Health over the Life Course: A Chain Graph Model of Inter-relationships among Socio-demographic, Societal and Lifestyle Factors [1]

by

Fernando Rajulton and Jianlin Niu
Population Studies Centre
Department of Sociology
University of Western Ontario
London ON N6A 5C2
Canada.

Contact: email: fernando@uwo.ca
jniu4@uwo.ca
Tel: (519)661-3693
Fax : (519)661-3200

[1] This paper is an off-shoot of a major project titled "The Family, Community and Health in the Context of Economic Change", funded by Health Canada.

Rationale and Background of the study

Since 1980s, health scientists have become aware that the health care system is only one of the inputs that affect population health levels and that the origins of many health problems are related to societal structure itself and therefore cannot necessarily be addressed by traditional health care (Attinger, 1985; Evans, Barer and Marmor, 1994). In spite of this general awareness, health is still seen as a problem of individuals and rarely as a population issue. The individual-centred approach in health research is still predominant, based on the idea of "case-control method" to discover how sick and healthy individuals differ. This type of research looks for "risk factors" which identify individuals with certain characteristics, and perhaps with certain types of behaviour, as being more susceptible to ill-health and disease. The concept of "relative risk" is the basic representation of aetiological force in such research. Sometimes these risk factors are also considered to be "causes" without methodological justification.

As Rose (1985) argued, one might have a very good understanding of why individuals differ, for example, in hypertension or levels of blood cholesterol and yet miss the most important question why hypertension and high levels of cholesterol are absent in certain populations. Because "what distinguishes the two groups is nothing to do with the characteristics of individuals, it is rather a shift of the whole distribution - a mass influence acting on the population as a whole" (p.34). Thus, the population-centred approach emphasizes that any further reduction in morbidity (and/or mortality) and future improvement in the health of a population lies mainly in improving the environment *and* moderating self-imposed risks.

Both the individual and population-centred approaches to research on health have their own advantages and disadvantages. Our aim is to combine their strengths in such a way that the "multicausal" nature of heath status (Dean et al., 1995) can be brought out. In a multicausal framework, our goal is not so much to predict the statistical effects of specific variables on health status of individuals (although they are interesting) as to a) understand how the sociodemographic, societal and behavioral influences that affect health are distributed in a population or in segments of a population over the life course, and b) to examine how the interrelationships among these influences change over the life course. The so-called risk factors of health are certainly different over the life courses of individuals in different segments of a population. The risk factors that have significant (or no significant) impact on health at one stage of the life course may lose (or gain) their significance at a later stage. It is therefore important to understand how the sets of risk factors (or "causes") moderate and change the relationships between health status and individual behavior or lifestyles over the life course, so that population health programs can be meaningfully targeted towards people at various stages of the life course.

Many of the individual, as well as societal and environmental, risk factors that influence the health status of a population are often neatly grouped under an umbrella term: Lifestyle. Certainly, lifestyles are not the same for all people. Lifestyles are culturally and socially determined variables. Lifestyles are shaped not only by values, customs, norms and beliefs cherished in different cultural settings but also by opportunities and constraints defined by specific social and economic situations, many of which are gender specific. If the theoretical perspective that lifestyles are patterns of living shaped in social and situational settings is valid, we should expect interactions among behavior variables to differ for various subgroups of a society defined, for example, by gender, social status, and immigrant character. In addition, since these social and situational settings may change over the life course, it is all the more important to examine them over the life course. Thus, for example, the initial advantage immigrants have in their health status over the native-born population (because of selective screening procedures at the time of immigration) and the "deterioration" in their health status over time (because of assimilation and acculturation into local practices and lifestyles) are well known in Canada and elsewhere (see for example, Chen et al., 1995). In the light of these implications, what are commonly called "individual" characteristics and behaviors are more than what they are; they are social situations. This is the perspective that this study likes to emphasize. Thus, the term "gender" does not merely refer to the individual characteristic of sex (which does not change over the life course in normal circumstances!), but it refers much more to the social and cultural constraints and controls placed on persons of specific sex.

There are specific lifestyle behaviors or practices that have been known to be statistically associated with health status either of individuals or of a population, such as for example, the eating habits and the types of food consumed. A limited number of lifestyle behaviors such as tobacco or alcohol consumption, physical exercise and dietary practices have also been studied extensively in the health literature. Yet, the findings are often inconsistent and sometimes even contradictory, leading to a confused state of our understanding as to how influential these behaviors are *in the presence of each other* (in other words, in a multicausal framework mentioned above). These inconsistent and sometimes contradictory findings may be due to the fact that statistical correlations can often hide serious contradictions; and, it is incumbent on researchers to clarify this point explicitly in their research.

To understand the importance/relevance of the study presented in this paper, it might be useful to restate some of the main ideas implied in the above paragraphs: 1) The impact of all health-related characteristics and/or behaviors are both interactive and cumulative over the life course. The fact that some of them that appear to be very important (or not important) at one stage of the life course may lose (or gain) their importance at a later stage helps us in a way to select into a model only those behaviors or characteristics that have *lasting* impact on health over the life course and drop unnecessary elements from model building. 2) Models based on simple parametric structures do not allow for elaboration of conditional and moderating influences. Standard analytical procedures generally ignore these conditional influences and complex interactions inherent in causal processes. We can no longer be satisfied with the mechanisms that merely statistically "control" for other (causal) factors and thus implicate simplistic and naive cause and effect relationships. 3) Our goal is also to shift the emphasis from simply predicting statistical effects of specific factors to studying patterns of interrelationships among the structural, behavioral and health variables. This requires analytic methods that elaborate direct and indirect relationships among different types of variables.

To achieve these objectives, the technique of Graphical Interaction modeling (and its special case, Chain Graphs) comes in handy. This technique was brought to the attention of researchers only during the late 1980s and has not yet caught up with researchers mainly because of the fact that the application of the technique usually entails a heavy investment of time and energy just to explore and screen the underlying relationships among the variables included in a model. Lack of proper software for a specific research context may also be another reason, although many are already available to researchers and more effort is being directed to improving on them. The problem associated with lack of proper software and investment of time is acute particularly with categorical variables (see below for some details). The next section presents only the main ideas underlying this technique; interested readers should consult the relevant literature cited in the next section.

