

Enhancing evidence based medicine: Twelve tips for conducting register-based research

Hannah L Brooke[2], Martin J Holzmann[3], Ola Olén[4], Mats Talbäck[2], Maria Feychting[2], Anita Berglund[2], Jonas F Ludvigsson[5], Rickard Ljung[2]

Corresponding author: Dr Hannah Louise Brooke hannah.brooke@ki.se

Institution: 2. Karolinska Institutet, Stockholm, 3. Karolinska University Hospital, Stockholm, and Karolinska Institutet, Stockholm, 4. Stockholm South General Hospital, Stockholm, and Karolinska Institutet, Stockholm, 5. Karolinska Institutet, Stockholm, Örebro University Hospital, Örebro, University of Nottingham, Nottingham, and Columbia University College of Physicians and Surgeons, New York

Categories: Medical Education (General), Research in Medical Education, Teachers/Trainers (including Faculty Development)

Received: 23/08/2016

Published: 24/08/2016

Abstract

With the increasing use of electronic health records for research, the need for clinicians to understand register-based research will become more important in the future. Based on our expertise we have compiled 12 tips for conducting register-based research. We aim to improve the subject knowledge of medical educators and those intending to critically appraise register-based studies. In addition, we aim to create a practical guide for those considering using register data in their own research. Our tips summarise characteristics of the research process specific to register-based research, from study design and ethical considerations, to validating the register and managing the data.

Keywords: Register-based research, Critical appraisal, Study design

Introduction

What is evidence based medicine?

Evidence based medicine, i.e. *"the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients"* (Sackett et al. 1996), is a fundamental concept in contemporary clinical practice (Mellis 2015). It has been ranked alongside vaccines, anaesthesia and contraceptive pills as one of the top 15 medical milestones since 1840 (Godlee 2007). To practice evidence based medicine, it is essential for medical students and clinicians to understand the design, analysis and limitations of original research (Windish et al. 2007;

General Medical Council 2009). Such understanding will allow them to critically appraise and interpret published research and to formulate research questions of their own.

Barriers to evidence based medicine

Most medical schools provide training in epidemiology and critical appraisal (Gillam & Bagade 2006). However, students have difficulty in understanding these subjects (Moffat et al. 2004). Beyond medical school, clinicians have limited knowledge of epidemiological concepts (Windish et al. 2007). For example, only 39% of a sample of medical residents in the US were able to recognise a case-control study and just 47% were able to recognise the definition of bias (Windish et al. 2007). Further improvements to medical education in clinical epidemiology are thus essential for the future of evidence based medicine.

Overcoming barriers to evidence based medicine in relation to register-based research

Content (subject) knowledge is one of the six core teaching competencies for medical educators (Srinivasan et al. 2011). To meet this competency, medical educators in the field of epidemiology must have detailed knowledge of a broad range of study designs and an awareness of the potential limitations of such designs. With the increasing use of electronic health records for research (Weiner & Embi 2009), the need for clinicians to understand register-based research will become more important in the future. It is therefore timely for us to use our experience in this field to compile 12 tips for conducting register-based research. We aim to improve the subject knowledge of medical educators and those intending to critically appraise register-based studies. In addition, we aim to create a practical guide for those considering using register data in their own research.

What is register-based research?

Outwardly, there may be few factors differentiating register-based research from traditional epidemiological research, except that data are obtained from registers rather than from surveys or clinical information (Thygesen & Ersboll 2014). Nonetheless, important differences between these data sources do exist (Thygesen & Ersboll 2014). The United Nations Economic Commission of Europe's (2007) report on register-based statistics in the Nordic countries identifies four key features that define a register:

1. A register is a systematic collection of unit-level data (a unit is often, but not always, an individual)
2. Data in a register are regularly updated so that changes in the data are recorded
3. Each unit in a register can be uniquely identified
4. Registers contain information on a complete target population (group of units), which is defined by precise rules

Using this type of data for research has unique strengths and limitations, which have been described previously (Olsen 2011; Thygesen & Ersboll 2014). In brief, registers provide a timely and cost-effective source of data that enable large-scale studies with long follow-up periods and low susceptibility to selection bias, recall bias and loss to follow-up. However, data are usually collected for administrative purposes so researchers have limited control over the type of information collected and the quality of that information. In addition, there are characteristics of the research process specific to register-based research, which we describe in more detail below.

