INSTITUTE OF HEALTH INFORMATICS

Graduate Programme in Health Data Science

Assessed Coursework

Student candidate number:	*****
Module:	CHME0014: Principles of Epidemiology Applied to Electronic Health Records Research
Date due:	Friday, 23rd August 2019, 12:00 midday
Word count: (excluding references, diagrams and appendices)	Word count: 2291
Disability or other medical condition for which UCL has granted special examination arrangements:	
My learning development:	On this assignment, I have been particularly focusing on....
	n/a
	In addition to general feedback, please give me feedback on .....
	General feedback is welcome.

Research Question:

Reducing antibiotic prescribing for acute respiratory infections is important for preventing antibiotic resistance. There is a concern however that reduced antibiotic prescribing may lead to more hospital admissions among those who are particularly at risk of complications. Design a study using Electronic Health Records to identify which groups of patients have the highest risk of hospitalisation following an acute respiratory infection to help GPs target antibiotic use.

Background

It has been reported that between 2005 and 2014 that primary care use of antibiotic prescriptions for patients with respiratory tract infections (RTIs) have decreased for male and female patients, [Gulliford M. 2016]. This reduction in acute RTIs (ARTIs) related antibiotic prescriptions comes at a time when there is an increase in antibiotic resistant infections (*e.g.* Methicillim-resistant Staphylococcus aureus (MRSA)), [Aslam B et al. 2018]. ARTIs have a number of potential causes, such as viral (*e.g.* influenza) [Bosch A A T M et al. 2013], bacterial (*i.e.* pneumococcal) [Khan S et al. 2015] or parasitic infections (*e.g.* interstitial pneumonia) [Shah P et al. 2016] and irritants (*e.g.* dust) [WHO indoor. 2007]. Common ARTI symptoms are shortness of breath, coughing, fatigue, increased temperature, nausea, respiratory tract inflammation and mucus discharge, [NHS Choices. 2018]. Treatment for ARTIs tends to focus on symptomatic relief using anti-inflammatories, anti-pyretics, decongestants, expectorants and in the case of bacterial infections, antibiotics, [NHS Choices. 2018].

ARTIs are divided into acute upper RTIs affecting the nasal passages, sinuses, pharynx and larynx and acute lower RTIs affecting the. trachea, bronchi and lungs, [HSE. 2011]. ARTIs are common across the UK population and are often associated with seasonal effects, age, along with other risk factors, [Ge X et al. 2018]. Typical risk factors for ARTI-related complications include gender [Falagas et al. 2007], age [Bovin G et al. 2002], asthma [Ramette A et al. 2018], COPD [Sethi S. 2010], bronchitis [Fisk W J et al. 2010], immunocompromised states [Schuetz P et al. 2017], influenza like illnesses [GBD 2017 Influenza Collaborators. 2018], and pneumococcal vaccination status [Luca D L et al. 2018].

ARTIs are generally managed initially in primary care and are recoded using Read codes at the GP practice [NHS datadictionary, 2019], while persistent illness or complications may lead to hospital referrals for secondary care. In the UK, hospital care is recorded using electronic health record (EHR) systems known as Hospital Episode Statistics (HES), [NHS HES, 2019]. CPRD combines Read codes and HES in to a linked data source, known as the HES-linked CPRD primary care data, [CPRD. 2019]. In addition, hospitals provide feedback to the primary care provider, via patient discharge summaries that can be used to identify hospitalisation (subject to the details being inputted to the GP practice records using Read codes), [NHS edischarge. 2019]. HES-linked CPRD primary care data will be used to explore the potential association between patient risk factors and hospital admissions in this study, in order to inform GP antibiotic prescribing procedures.

Aims and Objectives

A retrospective (historical) cohort study using electronic health records and statistical techniques such as univariate and multivariate analysis will be used to determine the risk factors leading to hospitalisation of patients with ARTIs, in order to inform GPs when defensively prescribing antibiotics. Using CPRD primary care EHR linked to Hospital Episode Statistics (HES) data, the sampling frame will comprise patients with relevant Read codes for conditions such as influenza like illnesses, acute upper respiratory tract infections (AURTIs) and acute lower respiratory tract infections (ALRTIs) (*e.g.* pneumonia with influenza (H2…00)) [Thorax BMJ. 2012]. This study will refer to these conditions collectively as ARTIs.

