2.4

Databases and registers as tools for disaster epidemiology

Authors

Philip J. Schluter and **Hyun M. Kim**, School of Health Sciences, University of Canterbury – Te Whare Wānanga o Waitaha, Christchurch, New Zealand.

2.4.1 Learning objectives

To understand the three major types of databases and registers available to disaster epidemiology researchers, and their associated strengths and weaknesses by:

1. Characterizing the salient differentiating features of these database and register types.
2. Providing case studies and examples to illustrate these and their usage.
3. Highlighting the strengths and weaknesses of each, and providing a global overview.

2.4.2 Introduction

The *exposome* is defined as "the totality of human environmental exposures", in other words, all the non-genetic exposures which shape individuals' life-course trajectories *(1)*. Quantification of this all-encompassing concept is challenging at a single point in time, and is even more complex over time – particularly in the context of health emergency and disaster risk management (Health EDRM) when people may be exposed to the risks or consequences of emergencies and disasters. An individual's exposure begins before birth and includes insults from multiple sources. In the normal course of events, genetics has been found to account for only about 10% of diseases, while the remaining causes appear to be from life histories and environment *(1)*. Significant insults from emergencies and disasters have even a greater impact.

A key factor in describing and understanding the exposome and a person's resultant life-course trajectory is the ability to accurately measure germane factors and exposures, and their effects. Databases and registers – due to their rapid evolution, availability, and the ability for them to be linked to other information sources – are increasingly being used by researchers to improve this understanding. Building on the discussion of disaster epidemiology in Chapter 2.1, this chapter outlines three major types of databases and registers that are useful for epidemiological investigations in the disaster context:

- Ongoing population-based databases and registers (typically comprising routinely collected administrative data);
- Pre-existing subpopulation databases and registers (often available from ongoing cohort studies initiated prior to the emergency or disaster event); and
- Post-disaster databases and registers (studies and databases initiated and established after the event and therefore containing no (or little) pre-event information).

2.4.3 Ongoing population-based databases and registers

Today's world is increasingly digitized with a vast amount of data produced daily. In 2018, it was estimated that 2.5 quintillion bytes of data were created each day, and this is rapidly accelerating *(2)*. Some 90% of the world's data were generated in the last two years alone *(2)*. Data are being routinely and more frequently collected from increasingly varied sources and archived. The promise of Big Data and machine learning and data science, then, is to map the exposome, and ascertain the contribution of events and exposures. However, much work remains to be done – although initiatives such as New Zealand's Integrated Data Infrastructure (IDI) *(3)* are helping. The IDI is a large national research database holding microdata about all New Zealand people and households. It links detailed data from health, education, justice, income and work, population and many other sources over time. Such databases can provide a holistic detailed baseline account and history of individuals in an emergency- and disaster-affected region, and the resultant effects on those who stay or flee, included on those who were unaffected. Because these data are prospectively collected and provide complete population coverage, they are likely to provide robust and less biased epidemiological estimates of factors and exposures before, during and after an emergency or disaster. However, they are limited by the scope and quality of the data that are actually collected *(4)* and administrative data typically lack important qualitative information. This is because administrative data collected by government agencies are generally for the purposes of registration, transaction, monitoring and record keeping, rather than for research or research-related objectives *(5)*. How these types might be used in Health EDRM research is discussed in Chapter 4.4.

Such broad-based, comprehensive, linked population-based datasets remain uncommon internationally, although this is changing rapidly. For example, in Republic of Estonia, an efficient, secure and transparent nationwide digit ecosystem has been built that includes integrated data from different healthcare providers to create a common record for every patient (https://e-estonia.com). Within the domain of health, medical databases are often massive repositories of routinely collected detailed information and may serve as a robust research tool *(6)*. For example, patient registries with complete nationwide coverage and individual-level linkage potential have existed in the Republic of Finland since 1969, Denmark since 1978, Sweden since 1987, the Republic of Iceland since 1999 and Norway since 2008 *(7)*. These health registers can be used to provide baseline information and to track the impact of emergencies or disasters. Case Study 2.4.1 provides one example in which routinely collected information from Christchurch Hospital in New Zealand is used to assess the impact of the 2010-2011 Christchurch earthquakes and a change in their healthcare service delivery model.