Chain Graph Model as a Multicausal Framework

The technique of Chain Graph modelling (Lauritzen and Wermuth, 1989; Wermuth and Lauritzen, 1990; Whittaker, 1990, 1993; Cox and Wermuth,1996; Wermuth,1993; Lauritzen and Richardson,2002; Edwards, 2000; Cox and Wermuth, 2004) enables us to use the mutlicausal framework proposed in this study to examine the impact of individual and social situations on health status. Since not many readers may be familiar with this methodology because of its recency, we present a brief outline of this methodology and some of its salient features that are relevant to this study.

The essential ideas of graphical interaction models come from the pioneering work of the geneticist Sewell

Wright (1921) who developed the now well-known technique of path analysis in the 1920s. However, the modern mathematical theory stems from the seminal paper of Darroch et al.(1980). The mathematical theory underpinning the technique of graphical modelling is probabilistic, particularly the notion of *conditional independence*. The graph, a quite handy tool in conveying statistical ideas as in path analysis, represents the pattern of multivariate associations and dependencies. In contrast to path analysis, the graphical models are concerned with partial correlations and partial covariance structures (in the case of interval and ratio measures) and partial associations (in the case of ordinal and nominal measures). The graphical models can also be considered as a subclass of log linear models (Vermunt and Georg, 2002), but built on the property of conditional independence to produce an independence graph.

As Whittaker (1993b) commented as a response to the paper by Cox and Wermuth (1993), “it is not so much the graphic display but the notion of conditional dependence and independence and the idea of a ternary relationship that X_1 affects (or is irrelevant to) X_2 *in the presence of* X_3, which constitutes the fundamental contribution of graphical models to statistical analysis” (p.273, italics mine). It is this basic idea that is of interest to the research objective of this study, because from among a multitude of factors that can influence the health status of individuals, it is a matter of commonsense and practical significance to select (and examine the interrelationships among) only those factors which clearly stand to explain health status, and drop all others which are unnecessary in the presence of the selected ones.

(Figure 1 about here)

To give a simple illustration, let us consider a hypothetical case of three variables: "health status" (Y), "smoking status" (X_1) and "employment status" (X_2). A research hypothesis might say that smoking and employment status affect health status. Such a simple hypothesis ignores a large number of ways in which this ternary relationship can be examined statistically. Some possible ways are illustrated in Figure 1, where the vertices represent variables (here, all three variables are assumed to be discrete)[2] and the connecting lines are called edges. These edges can be undirected (implying no cause-effect relationships) or directed (implying cause-effect relationship, and hence denoted by an arrow head). The first diagram A implies that X_1 affects X_2 which in turn affects Y, thus X_1 having an indirect relationship with Y. This is quite different from what is usually inferred from separate univariate regressions of X_1 on Y and X_2 on Y, shown in the second diagram B or from a multivariate regression of (X_1, X_2) on Y, shown in the third diagram C. Diagram A implies the notion of conditional independence: X_1 and Y are conditionally independent in the presence of X_2. This means that the information provided by X_1 is unnecessary as long as X_2 is there; it is X_2 which matters more in explaining Y. Such conditional independence statements are tested with statistical measures like gamma coefficients, chi-square values, partial correlation coefficients and edge exclusion deviance (Whittaker, 1993a:167), etc. depending on the measurement levels.

In graphic models, the idea of a "block regression" is used which is portrayed in the diagrams C through E. There are various ways in which the original hypothesis can be represented; what is shown in Figure 1 are some possible relationships. For instance, diagram D says that X_1 affects both X_2 and Y, which implies a conditional independence of Y and X_2 in the presence of X_1. In diagram C, X_1 and X_2 affect Y but there exists no relation between them. And, in the last diagram E, there is also an edge between X_1 and X_2 (without an arrow), thus implying a dependence between X_1 and X_2. In this case, the influence of X_1 (or X_2) on Y is modified by that of X_2 (or X_1) respectively, which s commonly understood as interaction effect.

Thus, the defining feature of a conditional independence graph is this: an edge connecting two variables is missing when they are independent *in the presence of a third variable*. As Whittaker (1993b:274) pointed out, “It is the *absence* of an edge which generates the graph. Admittedly this is a subtle point and choosing to visually represent a defining feature by a blank space is perhaps unfortunate”.The same idea is extended to the presence of several variables, in which case the defining relationship of the independence graph is that of pair-wise independence conditioned on *all* the remaining variables.

When several variables are examined in a research work, the notion of conditional independence implies some important theoretical properties, one of which is worth mentioning here: There is no better predictor for one variable from all the other variables than the ones that are its “nearest neighbours” in the graph. This point, which was explained when we discussed diagram A, should help avoid the confusion raised by ignoring the multicausal nature of health/disease status.

The different diagrams in Figure 1 address the same research hypothesis, namely a ternary relationship that exists between X_1, X_2 and Y. Yet, these illustrations clearly show how differently the same hypothesis can be examined and interpreted. The most important point that we learn from this is: It is theory and knowledge of the substantive domain that must guide the researcher using graphical models. The specification of the graphic structure (hence, the structure of influence of variables) will depend on the research question and the

---

[2] By convention in graph theory, discrete variables are represented by dots (bullets), while continuous

theoretical framework guiding the analysis.

The important advantages of the graphical model therefore lie in its ability a) to concentrate on meaningful relations among variables that constitute causation in the real world, and b) to move beyond the limitations of parametric models (or "relative risks") which are often plagued by undiagnosed and unrecognized problems of multicollinearity.

