dello studio: se pubblicato si trova altrimenti no. Generalmente studi pubblicati sono quelli importanti e pesanti, numerosi, che finiscono comunque in una pubblicazione. E poi pubblicati quelli che danno risultato interessante; studi negativi spesso non pubblicati.

Distribuzione degli studi trovati. In questo caso ogni studio è rappresentato da incrocio tra valore del risultato e della numerosità. Media dei risultati ottenuti altrimenti attenzione maggiore vs studi con risultato positivo. Si usa un triangolo oppure due iperbole; anche precisione invece della dimensione.

Quello visto è metodo di analisi con procedure a effetti fissi; c'è stimatore vero della popolazione e tutti gli studi come se fossero campioni dalla popolazione. Ogni studio ha proprio risultato che è risultato più errore casuale, con segno + o -. In questo modo mettendo insieme studi diversi, si ha riduzione del peso dell'errore casuale, distribuito in parte con segno + e in parte meno, con stima complessiva vicino al valore vero da stimare. La distribuzione è gaussiana, si assume distribuzione errori sia gaussiana e quello che differenzia errore casuale da valore complessivo è epsilon.

Fare una media, possiamo farlo come media semplice ma non ci conviene. Ci sono studi vicini al valore migliore e sono anche gli studi più pesanti; gli studi più lontani sono quelli meno pesanti. Pesare contributo di ogni studio per qualcosa corrispondente alla importanza dello studio, che è data dalla probabilità. Media pesata che valorizza valore più pesante.

Struttura della formula della media pesata. Il peso è qualunque cosa che associamo allo specifico studio. Al denominatore somma di tutti i pesi.

Nel caso delle metanalisi la procedura è usare inverso della varianza, calcolato in modi diversi a seconda del tipo di dato. Se variabili continue (es peso, albumina etc) la varianza è calcolata per ogni studio esattamente usando la formula della varianza (sommatoria al quadrato degli scarti tra ciascuna osservazione e la media, diviso per n osservazioni meno 1). Dopo questo calcolo si può sulle stesse informazioni si può calcolare la stima di precisione; la varianza dello stimatore è pari all'inverso della somma dei pesi e errore standard (deviazione standard della ripetizione del campionamento di uno stimatore statistico; funziona come la deviazione standard del singolo dato) è radice della varianza diviso dimensione campionaria.

Calcolo dell'IC (stimatore metanalitico ± 1.96 ES). Si usa distribuzione standardizzata perché errori casuali, distribuiti con gaussiana. Il test statistico è il test z come stimatore metanalitico diviso ES.

Esempio: metanalisi di sei studi fino alla costruzione del diamante metanalitico. In orizzontale sul diamante abbiamo IC.

ES in questo caso non sta per errore standard ma per effect size, stimatore metanalitico. Peso è inverso della varianza. ES per il peso. Stimatore complessivo è sommatoria di risultato individuale per peso diviso per qualcosa. Varianza complessiva ed errore standard complessivo.

Questo vale per gli studi con variabili continue; spesso per variabili sono delle frequenze, dei RR o degli OR. Studi caso-controllo sono studi in cui stimatore è fornito in base a OR.

Studi strutturati in modo diverso, con diverso peso. Ogni studio è sintetizzato in una tabella 2x2 (non solo caso-controllo). Riconducibile a delle proporzioni e tutti soggetti seguiti per il tempo predefinito; no coorte in cui soggetti che entrano ed escono, una coorte chiusa va bene con soggetti seguiti per tutto il tempo necessario.

Ogni studio ha un prodotto che corrisponde all'OR (quanto + alta probabilità di esito favorevole con trattamento che favorisce).

Se diversi studi sintesi allo stesso modo. Questo presuppone che numeri di soggetti di cui si ha bisogno. Facile che in uno studio ci siano dati presentati diversamente (es indicato solo OR con IC). Da qualche parte il num complessivo di soggetti c'è. Scomposizione di ogni studio e calcolo OR metanalitico, con procedura scritta da due studiosi (ORMH): c'è un peso (inverso della varianza) x stimatore specifico diviso sommatoria pesi.

Quale è il peso? Non la varianza ma la stima della varianza nelle tab 2x2 è num complessivo tabella diviso prodotto della diagonale (soggetti che vanno contro ipotesi di lavoro che è che i trattati vadano bene e i non trattati vadano male). Trasformazione in peso facendo 1 diviso varianza.

Ricalcolo dell'OR metanalitico con contributo di ogni studio.

Quando non tutte queste informazioni soprattutto varianza facile che studio abbia IC ed effect size; girando la formula si risale all'ES e alla numerosità.

