

**1673 CHALLENGES IN DATA POOLING**

<sup>1,2</sup>Ann C Olsson\*, <sup>1,2</sup>Saloshni Naidoo. <sup>1</sup>International Agency for Research on Cancer, Lyon, France; <sup>2</sup>Department of Occupational and Environmental Health, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Congella, Durban, South Africa

10.1136/oemed-2018-ICOHabstracts.350

**Aim** Studying the effects of rare exposures in relation to rare diseases is challenging. In this session we will discuss problems and benefits with pooling data in epidemiological studies. We will show different examples of different types of pooling projects as well as results from those, and point to the contrasts with meta-analyses and multi-centre studies.

Eero Pukkala<sup>1</sup>, Deborah Glass<sup>2</sup>, Maria Leon Roux<sup>3</sup>, Hans Kromhout<sup>4</sup>, Ann Olsson<sup>3</sup>, Ingrid S. Mehlum<sup>5</sup>

<sup>1</sup>Finnish Cancer Registry – Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland

<sup>2</sup>Monash University Centre for Occupational and Environmental Health, Department of Epidemiology and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Melbourne, Australia

<sup>3</sup>International Agency for Research on Cancer, Lyon, France

<sup>4</sup>Environmental Epidemiology Division, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

<sup>5</sup>National Institute of Occupational Health (STAMI), Oslo, Norway

**1673a ENTIRE POPULATIONS OF FIVE COUNTRIES AS AN OCCUPATIONAL STUDY COHORT – DOES IT WORK?**

E Pukkala. Finnish Cancer Registry – Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland

10.1136/oemed-2018-ICOHabstracts.351

The Nordic Occupational Cancer (NOCCA) project was the largest and in many aspects, also qualitatively, the most unique research study ever done on occupation and cancer incidence. It consisted of a follow-up study on the entire working populations of Denmark, Island, Finland, Norway and Sweden with three million cancer cases diagnosed 1961–2005. It described risks of 84 cancer types in 54 occupational categories (astra.cancer.fi/NOCCA) and developed a Nordic Job Exposure Matrix (NOCCA-JEM) that converts the individual job histories of all Nordic people to quantitative estimates of exposure to potentially cancer-related factors. Many of the results on dose-response associations between exposures and cancers have been novel findings or have confirmed (or not) findings from earlier smaller studies. The NOCCA network still produces about one new publication each month from the old data. Still, not only the NOCCA researcher team but also institutions such as Nordic Minister Council and occupational health professionals have stressed that it would be important to continue NOCCA research work with updated data. New features in the work life, such as effects of sedentary work, could not be fully studied in the original NOCCA data. There is also an urgent need to follow effects of work safety regulations and operations that were started after the harmful effect of work carcinogens were recognised. For example, restaurant workers who had the highest risk of many cancers are no longer exposed to tobacco smoke in their work environment; this

should be seen as a rapid decrease in their cancer risk. It has been proven that combining five entire national populations as a study cohort about 10 years ago was feasible and produced important results. The researchers of the NOCCA network are eager to continue and update the activity, provided that challenges related to funding and data access issues can be won.

**1673b THREE NESTED CASE CONTROL STUDIES FROM COHORTS OF PETROLEUM INDUSTRY WORKERS**

DC Glass. Monash University Centre for Occupational and Environmental Health, Dept. of Epidemiology and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Melbourne, Australia

10.1136/oemed-2018-ICOHabstracts.352

We pooled data from three nested case control studies from cohorts of Canadian, British and Australian petroleum industry workers and were able to examine relatively rare outcomes such as myelodysplastic syndrome (MDS) and myeloproliferative disease (MPD). The exposure assessment methodology was similar for all three studies. We examined the base estimates attributed to specific jobs before pooling the data and made some adjustments, notably to the background exposures attributed to individuals. In some cases the differences in exposure were explained by different circumstances, e.g. legal requirements. These inter-study comparisons were made blind as to case-status and before the data were pooled and the analyses were carried out. Multivariate analyses were used to identify main predictors of exposure intensity in ppm benzene. Site type (terminal, refinery, upstream) job (e.g. office terminal operator craft worker) and decade were strong predictors of exposure. Study was also a predictor of exposure which was of concern but the three studies covered different industry sectors (and hence jobs) and different decades. We therefore sought site type/jobs and decades which were represented in all three studies. There was little overlap with sufficient numbers to carry out such a comparison, with the exception of terminal workers in four job categories. We showed that job and era were significant predictors of exposure intensity but study was not a major predictor although the interaction terms study X job and study X decade were significant predictors of exposure intensity. Data pooling is more powerful than meta-analyses, allowing analyses by new metrics or new groupings perhaps derived from findings from one study which can be tested in a larger setting. Careful comparisons of data sets before they are pooled are essential to provide reassurance in the quality of the pooled data set and aid interpretation of pooled analyses.