Study Design

A retrospective cohort study design will use existing HES-linked CPRD primary care data on patients who presented to the GPs with an ARTI. HES data will be used to ascertain patient hospitalisation within one month of ARTI diagnosis by searching with the relevant ICD11 codes. Read codes are a list of standardised and hierarchically structured clinical codes used by NHS GPs to input and group clinical observations, procedures, diagnosis, treatments and outcomes of a patient in either primary and secondary care, [NHS Digital. 2019]. A Read code search utility can be used to search for read codes by inputted relevant search terms, [TailorMade. 2019]. For HES purposes cross mapping tables contained within Read codes are used to generate ICD11 codes, [NHS datadictionary, 2019].

International Statistical Classification of Diseases (ICD) codes are maintained by the World Health Organization (WHO) and are used to provide standardized epidemiological and clinical diagnostic codes for disease classification for HES purposes, (i.e. (using ICD10 codes), Viral pneumonia, J12.0 - J12.9), [Wikipedia. 2019], [ICD10. 2014]. ICD codes contain clinical details, such as 'signs, symptoms, abnormal findings, complaints, social circumstances and external causalities', [Wikipedia. 2019]. ICD11 codes were released on 18 June 2018 by the WHO, [WHO. 2019]. Relevant ethical approval will be sought, [CPRD isac. 2019].
The study will use data from the preceding one year split in to four equal time periods for seasonal comparison purposes. Each covariate will be checked for association with the hospitalisation outcome. The data frame will be checked for duplicates, missing data and outliers. Variables with significant association with the outcome will be used in multivariate analysis. Using either odds ratios or risk ratios, the likelihood of patient hospitalisation based on multivariate risk factors of significance will be explored. In this study, logistic regression will be used with the results presented as odds ratios.

Inclusions and Exclusions
Each ARTI patient will be included in the sample population calculations only once in the one-year study period. Patients will be excluded if they died within the 30 days period after exposure but before hospitalisation or if their ARTI resolved within the 30 days after exposure. Participants hospitalised after the 30-day period will be excluded from the study. Patients will be excluded if they presented themselves to hospital directly without first presenting to their GP. Missing data values will be imputed using the STATA ICE software package for multiple imputations, [Parker C et al. 2010].

Study Population
Descriptive analysis will investigate the patients from England and Wales most at risk of hospitalisation and will involve stratification by age-group (in 5 years groups), gender, BMI, practice location (for differences between urban and rural participants), for SES quintile by postcode (for differences in presumed prosperity based on postcode and associated access to adequate heating and ventilation), ethnicity, seasonality and by clinical risk factors (as defined by literature). The sample population will be as large as possible to improve the statistical power of the study, however this will depend on the data available.

