

Modern Approaches to Data Collections in Neuroepidemiology

ABSTRACT

Background

The use of administrative databases has been implemented in epidemiological fields.

Record linkage systems match data across administrative health databases supplying large samples for long-term observational studies, the preferred non-experimental approach to test hypotheses.

The present project shows the application of modern methodological approaches to the study of neurological diseases, specifically: I. to investigate the association between exposure to breastfeeding and the occurrence of MS in adulthood using prospectively collected community-based data; II. to assess the effect of antibiotic exposure on the risk of developing MS in the Emilia-Romagna region (RER), Italy; III. to assess the accuracy of death certificates (DC) in the identification of subjects with amyotrophic lateral sclerosis in the Limousin region, France.

Materials and Methods

Study I, a community-based cohort study, linking prospectively collected information on breastfeeding from the Cohort of Norway community-based surveys on health status with the Norwegian MS Registry for MS clinical information, and the population-based Medical Birth Registry of Norway. Associations between breastfeeding and MS onset were estimated.

II: A population based nested case-control study in the Emilia Romagna region. All patients with MS diagnosis seen at MS-centers (2015-2017) aged ≥ 18 years were eligible. For each patient demographic and clinical information were collected and 5 controls, selected among RER residents, were matched on age, sex, place of residence and index year. Information on antibiotic prescription was obtained through the linkage with the RER drug prescription database (ATC code "J01").

III: The French register of ALS cases in the Limousin region (FRALim) was used as gold standard for the population-based validation study. DC information of patients in the FRALim register and deceased within 2000-2011, including ICD10 codes for specific cause of death, were provided. Sensitivity, positive predictive value (PPV) were calculated.

Results

I: No association was found between having been breastfed at least 4 months and MS risk, adjusting for various maternal factors (HR = 0.90; 95%CI 0.68-1.19). The estimates did not change for different durations of breastfeeding or adjusting for perinatal factors.

II: During the study period 877 subjects with MS and 4,205 controls were included. Exposure to any antibiotic in the three years before index was associated with MS (OR= 1.52, 1.29–1.79). An exposure in the 8 and 13 years before showed increased effects size (OR=1.95, 1.44-2.63 and OR=3.04, 1.07-8.68, respectively). Similar results were found for different classes of antibiotics or time lag exposures. No dose-response effect was found.

III: DCs were available for 197 ALS patients, of whom 185 (93.9%) were correctly identified by the ALS ICD10 code. The overall sensitivity was 93.9% (95% CI 89.6-96.8)

and the PPV was 64.9 (59.1-70.4). Stratification for sex, age and year at death did not show difference in accuracy.

Discussion and Conclusion

The studies presented showed most of the values and issues in applying epidemiological methods to large collections of data. The different risks of bias in our studies were acknowledged and, whenever possible, overcome.

The most frequently reported were information bias, with misclassification of exposure or outcome, and the lack of accuracy of measurement methods. Selection bias was avoided using population-based samples. The lack of covariates in most situations can only be reduced using appropriate surrogate variables.

The application of neuroepidemiology to healthcare databases should therefore be an effort to maximize the signal over the noise, since the large amount of data may amplify the risk of bias. The use of this information need to be considered carefully and its validity should be assessed to correctly interpret the findings.

Approcci moderni alle raccolte di dati in Neuroepidemiologia

ABSTRACT

Background

L'utilizzo di dati amministrativi è stato implementato in campo epidemiologico. Infatti i sistemi di record linkage permettono di collegare diversi database amministrativi fornendo campioni per studi osservazionali a lungo termine, il migliore approccio non-sperimentale nella verifica di ipotesi.

In questo progetto applichiamo moderni approcci metodologici allo studio di malattie neurologiche; nello specifico per: I. indagare l'associazione tra allattamento al seno e insorgenza di sclerosi multipla (SM) in età adulta, usando dati prospettici community-based; II. valutare l'effetto dell'uso di antibiotici sul rischio di SM nella regione Emilia-Romagna (RER); III. determinare l'accuratezza dei certificati di morte (DC) nella diagnosi di sclerosi laterale amiotrofica (ALS) nella regione Limousin, Francia.