The above ideas are generalized to variables that are continuous or discrete, interval, ordinal or nominal, or any mixture of them. Variables can be seen either to embody a response-explanatory (or causal) structure or, for lack thereof, a simple "*equal footing*" structure. Most research objectives would call for a mixture of continuous and discrete variables that incorporate both response-explanatory and equal footing structures. This can be done by dividing the variables into two or more blocks, one of them containing only the set of response variables (for example, in this study, we consider health status and chronic illness), another containing the set of purely explanatory variables (for example, socioeconomic and social network variables) and the third containing the set of intermediate variables that are both explanatory and response variables (for example, lifestyle variables). In such cases, to reflect the different types of independence statements, lines are used for *intra-block undirected edges* (within both explanatory and response sets) and arrows are used for *interblock directed edges* (between response and explanatory sets).

When there are several blocks, the technique gives rise to what is known as chain graphs which describes the more complicated patterns of dependence between variables. In fact, a chain graph is not a mere statistical mode; it is (or must be) viewed as a "substantive research hypothesis" (Wermuth and Lauritzen, 1990) about direct and indirect relations among variables. An important difference of this technique from all other techniques commonly used in health research is that it enables us to investigate the strength, direction or lack of associations not only for the response variables but also for the explanatory variables. And, an important reason for examining the relations among explanatory variables is to identify any moderating or confounding influence among them. In the case of contingency tables, this is often referred to as the Simpson Paradox - a reversal in the direction of dependence when marginal distributions over other explanatory variables are taken into account. Such unexpected findings would point to either systematic errors in the data or to selection effects.

---

variables are represented by circles.

Chain graph models have been found useful in many empirical applications in diverse fields such as political science (Evans and Andersen, 2001), psychiatric epidemiology (Biggeri et al., 2001), life course research (Borgoni, 2004), health of the elderly (Didelez et al., 2002), sociological evaluation of graduate programs (Caputo et al., 1999), pathological disorders (Clelia and Biffi, 2004), and heart study (Klein et al., 1995). However, like any other technique, the chain graph should not be considered as a panacea for all analytical problems in research. There are many practical problems in the applications of this technique, such as, for example, number and choice of variables, choice of response and explanatory variables, model selection, model fitting and diagnostic procedures (see below for some details).What Wermuth (1993:201) said more than a decade ago still holds true. Although considerable amount of work on different aspects of models for multivariate dependencies and associations has been published, much more needs to be done. In particular, more empirical work will throw light on the usefulness of this model. This paper is a contribution towards that end and points out what can be done in the future.

To summarize before moving on to the next section, Graphical Chain Models provide a method for assigning a theoretically based structure to the analysis. In this study, socio-demographic and social situational variables are placed in the first block containing pure explanatory variables, followed by lifestyle variables in the second block containing both explanatory and response variables, with the third block containing two health variables considered as pure responses. As discussed in the previous section, the directional influence of variables may be time bound and may change over the life course, which adds more complexity to the type of analysis that can be done with longitudinal data (see for example, Borgoni et al., 2004). The strength of Chain Graphs will be in identifying varying interrelationships among the selected variables that have a lasting influence on the health status over the life course. This is what we shall aim in the following section.

Data and Analytical Framework

To achieve the objectives outlined in the last two sections, we need longitudinal data. Such data on health and life courses are available in Canada. The longitudinal National Population Health Survey (NPHS) conducted since 1994 has a wealth of information on health and other variables related to health. Data needed for analyzing the multicausal framework described above are available through access to confidential longitudinal data at the Statistics Canada Research Data Centre (RDC). So far, the NPHS has collected information over five waves (repeated every two years since 1994) on socio-demographic characteristics, lifestyle behaviors, life courses, and social support networks that we are using in this study. [For more specific details on sampling procedures, see Statistics Canada's NPHS Public Use Microdata Documentation.]

Since our main aim in doing this study is to examine health over the life course, we do separate analyses by age groups that slice the population into young adulthood, adulthood, middle age, and old age, that is, 20-34, 35-49, 50-64 and 65+ in 1994 when the first wave of the survey was done and follow these age groups over the five waves until 2002. Although by the fifth wave, some individuals in one age group might have moved to the next, we do not reclassify them into the higher age group, with the intention of following the individuals over the life course. This however may raise some problems for analysis because some questions asked of individuals of certain age in one wave were not asked of them again at higher ages, and thus giving rise to missing information at later waves. This is however not a serious problem since information obtained in earlier waves can always be used in the model to examine its long-term impact over the life course. And, although we have done our analyses for all the four age groups and our original intention was to present the results for all the four age groups, for lack of space and for simplicity of presentation, we provide the results in this paper only for two age groups 35-49 and 50-64, ages at which most individuals experience a change in their health status.

As discussed earlier, gender is one of the most fundamental determinants of life situational and lifestyle differences, and hence of the life course analyses (McMullin, 1995). All analyses are therefore conducted separately for men and women in order to increase the validity and meaningfulness of the findings. We have thus four groups for comparison, and their longitudinal (weighted) sample sizes are as follows:

	Age group in 1994	
Sex	35-49	50-64
Men	1332	900
Women	1543	1103

We include in our analysis another fundamental influence on lifestyle, namely social status. Many health studies repeatedly confirm that social status is negatively associated with morbidity and mortality in all countries. However, the nature of social status inequalities in health is not yet clearly understood except the fact that they do exist. We shall examine this issue by using the *education* and *income adequacy* components of social status. The education variable (denoted by *edu*) has four categories: less than secondary, secondary, post-secondary and college or university degree, the last used as the reference category. The variable “income adequacy” (denoted by *inc*) is derived from the household income adjusted for household size. It has been reclassified into three categories: lower than middle, middle, higher than middle, the last serving as the reference category. Basically, “lower than middle” denotes less than $15000 for households with one or two persons, less than $20000 for households with 3 or 4 persons, and less than $30000 for households with 5 or more persons. "Higher than middle" denotes $30000 or more for households with one or two persons, $40000 for households with 3 or 4 persons and $50000 or more for households with 5 or more persons. And, the "middle" denotes $15000 - 29999 for households with one or two persons, $20000 - 39999 for households with 3 or 4 persons, and $30000 - 59999 for households with 5 or more persons.

