# TIPOLOGIA di ANALISI sul DB

**Preparazione per esecuzione Query su DB**

* L’esecuzione di Query finalizzate alla caratterizzazione dell’intervento terapeutico mediante trattamento antibiotico su affezioni che colpiscono specifici apparati, richiede una preliminare elaborazione di alcuni dei campi interessati:
* Classificazione dei Patogeni Apparato Specifici: creare la lista degli agenti microbici responsabili di infezioni mirate a specifici organi in infezioni degli apparati (es. Infezione da Chlamydia🡪 Uretrite in infezioni dell’Apparato Urogenitale). = Suddivisione Codici ICD-9 per Agenti Infettivi
* Classificazioni delle Patologie Apparato Specifiche: creare la lista delle patologie presenti nel DB che affliggano i diversi apparati sottoposti ad investigazione richiede un approfondimento relativo alla nomenclatura/definizione di tutte le patologie presenti nel DB (l’associazione tra la patologia/definizione “Salpingite” di origine batterica è associabile nel 90% dei casi circa ad infezioni delle Tube di Falloppio (App Urogenitale), e nel restante 10% alle Trombe di Eustacchio(App Masticatorio-Uditivo)=Suddivisione Codici ICD-9 per Patologie

**Esecuzione Query su DB**

* Query relative alla Distribuzione delle terapie antibiotiche per fascia di Età e Genere (Uso dati di anagrafica del paziente incrociati con terapia)
* Query relative alla epidemiologia locale delle patologie (con risoluzione a livello regionale, provinciale o di regione postale (CAP))
* Query relative alle abitudini prescrizionali locali delle patologie (con risoluzione a livello regionale provinciale o di regione postale (CAP))
* Query inerenti alla Frequenza delle Infezioni Apparato Specifiche (Uso delle Classi ICD-9)
* Query relative alla Frequenza annuale di Recidive Patologia-Specifiche (Uso delle Classi ICD-9)
* Query concernenti la Frequenza dei Pazienti con difficoltà di eradicazione o evento di recidivazione dell’infezione, sottoposti quindi a terapia antibiotica sequenziale:pz con una prescrizione, con 2 prescrizioni consecutive, con 3 prescrizioni consecutive entro l’anno🡪 L’inervallo di tempo trascorso tra una e l’altra permette una valutazione della tipologia di difficoltà terapeutica in cui si incorre (Uso delle Classi ATC - Uso delle Classi ICD-9)🡪 Possibili ipotesi di Batterioresistenza
* Query per il confronto delle abitudini prescrizionali per tipologia infetiva
* Query relative l’appropriatezza prescrizionale antibiotica (misuso) mediante identificazione delle patologie solitamente non trattate con antibiotico nella pratica clinica
* Query relative all’eccesso di prescrizioni antibiotiche caso specifiche (abuso) mediante l’identificazione di pazienti e relative patologie per le quali siano state contravvenute le eventuali raccomandazioni o linee guida terapeutiche.

Nel corso delle analisi e mediante opportune valutazioni sarà possibile ampliare la tipologia di analisi per ottenere nuovi risultati dal DB

# VALUTAZIONE QUALITATIVA DEI DATI CONTENUTI NEL DB

## DATI di ANAGRAFICA del PAZIENTE

Esempio da Cartella “pazienti”: Possibili combinazioni di dati mancanti (?) o errori immissione dato (grassetto rosso) tra i campi di maggiore interese per le specifiche del paziente

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **codice (ID Paziente)** | **data\_open** | **userid (ID Medico)** | **nascita** | **sesso** |
| 0=%,4JJFQ#/%7H:%.M@}+O | 08/04/1998 | SGQYWS8 | 15/01/1908 | M |
| 2#|G=?E^F->(N$U;}S1{CA | 20/11/2008 | WQESQXE | 16/03/1933 | F |
| 3C>1&20G3{WY7LYS?1\_1K/ | 31/07/2001 | ? | 13/12/1965 | F |
| 0^F67/,7LW}.S<6NWM\_I)V | 19/01/2001 | ? | 09/04/1962 | ? |
| 1LP\*W\*8038(17KÂ¶K,NA53# | 28/11/2000 | ? | ? | F |
| 0=%,4JJFQ#/%7H:%.M@}^4 | 08/04/1998 | SGRYWS4 | ? | M |
| 1[4<{#9Â¶?]{.59MH\_P|2?Q | 24/11/2008 | WQESQXE | ? | ? |
| 1)S{:Y0GG){.Q3GV7{I;=6 | 23/02/2011 | SQUEWPC | **12/03/1900** | M |
| 10>>.-WD\*-YÂ¶7JP!4:F2&E | 22/07/1999 | NQU5T4AA | **21/04/1900** | ? |
| 1I[6HO\_\*/B/KDP&W|#F!)V | 18/12/2006 | LQESQQE3 | 27/01/1938 | **T** |