However, routine databases and registers are often not appropriately designed for specific disaster research purposes or do not lend themselves to this. At times, they absorb considerable resources for very little scientific gains *(4)*. Furthermore, the precise exposures or confounders that researchers wish to explore or account for are frequently absent from these databases *(10)*. This, in part, motivated the development of the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) statement to aid transparency and improve research reporting (www.record-statement.org). Although, for hazards that have slow onset but long duration, such as deforestation and air pollution, the ongoing population-based databases designed with long term broad-based measures may be advantageous over post-disaster tailored databases.

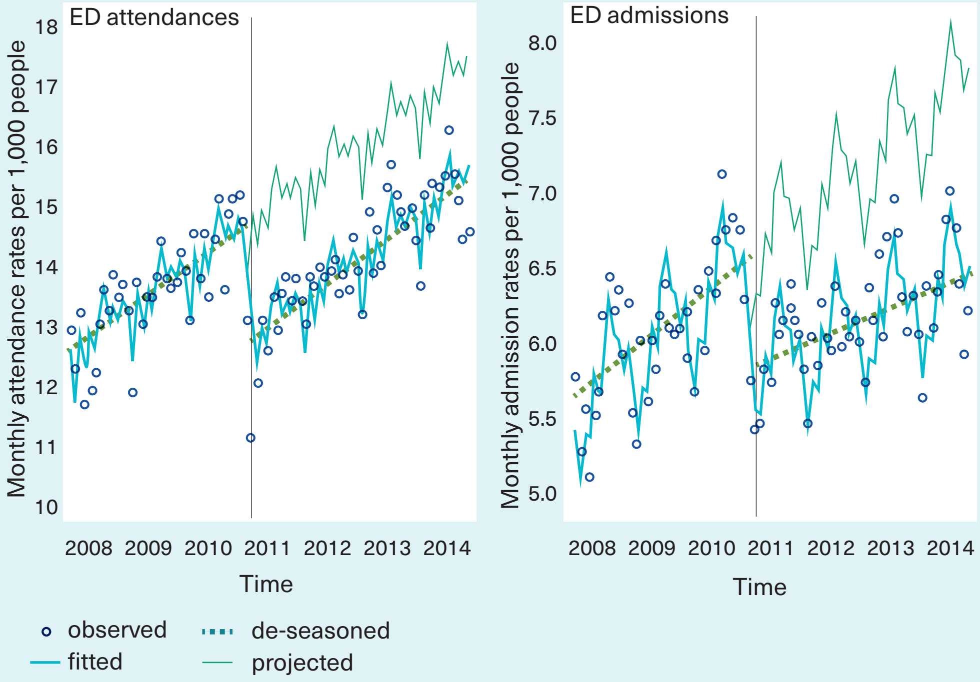
Case Study 2.4.1
Measuring the impact of integrated health system changes on emergency department attendances and acute admission, precipitated by an earthquake *(8)*

Hospital systems routinely collect data on a number of activities, including emergency department (ED) attendances and acute admissions. These attendances and admissions are costly and often preventable. Moreover, in many countries, the healthcare service provision is increasingly recognized as being unsustainable. In response, the Canterbury District Health Board initiated a shift to an integrated person-centred healthcare model *(9)*. However, the 2010–2011 Christchurch earthquakes and aftershock series (the most devastating of which occurred on 22 February 2011, resulting in 185 deaths, more than 6500 injuries, an estimated 10 600 people relocating to outside of Christchurch, and costing NZ$ 40 billion – or 19% of New Zealand's Gross Domestic Product) compromised infrastructure and disrupted services, so that this new healthcare delivery model was rapidly implemented. While conceptually appealing, the evidence base for such a service model is relatively weak, and the empirical impact it had within the Canterbury District Health Board was unknown. By interrogating the routinely collected ED attendance and admission records for Christchurch Hospital, the single tertiary hospital in the region serving approximately 500 000 people, one important component of the earthquake impact and change in service delivery model could be measured.

Figure 2.4.1 presents the observed, fitted, de-seasoned and projected standardized population emergency department attendance and acute admission rates, derived from models using Bayesian change-point methods. The 'projected' line gives the predicted rates based on pre-earthquake and pre-existing healthcare delivery model, while the 'fitted' line gives the actual rates derived from the routinely collected data. The demonstrative change post-earthquake, together with the significantly decreased rate of growth in emergency department admissions is also depicted. These findings support the conclusion that, after the earthquake, the Canterbury District Health Board's integrated health systems transformations have resulted in a dramatic and sustained reduction in emergency department attendances and acute hospital admissions.

