Second International Network Sampling Workshop (INSW)
2 – 4 March, 2000

Hotel Beaumont, Wycker Brugstraat 2, 6221 EC Maastricht, the Netherlands

This workshop is organised by Marinus Spreen (University Maastricht), Ove Frank (Stockholm University) and Tom Snijders (University of Groningen). Hotel Beaumont, Wycker Brugstraat 2, 6221 EC Maastricht, the Netherlands. The hotel is located in the inner city of Maastricht, about 500 m. from the Central Railway

---

Estimating the effective rate of sex partner change from individuals with sexually transmitted diseases

Nico Nagelkerke
E-mail: nico.nagelkerke@rivm.nl
National Institute of Public Health and the Environment, The Netherlands

The effective rate of partner change is an important variable for determining whether an epidemic of a sexually transmitted disease (STD) occurs. It is essential for understanding the transmission dynamics of STDs. The study addressed the problem of estimating the effective rate of partner change. Under certain conditions the mean rate of partner change in STD patients was found to estimate the effective rate of partner change in the general population. The study concluded that to estimate the effective rate of partner change in a population, the use of surveys of patients with STD provides a simpler and less costly alternative to population surveys of sexual behavior.

Estimation with Link-Tracing Sampling Designs: A Bayesian Approach

Mosuk Chow and Steven K. Thompson
E-mail: mchow@stat.psu.edu
The Pennsylvania State University

In link-tracing designs, social links are followed from one respondent to another to obtain the sample. For hidden and hard-to-access human populations, such sampling designs are considered the most practical way to obtain a sample large enough to study. In this paper, we propose a Bayesian approach for the estimation problem. For problems using link-tracing designs, it is quite often that there are prior information on the characteristics that one wants to estimate. Thus, using this information effectively via a Bayesian approach should yield better estimators. Also, under the Bayesian setup, obtaining interval estimates and assessing the accuracy of the estimators can be done without much added difficulties whereas such tasks would be very difficult to perform using the classical approach. In general, a Bayesian analysis yields one distribution (the posterior distribution) for the unknown parameters, and from this a large number of questions can be answered simultaneously.

Comparisons of Network Information Collected with Specific versus Non-specific Network Instruments in an European Cohort of TC ("Therapeutic Community") Clients

Sébastien Reichmann
E-mail: reichman@iresco.fr
LASMAS - CNRS, Paris, France

Comparisons between the French sample (n=72) and the whole European sample (n=724) show similar distributions for several network items, when using a non-specific network instrument, EuropASI (i.e., the European version of the Addiction Severity Index). For example, at the question "With whom do you spend most of your free time?", 43% of the French sample (FS) vs 45% of the European sample (ES) answered drug/ alcohol user friends, about 10% of the FS vs 12% of the ES answered non-user friends, and 5.6% of the FS vs 5% of the ES answered family members drug/alcohol users. Moreover, when comparing the patterns of answers from the other European countries with the whole present working sample, several countries show patterns close to those of the French sample (Italy, Spain, Scotland), while other countries have a different pattern (clients from Sweden and Germany report lower percentages of friends with current alcohol or drug problems, while clients from Norway and Belgium report higher percentages of the same category of network members). On another hand, the network information collected with the MSNA (Maastricht Social Network Analysis), a specific network instrument, validates the network information from EuropASI concerning the size of the friends sector, for a majority of clients (i.e., clients who report spending their free time with drug/alcohol user friends). More specifically, from a network sampling prospective, for the same category of clients, there seems to be a trend of counting most of their user friends as "close friends".

Social Network Analysis Using Sample Data on Perceived Structure and Reported Composition of Personal Networks

Ove Frank
E-mail Ove.Frank@stat.su.se
Stockholm University Stockholm, Sweden

A social network is considered which is too large to observe in its entirety. The network consists of ties between individuals of k different categories. The number k is given but the categorical frequencies are unknown. The numbers of ties within and between the categories are also unknown. A sample of individuals is selected to report on the categorical composition and the perceived tie structure of their personal networks. Practical matters of observation and identifiability of network members are noted, and various ways of handling statistical inference are discussed. It is argued that in many practical situations there is a need for both a probabilistic sample selection modelling and a probabilistic network modelling. The ideas are illustrated by discussion of the statistical analysis of a large number of personal networks collected in a health survey on sexually transmitted diseases.

