O19 - Comunicación Oral/Oral communication

Métodos II

Methods II

Viernes 3 de Octubre / Friday 3, October
9:00:00 a/to 11:00:00

Moderador/Chairperson:
Santiago Pérez Hoyos

244

REFUSERS IN A DISABILITY PREVENTION TRIAL IN OLDER ADULTS ARE LESS HEALTHY: BASELINE AND FOLLOW-UP ANALYSIS

Minder Christoph E.*, Stuck Andreas E.**, Gillmann Gerhard***
**Dept of Social and Preventive Medicine, University of Berne, Bern, Switzerland. **Geriatric Research Unit, Spital Bern Ziegler, Bern, Switzerland. ***Dept of Social and Preventive Medicine, University of Berne, Bern, Switzerland.*

Objectives: This study explored differences in health and survival between refusers and participants in a longitudinal study with extensive baseline and follow-up information.
Methods: Results of a trial comparing 791 participants and 401 community-residing older adults who refused to participate in a study concerning preventive home visits were examined. Information was collected from interviews, insurance records, and government files.
Results: Despite similarities in terms of age, sex, and self-perceived health at baseline, refusers had a 1.53-fold higher risk of dying over 5 years than participants (95% CI: 1.20 - 1.94). Risk of refusers compared to participants for entering a nursing home over a 3-year follow-up was 1.58 (95% CI: 1.05 - 2.36). Risks of dying and of nursing home admission varied between refusers and participants when refusers according to self-reported reason for refusal (too ill, too healthy, no interest, and other reasons).
Conclusions: Future studies should include follow-up data to allow comparisons between refusers and participants and should address the presence of multiple subgroups of refusers with varying health prospects.

245

PUBLISHED DRUG EFFECTIVENESS STUDIES USING SECONDARY DATA SOURCES

Judith Strobl, Tom Walley
Department of Pharmacology & Therapeutics, University of Liverpool, Liverpool, UK.

Background: Increasingly, clinical patient-level data are collected routinely in patient registries or disease databases in primary or secondary care. These data might offer a ready means to assess drug effectiveness in a naturalistic setting. Potential hurdles for effectiveness studies using secondary data sources include data access and data protection issues, the quality and applicability of the available data, potential biases, and the need for complex analysis methods.
Objectives: To identify and review published studies assessing the effectiveness of a specific drug treatment by using exclusively existing data from patient registries, databases, or electronic patient record systems.
Methods: We searched Medline [1966-2002] restricted to journals included in the Index Medicus. We included studies evaluating a particular drug using patient-level data collected routinely as part of an electronic patient record system, a registry or database, and not for the exclusive purpose of addressing the objectives of the reported study or another single research hypothesis. Studies had to involve a comparison group.
Results: Forty-two published "database studies" met our inclusion criteria. Eighteen studies investigated treatments or preventative interventions in chronic conditions, and a further four studies evaluated HIV/AIDS treatment. Twenty-five studies defined their primary outcomes through clinical information (including hospitalisation episodes), five studies defined primary outcomes based on drug treatment data. Only few studies took dose and continuity of the investigational treatment into account. Thirty-eight studies (90%) were retrospective cohort studies. Ten studies are known to have been approved by a research ethics committee. Published reports rarely clarify whether fully anonymised data has been used, but in most studies this is unlikely to be possible. Only one study explicitly reported seeking patient consent. Seventeen studies (40%) have at least been partly funded by industry sources. Seven studies report having undertaken data validation against other data sources; four further studies have used data sources previously validated.
Conclusion: Drug effectiveness evaluations using secondary data seem to be particularly common in chronic disease areas, with some studies reporting follow-up periods well exceeding those usual in randomised controlled trials. Published reports vary considerably in their attention to data description, sampling, data quality, and data protection and confidentiality issues. Further work should focus on (1) recommendations for best (reporting) practice, (2) the rationale for the choice of study design and data source for the evaluation of a particular treatment, and (3) the role of "database studies" in health technology assessment.