Materiali e Metodi

I: Studio di coorte community-based che incrocia informazioni su allattamento al seno dai questionari sulla salute "Cohort of Norway" con il "Norwegian MS Registry" per le informazioni cliniche sulla SM, e con il "Medical Birth Registry of Norway". È stata valutata l'associazione tra allattamento e SM.

II: Studio nested caso-controllo di popolazione (RER). Tutti i pazienti con diagnosi di SM, visti nei centri-SM (2015-2017) e ≥ 18 anni, erano elegibili. Per ciascun paziente sono state raccolte informazioni demografiche e cliniche e selezionati 5 controlli, appaiati per età, sesso, residenza e anno index, tra i residenti RER. L'informazione sull'uso di

antibiotici è stata ottenuta dal database delle prescrizioni di farmaci della RER (codice ATC “J01”).

III: Studio di validazione di popolazione che utilizza il “French register of ALS cases in the Limousin region” (FRALim) come gold standard. Le informazioni del DC su pazienti segnalati nel registro FRALim e deceduti nel periodo 2000-2011 sono stati forniti, inclusi codici ICD10 per le cause di morte specifica. Sensibilità e valore predittivo positivo (PPV) sono stati calcolati.

Risultati

I: Non è stata trovata associazione tra l'essere stati allattati per almeno 4 mesi e rischio di SM, aggiustando per diversi fattori materni (HR=0.90; 95%CI 0.68-1.19). Simili valori sono stati stimati anche per diverse durate di allattamento o aggiustando per fattori perinatali.

II: Sono stati inclusi 877 soggetti con SM e 4,205 controlli. L'esposizione ad un antibiotico nei tre anni precedenti l'anno indice è associata a SM (OR= 1.52, 1.29–1.79). L'esposizione negli 8 e 13 anni precedenti l'indice mostrava valori aumentati (OR=1.95, 1.44-2.63 e OR=3.04, 1.07-8.68, rispettivamente). Risultati simili sono stati trovati per diverse classi di antibiotici. Non si è evidenziato un effetto dose-risposta.

III: I DCs di 197 pazienti con ALS sono stati considerati, di cui 185 (93.9%) correttamente identificati con il codice ALS su ICD10. La sensibilità complessiva era 93.9% (95% CI 89.6-96.8) e il PPV 64.9 (59.1-70.4). La stratificazione per sesso, età e anno di morte non ha mostrato differenze nell'accuratezza.

Discussione e Conclusioni

Gli studi presentati mostrano molte delle caratteristiche e dei limiti dell'applicare metodi epidemiologici a grandi raccolte dati. Nei nostri studi il rischio di bias è stato riconosciuto e, quando possibile, ridotto.

I limiti più di frequente riportati sono i bias di informazione, con misclassificazione di esposizione o outcome, e la mancata accuratezza nei metodi di misurazione. I bias di selezione sono stati evitati utilizzando campioni di popolazione, mentre la mancanza di covariate in molte situazioni può essere solo contenuta, utilizzando appropriate variabili surrogate.

L'applicazione della neuroepidemiologia a database sanitari dovrebbe costituire uno sforzo per massimizzare il segnale sul rumore di fondo, dal momento che il grande volume di dati potrebbe amplificare il rischio di bias. L'utilizzo di queste informazioni e la loro validità devono quindi essere attentamente ponderati per interpretare correttamente i risultati.

Approches modernes pour la collection des données en Neuroépidémiologie

Résumé

Contexte

L'utilisation de bases de données administratives a été mise en œuvre dans l'épidémiologie.

Les systèmes informatisés correspondent aux données administratives sur la santé, fournissant de grands échantillons de patients pour des études observationnelles à long terme, c'est l'approche non expérimentale privilégiée pour tester les hypothèses.