Valore stimatore puntuale $\pm 1.96 \times ES$; ES dipende da DS e da radice della dimensione campionaria.

Cambio incognita

Quali sono i problemi: legati alla eterogeneità tra gli studi, che spesso hanno margini di differenza che sono abbastanza importanti, da identificare e decidere cosa fare. Non modifica dei risultati ma decidere quali strategie seguire nel mettere insieme i diversi studi in unica analisi.

Qualcosa che misura eterogeneità: non a occhio, anche se si vede. La procedura è calcolo indicatore (q di Cochrane) che è indice di omogeneità ma in realtà tanto più aumenta, tanto più è grande eterogeneità.

Q_c : si calcola lo scarto di ogni studio che si scosta dal risultato della metanalisi; qualche studio più vicino e qualcuno più lontano. Serve quando per analisi della varianza si calcola la differenza tra gruppi.

Scostamento elevato alla 2 x peso dello studio; costruzione dell'indicatore che senza eterogeneità va vs 0 (se studi allineati gli scarti sono tutti 0); crescendo eterogeneità valore degli scarti aumenta progressivamente, caratteristica della maggior parte degli studi (in assenza di differenza dall'ipotesi nulla il risultato del test statistico deve dare risultato piccolo e viceversa). Con aumento valore si ha discostamento della ipotesi nulla che diventa meno probabile.

Il risultato complessivo letto come distribuzione chi quadrato, di probabilità (valutare scostamento tra osservazioni e un valore atteso). Valore complessivo e chi^2 ha gradi di libertà che dipende dal num di studi.

Regola spannometrica: atteso del chi^2 è gradi di libertà in assenza di scostamento. Valore probabilità di osservare queste differenze tra studi se realtà di distribuzione perfettamente omogenea. Tabelle di trasformazione chi^2 .

Nel caso di analisi con dati di frequenza, con calcolo OR, si procede allo stesso modo, ma ricorda che OR non ha distribuzione simmetrica, ma va da 0 a infinito, no valori negativi e valore nulla in assenza di associazione è 1. Utilizzare distribuzione normale con i log dell'OR, come nel calcolo dell'IC dell'OR. Calcolo log OR metanalitico, OR di ogni studio, scarto tra i due OR trasformati, al 2 x peso e si arriva a un valore analogo a quello a cui si arriverebbe con studio con dati num quantitativi.

Esempio: 7 studi, di cui 5 simili e vicini alla ipotesi nulla e uno abbastanza lontano (OR di circa 5) e uno molto lontano (OR circa 11).

In questi casi si trasforma valutazione visiva in test di eterogeneità. Se rifiuto ipotesi nulla che i test sono omogenei, devo vedere chi è responsabile della eterogeneità (quale studio e quale motivo). Il Q_c scende, valore probabilità di risultato come questo quando vera ipotesi nulla 99% e insieme degli altri omogeneo, quando tolto quello diverso.

Problema OR_{mh}: come possibile 0,5 in caso-controllo? Non calcolabile il risultato se c'è 0 nella cella o se uno degli OR è 0 o infinito. Normalmente si sostituisce 0 con valore molto piccolo (es 0,5).

Quello che troviamo di variabilità quanto è attribuibile al caso e quanto alla eterogeneità tra gli studi esclusi; anche quando tolti alcune differenze ci sono. Se eterogeneità residua capire cosa togliere, rifare analisi etc.

Da qualche aa indicatore I^2 (I quadro) che funziona come R^2 della revisione multipla (se piccolo ok, altrimenti introduzione altre variabili); I^2 dice quale parte eterogeneità non spiegata come variabilità casuale. Ottenuta sulla base del Q e del num di gradi di libertà. Espressione in %.

Motivi per cui abbiamo variabilità tra studi e che possiamo controllare: probabilità differenze tra studi per effetto del caso. Errore casuale, errori sistematici (cercare), può esserci confondimento, bias.

Esempio: studio di valutazione del b-carotene come prevenzione del tum polmone tra fumatori, con risultati disomogenei. Distribuzione sorgenti di disomogeneità. Studi osservazionali di coorte mostravano beneficio mentre trial (randomizzazione soggetti trattati e non trattati) non mostravano miglioramento ma anzi aumento del rischio che per ogni studio non significativo ma mettendoli insieme risultato ss. Metodologia indagine e problema degli studi osservazionali hanno portato a diversi risultati; anche metodologia raccolta informazioni. Dalla storia dietetica ad es si ha situazione con minore evidenza di associazione, se invece ricostruzione nelle 24 h precedenti risultato protettivo che raggiunge la ss, se usato altro indicatore

(frequente assunzione specifici cibi indice) valutazione ancora più distante dal valore nullo, quindi tre metodiche non confrontabili e unica metanalisi no servizio buono.