Human contact networks and the spread of airborne infectious diseases.

Jacco Wallinga
E-mail: Jacco.Wallinga@rivm.nl
National Institute of Public Health and the Environment, The Netherlands

Airborne infectious diseases spread through human contact networks from person to person by respiratory droplets. Many of these diseases, such as measles, mumps and rubella primarily affect children, which means that age is an important variable that indicates the risk for acquiring the disease. Therefore modelling studies of these diseases require a proper estimation of age-specific contact rates. The contact rates can be estimated directly if the relevant at-risk events for an infection can be defined and quantified. One such at-risk event might be a conversation, as if two people are close enough to converse with each other then they are probably close enough to pass on infections. In this paper we present observed age-specific conversation rates as reported by an age-stratified sample from a population. We explore the following questions: 1.What statistical model describes adequately the self-reported contact rates? 2.To what extend can the self-reported contact rates explain the observed age-specific immunity (as estimated by presence of specific antibodies determined in blood samples taken from the study population)? 3.What vaccination coverage is needed to reach elimination of the disease?

Estimation of Population Characteristics from One-wave Snowball Samples in Structured Populations

Tom A.B. Snijders
E-mail: t.a.b.snijders@ppsw.rug.nl
University of Groningen, The Netherlands

We consider a one-wave snowball sample in a population composed of individuals of two kinds, say, A and B (e.g., drugusers with, and without, a home). It is assumed that relation frequencies between A-A, A-B, and B-B individuals can be different, and that the initial sample has different sampling fractions in both subpopulations. We derive model-based and design-based estimators for population means as well as population totals of a quantitative variable in this population. This extends our earlier work on the estimation of the size of the population.

Partitioning of directed networks

Claudie Hassenforder-Chabriac, Louis Ferré, Bertrand Jouve
e-mail: jouve@univ-tlse2.fr
GRIMM, University of Toulouse 2, France

We present an efficient method for revealing the hidden structure of directed networks. An Euclidean distance d is define on the vertex set of the underlying abstract digraph D. That set is then isometrically embedded in Rn by positioning the vertices with respect to the distance d between them with the help of Principal Component Co-ordinates. Finally, that embedding is partitioned by iterative processes. The choice of the distance is fundamental since it influences the shape of the embedding. The square of the distance d we use is the half sum of the Czekanovsky-Dice coefficient respectively applied to the predecessors and successors boolean matrices. The choice of that coefficient is guided by previous results obtained by De Fraysseix and Kuntz on non-oriented graphs. The separation between predecessors and successors is a way to take into account the orientation in the expression of the distance. A hierarchical cluster analysis with Ward strategy may be performed to partitioned the points of the Euclidean embedding. That method is known to be very efficient for Euclidean distance table. But for large networks, a reduction of dimension may be suitable before partitioning, in order to lower the calculation time. To empirically evaluate the effect of such a reduction, hierarchical cluster analysis on the Euclidean distance table is compared with a similar analysis on a configuration of lower dimension. We present some results for different networks.

Pupils' networks in high schools

Chris Baerveldt
E-mail: c.baerveldt@fss.uu.nl
Utrecht University, The Netherlands

A study in Dutch high schools in 1987-1990 showed that the educational climate of the schools had little impact on the petty crime of their pupils. However, analysis of the impact of school friends, stressed the impact of pupils' networks: the spatial correlations between best friend relationships and petty crime were high. Because of the limits of the study a new study was conducted in 1995. All pupils of the third grade of an intermediate level of education in 23 schools filled in a written list about delinquent behaviour and eight different kinds of relationships with fellow pupils. The procedure was as follows: each pupil was given a personal code and a code list of all fellow pupils in the same year. The pupils filled in their own identification code on the form. For the network items all subjects were instructed to fill in the codes of the other pupils in question per item. In 1996, all pupils of the fourth grade in 19 of the 23 schools filled in the same delinquency and network items, thus providing two wave data of 990 pupils in 19 networks. The procedure has several important merits. One of the advantages is that it is possible to analyse on the dyad level, for instance symmetry in personal relationships can be asserted, and the association between individual behaviour and relationships can be analyzed directly. Another merit is that complete networks can be studied, for instance their density or segmentation can be compared. Furthermore, multi level hypotheses can be tested about the influence of network structure on processes on the dyad or individual level. However, there are also some new problems. The first problem is that the importance of pupil's networks can differ between pupils: some have their most important friends outside school, other see all their important peers within the same classroom. To control for this, new statistical procedures are needed. These procedures are scarce, as is the software to analyze effectively and efficiently.