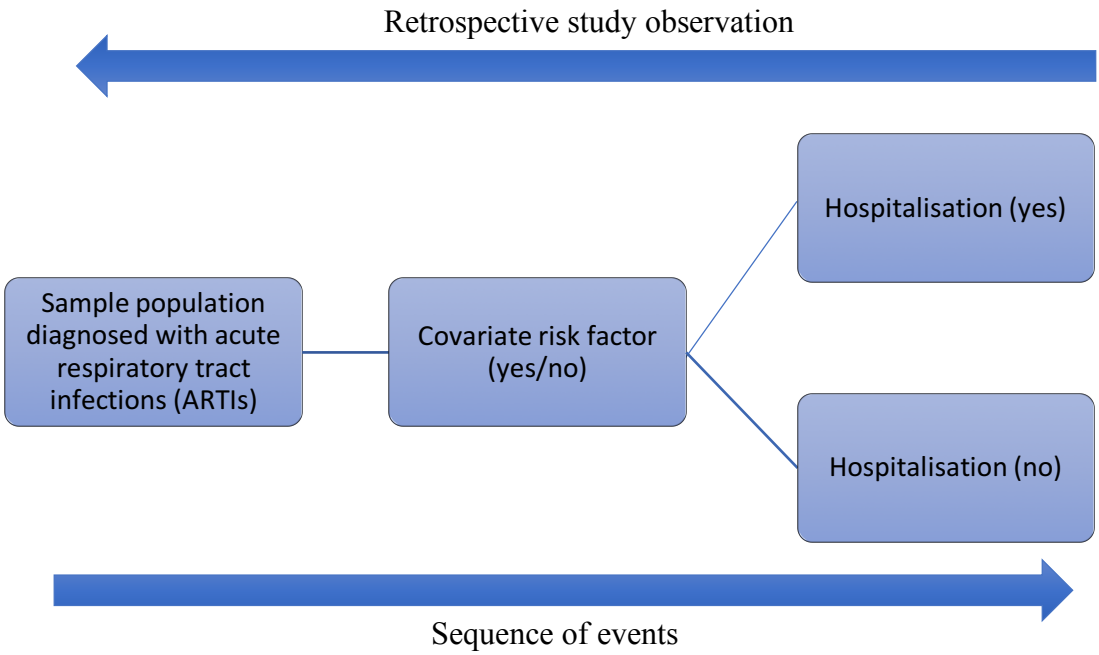

Fig 1. Cohort study sequence of events and retrospective study observation for an ARTI diagnosed sample population exposed to risk factors in order to examine their association with the likelihood of patient hospitalisation.

[i] Assignment word count = 2291 (excluding headers, footers, figures, tabulations and references).

*****

Analysis

STATA (v.14) statistical software package will be used to examine the relationship between the various risk factors (exposures) and outcome. Risk factors will be determined from the primary care records while patient hospitalisation will be determined using linked HES data. The combined data will then be reviewed and statistically analysed with STATA. Univariate and multivariable logistic regression will be used to assess the precision and statistical significance of the findings and results will be presented as odds ratios, 95% confidence intervals (CI) and p-values (OR, CI, p-value). Where the statistical probability is <= 5% (p<=0.05) it will be deemed to be statistical significant. Multivariate analysis will be used to check for possible confounding and effect modification using the Mantzael-Haensen (MH) method (comparing the crude and the adjusted odds ratio). Similar crude and adjusted odds ratios indicate a lack of evidence of confounding, while the opposite would indicate confounding. A test of homogeneity will be used to indicate effect modification with a p-value of <=0.05 indicating a statistically significant interaction. The percentage of hospitalised patients compared to patients with non-hospitalised outcomes will be presented in both relative and absolute terms in order to assess the true scale of significant risk factors amongst those patients hospitalised.

(A)

	Outcome		
Exposure	**Yes**	**No**	**Total**
Yes	A (%)	B (%)	A + B (%)
No	C (%)	D (%)	C + D (%)
Total	A + C (%)	B + D (%)	A + B + C + D (%)

(B)

Odds ratio	(AD/BC)

Table 1. (A) Two by two table illustrating the relationship between the exposures and the outcomes for odds ratio calculations (in numbers and percentages) and (B) examples of odds ratio formula.

The results will be used to describe the likelihood of hospital admission with and without relevant risk profiles for patients with ARTIs.

Brief Analysis Plan

1. The total sample population distribution will be plotted (using STATA) in patient numbers and percentages by age group and gender across a one year follow up period.
2. A summary of the distribution of individual ARTI risk factors in terms of patient numbers and percentages of the total sample population will be presented.
3. The association of risk factors, such as, age, gender, smoking status, asthmatic status, previous COPD or pneumonia symptoms, immunocompromised states, influenza like illnesses and pneumococcal vaccination history as typical risk factors leading to patient hospitalisation will be examined in terms of numbers, percentages and p-values for the total sample population.
4. Controlling for each risk factor the association with hospitalisation will be investigated in terms of patient numbers and percentages for their crude odds ratio, 95% confidence interval and p-value (for statistical significance of each risk factor).
5. A multivariable model will be executed by controlling for risk factors which are statistically significant in step 5 and results presented using adjusted odd ratio, 95% confidence interval, p-value. A test of homogeneity will be conducted (to test for effect modification).
6. The presence of confounding will be assessed by comparing the crude and adjusted odds ratios.
7. A comparison of crude and adjusted odds ratios of patients, alongside the odds ratios will be summarized (as per points no. 6 and 7).
8. The association between risk factors with statistically significance and the risk of hospitalisation for ARTI patients will be compared in number and percentage terms.

Confounding and Effect Modification

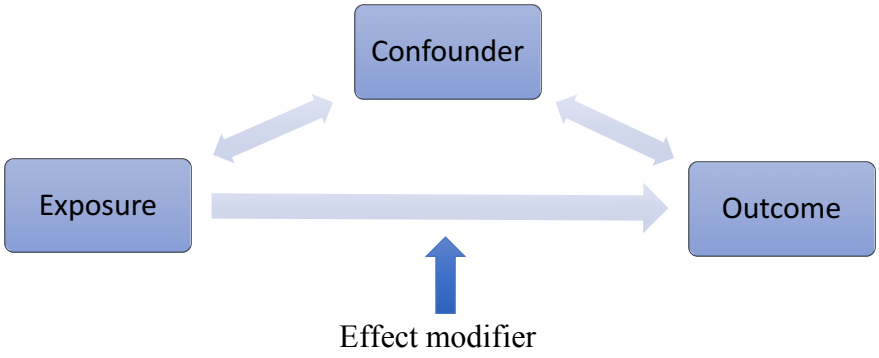

Fig 2. Illustration of the confounder and effect modifier (Interaction) relationship on the exposure to outcome pathway.