**1673c DATA AVAILABILITY AND DATA POOLING EFFORTS IN THE AGRICOH CONSORTIUM**

ME Leon Roux. International Agency for Research on Cancer, Lyon, France

10.1136/oemed-2018-ICOHabstracts.353

AGRICOH is an international consortium of agricultural cohort studies co-coordinated by the International Agency for Research on Cancer in France, and the National Cancer Institutes in the USA, to encourage and support data sharing to study disease-exposure associations for which data pooling

represents a significant gain in statistical power compared to analysis of individual cohorts. Cohorts in the consortium study a variety of health outcomes, including cancer, respiratory, neurologic, reproductive, allergic disorders, injuries, and overall mortality in association with a wide array of agricultural exposures, including potential health hazards and protective factors of relevance to agricultural populations. AGRICOH, then, offers a rich framework to conduct pooling projects on determinants of health and disease in farmers, farm workers and their families. As of May 2017, 29 cohorts from 5 continents are participating in AGRICOH. The studies, varying in size to address different objectives, are from Australia (2), Canada (2), Chile (1), Costa Rica (1), Denmark (1), France (6), New Zealand (2), Norway (3), Republic of Korea (1), South Africa (2), the UK (1), and the USA (7) (<http://agricoh.iarc.fr/>). Conceived independently, these cohorts adopted varying definitions of the populations and farm types they study, the approaches to select cohort members from the target populations and the extent of documenting farming related activities and exposures. In the first pooling project in AGRICOH, on application of pesticides and risk of myeloid and lymphoid malignancies, harmonisation of pesticide exposure and cancer data has made it possible to generate combined estimates of cancer-type and individual pesticide associations based on over 3 00 000 farmers and farm workers enrolled in three cohorts from France, Norway and the USA. This project illustrates challenges encountered and solutions adopted to address the study of a ubiquitous exposure suspected of being associated with excess cancer risk in this setting.

**1673d ASSESSING EXPOSURE TO OCCUPATIONAL CHEMICALS IN LARGE-SCALE EPIDEMIOLOGICAL STUDIES ON OCCUPATIONAL CANCERS**

H Kromhout. *Environmental Epidemiology Division, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands*

10.1136/oemed-2018-ICOHabstracts.354

When pooling data across community-based case-control studies exposure assessment has to be redone in order to overcome the issue of the lowest common denominator. Also when exposure assessment is based on expert judgement, no straightforward approaches exist for calibration and pooling of these exposure estimates. Often though, complete and detailed occupational histories of study participants will be available which can be used for a standardised approach across centres/studies. For instance in the SYNERGY project we successfully managed to collect actual exposure measurements across Europa and Canada covering almost 4 decades (1970–2010). Based on this wealth of exposure data a JEM could be elaborated with quantitative estimates of level of exposure by job, year and region. Combining this JEM with occupational histories of cases and controls resulted in quantitative exposure histories which allowed derivation of quantitative exposure response relationships for amongst others silica and asbestos. In industry-based cohort studies exposure assessment can often be performed at a more detailed level by ascertaining detailed occupational histories and collecting production characteristics in multiple companies enrolled in a cohort study. By collecting

industry-specific measurements with detailed auxiliary information very detailed exposure models can be derived. Consequently these models will allow for quantitative exposure estimates at the detailed level of exposure scenario (rather than at the level of a job). For instance, in the European Asphalt Workers study we were able to estimate quantitatively workers exposure to bitumen fume, organic vapour, and benzo(a)pyrene. Standardisation of exposure assessment tools, approaches and empirical modelling are needed in this day and age where big data will be the norm and will be needed to discern undetected health risks. However, availability of actual measurements of workers' exposure will stay a prerequisite in order to calibrate and validate exposure assessment methods employed in large-scale epidemiological studies on occupational health risks.

**1673e POOLED ANALYSIS OF CASE-CONTROL STUDIES ON THE JOINT EFFECTS OF OCCUPATIONAL CARCINOGENS IN THE DEVELOPMENT OF LUNG CANCER**

AC Olsson. *International Agency for Research on Cancer, Lyon, France*

10.1136/oemed-2018-ICOHabstracts.355

Lung cancer is the most common cancer globally, and tobacco smoking is well established as the main cause. Yet occupational exposures play an important role among exposed workers, especially jointly with smoking. The SYNERGY project was established in 2007 to estimate joint effects of asbestos, respirable crystalline silica, polycyclic aromatic hydrocarbons, chromium/nickel and smoking in the development of lung cancer. Sixteen case-control studies conducted between 1985 and 2010 from Canada, Europe, New Zealand and China were pooled, including 19 370 lung cancer cases and 23 674 controls with detailed information on tobacco habits and lifetime occupations. Controls were recruited from the general population (81%) or hospitals (19%), and were individually or frequency matched to cases by sex and age. Information was predominantly collected by interviews with the study participants themselves, though next-of-kin respondents were accepted in five of the studies if subjects were unavailable (9.1% of cases, 6.6% of controls). The database comprises around 14% never smokers, whereof 822 cases. Women represent around 20% of the study population. A quantitative job exposure matrix (SYN-JEM) was created based on exposure measurements from multiple countries together with auxiliary data, covering a time period of more than 50 years. SYN-JEM is based on statistical models that predict job-, time-, and region-specific exposure levels. Cumulative exposures (e.g. ff/ml-years) were calculated for each subject by linking SYN-JEM with individual occupational histories. Unconditional logistic regression models were used to estimate odds ratios (OR), 95% confidence intervals (CI), and trends. The strengths of SYNERGY includes bringing together epidemiologists and exposure assessment experts from around the world to advance occupational cancer epidemiology, 2) power to study small risks, 3) providing quantitative exposure estimates for population-based case-control studies, and 4) allowing subgroup analyses, e.g. by gender, histology and smoking status.