Altra situazione è quella che in caso-controllo è critico atteggiamento dell'intervistatore, che se diverso (suggerimento delle risposte, disinteressato etc) crea premesse per associazione e differenza tra gli studi con blinding dell'intervistatore e studi in cui intervistatore sapeva che intervistava, se caso o controllo.

Analisi congiunta porta a risultato anche ss ma non giusto.

Problemi ricostruzione retrospettiva della esposizione.

Studi di coorte su volontari seguiti per molti aa, con info all'inizio dello studio.

Disegno dello studio è critico nel dare risultato.

Esempio: valutazione colme scelta di diversi indicatori degli eventi e di stimatori statistici complessivi porta a risultati apparentemente diversi.

In generale si fa approfondimento e si cerca di suddividere. Quando eterogeneità riconducibile a due o tre gruppi di studi è meglio separarli e analizzarli separatamente. Altrimenti metodiche statistiche più complicate.

Tutto questo anticipato prima delle analisi; raccoglimento dati secondo protocollo che prevede cosa fare se eccesso etero. E se c'è etero capire quali cause

Modello a effetti fissi: unico estimatore

Modello a effetti casuali: no unica misura ma ogni studio corrisponde a misura diversa dagli altri; non solo variabilità casuale ma componente di variabilità casuale, differenza tra studi. Calcolo complicato. Modello scompone differenza tra studi in differenza parte dovuta all'errore casuale e parte dovuta a differenza intrinseca tra studi. Tutto questo pesa poco di solito su effetto complessivo ma grande aumento varianza e influenza su ES e precisione IC e valore test di significatività. Aumento IC e il valore probabilità corrispondente diminuisce. No questo modello se no bisogno.

Analisi della sopravvivenza (lezione 4 – 04.05.2018)

Dal punto di vista tecnico è modalità di analisi da una coorte di soggetti in cui si valuta p che si verifichi evento considerando sviluppo nel tempo di osservazione. Studio riguarda spesso mortalità dei soggetti ma in realtà qualsiasi vento da valutare la p di comparsa nel tempo è trattabile.

Non nate in medicina ma nate in applicazione di statistica in problemi ingegneristici. Applicazione alla medicina con 2 metodologie.

Valutare come compare nel tempo e come cambia p di avere un evento. Con una condizione: evento deve comparire una volta soltanto; nel caso della morte no problema, nel caso della diagnosi o della guarigione possibile che fase di miglioramento con rappresentazione della malattia e in questo caso separare le due analisi, della prima e della seconda remissione.

Obiettivo ragionare sul tempo che intercorre tra esposizione (a nuovo trattamento, l diagnosi di una pato, esposizione ad agente con effetti avversi etc) e il prodursi dell'evento.

Sull'evento ci sono definizioni diverse da cui arrivano analisi specifiche (mortalità specifica per cause, sopravvivenza priva di malattia etc). Ragioniamo sulla sopravvivenza complessiva; spesso analisi su eventi che intercorrono prima della mortalità. Capitolo concettuale degli end point surrogati.

Una definizione dei criteri in onco per classificazione della risposta e per definire end point da misurare in analisi di sopravvivenza in associazione per la ricerca; criteri di classificazione (recist).

Risultato si una AS: prodotto di una analisi secondo metodo di cm; confronto sopravvivenza in diverse coorti per periodo di diagnosi in bambini affetti da leucemie infantili. Come si presenta e come si possono valutare i risultati.

Due assi: p cumulativa di sopravvivenza e tempo, che parte da momento comune (es giorno della diagnosi). Asse verticale dà p cumulativa di sopravvivenza, cioè ci dice combinata con il tempo quale è per un soggetto in queste condizioni la p di essere ancora vivo a certo tempo dalla diagnosi.

A che tempo si raggiunge valore predefinito della p ma si può anche dire qual è la p a un tempo specifico.

Le date si riferiscono al gruppo di pz che hanno avuto la D in determinati aa.

Scalini, modificazione della sopravvivenza; quando accade evento il valore della sopravvivenza cambia e si crea uno scalino nella curva.

Si può fare per ogni gruppo di soggetti o per soggetti divisi in singole categorie (es sesso, età alla D, GB alla D etc).

Si può calcolare test statistico che dice qual è p di avere solo x effetto del caso in gr che sono estrazioni casuali, di avere curve come quelle osservate. P non relativa a specifico istante quindi analisi utile perché consente valutazione delle differenze e quando curve diverse che si avvicinano è test che neutralizza le differenze.

Effetto del caso.