Where both the exposure and the outcome share a relationship with one or more other variables, this is known as confounding. Effect modification is where one or more other variables interact with the exposure to outcome pathway leading to a possible change in the real association between the exposure and the outcome.

Potential confounders for this study could be socio-economic status, animal exposures, allergies [Cogswell J J et al. 1987], occupational exposures and indoor exposures (e.g. smoking), where the latter was found to have an effect in pre-school children that was inversely proportional to age, [Witorsch P et al. 1993].

- Confounders will be controlled for using statistical methods like stratified analyses and MH method to avoid over-matching.
- A large sample size increases the statistical power of the study and so minimises type 2 errors.
- Comorbidities in the ARTI affected patients will be adjusted for as previously mentioned in the Inclusions and Exclusions section.
- Identification of possible confounders and effect modifiers will be considered before undertaking statistical analysis.
- Effect modification will be investigated using the MH method (testing for homogeneity).

Study limitations

Strengths of the study design

Minimising bias and confounding will lead to improved internal validity, which is a precondition for generalizability across the wider population. EHR based retrospective cohort studies are cost effective compared to prospective cohort studies and are based on existing CPRD data that reflects the general population. The at-risk population size can be readily estimated and risk factor exposure is known prior to the outcome [Sedgwick P. 2014]. Recall bias are minimised as this study uses existing datasets and the GP and participants are unaware of the study at the point of primary consultation due to the studies' retrospective nature. There is an understanding that statistically significant measured values need to be clinical significant and therefore the clinical context will factor in the interpretation of the measured values. Misclassification bias in hospitalisation status, may affect the odds ratios, however, it is anticipated that low numbers of participants would be affected in this way due to the comprehensive recording of hospitalisations by using the linked data source.

The statistical power can be used to estimate the sample size needed for the study hypothesis to be true. Representativeness of the participants can be impacted upon by the population represented by the data source, such as regional representation, which has an effect on access in rural locations and the possible impact on hospitalisations. GPs may be prone to erring on the side of caution when referring certain patient groups, such as and very young, asthma sufferers and the old to hospital, leading to an over representation of those study participants. As the EHR data will already exist the sample population size can be adjusted to suit the study. They will also be more cost effective, require less time to setup and complete than a prospective cohort study,

while allowing for the effective study of low incidence conditions using already identified participants, [Wikipedia cohort study. 2019].

Weaknesses of the study design

While CPRD primary care and HES datasets are useful sources of existing data, it will involve using datasets not original designed for research purposes and which represents ~15% of primary care practices in the UK, [CPRD. 2019]. Not all GP practices are linked to HES data and there can be a lag in HES records becoming available, [CPRD. 2019]. A limitation in using EHR data is that only risk factors that are captured can be assessed, there may also be details in free text, not captured in the EHR data used in this study. Residual confounding is an issue with observational studies. There may also be inconsistencies in how the data will be inputted by various GPs over the time frame of the retrospective study. While biases have been minimised, observer and selection bias, where it occurs may have a large affect. A potential source of misclassification for future retrospective cohort and other types of studies is the change to the Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED CT) code use in electronic health records (EHR) by 2020, which may affect how certain conditions are classified, [Kristjánsson S. 2017]. Missing data can be dealt with by using multiple imputation or by using a complete case analysis of the data. Causes of missing data vary from not being initially input by the primary or secondary care staff, being incorrectly inputted (thus rendered unusable) to covariate data being missed at follow-up, [Karahalios A. 2012]. The cohort study conclusions are based on statistical calculations and therefore should not be taken at face value without considering real world effects and possible reverse causality. Reproducibility of the study depends on the study design, mitigating bias and adjusting for confounding factors.