12 Tips

Tip 1: Consider whether register data is suitable to address the research question at hand

The process to request and obtain access to register data already begins when formulating the research question. As with all research, it is crucial to *á priori* state the specific problem including hypotheses about biological mechanisms and relevant pathways, as well as potential subgroup analysis, because this will have direct consequences for the most appropriate choice of study design (Sackett & Wennberg 1997; Danaei et al. 2013). Conducting a systematic literature review and meta-analysis is useful to identify what is known and what is not yet known about a subject – the gap of knowledge (Brusselaers 2015). It identifies uncertainties and discrepant results, and potential explanations for these. The research question you want to address directs the review, which in turn guides further development of the research question. A scientifically based well-defined research question is crucial to avoid potential pitfalls like confounding or reverse causation.

Registers are particularly useful for large-scale cohort studies, especially when long follow-up is needed, but can also be used for case-control studies, case-cohort studies and as a base for register-based randomised clinical trials (Frobert et al. 2013). However, registers may not be suitable for all research questions. For example, it may not be possible to address problems that involve factors that are not routinely collected. Similarly, there may be data lacking for potential confounders of interest. In these situations, a judgment must be made whether the advantages of using register data outweigh the limitations of possible residual confounding. Of note, register-based data are not limited to the government-administered registers. It is often possible to access information on factors such as life quality, smoking and body mass index through quality registers (Emilsson et al. 2015; Norgaard & Johnsen 2016). Quality registers are most useful when conducting follow-up analysis using the cases as a cohort. However, when information is gathered from registers to which reporting is not mandatory, such as disease quality registers, the material used for analyses is no longer population based, so selection bias may have been introduced if the likelihood of being included in the analysis is related to both the studied outcome (i.e. the dependant variable) and the exposure (i.e. the independent variable) of interest.

Tip 2: Evaluate if minimum requirements for register-based research are available

There are three minimum requirements for register-based research. Firstly, each individual in the register must be assigned a unique identification code. Where this code is consistent between multiple data sources the sources can be linked. The ability to link information from multiple registers greatly enhances the scope of research questions that can be addressed. Secondly, a legal framework for using the data and combining data from multiple sources must be in place. Thirdly, the register should have a reasonable level of quality and completeness. Sweden is set up with unique identification codes of the entire population (described previously (Ludvigsson et al. 2009b)) and legal processes that correspond to most research situations (Nordic Council of Ministers 2014; Ludvigsson et al. 2015). Not all countries may have these advantages and advice on legal requirements should be sought before embarking on register-based research. Issues of register data validity will be discussed in tip 7; however, low quality and largely incomplete registers prevent useful research from being conducted.

Tip 3: Build a research team with the necessary expertise to successfully complete the project

The optimal team for performing high quality register-based clinical studies typically includes clinical, epidemiological, biostatistical, sociological and pathophysiological competence. Clinicians are needed to formulate the most clinically relevant questions. Clinicians will also help assessing the validity of different codes in the registers, based on their use in clinical practice. Note, for example, that the International Statistical Classification of Diseases and Related health Problems (ICD)-codes have changed over the decades, so knowledge about the different codes used throughout the whole period you are studying is important. Epidemiological competence is important, since in a typical register-based project there are an abundance of possibilities to go astray in the veritable jungle of study design options. Ideally, you should try to find collaborators who have already published high quality studies

using a specific exposure or outcome, and who have already gone through the review process where most of the weaknesses for a particular design or main variable have already been discovered. To carry out a project efficiently, you will need biostatistical and data management competence. The datasets that you work with in register-based research are generally large and complex. One of the advantages of large-scale register data is the possibility to examine socioeconomic inequalities in health, treatment and prognosis. Sociological competence is beneficial for adding this additional layer to your study. Finally, you will need a high pathophysiological competence relevant to the research question. Otherwise the paper will risk lacking an in depth discussion of the possible biochemical background to the findings presented.