Le projet actuel montre l'application d'approches méthodologiques modernes pour l'étude des maladies neurologiques, en particulier: I. Étudier l'association entre l'exposition à l'allaitement maternel et la survenue d'une Sclérose en Plaque (SEP) à l'âge adulte en utilisant des données communautaires collectées de manière prospective; II. Évaluer l'effet d'une exposition aux antibiotiques sur le risque de développer une SEP dans la région Émilie-Romagne (RER), Italie; III. Évaluer la précision des certificats de décès dans l'identification des sujets atteints de la Sclérose Latérale Amyotrophique dans la région Limousin, France.

Matériel et Méthodes

Pour l'étude I, une cohorte a regroupé les informations sur l'allaitement maternel collectées de manière prospective lors des enquêtes communautaires de la Cohorte de Norvège, le registre norvégien de la SEP pour les informations cliniques, puis le registre médical des naissances de la Norvège. Les associations entre l'allaitement maternel et l'apparition de la SEP ont été estimées.

L'étude II était une étude cas-témoins niché en population générale dans la RER. Tous les patients diagnostiqués d'une SEP dans les centres-SEP (2015-2017), ≥ 18 ans étaient éligibles.

Pour chaque patient, les informations démographiques et cliniques ont été collectées puis 5

contrôles, sélectionnés parmi les résidents de la RER, ont été appariés en fonction de l'âge, du sexe, du lieu de résidence et de l'année d'indexation. Des informations concernant la prescription d'antibiotiques ont été obtenues grâce au lien avec la base de données des prescriptions de médicaments RER (code ATC « J01 »).

Pour l'étude III, le registre français de SLA dans la région Limousin (FRALim) a été utilisé comme gold standard pour l'étude de validation en population générale. Les informations de décès des patients du registre FRALim et décédés entre 2000 et 2011, y compris les codes ICD10 pour une cause spécifique de décès, ont été collectées. La sensibilité et la valeur prédictive positive (VPP) ont été calculées.

Résultats

Etude I: L'association entre le fait d'avoir été allaité au sein pendant minimum 4 mois et le risque de SEP n'a pas été significative, après ajustement sur divers facteurs maternels (HR=0.90; IC95% 0.68-1.19). Après divers ajustements sur différents facteurs périnataux et durées d'allaitement, l'association reste non significative.

Etude II: Au cours de la période d'étude 877 sujets atteints de SEP et 4,205 témoins ont été inclus. L'exposition à un antibiotique au cours des trois années précédant l'indexation était positivement associée à la survenue d'une SEP (OR=1.52, IC95%1.29-1.79). Une exposition au cours des 8 et 13 années précédentes a montré une augmentation de la taille de l'effet (OR=1.95; 1.44-2.63 et OR=3.04; 1.07-8.68 respectivement). Des résultats similaires ont été trouvés pour différentes classes d'antibiotiques ou durée de période d'exposition. Aucun effet dose-réponse n'a été trouvé.

Etude III: Les données de décès étaient disponibles pour 197 patients SLA, dont 185 (93.9%) étaient correctement identifiés par le code SLA ICD10. La sensibilité globale était de 93.9%

(IC95% 89.6-96.8) et la VPP était de 64,9 (IC95% 59.1-70.4). Une stratification sur le sexe, l'âge et l'année de décès n'a pas montré de différence de précision.

Discussion et conclusion

Ces études ont montré des bénéfices mais également des difficultés lors de l'application de méthodes épidémiologiques pour une large collection de données. Les différents risques de biais dans nos études ont été reconnus et surmontés dans la mesure du possible.

Les biais d'information étaient les plus fréquemment signalés, avec une erreur de classification de l'exposition ou de l'outcome et un manque de précision des méthodes de mesure. Le biais de sélection a été évité en utilisant des échantillons en population générale. Le manque de covariables dans la plupart des situations ne peut être réduit qu'en utilisant des variables de substitution appropriées.

L'application de la neuroépidémiologie avec les bases de données de santé devrait donc être un effort pour maximiser le signal sur le bruit, car une grande quantité de données peut augmenter le risque de biais. L'utilisation de ces informations doit être consciencieusement étudiée et leurs validités doivent être évaluées de manière à aboutir à une interprétation correcte des résultats.