Caratteristiche dei dati: caratteristiche che ci impediscono di usare ad es la media, che sarebbe indice di sopravvivenza. Media non applicabile perché risultati non distribuiti con curva gaussiana. **Dati di sopravvivenza.**

Conosciamo il **tempo di sopravvivenza (Ts)** ma anche in queste curve sviluppate per molti aa dalla D c'è proporzione molto alta e nella condizione peggiore oltre 50% che non hanno raggiunto evento misurato; oltre 50% soggetti per cui tempo di sopravvivenza è sconosciuto, soprattutto per le forme meno aggressive di leucemia.

Il problema per calcolare una media è che dato metto per soggetto non ancora morto come tempo di sopravvivenza; almeno il tempo in cui posso seguire il pz ma sicuramente sarà maggiore quindi non applicabile se non nel caso limitato in cui tutti soggetti morti per evento di interesse.

Dati che non sono distribuibili in modo gaussiano per due motivi: abbiamo il Ts noto e abbiamo anche il fatto che sottostimiamo il Ts per le persone seguite per meno tempo.

La frequenza dei tempi di osservazione è quella che abbiamo; per la maggior parte dei soggetti no stima Ts.

Osservazioni troncate: tempo di osservazione per quelle persone che non hanno sviluppato l'evento

Modalità di analisi e diverse funzioni (noi vediamo funzione cumulativa di densità).

Metodo di Kaplan-Meier; ce ne sono anche altri. Questo metodo no se tantissime persone.

Calcolo della p di morire e della p di sopravvivenza

Esempio: grafico con coorte su asse verticale e mesi su orizzontale. Evento può coincidere con fine dello studio (evento non si verifica e soggetto vivo alla fine dello studio) oppure soggetto può abbandonare prima della fine dello studio oppure evento si verifica. Calcolo del rischio in relazione al tempo, con approccio tipo studio di coorte (rischio cumulativo, che non aggiunge niente; si contano gli eventi, le persone studiate e la p è evento su persone) oppure calcolare tassi di incidenza per cui per ogni intervallo si calcola freq dell'evento, ma questo corrisponde a metodi già visti; oppure disegnare la curva di sopravvivenza.

Il primo step è passare da tempo di calendario definendolo come inizio dello studio e allinearli a tempo 0, definito su una scala uguale per tutti es tempo dalla D, dall'inizio della esposizione etc. Questo ordinamento

casuale dei soggetti, per cui alcuni hanno molti mesi di osservazione e altri pochi e tutti sono vivi alla fine della osservazione e altri soggetti hanno evento o scompaiono.

Allineamento e ordinamento poi analisi.

Operazione del calcolatore e la seconda si può fare a mano o con excel. Definire p di avere un evento per ciascuna delle unità di tempo in cui suddivisione dell'asse temporale. Alla prima tacca (1 mese) quanti eventi verificati? In questo caso nessun evento e p di avere un evento al primo mese è 0 eventi su tot soggetti osservati (15). **Probabilità di morire: num eventi su soggetti osservati.**

Al 4 mese prima diminuzione delle dimensioni del gruppo di studio; fino al 4 mese sono 15 e al 4 mese diventano 14. Al 5 mese si verifica il 1 evento ($p = 1$ evento su 14 soggetti). Calcolo della densità di sopravvivenza a ciascun tempo, come un calcolo della p di sopravvivenza la tempo 1, 2, 3, 4 e per ciascuno calcolo quanti eventi e come cambiano le dimensioni del gruppo.

Per la densità di sopravvivenza a ciascun tempo KM non vede intervallo del mese ma vede tipo corsa a ostacoli in cui a ogni mese c'è un ostacolo; la p di evento è num eventi diviso il num di soggetti che si sono presentati davanti all'ostacolo.

Stima p di morte istantaneo: num decessi diviso num soggetti a rischio. P evento considerato alo scadere di un tempo è 1 evento diviso per 14 persone sotto osservazione (5 mese).

P complementari per ricavare **probabilità di sopravvivenza**, complemento 1 della p di morire. **1- p di morire** dà la p di sopravvivenza. Vale per ciascun tempo. Quando combinare sviluppo di tempi diversi trattiamo p come p indipendenti tra loro quindi p condizionate. La p di sopravvivere a 5 mese esseno arrivato al 5 mese e così via.

La p cumulativa o complessiva è prodotto delle p condizionate calcolate in questo modo. La p cumulativa di sopravvivenza è la produttoria delle p condizionate che danno la p di sopravvivenza a ciascun tempo.

In KM il tempo è fissato come scadenza unica ma ci sono analisi guidate dalla distribuzione dei dati; se non succede nulla e poi in tre sett 3 eventi e per 1 mese 0 eventi, KM ignora intervalli in cui non succede nulla e si concentra sulla data esatta dell'evento. In KM conviene utilizzare dettaglio più fine possibile per valutare lo sviluppo del tempo. Se tempo, stima della sopravvivenza in giorni.

