The statistics of health

Professor Rhonda Rosychuk discusses her biostatistics research laboratory and the challenges and rewards of studying the numbers behind health outcomes

Firstly, could you begin by explaining how you came to establish a biostatistical research laboratory?

I received excellent doctoral training as a PhD student in statistics at the University of Waterloo and was recruited to the University of Alberta to join the Department of Pediatrics as a biostatistician. In this novel position, I was able to secure salary awards from external funding agencies, develop my own statistical research programme, and establish collaborations and consultations with medical researchers and trainees. I immediately set about securing grant funding for personnel and equipment and formed a biostatistical laboratory to meet the requirements of my studies. As my research programme evolved, I became more interested in the geographic variation of diseases and illnesses and decided to develop a grant that would build on my strengths in disease cluster detection, my experience with administrative health databases, and my collaborations with medical experts.

What are some of the many complexities inherent in the analysis of spatial data?

Adequately modelling spatial correlation is a challenge. The key idea is that data observed closely together are more likely to be similar than data observed less closely together. This spatial correlation needs to be incorporated in statistical models and analyses to be able to draw appropriate conclusions. The addition of time can also add to this complexity.

Further difficulties include the fact that datasets are large and computations may be an issue; geographic boundaries may change over time-making spatio-temporal analyses more complicated; and sources of data may be available from different agencies which do not reference their data to the same geographic unit: for example, population may be aggregated by census tract but cancer cases are aggregated by health authority. Statistical methods provide tools to answer questions, and users need to know the properties and features of these methods to ensure valid conclusions are drawn from the data.

What challenges have you faced in developing new methodologies and how were these overcome?

Methodological developments are like driving to a destination without a detailed road map. You know the starting point and where you want to go but the points along the journey are not known in advance. Sometimes you come to points which cause you to take a different route and find new solutions you could not have anticipated. In addition, statisticians are in high demand and the recruitment of research staff, the accessibility of administrative data, and computational requirements have presented some challenges, but I work hard to find solutions to surmount these challenges.

Why is spatial statistics a critically important emerging field in health research? Why should students consider a career in this area?

The data available in the world today is expanding rapidly and extracting information from it requires specialised statistical skills. The improvements in computer power and available methods in software mean that spatial analyses are becoming increasingly popular. There is also greater recognition that health outcomes are related to a multitude of factors including environment and lifestyle that are often clustered geographically. With all of these factors to consider, spatial statistics is poised to make substantial contributions to health research. The growth in this area provides students with excellent opportunities to specialise in spatial statistics and they will have many career options when they enter the work force.

What excites you most about your work?

I am excited by the prospect of discovering something new. Problem solving is tremendously satisfying and knowing that the solution may help answer a health question is gratifying. I enjoy the challenge that a research career provides and I value the opportunity to share my knowledge and expertise with trainees and collaborators, and to see them succeed. There are more potential projects than available time and one of the keys to my success is that I try to focus on researchers that share the same enthusiasm for discovery.

What other projects have resulted from this research grant? Where do you hope to direct your efforts in the future?

We have added features to our tests and models such as adding a temporal component to examine clustering over time and alternative distributions and approximations to improve computation times. I intend to build on my strengths and knowledge by developing a marked point process model approach for spatio-temporal modelling and analysing spatial and environmental factors for several cardiac conditions.

Mapping disease

A team at the University of Alberta in Canada is working to understand geographical variation in order to provide invaluable information for the study of causes and factors behind diseases, and present health authorities with essential data about risks to the public

A DISEASE CLUSTER is the occurrence of a particular illness affecting an unusually large number of people within a certain geographic area or period of time. Without disease surveillance systems, health authorities may be alerted to clusters by clinicians perceiving an increase in a certain disease, or by members of the public who have observed an illness afflicting people they know. Health authorities are obliged to respond to enquiries in a timely fashion, either to reassure the public that no clustering exists or to warn the public as appropriate, and to initiate further investigation into the cause of the cluster.

The identification and measurement of patterns of disease are important goals in public health. Disease maps attempt to enhance the understanding of disease processes by showing the spatial distribution – or the geographical range – of diseases and identifying these clusters. These patterns are not restricted to distributions of disease in geographical areas, but may also arise in workplaces, occupations, families or other structures. With a better understanding of these spatial distributions, health researchers can formulate aetiological hypotheses, conduct further studies to assess these hypotheses, and identify exposures and characteristics of increased risks to the disease.

NEW STATISTICAL METHODOLOGIES

Professor Rhonda Rosychuk is a biostatistician in the Department of Pediatrics at the University of Alberta, Canada. One of her research interests is the development of new statistical methodologies in disease surveillance and cluster detection to reveal statistically significant disease clusters in regions with diverse population sizes and around sources of possible contamination. Her team was awarded a Canadian Institutes of Health Research (CIHR) grant to create and assess new statistical methods and the application of these methodologies to important health datasets extracted from population-based databases in Alberta.

The objectives of the research are fourfold, and involve developing a spatial scan for the detection of clusters of disease-related, or correlated, events; investigating the power of cluster detection methods for disease-related events; developing a spatio-temporal model that includes seasonal effects for different geographic areas and time to capture extra variability; and finally, applying the new methods to all presentations – or visits – to Alberta emergency departments (EDs) for various medical conditions and the Alberta Cancer Registry for all childhood cancer diagnoses.

Cluster detection methods identify geographic areas with greater than expected cases of disease, but further epidemiological investigations are required to ascertain if a suspected cluster is indeed genuine or not. Without confirmatory evidence, local reports of incidence of disease may be false alarms. Similarly, coincidental observations may not pick up the true extent of a real disease cluster.