Tip 4: Assess additional data sources that may complement standard administrative data

Registers can often be linked to other sources of information in order to address specific research questions. For example a birth cohort or a randomised controlled trial can be enriched by linkage to registers in order to get complete long time follow up data, which would otherwise be hard to attain (e.g. antibiotic exposure during childhood (Uusijarvi et al. 2014) or long term outcomes after an intervention (Dalen et al. 2015)). Administrative databases within hospitals may be a way to study patient populations with a certain disease, or symptom. Furthermore, information on blood tests, blood cultures or other tests may be retrieved from laboratory (Selmer et al. 2014) or histopathology databases (Ludvigsson et al. 2009a). Such information is not commonly available in other registers. Figure 1 depicts how a hospital administrative database was used to conduct a study in 2014 (Bandstein et al. 2014). The study population was identified from the local register at the Department of Emergency Medicine, which was linked to laboratory data and then to national registers containing additional information. Although the resulting study would not have been possible based only on nationally held register data, the process of linking many data sources together can be time-consuming. As such, we recommend that only experienced research teams take such approaches.

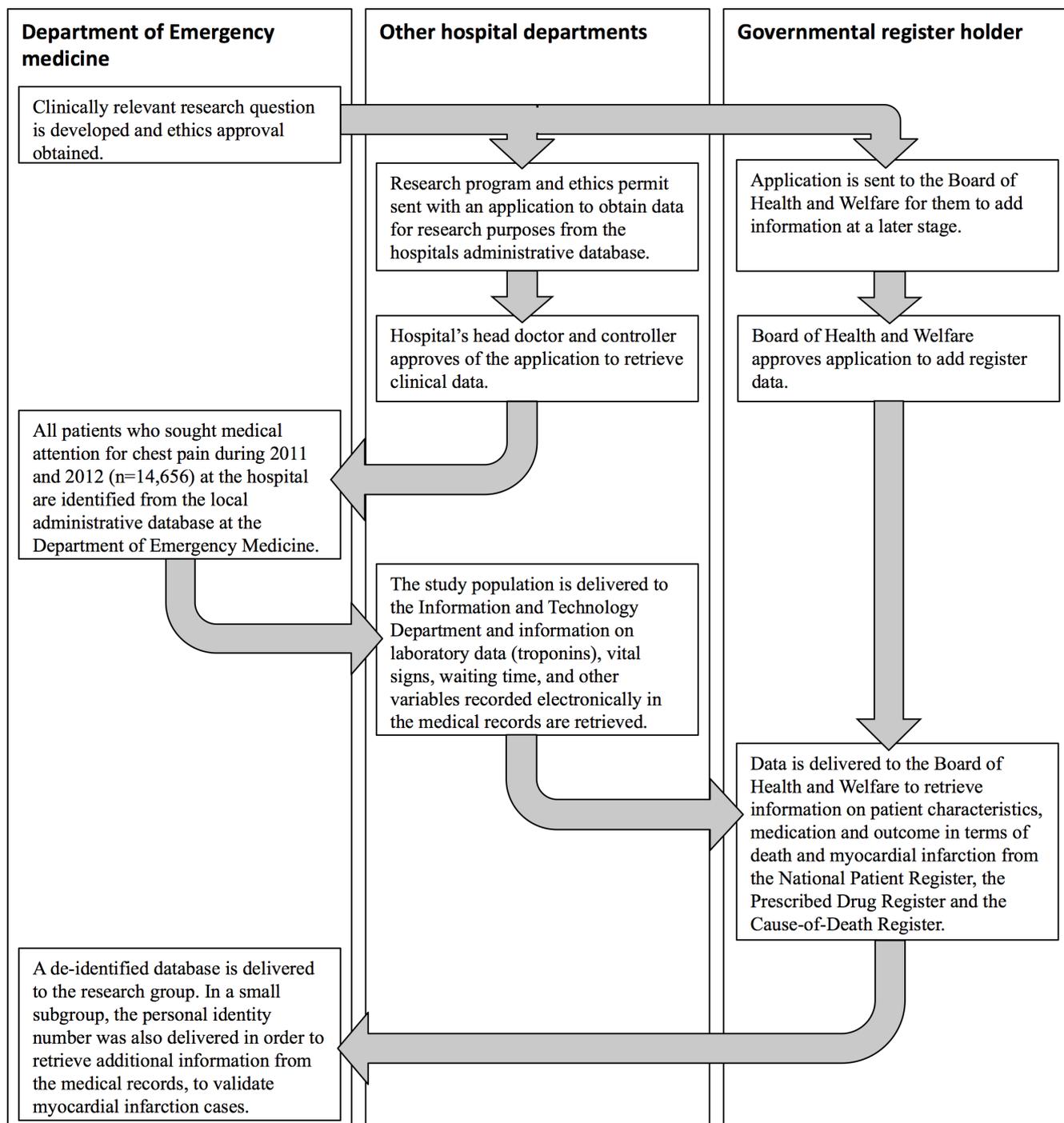


Figure 1. An example of how a hospital administrative database was combined with data in national registers to conduct a study in 2014 (Bandstein et al. 2014).

Tip 5: Consider collaborating with other countries or regions with similar registers and pooling data

Pooling data is much more complex than performing a single register-based study. As such, it has to be worth it, i.e. you usually only pool data for uncommon exposures, outcomes or important sub-analyses with few events that really need the extra statistical power you gain from pooling. Your ethical permits must cover pooling of data from several centres and you should check with the register holders the extent to which transferring data between centres is allowed. In many cases data are not allowed to leave the country, and all analyses have to be made on local servers