Si riporta in una tabella che dice al tempo 1 quanti soggetti totali, eventi e quanti soggetti che vengono persi (che escono dalla osservazione senza evento); calcolo p di superare questo tempo. Supponiamo che tempo minimo è il mese. Al mese 1 avevamo 15 soggetti ma 0 eventi con p di morire 0 e p di sopravvivere 1-0 (1). Al tempo 5 la p di morire è $1/14$ mentre p di sopravvivenza $1-1/14$. Al tempo 6 il gruppo è diminuito di un soggetto che è morto e ulteriore perdita di un soggetto ma p non diminuisce perché num eventi è 0. Al tempo 7 diminuzione soggetti e così via fino all'evento successivo.

La p cumulativa di sopravvivenza cambia al momento in cui si verifica un evento, non prima; è corretto che in KM rappresentazione grafica preveda degli scalini perché p di sopravvivenza costante fino alla verifica dell'evento. Bacchette che dicono dove uscita di un soggetto dallo studio, che non determinano cambiamento della p di sopravvivenza.

Fine quando completato il tempo previsto di analisi dei dati o morte dell'ultimo soggetto o uscito ultimo soggetto.

Funzione cumulativa di sopravvivenza: valore fraz cumulativa sopravv al tempo immediatamente precedente x funzione sopr di densità sopravv al tempo analizzato.

Confronto di due gruppi, trattati ad es modo diverso. Gruppo con trattamento standard, con soggetti ordinati; secondo gruppo con trattamento sperimentale, non succede nulla in questo gruppo fino al tempo 6 e poi ci sono eventi. I due gruppi all'inizio avevano uguale dimensione ma al tempo 6 nel gruppo standard ci sono 14 persone mentre in quello sperimentale 21. Cambiamento della numerosità relativa nel tempo tra gli studi.

Calcoli: p istantanea e p cumulativa per entrambi i gruppi. Se mettiamo su un grafico: valutazione che non è sufficiente per dare risultato. Come calcolare test considerando che num soggetti a ciascun intervallo cambia nei due gruppi in modo diverso con diverse proporzioni

Logrank test: Distribuzione num eventi osservati tra due gruppi a ciascun tempo di osservazione secondo ipotesi nulla (trattamenti sono uguali); è solo il caso che decide se evento nei due gruppi. Calcolo num veti attesi se fosse vera la equivalenza dei due gruppi. Simile al χ^2 . Decido se distribuzione osservata simile a quella attesa o diversa.

Ipotesi nulla è che tra i gruppi no differenza quindi num eventi distribuito nei due gruppi a seconda della consistenza num dei gruppi.

Immaginiamo 1 gr con 50 persone e secondo gr con 25 persone. A tempo x ci sono 4 eventi. Come distribuirli? Atteso maggiore per il primo gruppo (formule). Ripetizione a tutti i tempi a cui compare un evento; è la parte lunga. Non sono tantissimi i programmi statistici di base che fanno analisi di sopravvivenza. Allo stesso momento costruire una tabella che dà funzione cumulativa sopravvivenza nel primo e nel secondo gruppo arrivando alle analisi e mescolando i due gruppi si costruisce tabella degli eventi osservati, ripartiti nei due gruppi con calcolo degli eventi attesi.

Analisi dei dati visti nelle tabelle e nel grafico. Avremo tot eventi osservati nel gruppo 1 e 2.

Test di Logrank è la somma dei quadrati degli scarti tra tot osservati e tot attesi diviso in num degli attesi del $gr\ 1 + 2$. Come leggere il risultato? Prendo tabella con distribuzione χ^2 e il num gradi di libertà è il num di gr a confronto meno 1. Si può fare fino a 3-4 gruppi. Dal punto di vista del confronto tra dati epidemio bisogna confrontare num maggiore di gr. Ripartizione dei gruppi in base alla consistenza num. Risultato si legge nella tabella del χ^2 con grado di libertà. Valore ad es 15 simile a 1×1000 di prob che non si trova neanche nella tabella.

Differenza tra le curve spiegabile con semplice effetto del caso con $p < 1 \times 1000$, che è molto $< 5\%$ e rifiuto in tranquillità la ipotesi nulla dicendo che i trattamenti sono diversi.

(La funzione di rischio è la p di avere evento istantanea. La funzione di densità di p tiene conto di intervallo di tempo la funzione di rischio è istantanea e dice p di morire immediatamente dopo istante. Funzione di rischio diversa perché può avere degli alti e bassi (ad es aumento dopo intervento chir).