Mortality research indicates that absence of social networks (lack of social and individual ties) is associated with an excess mortality for both men and women (see for example, Kawachi et al., 1996; Iwasaki et al, 2002; Melchior et al., 2003). This association has been found to be independent of the self-reported physical health and socioeconomic status, as well as of health practices such as smoking, alcoholic beverage consumption, physical activity, utilization of preventive health services or a cumulative index of health practices. In the Seattle Longitudinal Study on the relationship between social environment, social networks, and health outcomes, Bosworth and Schaie (1997) found lower levels of perceived social environment and social networks were associated with increased number of health problems and hospital visits. See also Pescosolido and Levy (2002) for more ideas on social networks and health. We hypothesize that a similar association can be found also with health status and chronic illness and that it can act through lifestyle variables. The NPHS has collected information on the social involvement of respondents (denoted by *ssi*), measured by two items that reflect the frequency of participation in associations and voluntary organizations and frequency of attendance at religious services. This information was collected however only in waves 1 and 2. A score is attached to each individual, ranging from 0 to 6, higher scores denoting greater involvement. We have recoded this score into three categories: No for score 0, Low for scores from 1 to 4, and High for scores 5+, the last used as the reference.

In addition to social involvement, we also use marital status of respondents (denoted by *mar)* as a proxy for social networks, because married persons have been found to enjoy better health than others not only because of their own interpersonal relationships but also because of the social networks these relationships automatically bring into their lives. Given the dramatic changes in the roles of marriage and the family in developed societies like Canada, we have included this variable to examine its impact on health status over the life course. The variable has three categories: married/common-law/partner, single, and widowed/separated/divorced, the last used as the reference.

Two variables that describe personal life-time experiences that can have serious consequences on health at a later time are also included in our analysis. The survey collected information on recent life events (denoted by *rle*), the term representing "negative events" experienced by the respondent or by someone close to the respondent in the past 12 months before the interview, such as physical abuse, unwanted pregnancy, abortion or miscarriage, major financial crisis, and problems at work/school. A score ranging from 0 to 6 is given for the number of such events experienced by individuals. We have recoded the values into yes-no format for simplicity, the yes category as the reference.

In addition to the "negative events", the survey also collected information on "traumatic events" experienced during childhood, adolescence or adulthood, such as divorce, unemployment, drug abuse, sexual abuse, etc. This also has a score ranging from 0 to 6, which again we have recoded into yes-no format, again the yes category serving as the reference (and denote it by *cas*, standing for childhood-adult stressors). Information on these two variables - recent life events, and childhood-adult stressors - was not collected in all the waves; the former only in waves 1 and 4, and the latter only in wave 1.

Three specific lifestyle behaviors, namely smoking, drinking alcoholic beverages, and body mass index (denoted by *smo*, *dri*, and *bmi* respectively), have been included in our analysis in order to highlight the main and interactive impact of these habits on health over the life course. As much as possible, we consider not merely the incidence of these behavioral habits (in yes-no format) but the range of these practices whenever measured. The variable "smoker type" has the following categories: daily smoker, occasional smoker, former smoker, and never smoked, the last serving as the reference. Similarly, the variable "drinker type has the following categories: regular drinker, occasional drinker, former drinker, never drank. The variable "body mass index" was calculated for persons 20 to 64 years old, excluding pregnant women. This variable has been recoded into 4 categories: Underweight, normal weight, overweight and obese, the last serving as the reference.

Two health variables are included in the analysis, namely health utility index and chronic conditions (denoted by *hui* and *chr* respectively).The latter was measured in terms of the number of chronic conditions (that is, conditions that have lasted or are expected to last 6 months or more) such as food allergies, asthma, arthritis or rheumatism, back problems, high blood pressure, migraine, bronchitis or emphysema, diabetes, epilepsy, heart disease, cancer, stomach or intestinal ulcers, stroke, and urinary incontinence. It is a highly positively skewed distribution, with most of the respondents having no chronic conditions. Therefore, the variable has been simply recoded into yes-no format, the former as the reference.

The health utility index (HUI) or health status index, developed at McMaster University's Centre for Health

Economics and Policy Analysis, synthesizes both quantitative and qualitative aspects of health and describes an individual's functional health, based on eight attributes: vision, hearing, speech, ability to get around, dexterity (use of hands and fingers), cognition (memory and thinking), emotions, and pain and discomfort. The HUI is a single numerical value for any possible combination of levels of these eight attributes, and ranges from -0.36 to 1. For example, an individual who is near-sighted but fully healthy on all the other seven attributes will have a score of 0.95. The index can also take negative values, these negative values being interpreted as "health that is worse than death". Examining the NPHS data tells us that about 1% of Canadians have negative scores on HUI. They have been retained in the analysis. Since most individuals have high scores on HUI, after checking its distribution, the HUI is treated as an ordinal variable with the following categories: Scores less than 0.8, 0.8 to 0.8999, 0.9 to 0.9499, and 0.95+, the last as the reference category. For a detailed explanation of the HUI, see Berthelot, Roberge & Wolfson, 1993:161-72.[3]

In our preliminary analyses and in screening procedures used for the chain graph model, we considered many other variables usually found in the health literature such as occupation, immigrant status, working status, social support, depression score, self-esteem score, and mastery score. All these were dropped in later analyses not only for the sake of simplifying the model but also for their small or no contribution to the betterment of the model. Besides, the rule of parsimony is an absolute must in chain graph modeling, particularly for examining the direct and indirect effects of the variables in the model. Table 1 presents the variables used in the Chain Graph model.