via remote access that only allow output to be exported. Alternatively, data can be aggregated before transfer so it no longer contains sensitive information. This facilitates the data transfer process. Before you come to the analyses of the data you have to build a sound dataset. Write detailed codebooks, where all variables from the different registers are compared. For example, many ICD-codes have slightly different versions in different registers and might also have been used in slightly different ways. You also have to identify which variables that are found in all of the participating registers, for which part of the study population and for which part of the follow up time.

Tip 6: Become familiar with the strengths and limitations of available register data - review available documentation regarding appropriate variables to request from register holders and talk to those who have used these registers before

In medical research either the outcome, exposure or the population under study is defined by some kind of medical condition or procedure. Hence, it is crucial that this medical condition, and its time of onset, can be correctly identified in the registers. Acute, well-defined conditions, needing inpatient care, such as myocardial infarction, usually have a higher reliability and validity compared to non-acute diffuse conditions with a gradual disease onset most often treated in outpatient care, such as gastroesophageal reflux. Furthermore, data may be missing in some registers for various reasons. For example, drugs that are administered as part of hospital treatment will most often be missing if data is obtained from a prescription register. It is important to understand the strengths and limitations of each register in order to select the most appropriate data source. Information about the data contained in each register is often available through the register holders. We recommend that this be thoroughly reviewed before deciding which data to request from which register. If a register-based definition has commonly been used and published, one should have very good reasons to choose another definition. In order to ensure transparency and reproducibility the ICD-codes and other criteria used to define your study population should be clearly stated in the manuscript. It should also be noted that although each register has extensive content at the government agencies, researchers are rarely allowed to access the full datasets but will be asked to define what variables they are interested in, and why.

Tip 7: Evaluate the quality/validity of data available and consider conducting a validation study