Figure 2.4.1 shows scatter plots of observed Canterbury District Health Board standardized monthly Emergency Department (ED) attendance rates (left) and Emergency Department admissions (right) per 1000 people (hollow circles), together with a superimposed fitted lined from the full time-series model (solid line), the estimated de-seasoned trend line (heavy dashed straight line) and the extrapolated projected line (grey line). The vertical line denotes the 22 February 2011 earthquake.

Figure 2.4.1 Scatter plots of observed Canterbury District Health Board standardized monthly Emergency Department attendance and admission rates per 1000 people.



2.4.4 Pre-existing sub-population databases and registers

Another rich source of exposome data arises from serendipitous pre-existing cohort or longitudinal studies that were already being conducted in an area affected by an emergency or disaster. These studies often contain pre-event information from multiple health, social and environmental domains; invariably using instruments with excellent research-orientated psychometric properties. Moreover, participants in these studies commonly have their data augmented by information collected from other sources. This reduces responder burden, and also harnesses a greater information landscape. One such example is the Avon Longitudinal Study of Children and Parents, which was established to understand how genetic and environmental characteristics influence health and development in parents and children *(11)*. Other examples include the nationally representative Demographic and Health Surveys which are regularly conducted in various African, Asian, European, Oceanian, Latin American and Caribbean countries *(12)*, the China Health and Retirement Longitudinal Study *(13)*, and the Nurses' Health Study in the USA *(14)*.

A similar study exists in Christchurch – the Christchurch Health and Development Study, which follows 1265 children born in 1977. At the time of the 2010-2011 Christchurch earthquakes, the Christchurch Health Development Study cohort participants were aged 34 years, with just over 50% exposed to the earthquakes and the remainder unexposed (forming a non-randomized control group). The comprehensive pre-event data, combined with the different earthquake exposure levels, provides a powerful mechanism to understand the disaster impact: the study presented as Case Study 2.4.2, for example, explored the role of peri-traumatic stress in predicting major depression symptoms. Pre-existing longitudinal studies with data linkage capabilities to population-based registers can also provide new possibilities for analysing peri- and post-traumatic stress symptoms following a disaster. For example, when combined with health service use data, a more comprehensive view of the impact of physical and mental trauma on individuals across a longer time span can be gained.

Case Study 2.4.2
Understanding the role of peri-traumatic stress and disruption distress in predicting symptoms of major depression following exposure to a natural disaster (15)

Few studies have examined the contribution of specific disaster-related experiences to symptoms of depression. This study investigated this among an existing cohort of individuals exposed to the 2010-2011 Christchurch earthquakes and associated major aftershocks. One of the perennial challenges associated with disaster epidemiology research is the availability of detailed pre-event data. However, Christchurch is home to the long-running Christchurch Health and Development Study, a birth cohort of 1265 children born in 1977. This cohort has now been studied repeatedly from birth to age 35 years, has maintained high retention (79% of those surviving) and the resultant database contains a large repertoire of life-course information. More than 50% of the study cohort were exposed to the earthquakes, and at age 35 years, those exposed were interviewed about their experiences of these earthquakes.

The strengths of this study include the availability of data from a well-studied cohort and the use of a model which tests for both peri-traumatic and post-event distress simultaneously. Pre-earthquake covariates included cognitive ability, prior history of mental disorder and familial socioeconomic status measures. Previous studies which report that major depression is related to post-event factors have not looked at confounders of this association. The study found that peri-traumatic stress is an under-recognized predictor of major depressive disorder following a disaster caused by natural hazards.

2.4.5 Post-disaster databases and registers

Pre-existing population-based databases and registers or research-based studies are often inadequate or insufficient to understand the health impacts and service gaps on a population following an emergency or disaster. In such instances, post-disaster databases or registers are needed. These are flexible and tailored to contain instruments and tools that are most pertinent to the specific population and situation. However, critical gaps in observational research instruments still exist, such as the monitoring of long-term mental health or psychosocial risk of people in both a clinical and community setting *(16)*. Moreover, the clear disadvantage of this approach is that predisaster information must be recalled or retrieved retrospectively, which can suffer from important biases, such as selection bias and information bias. Practical and ethical considerations are also paramount. These include interference with emergency responses or recovery, participant safety and sensitivity and ensuring that truly informed consent can be obtained (see also Chapter 3.4). Nonetheless, this is a common and important approach taken by researchers and agencies alike. Examples include the World Trade Center Health Registry *(17)* described in Case Study 2.4.3, the 1995 Oklahoma City Bombing Injuries Database *(18–20)* and the Canterbury Earthquake Recovery Authority Wellbeing Survey *(21)*.