Se disegniamo funzione di rischio abbiamo andamenti a onde, con diminuzione ed aumenti. La scala non parte da valore predefinito, come per le altre funzioni. Quando andamento a onde è incompatibile con funzione cumulativa di sopravvivenza.)

No equivalente di IC, che dovrebbe dare estimatore comune che dice quale differenza tra le curve; non può essere calcolato perché da un momento all'altro cambiano le differenze e no procedura per il calcolo del rapporto tra rischio di morte in un gr e nell'altro. Per calcolare questo rapporto analisi multivariate con metodo di Cox.

Modello di cox è multivariato per tenere conto di diverse variabili che influiscono sulla funzione di sopravvivenza. Otteniamo HR, un rapporto tra due rischi di morte; il rischio di morte nel gr trattato e di riferimento. Interpretazione come OR o RR. Se no differenza vale 1, con escursione vs infinito asimmetricamente.

Risultati di analisi con metodo che porta alla stima HR si trovano in diverse forme. Stessi soggetti visti per analisi iniziali su LL considerati insieme diversi possibili fattori; relazione con GB, relazione con immunofenotipo, anche mettere insieme in una analisi per correggere risultato di una analisi per eventuale effetto confondimento di un'altra variabile.

Sono analisi che danno confronto all'interno dei gruppi, relativamente al conteggio dei GB aumento passando tra categorie determina aumento del rischio di un tot; rischio di morte.

Anche calcolare IC e vedere come crescendo GB in categorie si ha aumento progressivo del rischio. Due IC sono tali per cui entrambi ss e differenze sia rispetto a baseline che all'altra categoria. Informazioni ricche e sintetiche rispetto alla semplice lettura del grafico e aggiustamento reciproco per le due variabili.

Problema: Il modello di cox può stimare correttamente la relazione tra le due curve se queste sono sempre "parallele", c'è sempre proporzionalità tra le distanze tra i valori delle due curve, su tutto lo sviluppo del tempo di osservazione. È il problema di questo modello, che dà un solo risultato di OR per tutta la osservazione; questo perché abbia un senso deve essere corrispondente a una costanza nella posizione relativa della funzione di sopravvivenza nei due gruppi. Se questo non vale si dà risultato non chiaramente definibile.

Applicando il calcolatore dà sempre un risultato ma in curve diverse cosa dà? Una media, un punto a un tempo? Non si sa. Quindi il risultato non è applicabile e interpretabile. Pesare quanto pesa lo spostamento. Verifica della adeguatezza, del fatto che HR sia sempre applicabile su tutto lo sviluppo temporale considerato.

Percorso di analisi di sopravvivenza parte da analisi della curva di KM; guardare e vedere se le curve si intersecano e se ciò avviene non si fa cox. Se non si intersecano o se si intersecano in fasi in cui molto vicine, si pensa di procedere con cox. Calcolo del log della funzione della densità di sopravvivenza e tabularlo rispetto al log del tempo di sopravvivenza. Lo fa il calcolatore. Vedere meglio se le curve sono parallele. Altra modalità per vedere se curve parallele tabulare log neg del log della sopravvivenza cumulativa che ha rispetto alla analisi della sopravvivenza cumulativa aiuta a leggere se le curve si intersecano o meno. Sempre diagnostica preliminare come questa.

Nel caso della regressione logistica o lineare è più semplice. Nel caso di cox le stesse cose vanno fatte con metodiche nel programma di calcolo di cox. Il programma dà modo di fare diagnostica necessaria.

Importante la parte grafica, ancora più del valore assoluto. Pacchetti statistici almeno medi, altrimenti difficile comprensione analisi di base. Quando passare da analisi semplice a multivariata attenzione alle complessità aggiuntive, che si guardano sul grafici di sopravvivenza.

Articolo Brentuximab (lezione 5 – 16.05.2018)

Trial multicentrico, importante dal punto di vista dei costi e delle pubblicazioni.

Obiettivi: confrontare il protocollo standard ABVD e nuovo protocollo con farmaco sperimentale Brentuximab A-AVD.

St di fase 1 partito per aggiungere al protocollo standard il brentuximab ma ci si è resi conto che tox polmonare già presente della bleomicina esaltata dalla interazione quindi in corso di fase 1 modificazione delle dosi e verifica presenza di interazione con modifica del protocollo tp che era stato pensato inizialmente.

Fase 2 è stata sostituita da osservazione prolungata dei pz (num grande) della fase 1. Obiettivo dire se passaggio alla fase 3.

Dettaglio ancora maggiore nel protocollo vero dello st (cioè la sintesi di tutto quello che può essere il percorso previsto e gli imprevisti e le reazioni a questi).