Respondent-Driven Sampling

Steve Thompson
e-mail: skt@stat.psu.edu
The Pennsylvania State University, USA

With respondent-driven sampling designs initial respondents are used to find additional research participants. Methods such as coupon based recruitment can provide substantial cost savings over those in which investigators need to find the additional members of a hidden population themselves. In some situations, a portion of the hidden population my be accessible only through respondent-driven methods. In this talk models and statistical aspects of respondent-driven sampling will be discussed.

Space and Social Support Networks of the Poor: Parameters and Sampling Problems

Maarten Loopmans
e-mail: Maarten.Loopmans@geo.kuleuven.ac.be
University of Leuven Belgium

The Institute for Social and Economic Geography at the KULeuven studies social exclusion via Polanyi's model of economic integration. As geographers, we look for the spatial influences on the economic integration of the poor. Research focuses on the neighbourhood level. The importance of reciprocity to economic integration of the poor evokes our interest in social networks. What we try to measure in order to establish the influence of space on reciprocity is twofold: first, the influences of physical and social characteristics of space on personal network structure; second, the influences of network characteristics on reciprocal survival strategies. The first aim demands a network sampling method that accounts of meeting places and local social context. The second aim implies the study of social support networks. The poor's networks are seemingly different from those of other social strata, and they aim at different kinds of support. This affects our choice of network parameters. Interviewing the poor about their survival strategies implies some methodological difficulties, since many strategies are illegal or taboo. Interviewees are not easily found and are not inclined to reveal much about their social network. Revealing this hidden information means using qualitative sampling methods instead of representative sampling.

Combining cluster sampling and link-tracing sampling for estimating the size of hidden populations

Martin Felix-Medina
E-mail: martin@stat.psu.edu
The Pennsylvania State University, USA

In this talk we will present a sampling methodology, which is still in progress, for estimating the size of a hidden population, like drug users, within a region, like a city or a town. We will assume that we can construct a sampling frame of public places, like bars, parks, city blocks, etc., where members of the target population can be found. However, we will not assume that the sampling frame covers the whole population of interest. Using an ordinary cluster sampling a sample of sites is selected, and by means of ethnographic methodologies the members of the target population who ``belong'' to each sampled site are identified. Then for each sampled site we trace links between the place and the members of the target population who do not belong to that particular site. We estimate the size of the portion of the target population located in the region not covered by the sampling frame by a likelihood estimator. We propose to estimate the size of the other portion by some estimators which can be seen like combinations of design-based and model-based estimators. Variance estimators, which take into account the sampling design, are also proposed.

Measuring peer influence in adolescent smoking research; Past, present and future perspectives

Kremers, S.P.J., Mudde, A.N. & De Vries, H.
e-mail: s.kremers@gvo.unimaas.nl
University of Maastricht, The Netherlands

Research on adolescent smoking has consistently identified a strong relationship between the behaviour of peers and individual adolescent behaviour. These relations are strong according to behavioural science standards, and are replicated in many studies. As a consequence, peer influence is widely and popularly believed to be the major determinant of adolescent cigarette use. However, the importance of peer influence as a cause of adolescent smoking behaviour can be questioned. Four issues should be considered when interpreting measures of peer influence. The first issue is the psychological concept of projection. Second, the majority of influence studies have not separated the effects of friendship selection from the effects of peer influence. The third issue can be ascribed to a lack of consensus regarding operationalisation of the smoking initiation continuum. Fourth, current sociometric adolescent smoking studies do not describe the personal friendship network of an individual. Prominent methodological tools that have been applied in the past to disentangle the peer influence paradigm with regard to adolescent smoking are described. Then, shortcomings of methodological tools currently applied in innovative smoking research will be argued. Finally, the usefulness of personal network analysis is discussed as a tool in solving the current theoretical and practical issues of measuring peer influences in adolescent smoking research.