Case Study 2.4.3

World Trade Center Health Registry (17) and the longitudinal determinants of depression among World Trade Center Health Registry enrollees, 14 to 15 years after the 9/11 attacks *(22)*

The World Trade Center Health Registry is now the largest registry in UnS' history to track the health effects of a disaster. It tracks the impact of the 9/11 attacks, a series of four coordinated attacks by the terrorist group al-Qaeda on 11 September 2001. The attacks killed 2996 people, injured more than 6000 others, and caused at least US$ 10 billion in infrastructure and property damage, with other dying of 9/11-related cancer and respiratory diseases in the months and years after the attacks. The World Trade Center Health Registry was established post-disaster, and enrolment was voluntary for people who lived, worked or went to school in the area of the disaster, or who were involved in rescue and recovery efforts. To enrol, participants completed a confidential "Wave 1" health survey in 2003 or 2004. More than 71 000 people enrolled, including 4000 survivors of the collapsed World Trade Center towers. Multiple measurement waves have followed, with surveys in 2007, 2011 and 2015. The results of these surveys help determine the extent to which physical and mental health conditions have persisted, and whether any new symptoms and conditions have emerged.

Another important goal is to identify and help address gaps in physical and mental health treatment. For example, in Jacobson and colleagues (2018) study, the longitudinal determinants of depression among different PTSD levels were examined for 21 258 enrollees who had completed four questionnaires over 14 years of follow-up. They found that 18.6% experienced depression, and it was more common among those who had ever experienced PTSD (56.1%) compared with those who had not (5.6%). These findings highlight the substantial burden of depression in a trauma-exposed population 14 to 15 years after the disaster, especially among those with PTSD. Many World Trade Center Health Registry research outputs have been published *(23)*. Moreover, like many bodies (such as the Integrated Data Infrastructure in New Zealand), the World Trade Center Health Registry welcomes proposals for new studies from external researchers. Upon approval, researchers can request de-identified survey data or request that the Registry facilitate recruitment of enrollees into a study.

2.4.6 Conclusions

Disaster epidemiology researchers are able to use a variety of health-related databases and registers when studying topics of relevance to Health EDRM. A broad overview of the important strengths and weakness typically associated with databases and registers is presented in Table 2.4.1. However, each specific dataset and scenario may have other important strengths and weaknesses and requires careful critique and evaluation before it is used in research.

Table 2.4.1 Important strengths and weakness typically associated with databases and registers used as tools for disaster epidemiology

Major register types	
Strengths	**Weaknesses**
Ongoing population-based databases and registers:	
Cost: usually relatively inexpensive; Coverage: usually population wide; Predisaster information available; Time: relatively quick to undertake.	Not designed for disaster research; Important instruments or variables may be missing or have poor psychometric properties; Database linking may be difficult or impossible; Selection bias may mean that those missing from the register are importantly different from those included; Data are often aggregated or grouped in ways that lead to findings suffering from the ecological fallacy; Big Data datasets require data storage systems, computation capacity and performance, and analytical techniques that are (currently) often beyond the scope of many individual researchers.
Pre-existing sub-population databases and registers:	
Cost: potentially inexpensive if ‘added-on’ to an existing study; Predisaster information available Typically cover subject matter in-depth; Capability to augment with qualitative information; Instruments normally designed for research purposes and often tested for psychometric properties and reliability.	Not originally designed for disaster research, so may miss important factors or exposures; Recruitment or retention to the existing study may limit the external validity of finding; Study participant sample size may lack statistical power.
Post-disaster database and registers:	
Designed and tailored for disaster and population of interest.	Cost: usually expensive; Predisaster information is limited; Potentially time consuming and resource or expertise intensive; Timely collection of data may be unethical; If a multi-agency, multi-sector research collaboration then competing interests may exist and hamper the scope.

2.4.7 Key messages

- **There are multiple and growing sources of data available for disaster epidemiology research. Knowledge of the exposome can be extended and developed by using and linking these data, and exploring how emergencies and disasters affect people's likelihood of mortality, morbidity and life-course trajectories.**
- **The expediency of using routinely collected data is often offset by the coverage, depth and quality of the variables available to researchers. This often requires initiation of a post-disaster study, that is both specifically and contextually relevant to the disaster and the population affected.**
- **As more better quality and richer data are collected, Big Data, machine learning and data science are likely to play an increasingly important role in disaster epidemiology research. However, possible avenues to augment these quantitative data with qualitative information still need to be explored.**

2.4.8 Further reading

Kreis IA, Busby A, Leonardi G, Meara J, Murray V, editors. Essentials of Environmental Epidemiology for Health Protection: A Handbook for Field Professionals. Oxford, United Kingdom: Oxford University Press. 2013.