(Table 1 about here)

Finally, a word about the software available for building the chain graph models. All the variables used in this study are categorical, and most of them are multinomous, which raises very specific problems in building the chain graph models. Among the software currently available to researchers, we tried with MIM (Edwards, 1987, 2000), DIGRAM (Discrete Graphical Modeling by Kreiner, 1987, 1992, 2003) and GRAPHFITI (Blauth, 2000, 2002). For a comparative description of these and other software, see Blauth and Pigeot (date unknown). The number and the type of variables are very crucial for using any software for chain graph models. Initial screening procedures with the 11 variables selected for analysis were running for hours, sometimes even for 12 hours, and at the end of it we received messages such as "out of memory". One specific

[3] Like many other health surveys, the NPHS also has measured the "perceived health status" of individuals. Most health studies have used this variable. Although we used this variable in our preliminary analyses, we do not find it as good a measure as HUI and for the sake of simplifying the model do not present it in this study.

problem associated with the data set used in this study (which many studies done so far do not seem to be concerned with) makes it highly desirable to address it in future versions of the software. That is the problem of sampling weights. Like all other Canadian national survey data sets, the NPHS data also have sampling weights associated with each individual, which should be used for generalization of results to the Canadian population. Except for MIM, no other software allows sampling weights, but MIM also could not handle it and ran out of memory after running for 12 hours! Because of all these problems, we finally decided to do the analysis piece-meal and stepwise, through log-linear models for checking on conditional independencies within blocks (undirected edges) and through multinomous or binary logistic regression models for estimating the impact of variables from one block to another (directed edges).

Since the models are built in this study separately for males and females and for two age groups and since the multinomous nature of many variables makes presentations of direct and indirect effects rather cumbersome, we finally decided to restrict the presentation here to the following (although we analyzed the data from all the five waves): a) conditional independence graphs within blocks for wave 1 only; b) directed edges from explanatory variables in wave 1 only to health variables in waves 1, 3 and 5. Why we have taken this decision will become clear with the preliminary results presented in the next section.

Results

a) Changes in sociodemographic, lifestyle and health variables over waves

(Figure 2 about here)

With the longitudinal information provided by the NPHS on health status and other relevant variables, it is useful to check the changes in the selected variables over time before going for any type of analysis. Figure 2 presents these changes over the five waves from 1994 to 2002. Some interesting observations from Figure 2 are:

a) The separated/divorced/widowed category among women aged 35-49 in 1994 has almost doubled over these eight years, from about 12% to 20%, but not for men. The trend continues with the older age group 50-64 in 1994, where the proportion of women separated/divorced/widowed increases from 20% to 30%, while for men the proportion increases from to 10% to 15% only..
b) Educational attainment shows no change over time, since only a few of these individuals aged 35-49 or 50-64 in 1994 go for further schooling.
c) Income adequacy increases over the life course, both for men and women aged 35-49 in 1994; the above middle category increases from 60% to 80%. But the proportion falling into this category decreases for older women aged 50-64 in 1994 to around 50%; for men, however, there is still some

increase from 60% to 70%.

d) Social involvement, measured only for the first two waves, shows virtually no change for women and men of both the age groups, with about 50% of women and 45% of men falling into low involvement category. At least 20% of men and women of both age groups are highly involved in social activities, with older women even more participating in social activities (30%) than younger women.

e) Recent life events, measured only at waves 1 and 4, show a steady drop for both men and women as they get older; the older the individual, the less is the proportion reporting negative events in their lives.

f) Childhood-adult stressors, measured only in 1994, show a larger proportion of women than men experiencing them.

g) Changes in body mass index over time are interesting (or disturbing?). There is a steady decline in the proportion of men and women having normal weight, and a steady increase in the proportions of men and women becoming overweight or obese over time. Most men (about 50%) are consistently overweight at all time points, while most women are of normal weight, but the proportion steadily declines from 50% to 40%, nearing the same proportion overweight among younger women. The same proportion overweight holds steady among older women as well. Overall, about 20% of men and women fall into the obese category irrespective of their age.

h) The daily smoker category for both men and women generally decreases over time and the former smoker category increases over time, somewhat spectacularly for men, steadily increasing over age groups as well - from 30% to 50% among younger men and from 50% to 65% among older men.

i) In contrast, the regular drinker category is the predominant one among women and men, with about 50% of women and 75% of men, irrespective of their age, being regular drinkers all through the five waves.

j) Chronic conditions are more prevalent among women, and they steadily increase over time, from about 50% to 70% for younger women, increasing further to 85% among older women. Men also experience more and more chronic conditions over age, increasing from 45% to 80%.

h) Finally, the health status index shows a non-linear pattern of change in almost all categories, especially in the very healthy category (0.95+) over time for the age group 35-49. It is heartening to see that about 50-60% of women and men in the younger age group fall into the highest health category, which not surprisingly drops to 40-50% among men and women in the older age group.

These observations on changes over time clearly tell us that health-related variables over the life course are different for men and women and for different age groups, and that certain variables will be more important for women than for men.

b) Conditional dependencies and independencies

Figure 3 portrays the conditional dependencies and independencies (or undirected edges) among the variables within each block for males and females and for age groups 35-49 and 50-64 in 1994. The results obtained from the partial associations through loglinear analyses of the variables have been used to plot these independence graphs. As these independence graphs show, the three lifestyle and two health variables are all conditionally dependent, which holds true over all waves. Striking differences between men and women, and between the two age groups, are to be seen only in the first block of six socioeconomic and social situational variables. It might have been interesting to see how this *pattern of conditional relationships* among the six variables would have changed from wave to wave, but unfortunately, as mentioned earlier, information on three of these six variables (namely, recent life events, childhood-adult-stressors, and social involvement) was not gathered in all the waves. The relationships among the remaining three variables (education, marital status, and income adequacy) can be expected to persist over time.

(Figure 3 about here)

For the age group 35-49 in 1994, marital status is important for women in the sense that it is conditionally associated with other variables in the block, but not for men in whose case marital status is conditionally independent of all other variables in the block. Social involvement, however, is completely conditionally independent of all other variables in the block for both men and women aged 35-49 in 1994. Other differences between men and women aged 35-49 in 1994 include, for example, the conditional independence between education and childhood-adult stressors among women but not so among men. It would be worth seeing whether and how these gender differentials play out in their long-term impact on men's and women's health status.