References

- Gulliford M et al. 2016. Safety of reduced antibiotic prescribing for self limiting respiratory tract infections in primary care: cohort study using electronic health records, British Medical Journal, 354(i3410), pp. 1-10.
- Howard P. 2015. Promoting appropriate antimicrobial prescribing in secondary care [online]. Available from: https://www.england.nhs.uk/patientsafety/wp-content/uploads/sites/32/2015/04/03-cdi-promoting-antimicrobial-prescribing-sc2.pdf [Accessed 20 November 2017]
- Pealing L et al. 2015. Risk of tuberculosis in patients with diabetes: population based cohort study using the UK Clinical Practice Research Datalink. BMC Medicine, 13, pp. 135.
- Bonita R et al. 2006. Basic epidemiology. World Health Organization, 2nd edition.
- Parker C et al. 2010. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. British Medical Journal, 341(c4245), pp. 1-10.
- Thorax BMJ. 2012. Data supplement 1 – Online supplement (Read codes). Available from: http://thorax.bmj.com/content/suppl/2012/11/11/thoraxjnl-2012-202374.DC1/thoraxjnl-2012-202374supp.pdf [Accessed 19 December 2017]
- Aslam B et al. 2018. Antibiotic resistance: a rundown of a global crisis. Infect Drug Resist. 2018; 11: pp. 1645–1658. doi: 10.2147/IDR.S173867
- Ge X et al. 2018. Epidemiology and Seasonality of Respiratory Viruses Detected from Children with Respiratory Tract Infections in Wuxi, East China. Med Sci Monit. 2018; 24: pp. 1856–1862. doi: 10.12659/MSM.908483
- HSE. 2011. Respiratory tract infection. Available at: https://www.hse.ie/eng/health/az/r/respiratory-tract-infection/causes-of-respiratory-tract-infections.html [Accessed 19 August 2019]
- Bosch A A T M et al. 2013. Viral and Bacterial Interactions in the Upper Respiratory Tract. Plos Pathogens. January 2013, Volume 9, (Issue 1), e1003057. doi: https://doi.org/10.1371/journal.ppat.1003057
- Khan S et al. 2015. Bacteria Etiological Agents Causing Lower Respiratory Tract Infections and Their Resistance Patterns. Iranian Biomedical Journal 19 (4): pp. 240-246 (October 2015). doi: 10.7508/ibj.2015.04.008
- Shah P et al. 2016. Parasitic infestation of lung: An unusual cause of interstitial pneumonitis. Lung India. 2016 Mar-Apr; 33(2): pp. 222–224. doi: 10.4103/0970-2113.177456
- WHO indoor. 2007. Indoor air pollution and lower respiratory tract infections in children. Available at: https://apps.who.int/iris/bitstream/handle/10665/43777/9789241595728_eng.pdf;jsessionid=0A30064540C04B4C0EE7CA54DB5562F0?sequence=1 [Accessed 19 August 2019]
- NHS Choices. 2018. Respiratory tract infections. Available from: https://www.nhs.uk/conditions/Respiratory-tract-infection/#lower-respiratory-tract-infections [Accessed 19 August 2019]
- NHS Digital. 2019. Read codes. Available from: https://digital.nhs.uk/services/terminology-and-classifications/read-codes [Accessed 19 August 2019]
- NHS HES. 2019. Hospital Episode Statistics (HES). Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics [Accessed 19 August 2019]
- NHS edischarge. 2019. eDischarge summaries. Available from: https://digital.nhs.uk/services/transfer-of-care-initiative/edischarge-summaries [Accessed 19 August 2019]
- Wikipedia. 2019. International Statistical Classification of Diseases and Related Health Problems. Available from: https://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_Related_Health_Problems [Accessed 19 August 2019]
- WHO. 2019. Classification of Diseases (ICD). Available from: https://www.who.int/classifications/icd/en/ [Accessed 19 August 2019]
- Meier C R et al. 1998. Acute respiratory-tract infections and risk of first-time acute myocardial infarction. The Lancet, volume 351(Issue 9114), pp. 1467-1471. https://doi.org/10.1016/S0140-6736(97)11084-4
- Ramette A et al. 2018. Longitudinal Associations Between Respiratory Infections and Asthma in Young Children. Am J Epidemiol. 2018 Aug; 187(8): pp. 1714–1720. doi: 10.1093/aje/kwy053
- Fisk W J et al. 2010. Association of residential dampness and mold with respiratory tract infections and bronchitis: a meta-analysis. Environmental Health 2010, 9:72. doi: https://doi.org/10.1186/1476-069X-9-72