Sampling polydrug users in an urban area: design and practical problems

Moniek Coumans & Marinus Spreen
e-mail: ivoozl@worldonline.nl
Institute for Addiction Research (IVO) Rotterdam The Netherlands

The Heerlen Drug Monitoring System is a system in which data are continually collected on drugs, drug use and networks of drug users within a certain region and over a certain period of time. The system is based on three pillars: community field work, the network of (key) informants and a random network sample. One of the main questions in Heerlen focuses on an estimation of the size of the drug using population and is answered by the sample survey. Data gathering for the survey was performed by six trained (drug-using) interviewers. The main goal of the social network part of the questionnaire was an estimation of the size of the drugusing population, but while applying Social Network Analysis additional information can be gained about several levels of inference: individual, relationship and network. The sampling frame consisted of a list from the institute for addiction care (CAD) containing the visitors from a low threshold service, Heroin Prostitution Project and methadone program. Selected respondents (egós) were

asked to nominate other drug users (alters). From these nominees a more or less a-select sample was drawn containing drug users not-in–treatment. As well in the sampling-stage as in the stages of approaching the respondents and conducting the interviews, several practical and methodological problems emerged: quality of the sampling frame, reliability of the interviewers and accessibility and co-operation of the selected respondents. It could be worthwhile to discuss implications of these limitations for the interpretation of the results.

A power law in reports of knowledge of subpopulation sizes

Peter Killworth, Christopher McCarty, H. R. Bernard, Eugene Johnsen, Gene Shelley Chris McCarty
e-mail: chrism@bebr.ufl.edu
University of Florida USA

A new power law is described. It connects the size of some subpopulation (a group of people defined by some criterion, e.g. those who are diabetics, or those whose first name is Robert) with what respondents say about their knowledge of this subpopulation. Two replicable versions of the power law exits. The first is that the mean reported number of members of a subpopulation reported as known to respondents varies as the square root of the size of the subpopulation. The second is that the mean proportion of respondents who report knowing anyone in a subpopulation varies approximately as the cube root of the size of the subpopulation. These findings are hard to explain from existing psychological theory. We discuss the effects this finding has on network scale-up methods which assume that the mean number reported would vary linearly with subpopulation size rather than as a square root. Adaptations of previous theory lead to some improvements in estimated sizes of subpopulations (when these are known and can be compared).

Random Walks Through Large Social Networks: What the real world actually looks like.

Alden S Klovdahl
Email: alden.klovdahl@anu.edu.au
Australian National University

As with populations of individuals, real-world networks are often too large to observe other than by sampling. There are various forms of network sampling but here the focus is on random sampling from the relevant sampling frames of ostensibly visible populations. The procedure discussed is the random walk sampling design, which typically involves an initial random sample from a conventional frame (eg, an areal

frame), obtaining data on those (network associates) to who each initial person is linked, followed by – for each initial node – the random selection of one of these (network associates) to be the next network node visited on a random walk of the desired length through a large network. Some data from five studies carried out using a random walk design (in Canberra (Australia), in Atlanta (Georgia, USA) and in San Juan (Puerto Rico) are considered. Also discussed are some issues in determining how the 'real world' actually looks.

Network Snapshots Created by Epidemiologically-Relevant Interaction in Infectious Disease Outbreaks: Quantifying the role of different actors in outbreak networks

Alden S Klovdahl
Email: alden.klovdahl@anu.edu.au
Australian National University

Sampling – some procedure whereby a researcher goes out and actually selects or recruits respondents – is often seen as the only way to obtain information about large networks in society. In some instances, however, networks become visible as a result of individual activities and their consequences. When pathogens (bacterial, viral, etc.) are transmitted in the course of human interaction, for example, network structures may be exposed by routine disease surveillance and control activities. Here the focus is on networks and tuberculosis. It is suggested that the nature of tuberculosis and its transmission (airborne, by droplet nuclei) is such that ordinary social network concepts are not sufficient. Accordingly, the concept of an 'outbreak network' is introduced. It is further suggested that most currently available measures of the importance of different actors 'playing a role' in an outbreak, such as centrality measures, have serious limitations when used to study infectious diseases. Hence, the suspected infections transmitted (SIT) index is introduced. Some data on a tuberculosis outbreak, for which DNA 'fingerprints (RFLP/IS6110) were available are considered. Implications for observing networks in society and for disease control are discussed.