For the age group 50-64 in 1994, the independence graph is slightly different from that for the younger age group, some noteworthy differences being a) marital status becomes important for men in their pre-retirement ages; b) social involvement becomes conditionally associated with education in the case of women but it is still conditionally independent of all other variables in the case of men; c) education and childhood-adult stressors become conditionally independent among men too.

c) Directed edges through logistic regression models

(Table 2 about here)

As a preliminary step, the significant main and interaction effects of the explanatory variables in the first two blocks on the response variables in the second and third blocks *within each wave* were examined through logistic regression models. These logistic regressions can be binary or multinomous, depending on the type of response variables. Only those variables that have significant effects on the response variables are shown in Table 2. As can be seen in this Table, some explanatory variables do lose their significance over time, and only very few stand out very clearly. For example, consider the life style response variable smoking type. Although for both men and women aged 35-49 in 1994 (see the first two panels in the Table), education, income adequacy, social involvement, marital status and two interactions have significant impact on smoking type in wave 1, just two variables (namely, marital status and income adequacy) maintain their significant impact over time. For the same response variable among men and women aged 50-64 in wave 1, only income adequacy stands out as a lasting and significant explanatory variable. As for the ultimate health outcomes - chronic conditions and health utility index -, a different story unfolds. In the "younger" age group 35-49, both the socioeconomic and lifestyle variables have their significant impact over time. But in the "older" age groups, health variables are more directly affected by life style variables than by socioeconomic variables.

(Table 3 about here)

It would be a cumbersome task to display all the directed edges from the results shown in Table 2. Instead, as mentioned earlier, for the sake of parsimony and easy interpretation, only the directed edges from the first two blocks of socioeconomic and lifestyle variables in wave 1 to health blocks in waves 1, 3 and 5 are presented in Figures 4 and 5. The directed edges shown in these figures are based on the *significant* logistic regression coefficients presented in Table 3. It should be emphasized here that the causal structure obtained through logistic regressions in Table 3 and presented graphically in Figures 4 and 5 has been obtained through stringent application of Bonferroni inequality; that is, the *p*-values associated with the significance of these coefficients are different from model to model depending on the number of variables used in the model.

(Figures 4 and 5 about here)

As Figures 4 and 5 show, the health variables themselves have a very neat and clear pattern of effects from one time point to another, earlier health status affecting later chronic conditions and earlier chronic conditions affecting later health status. The graphic causal structure obtained for males is quite simple and straightforward. Among the socioeconomic variables, only income adequacy has far reaching impact on the health status of men in both age groups, 4 years later for the older age group and eight years later for the younger age group. In contrast, all three lifestyle variables have their impact on chronic conditions as well as on health status. Body mass index and drinker type, both in 1994, have significant impact on health status four years later among men aged 35-49 in 1994, while their smoking behavior in 1994 has significant impact on their chronic conditions eight years later. Among men aged 50-64 in 1994, their body mass index in 1994 seems to have significant impact on their chronic conditions eight years later, while their drinking habits have significant impact on health status as well as on chronic conditions.

For women, however, the picture gets more complicated. In contrast to men's, women's health variables seem to be affected not so much by life style variables as by socioeconomic and social situational variables. For the younger women aged 35-49 in 1994, all the socioeconomic variables except education seem to have lasting impact on either their health status or chronic conditions; in contrast, only their body mass index in 1994 has lasting impact on their chronic conditions. It is particularly noteworthy that not only women's marital status and childhood-adult-stressors have their impact on their chronic conditions later in life, but their social involvement and recent (negative) life events also have impact on their health status in later life. The impacts of socioeconomic and social situational variables are all the more accentuated in the case of older women aged 50-64 in 1994. In addition, two lifestyle variables, smoking and drinking come out with their significant effect for these women.

Apart from the overall and general causal structure that emerges in Figures 4 and 5, it may be worth examining the magnitude of the impact of these variables over different waves. The associated coefficients from the logistic regressions are presented in Table 3. Judging from the goodness-of-fit statistics (GOF) and Nagelkerke's pseudo-R^2 values given in this table, most models are good fits and explain 20% or more of variation in the response variables, in spite of using only a handful of variables in the model. The coefficients are to be interpreted in the same manner as with any logistic regression coefficient: a positive coefficient implies a greater likelihood of falling into a specific response category and a negative coefficient a smaller likelihood, both in comparison to the reference categories. We shall leave this exercise to those interested. Instead what the authors would like to emphasize here is the importance and greater relevance of examining both the direct and indirect effects revealed in the causal structures shown in Figures 4 and 5. For instance, consider women aged 35-49 in 1994. The "negative events" experienced by these women have not only direct effect on their health status four years later, but also have indirect effects through childhood-adult stressors on their chronic conditions four years later. A similar picture also emerges in the case of women aged 50-64 in 1994. All these insights will be completely missed if the graphic causal structure is not used in one's investigation. Such insights are not wanting in the case of men, for whom lifestyle variables are more important than socioeconomic variables, and we could always draw the directed edges from the socioeconomic block to lifestyle block to obtain similar insights (not shown here).

In the context of discussing indirect effects, one specific point needs researchers' (specifically computer programmers') attention in the future. While it is much easier to compute relevant indirect effects in the case of interval measures (as done in path analysis), it is not that easy to compute them in the case of ordinal or nominal measures, especially with multinomous categories. Some break through has been made towards this end in recent times. Eshima and Tabata (1999) and Eshima et al. (2001) provide good illustrations of effect analysis of recursive causal systems of categorical variables; they have contributed some formulations towards computing indirect effects with the logistic regression coefficients, although for dichotomous categories. We are planning to work further on this and write a program that will be useful in analyses of the type presented in this paper.

Discussion

Health over the life course is a study of complex set of dependencies and independencies among variables that are thought to have impact on individual health. It is worth studying the *patterns and distributions* of such dependencies in a population or in comparative segments of a population, either for a specific point in time or various points in time as done in this study. Without a clear understanding of these distributions, the long-term impact of these variables on health cannot be adequately understood; much worse, only inconsistent and often contradictory findings will be the outcome. This study has focused attention on this specific point - to bring out the long-term or lasting effects of a handful of variables normally postulated to affect the health status of individuals.