The quality and validity of the reported condition or variable of interest can vary between registers. Completeness of outcome reporting may vary over time, between geographic regions, and with age of the patients, both as a result of a gradual introduction of new diagnostic techniques, changes in clinical practice and disease coding routines, or through administrative changes. Administrative reasons are nearly always behind sudden changes in disease occurrence etc. Hence it can be valuable to combine information from several registers or demand multiple recordings of inpatient or outpatient visits in order to classify the patient as having the condition of interest with a high degree of certainty (Bauer et al. 2015). In Sweden, the National Patient Register has nationwide coverage and a high completeness of hospitalizations, diagnoses and surgical procedures (Ludvigsson et al. 2011); however, it lacks more detailed clinical data such as disease activity indexes, indications for surgery or discontinuation of a treatment. Thus adding clinical information from disease-specific quality registers (Emilsson et al. 2015) can be useful (as discussed in Tip 1). Hopefully the diagnosis is already validated (Ludvigsson et al. 2011; Razavi et al. 2011; Tao et al. 2015). If not, it may be necessary to validate the information available before the start of a major research project. The most adequate way of validating the data in the register is to conduct a manual review study. A random sample of hospitalisations or visits reported to the register is drawn and the corresponding medical records are retrieved for manual review. A positive predictive value (Loong 2003) of the register reported condition can be calculated based on the proportion of reported conditions that after manual review of the medical records were judged to be correctly classified. Depending on the condition of interest around 300-1000 medical records usually have to be retrieved and manually reviewed.

Tip 8: Become familiar with register specific ethics procedures

While one may assume that ethical issues are absent in register-based research, it is not the case. An ethics committee must approve register-based research in most countries, as shown by an overview of the ethical aspects of register-based research (Ludvigsson et al. 2015). Although informed consent is usually not required for register-based studies, using data from registers is an enormous responsibility and it is only possible with the trust of the citizens. Precautions should be taken to avoid secrecy and integrity breaches. Hence, register holders usually replace possibly traceable individual identifiers with random numbers before the data is delivered to the researcher. This still allows linkage of data from the same person within and between the delivered data files, but removes the possibility to directly identify the individual. However, these records are, in most cases, still regarded as sensitive individual data. Other ethical aspects that may be considered are the participation of vulnerable subjects (e.g. children or unconscious patients), and the re-use of data (the latter makes the data collection more efficient and is an argument for register-based research).

Tip 9: Ensure necessary IT and computing infrastructure is in place

The secure storage of the data is of prime importance; data should be stored on an encrypted server in a locked room. The large volume of data contained within registers may mean that computer processing capacity will need to be increased. However, the analytical approach can be designed to increase analysis speed without upgrading the infrastructure. For example, it is possible to use Poisson regression in data that has been aggregated across individuals according to key characteristics, thus reducing the number of observations in the data set. Conducting a case-control or case-cohort study on a sub-sample of the study population can also increase computational efficiency. Analytical syntax for register-based research can be shared via platforms such as GitHub (<https://github.com>); this may provide a useful starting point for those new to register-based research.

Tip 10: Spend time understanding your data and creating an operational dataset that is easy to work with

When raw data are received from the register holders, some pre-processing is usually necessary before it is in a format that can easily be used for analyses. The extent of pre-processing depends on the complexity of the research question and the number of registers from which data must be merged. Sufficient time should be taken to understand the data in depth so as to avoid mistakes or overlooking important information. It is useful to create a codebook with detailed definitions, which can be used by all working in the project. Furthermore, it is important to record any changes made to the original data for documentation purposes and so that steps can be retraced and errors corrected at a later stage.

Tip 11: Conduct in depth evaluation of the data you receive to check for unexpected values, trends and patterns

Unexpected trends and patterns may indicate changes in coding practices, or policies over time. Engage in dialogue with register holders to gain further insight in such situations. Unexpected values may be a result of coding errors. For example, coding should be questioned if a patient with repeated diagnoses of urticaria (ICD-10 L50.9) is assigned a diagnosis of angina pectoris in the dermatology clinic (ICD-10 I50.9). In a healthcare setting, misdiagnoses or a lack of differentiation between suspected and confirmed diagnoses can lead to further misclassification. Which records and codes to include in analyses is vital as different combinations may result in an under- or over-estimation of the number of cases. Ensure that the number of records or individuals correspond to national publications for similar data, e.g. figures on disease incidence, mortality or prevalence. Similarly, check that the age and sex distribution is reasonable. Also check for data for reused personal identity numbers, individuals who die more than once, events after death, women who have been classified as primiparous at more than one birth

(especially if the births occur with many years in between) and other such irregularities.