246

DE-IDENTIFICATION OF EPIDEMIOLOGIC DATA: PROTECTING THE CONFIDENTIALITY OF STUDY PARTICIPANTS IN RELEASED DATA SETS THROUGH STATISTICAL DISCLOSURE RISK ASSESSMENT AND LIMITATION METHODS

Daniel Barth-Jones
Center for Healthcare Effectiveness Research, Wayne State University School of Medicine, Detroit, Michigan, USA.

Introduction: Growing concerns about the privacy of medical data have led to regulatory responses such as the European Union's Privacy Directive or the United States' HIPAA Privacy Standards. It is essential that epidemiologists conducting research with medical data sets be aware of these regulations and the impact they have on the conduct of epidemiologic research. Furthermore, as a practical solution to such privacy concerns, epidemiologists must develop competency in the areas of statistical disclosure risk assessment, and disclosure limitation methods that can de-identify epidemiologic data. The demographic and geographic characteristics of the individuals in epidemiologic data sets are important determinants of unintended disclosure risks for confidential medical information by record linkage methods.
Methods: Statistical disclosure risk measures were calculated for the entire United States population using publicly available data from the 2000 census obtained from the U.S. Census Bureau. Three potentially identifying characteristics commonly found in epidemiologic data sets were evaluated for associated disclosure risks: 1) Date of Birth/Age categorization, 2) Gender, and 3) Zip code or geographic classification. The extent of disclosure limitation achieved by global recoding for combinations of variables was analyzed. Changes in the information utility of the data caused by the disclosure limitation methods were expressed in terms of bias and precision for samples drawn from this population.
Results: Disclosure risks are demonstrated to be functions of underlying population density and the cross-classification structure of the demographic variables. By recoding the levels of detail retained for these common demographic variables, the epidemiologic and statistical utility of these characteristics can be retained, while simultaneously limiting the disclosure risk potential quite considerably. Analyses demonstrate that such disclosure control methods result in tolerable losses in precision without incurring significant bias.
Conclusions: Relatively simple disclosure limitation methods can substantially limit disclosure potential in epidemiologic data sets, and can assure that the risk of re-identification for individuals is very small. Where more refined geographic detail is needed, the trade-off between geographic resolution and demographic detail must be balanced to assure that acceptable disclosure risks are maintained in epidemiologic data sets.

247

META ANALYSIS: DEALING WITH HETEROGENEITY AND DISCRETE DISTRIBUTIONS

Ramon Clèries*, Josepa Ribes*, Víctor Moreno*, José Miguel Martínez**, F. Xavier Bosch*
**Servei d'Epidemiologia i Registre del Càncer, Institut Català d'Oncología, Hospitalet de Llobregat, España. **Unitat de Recerca Laboral, Universitat Pompeu Fabra, Barcelona.*

Introduction: There are several cohort studies among Hepatitis B Virus (HVB) carriers worldwide, which show association between HBV infection and Hepatocellular Carcinoma (HC). The relative risk estimated in each study indicates differences between countries and also, depending on the control group, within countries. Our main issue is to estimate how these studies can be grouped, and then modelling the relative risk in terms of covariates measured in each country. These variables are: alcohol consumption, hepatitis viruses prevalence, size of study, follow up and control group (Blood donors or general population). We have selected 12 studies which comply with the requirements to be included and which we have measured the covariates.
Methods: The statistic used to estimate the relative risk has been the Standardized Mortality Ratio (SMR). The SMR is the ratio between observed and expected number of liver cancer cases in each study. To determine the subgroups or clusters it has been used two methods: classical meta analysis and a maximum likelihood nonparametric estimator (MLNPE, Böhning-1996). Once determined the number of clusters and the SMR estimated for each one of them, we proceed to calculate the probability that a study belongs to each cluster. These estimation are compared with the estimation obtained by a compound distribution: poisson stopped binomial (PSB, Douglas-1980). For modelling the SMR by the risk factors we have used a generalized linear mixed model which accounts for geographical heterogeneity.
Results: The global SMR has been estimated around 18 (IC95%:10.7-32.1), and there's 3 clusters detected by the MLNPE.: Two groups: Taiwan-China and Japan-Europe-USA. Each one of those two groups have also two subgroups determined by the control group.
Conclusions: In the classical meta analysis method it is not possible to determine clusters by means of SMR estimates and its confidence intervals, and only residuals analysis helps to indicate a possibly cluster. On the other hand, clusters can be estimated by the MLNPE, but variance within the cluster it is difficult to compute. To solve this problem we propose first to estimate the possibles clusters by MLNPE and then estimate the SMR for each by means of the PSB.