For simplicity of presentation, this study considered only two major segments of the Canadian population defined by age and gender. Other major segments of the population can be considered in the future. In particular, a fundamental aspect of lifestyle differences in Canada is the immigrant character of individuals. Immigration procedures in Canada select healthy individuals for entry into Canada, and thus the more recent immigrants are less of a burden on the country's health care. However, as studies have revealed, the immigrants who have lived in Canada for more than ten years resemble the Canadian-born in their health conditions and health-related behaviors (Chen et al., 1995). The determinants that shape the health status of recent immigrants need to be studied along with those that shape the health status of long-term immigrants who have adapted themselves to the lifestyles and customs of native Canadians. Although we tried to use this variable in our study, the number of cases is small for any disaggregation of the sample. One may need to have access to the Longitudinal Survey of Immigrants into Canada for a similar study.

This study on health over the life course has highlighted some important points for consideration either for future research or for framing health policies. Recent studies have delved into the importance of the role of social involvement and social networks in health status of men and women (see for example the references cited in an earlier section). Our study, making use of the graphical causal structure confirms this hypothesis, especially for women. What is an interesting, and important, insight from this study is that social involvement stands alone on its own right. It doesn't get "mixed up" with other socioeconomic variables, as shown in the independence graph. It exercises an independent influence on health, having a longer time effect. As observed in Table 3, its effect has been found to be significant for women only because we have used a rather stringent criterion of *p*-value based on Bonferroni inequality. Otherwise, the same inference should hold for men as well.

The longitudinal change in body mass index among both men and women, as pointed out earlier, is an interesting, yet somewhat disturbing, phenomenon. A constant proportion of Canadians (about 25%, of all ages together from the NPHS data) are obese, and an additional 50% become overweight, over their life course. That obesity is strongly influencing the health status as well as chronic conditions over the life course is not a new message, but this message needs to be communicated to the general populace. How acute the obesity problem will become with future generations of Canadians needs to be examined carefully in future studies, given the fact that 58% of men and 36% of women in the young adult age group 20-34 in 1994 (not presented in this paper) are already overweight or obese by the fifth wave in 2002. As the section on changes in body mass index reported, women seem to slip into overweight and obesity only during the adulthood and older ages. Our study indicates that obesity will be a serious health problem in the future, being especially associated with chronic conditions in adulthood and middle age.

The conditional dependencies and independencies as well as the chain graphs gave some insights into the way the selected variables influence each other as well as the health status over the life courses of men and women examined in this paper. As was postulated in the introductory section of this paper, it is clear that not all variables considered as "health-related" in the literature retain their influence over the life course changes of individuals. If we concentrate our attention only on the ultimate health outcomes, we discovered that while both the socioeconomic and lifestyle variables exercise their significant impact on health over time, it happens so only for the adult age groups (35-49 in 1994). In contrast, socioeconomic variables lose their significant impact on health over time for the older age groups (50-64 in 1994), in whose case lifestyle variables gain more prominence over time. The technique of chain graph modeling brings this out clearly by grouping the variables into blocks and studying the inter-relationships among the variables *within blocks as well as between* *blocks*. Here lies the power of this technique. The “causal structure” brought out by chain graph technique is not to be seen in most other studies which simply lump together all variables and examine them all in a “multivariate” framework. Thus, for example, one often reads statements like: “Among middle-aged adults aged 45 to 64, socioeconomic characteristics such as the education level and the household income are more important determinants of healthy aging than healthy behaviors” (Statistics Canada’s Daily, May 9, 2005). Results from the chain graph modeling on the other hand clearly show that such a statement is true only for women in that age bracket, not for men whose health status depends more on lifestyle variables, provided we ignore all intermediate changes in these behavioral variables over the life course. Something more than a mere multivariate framework is needed for studying “genuine” effects or causal structures.

However, as was remarked in an earlier section, the technique itself should not be considered as a panacea for all analytical problems in research. While conducting this study, we encountered many practical problems in the selection of the number of variables, choice of categories used in the response and explanatory variables, screening and diagnostic procedures, and in model fitting. And, not the least, the problem of dealing with sample weights and the problem of interpreting the “causal” effects in terms of direct and indirect effects in the context of categorical variables. Much work still needs to be done regarding these practical problems, and only more empirical work will throw light on the usefulness of this model for examining the causal structure of health related variables.

References

Attinger, E.O., 1985. Dynamic modeling and health policy research. In World Health Organization, *Research Methods for Health Development, Methodological Foundations for Research on the Determinants of Health Development*. Pp.56-73.

Berthelot, J., R. Roberge, & M. Wolfson. 1993. The calculation of health-adjusted life expectancy for a Canadian province using a multi-attribute utility function: a first attempt. Montpellier, France: Colloque Inserm/John Libbey Eurotext Ltd.

Biggeri, A., E. Stanghellini, and M. Ruggeri. 2001. Regression analysis of multiple outcomes. *Epidemiological Preview*, 25(2): 83-89.

Blauth A. and I. Pigeot. Date unknown. GRAPHFITI - A computer program for graphical chain models. Downloadable from http://www.billingpreis.mpg.de/hbp00/blauth.pdf.

Blauth, A., I. Pigeot, and F. Bry. 2000. Interactive analysis of high-dimensional association structures with graphical models. *Metrika*, 51: 53-65.

Borgoni, R., A.M. Berrington and P.W.F. Smith, 2004. Selecting and fitting graphical chain models to longitudinal data. Social Statistics Division and Southampton Statistical Sciences Research Institute (S^3RI) Methodology Working Paper M04/05

Bosworth H.B. and K.W. Schaie. 1997. The relationship of social environment, social networks, and health outcomes in the Seattle Longitudinal Study: two analytical approaches. Journals of Gerontology B Psychological Science and Social Science. 1997 52(5):p197-205.

Caputo A., A. Heinicke, and I. Pigeot. 1999. A graphical chain model derived from a model selection strategy for the Sociologists Graduates Study. *Biometrical Journal*, 41(2): 217-234.

Chen, J., E. Ng and R.Wilkins, 1995. The health of Canada's immigrants in 1994-95. In Statistics Canada, Health Reports,7(4)33-45.