*****

- Witorsch P et al. 1993. Review: Analysis of Potential Confounding Variables in Epidemiologic Studies of Parental/ Household Smoking and Respiratory Health in Preschool Children. Indoor Environ 1993; 2: pp. 71-91. doi: https://doi.org/10.1177/1420326X9300200203
- Pirez M C et al. 2011. Impact of universal pneumococcal vaccination on hospitalizations for pneumonia and meningitis in children in Montevideo, Uruguay. Pediatr Infect Dis J. 2011 Aug; 30(8): pp. 669-74. doi: 10.1097/INF.0b013e3182152bf1
- Luca D L et al. 2018. Impact of Pneumococcal Vaccination on Pneumonia Hospitalizations and Related Costs in Ontario: A Population-Based Ecological Study. Clinical Infectious Diseases 2018; 66(4): pp. 541–7. doi: 10.1093/cid/cix850
- Cogswell J J et al. 1987. Parental smoking, breast feeding, and respiratory infection in development of allergic diseases. Arch Dis Child 1987; 62: pp. 338-344.
- Kristjánsson S. 2017. Changing diagnosis codes for lower respiratory tract infections can confound longitudinal studies. Acta Pædiatrica ISSN 0803-5253. doi: https://doi.org/10.1111/apa.14017
- NHS datadictionary, 2019. Read Coded Clinical Terms. Available from: https://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/read_coded_clinical_terms.asp?shownav=1 [Accessed 19 August 2019]
- Schuetz P et al. 2017. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. Cochrane Database Syst Rev. 2017 Oct; 2017(10): CD007498. doi: 10.1002/14651858.CD007498.pub3
- Sethi S. 2010. Infection as a comorbidity of COPD. European Respiratory Journal 2010 35: 1209-1215. doi: 10.1183/09031936.00081409
- GBD 2017 Influenza Collaborators. 2018. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. The Lancet, volume 7 (Issue 1), pp. 69-89. doi: https://doi.org/10.1016/S2213-2600(18)30496-X
- Moore M et al. 2017. Predictors of pneumonia in lower respiratory tract infections: 3C prospective cough complication cohort study. Eur Respir J 2017; 50: 1700434. doi: 10.1183/13993003.00434-2017
- Bovin G et al. 2002. Virological Features and Clinical Manifestations Associated with Human Metapneumovirus: A New Paramyxovirus Responsible for Acute Respiratory-Tract Infections in All Age Groups. *The Journal of Infectious Diseases*, Volume 186, Issue 9, 1 November 2002, Pages 1330–1334. doi: https://doi.org/10.1086/344319
- Falagas et al. 2007. Sex differences in the incidence and severity of respiratory tract infections. Respiratory Medicine. Volume 101, Issue 9, September 2007, Pages 1845-1863. doi: https://doi.org/10.1016/j.rmed.2007.04.011
- TailorMade. 2019. TailorMade Info Technology Solutions Ltd. Easy Read code Search Utility. Available from: https://www.tailormadeit.co.uk/readcode_search.html [Accessed 19 August 2019]
- Ramirez-Santana M. 2017. Limitations and Biases in Cohort Studies. IntechOpen. Open access peer-reviewed chapter. February 28th 2018. doi: http://dx.doi.org/10.5772/intechopen.74324
- Szklo M. 1998. Population-based Cohort Studies. Epidemiologic Reviews Vol. 20, No. 1.
- Brilleman S L et al. 2010. The impact of attrition on the representativeness of cohort studies of older people. BMC Medical Research Methodology 2010, 10:71. doi: http://www.biomedcentral.com/1471-2288/10/71
- Sedgwick P. 2014. Retrospective cohort studies: advantages and disadvantages. BMJ 2014; 348: g1072. doi: 10.1136/bmj.g1072
- Karahalios A. 2012. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Medical Research Methodology 2012, 12:96. doi: http://www.biomedcentral.com/1471-2288/12/96
- Wikipedia cohort study. 2019. Retrospective cohort study. Available from: https://en.wikipedia.org/wiki/Retrospective_cohort_study#Advantages [Accessed 19 August 2019]
- CPRD. 2019. CPRD linked data. Available from: https://cprd.com/linked-data [Accessed 19 August 2019]
- CPRD isac. 2019. Research applications. Available from: https://cprd.com/research-applications [Accessed 19 August 2019]
- ICD10. 2014. Coding Common Respiratory Problems in ICD-10. Available from: https://www.aafp.org/fpm/2014/1100/p17.html [Accessed 19 August 2019]

*****