248

CHALLENGES OF THE HUMAN GENOME: HOW TO IMPROVE SUCCESS RATES FOR GENETIC ASSOCIATION STUDIES

Krina Zondervan, Lon Cardon
Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Introduction: To date, studies aimed at finding associations between genetic variants and multifactorial diseases have been notable by their lack of success and replicability. This is partly attributable to lack of good epidemiological practice such as inconsistent case definition, unsuitable control groups and insufficient sample sizes. However, the structure of the Human Genome introduces an important set of other conditions. Genetic association studies rely on variants under study (marker alleles) close to the causative allele being co-inherited more often than expected, termed linkage disequilibrium (LD). Using examples of known disease-gene associations, we explored how the power of case-control studies and effect sizes are related to the extent of LD between marker and (unobserved) causative alleles, and differences in their population frequencies.
Methods: We took five examples of disease-gene associations, two with rare (~0.05) causative allele frequency (Crohn's disease & *NOD2*; DVT & *Factor V Leiden*); one with low-medium (0.15) frequency (Alzheimer's & *APOE*); and two with high (0.7-0.85) frequency (Bladder cancer & *GSTM1*; NIDDM & *PPAR|&A*). D' was used as a standardized measure of LD between causative and marker alleles (0=no LD; 1=maximum LD). Assuming a candidate-gene approach, power to detect a significant association between markers and disease (using chisquare/Fisher's exact test; ¨ç=0.05) was calculated from samples of 1000 cases & 1000 controls for Crohn's, DVT, and Alzheimer's, and 5000 cases and 5000 controls for Bladder cancer and NIDDM. Effect sizes were calculated as allelic and genotypic odds ratios (ORs).
Results: For Crohn's, DVT, and Alzheimer's, ORs for the causal variant were substantial: 3.3-4.6 (allelic ORs), 3.0-4.0 (heterozygote ORs), and 11.7-40.0 (homozygote ORs). When using markers, allelic ORs were considerably lower but remained detectable with 80% power using differing marker frequencies and D' down to 0.5-0.6. For NIDDM and bladder cancer, effect sizes of the common causative alleles were much lower (allelic ORs: 1.2-1.3; genotypic ORs: 1.0-2.0). Using markers, therefore, no association would be detected for D' values < 0.8. For D' values > 0.8, marker frequencies had to closely resemble disease allele frequency to allow detection with 80% power. Detecting an association with a D' > 0.5 and causative allele/marker frequency up to ~0.3 would require 9000 cases and 9000 controls. In the case of a rare causative allele with low effect size, given the same marker parameters, a sample size in excess of 60,000 cases and controls would be needed.
Conclusions: Association studies using marker variants should take into account: *1)* whether candidate genes are investigated or a genome-wide scan is conducted; *2)* possible population frequency of the disease allele and effect size; *3)* population frequency of the marker alleles; *4)* local pattern of LD. The development of the HapMap aimed at providing a genomic map of LD may provide help in study design.

249

SELECTION BIAS/INTRACTABLE CONFOUNDING IN EPIDEMIOLOGIC STUDIES IN ELDERLY POPULATIONS

Roberto Pastor-Barriuso*, Eliseo Guallar**, Ana Navas-Acién**, Javier Damián*

**Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Madrid, Spain. **Department of Epidemiology and Welch Center, Johns Hopkins Medical Institutions, Baltimore, USA.*