Le frequenze dei dati presenti relative alle caratteristiche del paziente:

|  |  |  |
| --- | --- | --- |
|  | Frequenza % | Freq Assoluta |
| **codice (ID Paziente)** | 100 | 1.200.312 |
| **data\_open** | 100 | 1.200.312 |
| **userid (ID Medico)\*** | 89.9 | 1.079.384 |
| **nascita** | 99.8% | 1.198.660 |
| **sesso** | 99.7% | 1.196.915 |

\* Identificativo del medico per ogni paziente -> 120928 pazienti senza ID medico

Il **Numero di Medici** coinvolti nella compilazione del DB sembrerebbe aggirarsi intorno a 1460 medici. Da notare è però la presenza di 260 medici con un numero di pazienti < 10, e un altrettanto poco probabile, o poco significativo dal punto di vista delle investigazioni, numero di 364 medici con un numero di pazienti < 20.

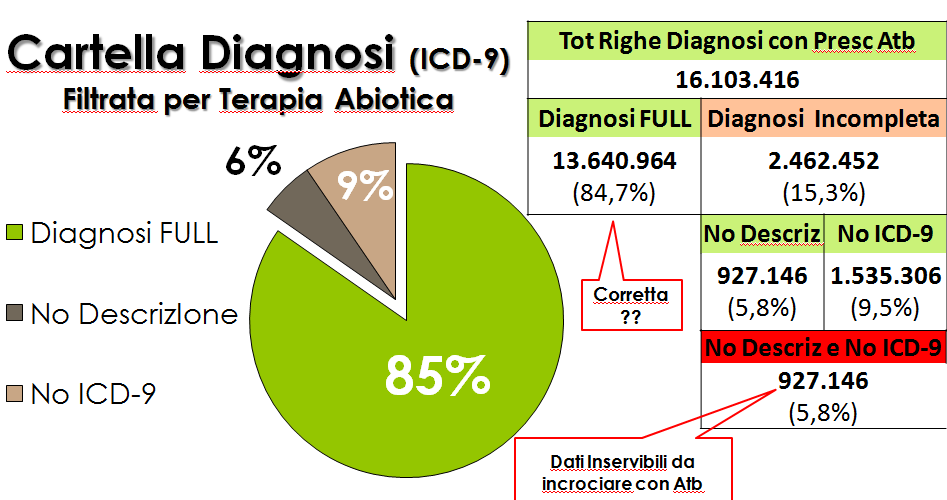
## QUALITA’ DATI DIAGNOSI al PAZIENTE

La qualità del dato diminuisce notevolmente quando si va a valutare i dati relativi alla Diagnosi effetuata ad ogni visità.

La diagnosi è composta da 2 campi:

1. ICD-9 : codice univoco assegnato alla patologia ( è organizzato gerarchicamente con un numero identificativo principale es 320 (meningite batterica) relativo ad una macroclasse di patologie, e una serie di sottoclasi per identificare patologie più specifiche come 320.0 (Meningite da Haemophilus), 320.1 (Meningite pneumococcica)....etc.
2. Descrizione Testuale: diagnosi descrita dal medico, comprendente terminologie speso non presenti nel codici ICD-9

Di seguito si riporta la valutazione della qualità dei dati relativi esclusivamente alle diagnosi alle quali è seguita come terapia la prescrizione di antibiotici:



La presenza del campo nell’85% dei casi non significa un altrettanta accurata descrizione della patologia, per la quale è stato spesso utilizzato un generico codice ICD-9 come il 799.9 o una incompleta descrizione testuale della patologia.

## CONSIDERAZIONI sulla QUALITA’ del DATO

## -Perdita progressiva delle informazioni-

La mancanza di singoli campi nel corso delle analisi, in particolare nel processo di Join delle tabelle (unione delle tabelle), per il quale è necessario che tutti i campi di riferimento siano presenti, porterà ad un aumento cumulativo della *perdita di informazioni* in quanto la mancanza del dato, ad esempio la data di nascita o il genere, comporterà l’eliminazione dell’intero paziente al quale manca il dato, nel caso in cui si vogliano effettuare valutazioni sulla distribuzione delle patologie infettive nelle fasce di età distinte per genere.

Il join è infatti un’operazione di combinazione del contenuto di due tabelle che consente di generare una tabella con le righe e le colonne di ciascuna delle due tabelle. Le tabelle vengono unite mediante dati condivisi delle due tabelle (condizione di join). Ad esempio nel join coinvolgente la tripla codice1, sesso1, nascita1, se essa è integralmente presente nella prima tabella ma non nella seconda, la corrispondente riga della prima tabella non farà parte della tabella risultato dell’operazione di join Quindi la tabella risultante avrà le colonne di entrambe le tabelle iniziali solo se esse avranno questi campi con un contenuto uguale.