Tip 12: Assess the extent of missing data on exposures and other variables and determine an appropriate strategy to handle missing data

Missing data exist in most registers and will occur for different reasons. For example, some variables may have only been collected for a limited period of time (e.g. placental weight in the Swedish Medical Birth Register) while individuals with protected identity may ‘disappear’ from the population register. It is important to know why data are missing in order to handle it in an appropriate way. If data are missing completely at random and if the register is large enough it may often be appropriate to drop records with missing data. Imputation techniques are available and may be appropriate to use in some situations (Sterne et al. 2009). In addition, it might be possible to use information or proxy variables from another register or from a different year. The possibility to use substitute data is higher if the information is likely to remain stable over time or is unlikely to change after a certain age for example, education. When using hierarchical variables, e.g. ICD-codes, ascertain the completeness of the required level of detail. If data are missing at the required level it may be an oversight of those providing the information, or this information may have never been recorded.

Registers are particularly useful for large-scale cohort studies, especially when long follow-up is needed, but can also be used for case-control studies, case-cohort studies and as a base for register-based randomised clinical trials (Frobert et al. 2013). However, registers may not be suitable for all research questions. For example, it may not be possible to address problems that involve factors that are not routinely collected. Similarly, there may be data lacking for potential confounders of interest. In these situations, a judgment must be made whether the advantages of using register data outweigh the limitations of possible residual confounding. Of note, register-based data are not limited to the government-administered registers. It is often possible to access information on factors such as life quality, smoking and body mass index through quality registers (Emilsson et al. 2015; Norgaard & Johnsen 2016). Quality registers are most useful when conducting follow-up analysis using the cases as a cohort. However, when information is gathered from registers to which reporting is not mandatory, such as disease quality registers, the material used for analyses is no longer population based, so selection bias may have been introduced if the likelihood of being included in the analysis is related to both the studied outcome (i.e. the dependant variable) and the exposure (i.e. the independent variable) of interest.

Concluding remarks

Medical educators with detailed knowledge of a broad range of study designs, and an awareness of the potential limitations of such designs, are essential for the future of evidence based medicine. To help address this need we have compiled 12 tips for conducting register-based research. Our tips summarise characteristics of the research process specific to register-based research, from study design and ethical considerations, to validating the register and managing the data. This paper will improve the subject knowledge of medical educators and those intending to critically appraise register-based studies. Moreover, this paper will act as a practical guide for those considering using register data in their own research.

Take Home Messages

Notes On Contributors

Hannah Brooke is a COFAS Marie Curie Post-doctoral Fellow at the Institute of Environmental Medicine, Karolinska Institutet. She specialises in using register-based research to study socioeconomic factors in relation to cancer incidence and survival.

Martin J Holzmann is a cardiologist/internist, and an Associate Professor of epidemiology. He has conducted a number of register-based studies, merging information from national health-care registers, quality-of-care registers, with administrative data as described above.

Ola Olén is a pediatric gastroenterologist at the Sachs' Children's hospital and a researcher at the Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet. His research is mainly focused on register-based studies of outcome and prognosis in inflammatory bowel disease.

Mats Talbäck has a PhD in epidemiology and works as a Statistician at the Institute of Environmental Medicine, Karolinska Institutet. He has extensive experience with population based-registers and register-based research and has previously worked at both Statistic Sweden and the National Board of Health and Welfare.

Maria Feychting is Professor of Epidemiology, and Head of the Unit of Epidemiology at the Institute of Environmental Medicine, Karolinska Institutet. She has extensive experience of register-based epidemiological research.

Anita Berglund is Assistant Professor of epidemiology, and coordinator for the Doctoral Programme in Epidemiology at Karolinska Institutet and for the Swedish Interdisciplinary Graduate School in register-based research.

Jonas F Ludvigsson is a paediatrician and Professor of clinical epidemiology. He has been the president of the Swedish Society of Epidemiology (2011-14), and of the Swedish Society of Pediatrics (2014-16).

Rickard Ljung is an Associate Professor and specialist in social medicine, at the Institute of Environmental Medicine, Karolinska Institutet. He has previously worked at the National Board of Health and Welfare and has long experience of using registers in public health epidemiology research.

