Chapter 4.6
Health-related risk modelling

Holly C.Y. Lam
Zhe Huang
Emily Y.Y. Chan

Learning objectives

To have a basic understanding of some modelling methods that might be applied in health emergency and disaster risk management (Health EDRM) research studies investigating:

- Short-term environmental health associations.
- Factors associated with the uptake of protective behaviours.
- Trends in influenza.
- Health-related vulnerability index.

Applicability of health-related risk models to Health EDRM

Health EDRM is important for reducing the public health impacts of disasters and emergencies, and research methods can require:

The use of statistical modelling to establish **mathematical associations between variables** related to health-related risks.

Models for evaluating short-term environmental health associations (1)

Improved weather forecasting and better understanding of the health risks of extreme environmental events can support the implementation of effective health protection plans and improvements in resource allocation. This uses modelling methods to evaluate short-term associations between environmental exposures and health outcomes.

For example, ambient temperature usually has a short-term association with health outcomes, with a delayed effect for the temperature-health association, which is not always linear.

Models for evaluating short-term environmental health associations (2)

The estimation of a temperature-health association is complicated and a time series design is the most common research method to reveal these short-term associations. Time series data are a **series of sequential records in equal time units**, such as the number of deaths and the average daily or weekly temperature within a specific time period.

The statistical analysis methods in time series studies include:
- Time stratified model.
- Periodic functions.
- Flexible spline functions.

Types of time series study - time stratified model

Time stratified model → exposure and outcome are associated in stratified time units.

- Time intervals are indicated by **indicator variables** instead of the true date record.

✓ Relatively easy to understand.

✗ The model includes many parameters and it cannot facilitate the calculation of the continuous effect from one time unit to another.

Types of time series study - periodic functions

Periodic functions (Fourier terms) → model exposure and outcome

- Periodic functions, such as **sine** and **cosine function**, are used to represent the periodic characteristics (such as calendar months).

✅ Creates smooth predictions.

❌ The period of the pattern is fixed, which might not be appropriate for representing trends that are mathematically complicated and do not have a fixed pattern.

Types of time series study - flexible spline function

Flexible spline function $\longrightarrow$ combines different polynomial curves

- This design is most commonly applied in assessing short-term associations between temperature and health outcomes.

✓ Allows the health outcome to be linked to a number of exposures with different non-linear associations at the same time.

✓ Helps control for long-term demographic factors at a population level (e.g. overall changes in the proportion of smokers and risk of hospital admission).

Incorporating the non-linear delayed effects of ambient temperature on health outcomes into the spline model

Distributed Lagged Non-linear Model (DLNM) → three-dimensional data analysis, considering the exposure, health outcome and the delay (time) dimension at the same time

- Spline functions can also be applied on the time dimensions.
- Combining flexible spline approach and DLNM tackles most of the concerns when evaluating short-term environmental health associations.

- Complicated because it involves one more dimension than other traditional time series designs.
- The problem of effect modification by other factors (such as age and gender) is present and needs to be handled separately.

Uptake of protective behaviours

- Applying appropriate protective behaviours during extreme events can lower exposure to hazards and hence reduce health risk.
- Sociodemographic factors and knowledge, attitude and practice are common examples of determinants of health behaviours.
- Research that identifies associated sociodemographic factors and helps with understanding knowledge, attitude and practice for a protective behaviour provides evidence to support health promotion policies.

# How to identify factors associated with the uptake of protective behaviours?

Logistic regression $\rightarrow$ a regression model characterized by one binary dependent variable (**outcome**) and multiple independent variables (**explanatory variables**).

- Used to investigate the association between the outcome variable and an explanatory variable with adjustment for other confounders.
- Widely used to identify factors (such as knowledge and gender) that might be associated with the likelihood of a group of people acting in a certain way and for comparing this to a reference group of other people.

General guide for including explanatory variables to reveal the real associations between factors in Health EDRM (1)

1. There should be at least 10 cases for each explanatory variable in each outcome group and statistical power increases with increasing numbers of cases.
2. To reduce the number of explanatory variables in a regression model a univariate analysis can be used.

> ➤ chi-square test (for categorical variables)
> ➤ t-test (for continuous variables)
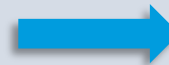
→ Quick assessment of the potential associating factors

General guide for including explanatory variables to reveal the real associations between factors in Health EDRM (2)

3. Explanatory variables showing potential association with the outcome in the univariate analysis, together with some core explanatory variables (supported by literature or hypothesis) are then entered into the logistic regression model.
4. Model selection:
   - removing non-significant variables from a full model.
   - adding variables and keeping those that are significant.

Case study 4.6.1: *Data collection by telephone survey*

- A population-based telephone survey among the Hong Kong population investigated their weather information acquisition pattern during an intense cold spell.
- The Chi-square test and a **logistic regression model** were used to identify independent associated factors in a two-stage analysis.
- Univariate analyses were used to identify potential associated factors with the outcome and factors (p-value <0.20) from the chi-square test were entered to the multiple logistic regression analysis, to assess their independent association with the outcome.

Case study 4.6.1: *Data collection by telephone survey*

- In the univariate analyses, educational attainment, age and marital status were significantly associated with current use of smartphone apps to acquire weather information.
- In multiple logistic regressions, only older age and lower education level remained significantly associated with lower smartphone app usage.