One promising method for identifying possible disease outbreaks developed by Rosychuk's team is to combine statistical disease cluster detection techniques with administrative databases. "This approach is not only cost-effective and efficient but also allows for timely cluster detection, because the necessary data are often already available by administrative area," she asserts. "Identification of suspected clusters can motivate subsequent

INTELLIGENCE

BIOSTATISTICAL METHODS FOR DISEASE CLUSTER DETECTION AND SPATIAL MODELLING

OBJECTIVES

To develop a spatial scan for the detection of clusters of disease-related (correlated) events; investigate the power of cluster detection methods; develop a spatio-temporal model that includes seasonal effects; and apply the methods to Alberta datasets (eg. emergency department presentations for asthma, childhood cancer diagnoses).

KEY COLLABORATORS

University of Alberta: **Dr Brian H. Rowe**, Professor, Department of Emergency Medicine, Faculty of Medicine & Dentistry • **Professor Amanda S Newton**, Assistant Professor, Department of Pediatrics, Faculty of Medicine & Dentistry • **Dr Hsing-Ming Chang**, Postdoctoral Fellow

Professor Mahmoud Torabi, former Postdoctoral Fellow, now Assistant Professor, University of Manitoba

FUNDING

Canadian Institutes of Health Research

CONTACT

Professor Rhonda J Rosychuk
AI-HS Health Scholar
Director, Biostatistics Consulting Group

Department of Pediatrics,
University of Alberta
Room 3-524, Edmonton Clinic Health Academy
11405 87 Avenue NW
Edmonton, Alberta
T6G 1C9, Canada

T +1 780 492 0318
F +1 780 248 5625
E rhonda.rosychuk@ualberta.ca

www.ualberta.ca/~rhondar
www.BiostatsConsultGroup.ualberta.ca

PROFESSOR RHONDA J ROSYCHUK has a BSc Honours in Pure Mathematics from the University of Alberta and an MSc in Statistics from the University of British Columbia. She completed her PhD in Statistics at the University of Waterloo, and has been accredited as a Professional Statistician by the Statistical Society of Canada and the American Statistical Association. Professor Rosychuk is Director of the Biostatistics Consulting Group, formerly known as Pediatric Statistical Consulting which she founded. She holds a Health Scholar salary award from Alberta Innovates-Health Solutions.

epidemiological investigations to search for possible causes, if the clusters are deemed important by the health authority".

EVENT CLUSTER DETECTION TESTS

Existing surveillance methods either evaluate areas of constant population and compare the number of cases of disease, or analyse the specified number of cases and compare underlying population sizes. However, Rosychuk's argues that 'cases' may not be the most relevant unit of analysis; rather, multiple disease-related events experienced by an individual, such as multiple ED presentations can provide additional information. For example, 10 people presenting to an ED is not the same as 10 visits by one person. Yet, current methods are not based on such correlated data. "I wanted to identify clusters of disease-related events such as ED presentations because I thought that events may provide additional insights over disease incidence and be helpful for health services research," she elucidates.

> Disease maps attempt to enhance the understanding of disease processes by showing the spatial distribution of diseases and identifying these clusters

In Rosychuk's view, a method that more closely captures the distribution of the data, and the correlation, would be an improvement over Poisson models, which are commonly used for the analysis of disease rates. Thus, the group has employed a compound Poisson approach: "This method seemed applicable for the number of ED presentations, whilst at the same time reflecting the fact that presentations are linked to patients," she explains.

Seeing a need for a standardised approach for detecting aggregations of disease-related events, the team has developed new 'event cluster detection tests'. Using known population counts and numbers of cases and events in each geographical sub-region in their study region, the investigators either examine each sub-region separately, combining the resulting information with that of neighbouring regions to reach a pre-specified threshold of events; or they consider many different combinations of regions and count the number of events. They then calculate the probability of observing the events based on the population size and the hypothesised distribution of the events in the absence of clustering. Their research has included specifying different hypothesised distributions and examining the performance of the methods.

TEMPORAL AND SPATIAL VARIATIONS

For complex systems, models may need to capture variations in time as well as geography. In the infectious disease context, malaria and influenza-related mortality have been noted to have spatio-temporal as well as seasonal effects. Spatial and spatio-temporal disease rate models are able to capture variations in true rates and better separate systematic variability from random noise, the latter of which usually overshadows crude rate maps. Since disease rates may differ substantially across geographical regions because of regional characteristics and risks, spatio-temporal analysis provides reliable estimates of the underlying disease risk by incorporating data, or 'borrowing strength' from neighbouring geographic sub-areas.

Rosychuk's group has recently proposed a new methodology to handle such effects. Random effect terms that are assumed to follow particular models such as conditional autoregressive models capture the spatial, temporal and interaction components. The study incorporates seasonal factors into the statistical model by including cosine and sine functions. The ultimate result is a regression-type model for the expected average count for a sub-area at a particular time point with fixed and random effects. Using a generalised estimating approach, the investigators were able to obtain the estimates for parameters in the model. "In some contexts, the underlying rates may change over seasons within a given year," notes Rosychuk. "A model that more closely represents the structure of the data will likely provide more appropriate inferences."

WATCH THIS SPACE

Work with Professor Mahmoud Torabi, a former postdoctoral fellow, and Dr Hsing-Ming Chang, a postdoctoral fellow, has included a spatial scan created for disease-related event cluster detection and an examination of cluster methodologies for the detection of childhood cancer cases in Alberta, as well as a hierarchical spatiotemporal analysis of childhood cancer trends. Crucially, additional research projects have resulted from this research grant, including an examination of two methods for incorporating temporal testing in a specific cluster detection test. However, it takes time to develop new methodologies, and some of the datasets already tested have become too old to publish in journals. The researchers are planning to apply the new methodology to more recent datasets to be able to present more current results in subject-specific journals, opening doors for substantial contributions to health research.