Protocollo preceduto da una sintesi che descrive il percorso in termini sintetici.

Trial clinico randomizzato controllato, non in cieco. Perché non in cieco? Difficile condurre uno st come questo con farm così pericolosi in cieco (la procedura in cieco, in caso di EC, sarebbe contattare il centro che risale a quale farm fanno i pz).

Dal punto di vista tecnico no tanta differenza tra i gr, che hanno comunque 4 farm. Tecnicamente avrebbe potuto essere in cieco dal punto di vista tecnico, dicendo al pz di stare un po' di più (uno dei due protocolli è somministrato più lentamente).

Pz: da un lato situazione brutta di tum (st 3-4), ma allo stesso tempo buon PS (che viene definito da diversi indici), condizione per quanto riguarda il quadro emato, funzionalità epatica e renale sufficiente.

Nell'ambito della patologia sono pz in condizioni abbastanza buone.

Criteri di inadeguatezza: neuropatia, gravidanza (in genere mai inclusione negli st), altro tumore (sia in atto che in APR), condizioni CV. Da ricondurre probabilmente agli effetti specifici dei farm.

Percorso: fase di screening per valutare criteri di inclusione ed esclusione, randomizzazione (sistema centralizzato probabilmente, con contatto telematico con il centro), trattamenti, verifica dello stato di malattia, che sulla base di valutazione su risultato della PET, porta alla continuazione del trattamento o alternativa tra continuazione trattamento o passaggio a tp alternativa (a scelta del pz). Fasi di valutazione e follow up.

Tempistica precisa per ogni atto nello studio (raccolta consenso informato, misurazione del peso a intervalli prestabiliti, ripetizione test di gravidanza etc). Percorso estremamente guidante, la partecipazione è percorso estremamente guidato.

Confronto con disegni più elastici; in questo modo riduzione variabilità su una serie di elementi e riduzione possibilità che ricercatore possa influire in qualche modo (es rallentando delle valutazioni) ma anche riduzione errore di fondo che ci può essere.

Outcome: entità che interessa misurare. Dall'outcome discende end point misurabile.

Modalità per misurazione outcome.

End point principale: modified progression free survival. È la progressione di malattia o la morte. Hanno aggiunto del fatto di tenere conto di situazioni in cui no esattamente progressione ma mancato cambiamento delle condizioni (modified).

I sotto stanti end point sono quindi la progressione di malattia, morte per qualsiasi causa e misura di una risposta non completa (punteggio alla PET > o uguale a 3, seguito da aggiunta CHT successiva).

Analisi principale è di sopravvivenza quindi riferimento a un certo tempo, in particolare il terzo end point (riferito alla data in cui è stato effettuato l'esame PET).

Condizione raggiungimento uno degli end point previsti e costruzione analisi di sopravvivenza

La sopravvivenza complessiva è end point secondario

Perché evento più importante è una delle componenti e non end point principale? In caso come questo scelta più sensibile dal punto di vista statistico, che prima riconosce differenze tra trattamenti, mettendo insieme evento più grave e meno freq con altri eventi neg per quanto riguarda la prognosi della pato.

Comitato di guida, di revisione e gr indipendente di revisione dei dati e di revisione della sicurezza del farmaco. Investimento da parte di questo aspetto impegnativo.

Braccio della sicurezza: verificare dati su EC e effettuare analisi ad interim. La analisi ad interim è analisi di futilità (quello che stimo facendo non serve e potrebbe essere peggiorativo). No valutazione statistica ma deciso a priori quando condurla e quali erano i criteri con cui applicare decisione. Condizione a un terzo del reclutamento e scelta se farm nuovo è inferiore (indicazione sicurezza e efficacia) allora lo si considera non adeguato e si interrompe lo st.

Protocollo prima di condurre lo st. Valutazione complessiva id sicurezza ed efficacia alla fine dello st.

Protocollo: se tasso di remissione completa (primo end point) e proporzione sogg, per trattamento sperimentale è almeno 5% < tp standard e se trend nella progression free survival e altri end point con efficacia inferiore; se questi due eventi: fine dello st.

Criteri di inclusione ed esclusione. Alla fine pz inclusi sono nella tab 1.

Sogg separati per braccio di tp, poi totale. Essenziale confronto due bracci tp e vedere se randomizzazione ha portato a due gr simili tra loro.

Nelle analisi supplementari qualche dettaglio in più. In queste c'è il risultato del flusso secondo procedura concordata tra principali riviste per avere modalità comuni di rappresentazione dei risultati dei trial.

La randomizzazione è stata effettivamente 1 a 1.