Webb P, Bain C, Page A, editors. Essential Epidemiology: An Introduction for Students and Health Professionals (4th edition). Cambridge, United Kingdom: Cambridge University Press. 2020.

2.4.9 References

1. National Institute for Occupational Safety and Health. Exposome and Exposomics. 2014 https://www.cdc.gov/niosh/topics/exposome/default.html (accessed 16 January 2020).
2. Marr B. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. 2018 https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#633e5df660ba (accessed 16 January 2020).
3. Stats NZ Tatauranga Aotearoa. Integrated Data Infrastructure. 2018 https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure (accessed 16 January 2020).
4. Curbelo RJ, Loza E, de Yébenes MJ, Carmona L . Databases and registers: useful tools for research, no studies. Rheumatology International. 2014: 34(4): 447-52.
5. Connelly R, Playford CJ, Gayle V, Dibben C . The role of administrative data in the big data revolution in social science research. Social Science Research. 2016: 59: 1-12.
6. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. Journal of Epidemiology and Community Health. 2014: 68(3): 283-7.
7. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clinical Epidemiology 2015: 7: 449-90.
8. Schluter PJ, Hamilton GJ, Deely JM, Ardagh MW. Impact of integrated health system changes, accelerated due to an earthquake, on emergency department attendances and acute admissions: a Bayesian change-point analysis. BMJ Open. 2016: 6(5): e010709.
9. Gullery C, Hamilton G. Towards integrated person-centred healthcare - the Canterbury journey. Future Hospital Journal. 2015: 2(2): 111-6.
10. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Medicine 2015: 12(10): e1001885.
11. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology. 2013: 42(1): 97-110.
12. The Demographic and Health Surveys Program [website]. https://dhsprogram.com/ (accessed 1 May 2020)
13. China Health and Retirement Longitudinal Study [website] http://charls.pku.edu.cn/index/en.html (accessed 1 May 2020).
14. The Nurse's Health Study [website] https://www.nurseshealthstudy.org/ (accessed 1 May 2020)

15. Bell CJ, Boden JM, Horwood LJ, Mulder RT. The role of peri-traumatic stress and disruption distress in predicting symptoms of major depression following exposure to a natural disaster. Australian and New Zealand Journal of Psychiatry. 2017: 51(7): 711-8.

16. Généreux M, Schluter PJ, Takahashi S, Usami S, Mashino S, Kayano R, et al. Psychosocial management before, during, and after emergencies and disasters - results from the Kobe Expert Meeting. International Journal of Environmental Research and Public Health. 2019: 16(8): e1309.

17. NYC 9/11 Health. World Trade Center Health Registry. 2019 https://www1.nyc.gov/site/911health/about/wtc-health-registry.page (accessed 16 January 2020).

18. North CS, Nixon SJ, Shariat S, Mallonee S, McMillen JC, Spitznagel EL et al. Psychiatric disorders among survivors of the Oklahoma City bombing. JAMA 1999: 282(8): 755-62.

19. Glenshaw MT, Vernick JS, Li G, Sorock GS, Brown S, Mallonee S. Factors associated with injury severity in Oklahoma City bombing survivors. Journal of Trauma. 2009: 66(2): 508-15.

20. Oklahoma City Bombing. 2019. Oklahoma State Department of Health. https://www.ok.gov/health/Protective_Health/Injury_Prevention_Service/Oklahoma_City_Bombing

21. Morgan J, Begg A, Beaven S, Jamieson K, Johal S, Johnston D et al. Monitoring wellbeing during recovery from the 2010–2011 Canterbury earthquakes: The CERA wellbeing survey. International Journal of Disaster Risk Reduction 2015: 14(Part 1): 96-103.

22. Jacobson MH, Norman C, Nguyen A, Brackbill RM. Longitudinal determinants of depression among World Trade Center Health Registry enrollees, 14-15 years after the 9/11 attacks. Journal of Affective Disorders. 2018: 229: 483-90.

23. Published Research/Publications In New York City 9/11 Health [website] https://www1.nyc.gov/site/911health/researchers/published-research-publications.page (accessed 1 May 2020).