Acknowledgements

Hannah Brooke is a COFAS Marie Curie Fellow with funding from Forte (grant registration number 2015-01228). Ola Olén was funded by the Bengt Ihre research fellowship in gastroenterology while working on this project. Martin J Holzmann was funded by a grant from the Swedish Heart-Lung foundation (grant registration number 20150603). This work was also supported by Karolinska Institutet. The funding bodies had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Bibliography/References

Bandstein N, Ljung R, Johansson M, Holzmann MJ. 2014. Undetectable high-sensitivity cardiac troponin t level in the emergency department and risk of myocardial infarction. *J Am Coll Cardiol.* 63:2569-2578.

<http://dx.doi.org/10.1016/j.jacc.2014.03.017>

Bauer MS, Lee A, Miller CJ, Bajor L, Li M, Penfold RB. 2015. Effects of diagnostic inclusion criteria on prevalence and population characteristics in database research. *Psychiatr Serv.* 66:141-148.

<http://dx.doi.org/10.1176/appi.ps.201400115>

Brusselsaers N. 2015. How to teach the fundamentals of meta-analyses. *Ann Epidemiol.* 25:948-954.

<http://dx.doi.org/10.1016/j.annepidem.2015.08.004>

Dalen M, Ivert T, Holzmann MJ, Sartipy U. 2015. Coronary artery bypass grafting in patients 50 years or younger: A Swedish nationwide cohort study. *Circulation.* 131:1748-1754.

<http://dx.doi.org/10.1161/CIRCULATIONAHA.114.014335>

Danaei G, Rodriguez LA, Canero OF, Logan R, Hernan MA. 2013. Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res.* 22:70-96.

<http://dx.doi.org/10.1177/0962280211403603>

Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. 2015. Review of 103 Swedish healthcare quality registries. *J Intern Med.* 277:94-136.

<http://dx.doi.org/10.1111/joim.12303>

Frobert O, Lagerqvist B, Olivecrona GK, Omerovic E, Gudnason T, Maeng M, Aasa M, Angeras O, Calais F, Danielewicz M, et al. 2013. Thrombus aspiration during ST-segment elevation myocardial infarction. *N Engl J Med.* 369:1587-1597.

<http://dx.doi.org/10.1056/NEJMoa1308789>

General Medical Council. 2009. Tomorrow's doctors : Outcomes and standards for undergraduate medical education. Available from: http://www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf

Gillam S, Bagade A. 2006. Undergraduate public health education in UK medical schools - struggling to deliver. *Med Educ.* 40:430-436.

<http://dx.doi.org/10.1111/j.1365-2929.2006.02444.x>

Godlee F. 2007. Milestones on the long road to knowledge. *BMJ.* 334 Suppl 1:s2-3.

<http://dx.doi.org/10.1136/bmj.39062.570856.94>

Loong TW. 2003. Understanding sensitivity and specificity with the right side of the brain. *BMJ*. 327:716-719.

<http://dx.doi.org/10.1136/bmj.327.7417.716>

Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, Heurgren M, Olausson PO. 2011. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 11:450.

<http://dx.doi.org/10.1186/1471-2458-11-450>

Ludvigsson JF, Brandt L, Montgomery SM. 2009a. Symptoms and signs in individuals with serology positive for celiac disease but normal mucosa. *BMC Gastroenterol*. 9:57.

<http://dx.doi.org/10.1186/1471-230X-9-57>

Ludvigsson JF, Haberg SE, Knudsen GP, Lafolie P, Zoega H, Sarkkola C, von Kraemer S, Weiderpass E, Norgaard M. 2015. Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol*. 7:491-508.

<http://dx.doi.org/10.2147/CLEP.S90589>

Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. 2009b. The Swedish personal identity number: Possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 24:659-667.

<http://dx.doi.org/10.1007/s10654-009-9350-y>

Mellis C. 2015. Evidence-based medicine: What has happened in the past 50 years? *Journal of Paediatrics and Child Health*. 51:65-68.