Descrizione dei pz rispetto alle diverse fasi della tp e agli eventi che hanno reso necessario cambiamento della tp, per evento avverso, punteggio insufficiente alla PET o per altri motivi. 14 pz e 9 pz nei due gr. No indicazione che sia successo qualcosa di inaspettato o tendenza a far cambiare il trattamento (i pz che hanno cambiato tp sono pochi e simili come num nei due gr); lo st alla luce di questo è stato condotto bene (altrimenti se grande differenza tra gr ad es per importanti EC inaspettati).

Risultati: dopo conclusione del trattamento.

Analisi dei dati. Possibile che qualche sogg non passi al set di analisi dei dati (es errore nella raccolta dati, come errore nel segnare data di evento; deve essere spiegato perché esclusione).

Risultati veri e propri riguardano analisi di sopravvivenza, che sono i grafici a pag 336. Analisi più incisiva che chiarisce meglio il risultato. Valutazione del comitato indipendente di revisione e valutazione dell'investigatore (il clinico che coordinava lo st).

Il risultato è differenza tra un trattamento e l'altro ed è maggiore nel grafico dell'investigatore. È un problema costante perché si investono moltissimi soldi, quindi concetto di interesse (in uno st come questo si investono 10mila a pz solo come partecipante, poi c'è tutto il resto). Ente regolatore chiede che uno st come questo sia valutato non solo dai clinici.

Si vede inversione delle curve ma è irrilevante, dovuto alla mortalità tardiva a 4-5 aa di 2-3 pz.

Test statistico (logrank), HR misurato con applicazione del modello multivariato di cox e riepilogo delle componenti della modified progression survival (e individuazione pz corrispondenti).

Potenza statistica st sopravvivenza no num sogg ma num eventi che si osservano; se fosse stato calcolato su mortalità, serviva uno st 4-5 volte più grande id questo per osservare con stessa ss il risultato, perché num eventi complessivi è un sesto del totale in entrambi i gr.

Scelta di usare criterio di valutazione composito invece della sola mortalità è perfettamente idoneo.

Tendenza di riduzione spazio nella rivista principale e lasciare spazio gestito come si vuole per tratti supplementari.

Se noi avessimo usato sugli stessi dati solo analisi di sopravvivenza (overall survival, riferita a evento morte) avremmo risultato che può dire che nuovo trattamento dà suggerimento vantaggio ma non ss, quindi risultato nullo.

Figura 2. Forrest plot. Usato come strumento grafico per far veder graficamente e numericamente i risultati, che qui sono HR (quante volte è più grande o piccolo il rischio di morte o di evento composito in un gr rispetto all'altro). HR 0,77 (inferiore quindi, num che si trovava anche nel grafico).

Il risultato può mostrare effetto interazione con altre caratteristiche? Effetto uguale per tutti i pz o caratteristiche che influenzano il risultato?

Sopra i 60 aa no vantaggio, che è tutto sotto i 60 aa. Se usiamo altri cut off il risultato non cambia, i più giovani hanno quindi un vantaggio. Si usano altri cut off di età.

Informazione principale funziona meglio tra i giovani.

St sono condotti in aree diverse geograficamente, quindi vantaggio essendo in USA (maggior capacità di seguire attentamente il protocollo) rispetto ad altri paesi. C'è qualcosa di diverso, potrebbe essere anche effetto di confondimento (es se pz EU e asiatici fossero tutti vecchi e in USA tutti giovani possibile confondimento per età).

Differenza effetto relativamente ad altre caratteristiche (stadio, IPS, sintomatologia, localizzazione extranodali, sesso). Farm molto efficaci sulle localizzazioni extranodali.

Informazioni che servono a decidere se pz, una volta che trattamento ha superato al selezione, conviene che sia trattato in un modo o nell'altro.

Qualunque sia il PS il farm nuovo è uguale a quello vecchio.

Queste analisi sono fatte con modello di regressione che non tiene conto della stratificazione; bisogna veder quali variabili di aggiustamento considerate nel modello.

Eventi avversi. Nuovi farm vanno meglio o peggio?

Distribuzione degli effetti avversi; quasi tutti i pz hanno avuto almeno 1 effetto avverso. Se guardiamo quelli severi sono stati visti i 83% pz sperimentali e 66% dei pz standard; stessa differenza anche in diversa classificazione degli eventi avversi.

Fra eventi avversi gravi tra doppio e aumentato di un terzo. Eventi avversi osservati tutti durante il trattamento; dopo cessazione del trattamento freq è la stessa.

Neutropenia, neuropatia periferica (aspetti sensoriali e motori) e altri eventi comunque abbastanza pesanti. Tutti leggermente più freq nel nuovo trattamento.