Introduction: Epidemiologic studies of chronic diseases frequently find an age-related decrease in the relative risk associated to most risk factors. A common explanation for this phenomenon is that survival selects subjects who are no longer susceptible to the risk factor of interest, although so far no gene or combination of factors has been identified to explain non-susceptibility among survivors. This study provides an alternative explanation based on the confounding induced over time by all other known or unknown risk factors related to the disease.
Methods: To explore the potential magnitude of this effect, a simulation study of a cohort of 10,000 subjects aged 35 years at enrollment was performed under the following conditions: (a) the dichotomous risk factor of interest was randomly assigned to each subject for a given baseline prevalence, and the relative risk of exposed to non-exposed subjects was assumed to be constant over time; (b) other risk factors were aggregated into a common risk score, randomly generated from a normal distribution and assumed to be independent of the main risk factor at baseline; and (c) each subject's annual hazard of death was assigned from a proportional hazard model according to the above covariate pattern and to an average annual hazard obtained from the US life table in 1999. Several scenarios were considered for different combinations of the baseline prevalence p, the true relative risk RR, and the proportion in the variability of baseline hazards k explained by the main factor of interest.
Results: There was a relatively higher death rate among those exposed subjects with higher risk score, so that with advancing age, exposed survivors had lower risk scores than non-exposed survivors. This selection mechanism induced a negative association between the main risk factor and risk score among survivors, and consequently relative risks of exposed to non-exposed were progressively biased towards the null. For $p = 0.40$, $RR = 2$, and $k = 0.05$, the estimated relative risks at ages 50, 65, and 80 years were 1.80, 1.59, and 1.38, respectively. The decline was even more marked with the true relative risk increased and the variability explain by the main risk factor decreased.
Conclusions: This selection bias provides an explanation for the decline in relative risks with aging that could potentially affect all risk factors in observational studies. Control of this bias may be seriously limited by the lack of information on most determinants of risk. In contrast to the susceptibility postulate, under which it may no longer make sense to intervene on risk factors in aging populations, the proposed explanation suggests that age-related declines may just reflect increasing confounding, and hence it will be still very important to intervene on risk factors in elderly subjects.

250

ALLOWING FOR INDIVIDUALS' RISK AWARENESS WHEN INVESTIGATING RELATIONSHIPS OF BEHAVIOURAL RISK FACTORS TO SUBSEQUENT CANCER AND HEART DISEASE IN COHORT STUDIES

David R. Boniface*, Margaret E. Tefft**

**Dept. of Epidemiology and Public Health, University College London, London, UK. **Dept. of Statistics, University of Hertfordshire, Hatfield, UK.*

Introduction: A number of health behaviours such as diet, exercise and smoking, have been implicated in both heart disease and cancer. When examining the relationships of these factors to subsequent heart disease and cancer mortality it is necessary to exclude those individuals whose behaviours are likely to have been modified due to their awareness of early warnings or risk of the disease. Examples are recommendations of special diet after discovery at screening of high serum cholesterol or after discovery of polyps at sigmoidoscopy screening. Such individuals will report low risk levels of the health behaviour but be at raised risk of developing the disease during the follow-up period. Such individuals, if not excluded from analysis, will obscure the true relationship of the behaviour to the disease where it may exist in the general population. Where the same behaviour is suspected of carrying risks for both cancer and heart disease it may be necessary, when investigating factors related to the development of cancer, to consider excluding individuals reporting heart disease or awareness of raised heart disease risk and vice versa.
Methods: A cohort of 9003 randomly selected men and women in Great Britain was interviewed for the Health and Lifestyle Survey in 1984-85 and monitored subsequently for 17.5 years for deaths. The interview covered health, health-related behaviours, physical measurements, socio-demographic details and a dietary questionnaire. There were 2604 men and 2644 women aged 40-75 with 270 and 248 cancer deaths respectively and with 287 and 184 CHD deaths. Alternative approaches were applied to exclusions before carrying out survival analyses relating behaviours to subsequent death.
Results and Conclusions: Findings from the Health and Lifestyle Survey from survival analysis for the relationship of behavioural risk factors for cancer death and for CHD death were shown to depend critically on decisions made about exclusions for participant awareness of raised risk not only for whichever was the target disease, but also for the other disease where it has the same behavioural risk factors.