Prediction and forecasting of trends in influenza

Influenza causes large burdens to global public health and economies.
Influenza viruses gradually mutate and when a new contagious strain emerges in a community without immunity, this may lead to an epidemic.
Accurate prediction of strain types and number of cases is important for primary prevention strategies, because it facilitates effective vaccine strain selection and resource planning for the healthcare sector.

Predictive models for vaccine selection and the forecast of influenza activity (1)

Antigenicity-stability fitness model
- Estimates the expected growth rate (fitness) of viral clades by input of a few years of genetic and antigenic data and is able to predict frequency of trajectory of clades for about one year ahead.

Epitope Clade Growth (genealogical tree)
- Estimates antigenic differences by extrapolating recent growth hemagglutinin clades seeded by epitope mutation.

Local Tree Shape (genealogical tree)
- Estimates recent clade growth from information stored in the local shape of a hemagglutinin genealogical tree.

Predictive models for vaccine selection and the forecast of influenza activity (2)

These models provide an estimate of the effectiveness of current vaccines for the emerging influenza strains and identify new antigenic variants at an early stage of expansion.

- Linking antigenic properties and genetic data
- Identifying proposed vaccine strains

> ways of inferring phenotypic properties

Predictive models for vaccine selection and the forecast of influenza activity (3)

Forecasts can be based on single or multiple measures:

- Peak periods (time).
- Peak and outbreak magnitude.
- Case counts by day or week.

Health-related vulnerability index

- Understanding risk in all its dimensions is essential for effective Health EDRM and the collection of large volumes of data presents opportunities to describe reality accurately (Chapter 2.4).
- However, if there are too many variables for a clear understanding of data this causes the "curse of dimensionality".
- A **dimension reduction** method maps the numerous original variables into fewer independent dimensions, based on their correlation to each other.
- A good quality composite index uses careful variable selection and appropriate use of the dimension reduction method, and can facilitate communication and policy making.

Case study 4.6.2: *Principal components analysis (PCA) to develop a Heat Vulnerability Index*

- Principal components analysis (PCA) was used to combine socioeconomic indicators into a Heat Vulnerability Index in London, United Kingdom.
- 9 variables were identified: households in rented tenure, households in a flat, population density (persons/hectare), households without central heating, population >65 years old, population with self-reported health status, receiving any kind of social benefit, single pensioner households and ethnic group.
- These were included in the PCA.

Conclusions

Risk modelling is well established and can:

- Help resource allocation in Health EDRM.
- Predict temperature-health associations, but consistent associations have not often been found for other climate-related topics such as rainfall or sea level rise.
- Be used in other contexts (such as complex emergencies) or between varying contexts (such as rural versus urban) to understand health-related impact of hazards and disasters.

Key messages (1)

- Time series analysis is widely used for establishing short-term associations between exposures and health outcomes.
- Health-related risk modelling is an important research method and approach to support Health EDRM development
- Factors associated with protective or preparedness behaviours can be identified with the multiple logistic regression method.

Key messages (2)

- Linking antigenic properties and genetic data, and identification of proposed vaccine strains can help with influenza vaccine selection. They estimate the effectiveness of current vaccines for emerging influenza strains and identify new antigenic variants at an early stage of expansion.
- In constructing a health-related risk index, dimension reduction approaches such as principal component analysis and factor analysis are widely used to simplify the display of multivariate data.

Further readings

Jackson JE. A user's guide to principal components, 1991.
Information on challenges, considerations, and methods to measure impact in Health EDRM.
Wood SN. Generalized additive models: An introduction, 2006.
Brief introduction to statistical models commonly used for health-related risk assessments.
Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011: 43(8): 1.
Brief introduction to statistical models commonly used for health-related risk assessments.
Vynnycky E, White R. An introduction to infectious diseases modelling, 2010.
Brief introduction to statistical models commonly used for health-related risk assessments.
McSharry P. Parsimonious risk assessment and the role of transparent diverse models. In Risk modeling for hazards and disasters, 2018. pp. 263-9.
Brief introduction to statistical models commonly used for health-related risk assessments.

References

This chapter: Holly C.Y. Lam, Zhe Huang, Emily Y.Y. Chan T. Chapter 4.6: Health-related risk modelling

Data collection by telephone survey (case study 4.6.1): Chan EYY. Public health humanitarian responses to natural disasters. Routledge. 2017.

Forecast Model - Simulation Optimization (SIMOP) (case study 4.6.2): Nsoesie EO, Beckman RJ, et al. A simulation optimization approach to epidemic forecasting. PloS ONE 2013: 8(6): e67164.

Principal components analysis to develop a Heat Vulnerability Index (case study 4.6.3): Wolf T, McGregor G. The development of a heat wave vulnerability index for London, United Kingdom. Weather and Climate Extremes 2013: 1: 59-68.

Factor analysis to develop a Health Vulnerability Index (case study 4.6.4): Chan EYY, et al. Health vulnerability index for disaster risk reduction: application in belt and road initiative (BRI) region. International Journal of Environmental Research and Public Health 2019: 16(3): 380.

Contact information

Professor Emily Ying Yang Chan
Professor and Assistant Dean, Faculty of Medicine, The Chinese University of Hong Kong
Shatin, Hong Kong SAR, China
Email: emily.chan@cuhk.edu.hk

Health EDRM Research Network Secretariat
WHO Centre for Health Development (WHO Kobe Centre)
Email: wkc_tprn@who.int