<http://dx.doi.org/10.1111/jpc.12800>

Moffat M, Sinclair HK, Cleland JA, Smith WC, Taylor RJ. 2004. Epidemiology teaching: Student and tutor perceptions. *Med Teach*. 26:691-695.

<http://dx.doi.org/10.1080/01421590400013537>

Nordic Council of Ministers, NordForsk. 2014. Joint Nordic registers and biobanks: A goldmine for health and welfare research. *Policy Papers*, ISSN 1504-8640; 5. Available from: https://www.nordforsk.org/en/publications/publications_container/joint-nordic-registers-and-biobanks-a-goldmine-for-health-and-welfare-research.

Norgaard M, Johnsen SP. 2016. How can the research potential of the clinical quality databases be maximized? The Danish experience. *J Intern Med*. 279:132-140.

<http://dx.doi.org/10.1111/joim.12437>

Olsen J. 2011. Register-based research: Some methodological considerations. *Scand J Public Health*. 39:225-229.

<http://dx.doi.org/10.1177/1403494811402719>

Razavi D, Ljung R, Lu Y, Andren-Sandberg A, Lindblad M. 2011. Reliability of acute pancreatitis diagnosis coding in a national patient register: A validation study in Sweden. *Pancreatology*. 11:525-532.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. 1996. Evidence based medicine: What it is and what it isn't. *BMJ*. 312:71-72.

<http://dx.doi.org/10.1136/bmj.312.7023.71>

Sackett DL, Wennberg JE. 1997. Choosing the best research design for each question. *BMJ*. 315:1636.

<http://dx.doi.org/10.1136/bmj.315.7123.1636>

Selmer C, Olesen JB, Hansen ML, von Kappelgaard LM, Madsen JC, Hansen PR, Pedersen OD, Faber J, Torp-Pedersen C, Gislason GH. 2014. Subclinical and overt thyroid dysfunction and risk of all-cause mortality and cardiovascular events: A large population study. *J Clin Endocrinol Metab*. 99:2372-2382.

<http://dx.doi.org/10.1210/jc.2013-4184>

Srinivasan M, Li STT, Meyers FJ, Pratt DD, Collins JB, Braddock C, Skeff KM, West DC, Henderson M, Hales RE, et al. 2011. "Teaching as a competency": Competencies for medical educators. *Academic Medicine*. 86:1211-1220.

<http://dx.doi.org/10.1097/ACM.0b013e31822c5b9a>

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. 2009. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 338:b2393.

<http://dx.doi.org/10.1136/bmj.b2393>

Tao W, Holmberg D, Naslund E, Naslund I, Mattsson F, Lagergren J, Ljung R. 2015. Validation of obesity surgery data in the swedish national patient registry and scandinavian obesity registry (SOREg). *Obes Surg*.

Thygesen LC, Ersboll AK. 2014. When the entire population is the sample: Strengths and limitations in register-based epidemiology. *Eur J Epidemiol*. 29:551-558.

<http://dx.doi.org/10.1007/s10654-013-9873-0>

United Nations Economic Commission for Europe (2007). Register-based statistics in the Nordic countries : Review of best practices with focus on population and social statistics.

Uusijarvi A, Bergstrom A, Simren M, Ludvigsson JF, Kull I, Wickman M, Alm J, Olen O. 2014. Use of antibiotics in infancy and childhood and risk of recurrent abdominal pain-a Swedish birth cohort study. *Neurogastroenterology and Motility*. 26:841-850.

<http://dx.doi.org/10.1111/nmo.12340>

Weiner MG, Embi PJ. 2009. Toward reuse of clinical data for research and quality improvement: The end of the beginning? *Ann Intern Med*. 151:359-360.

<http://dx.doi.org/10.7326/0003-4819-151-5-200909010-00141>

Windish DM, Huot SJ, Green ML. 2007. Medicine residents' understanding of the biostatistics and results in the

medical literature. JAMA. 298:1010-1022.

<http://dx.doi.org/10.1001/jama.298.9.1010>

Appendices

Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons "CC BY 4.0" (<https://creativecommons.org/licenses/by-sa/4.0/>)