Clelia, D.S. and S. Biffi. 2004. Graphical chain models and mental disorders: An application to pathological gambling. *Biometrical Journal*, 46(2): 273-383.

Cox, D.R. and N.Wermuth, 1993. Linear dependencies represented by chain graphs. *Statistical Science*, 8(3)204-283. This includes comments on graphical models, causality and intervention from page 266 onwards.

Cox, D.R. and N. Wermuth. 2001. Some statistical aspects of causality. *European Sociological Review*, 17(1): 65-74

Darroch, J.N., S.L. Lauritzen, and T.P. Speed. 1980. Markov fields and log-linear models for contingency tables. *The Annals of Statistics*, 8: 522-539.

Dean, K., C. Colomer and S. Perez-Hoyos, 1995. Research on lifestyles and health: Searching for meaning.

Social Science Medicine, 41(6):845-855.

Didelez, V., I. Pigeot, K. Dean, and A. Wister. 2002. *Biometrical Journal*, 44(4): 410-432.

Edwards, D. 2000. *Introduction to Graphical Modelling.* 2nd edition, N.Y: Springer.

Eshima, N. and M. Tabata. 1999. Effect analysis in loglinear model approach to path analysis of categorical variables. *Behaviormetrika*, 26(2): 221-233.

Eshima, N., M. Tabata, and G. Zhi. 2001. Path analysis with logistic regression models: Effect analysis of fully recursive causal systems of categorical variables. *Journal of Japan Statistical Society*, 31(1): 1-14.

Evans, R. G., M. L. Barer, and T. R. Marmor, 1994. *Why Are Some People Healthy and Others Not? The Determinants of Health of Populations*. New York: Walter de Gruyter, Inc.

Evans, G. and R. Andersen. 2001. Endogenizing the economy: Political preferences and economic perceptions across the electoral cycle. Working paper No. 88, September. Centre for Research into Elections and Social Trends (CREST). Department of Sociology, University of Oxford. U.K.

Hagey, J., W.Millar, M.P.Beaudet and S.Hood, 1995. The national population health survey: what have we learned? An overview of preliminary findings. In *Towards the XXIst Century: Emerging Socio-demographic Trends and Policy Issues in Canada*, Proceedings of the 1995 Symposium organized by the Federation of Canadian demographers. St. Paul University, Ottawa, pp.103-110.

Iwasaki, M., T. Otani, R. Sunaga, H. Miyazaki, L. Xiao, N. Wang, S. Yosiaki, & S. Suzuki. 2002. Social networks and mortality based on the Komo-Ise cohort study in Japan. *International Journal of Epidemiology*, 31:1208-1218.

Kawachi, I., G.A. Colditz, A. Ascherio, E.B. Rimm, E. Giovannucci, M.J. Stampfer, & W.C. Willett. 1996. A prospective study of social networks in relation to total mortality and cardiovascular disease in men in the USA. *Journal of Epidemiology and Community Health*, 50, 245-251.

Klein, J.P., N. Keiding, and S. Kreiner. 1995. Graphical models for panel studies, illustrated on data from the Framingham Heart Study. *Statistics in Medicine*, 14(12):1265-1290.

Kreiner, S. 1987. Analysis of multidimensional contingency tables by exact conditional tests: techniques and strategies. *Scandinavian Journal of Statistics*, 14: 97-112.

Kreiner, S. 1992. Notes on DIGRAM, Version 2.10. The Danish Institute of Educational Research, Copenhagen, Denmark.

Kreiner, S. 2003. Introduction to DIGRAM. Department of Biostatistics, University of Copenhagen, Denmark.

Lauritzen, S.L. and T.S. Richardson. 2002. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B,* 64(3):321-361.

Lauritzen, S.L. and N. Wermuth, 1989. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, 17(1)31-57.

McMullin, J. 1995. "Theorizing Age and Gender Relations," in *Connecting Gender and Ageing: A Sociological Approach*, edited by Sarah Arber and Jay Ginn. Philadelphia: Open University Press.

Melchior, M., I. Niedhammer, L. F. Berkman, & M. Goldberg. 2003. Do psychosocial work factors and social relations exert independent effects on sickness absence? A six year prospective study of the GAZEL cohort. Journal of Epidemiology and Community Health, 57:285-293

Pescosolido, B. And J.A. Levy. 2002. *Social Networks and Health*. Series No.8 in Advances in Medical Sociology. London: Elsevier.

Roberge, R., J.M. Berthelot and M. Wolfson. 1995. The health utility index: Measuring health differences in Ontario by socioeconomic status. *Health reports,* Vol. 7, No.2, 25-32. Ottawa: Statistics Canada.

Rose, G., 1985. Sick individuals and sick populations. *International Journal of Epidemiology*, 14(1)32-38.

Vermunt, J.K. and W. Georg. 2002. Longitudinal data analysis using log-linear pathmodels with latent variables. *Metodologica de las Ciencias del Comportamiento*, 4(1): 1-17.

Wermuth, N. and S.L.Lauritzen, 1990. On substantive research hypotheses, conditional independence graphs and graphical chain models. *Journal of Royal Statistical Society, B*, 52(1)21-50. Also, discussion of this paper in the same issue, pp.51-72.

Wermuth, N., 1993. Association structures with few variables: Characteristics and examples. In Dean, K.(ed.) *Population Health Research, Linking Theory and Methods*, pp.181-203.

Whittaker, J., 1990. *Graphical Models in Applied Mathematical Multivariate Statistics*. John Wiley:N.Y.

Whittaker, J., 1993a. Graphical Interaction Models: a new approach for statistical modelling. In Dean, K.(ed.) *Population Health Research, Linking Theory and Methods*, pp.160-180.

Whittaker, J. 1993b. Comment. Response to Cox and Wermuth's paper "Linear dependencies represented by chain graphs". *Statistical Science*, 8(3)273-276.

Wright, S. 1921. Correlation and causation. *Journal of Agricultural Research*, 20: 162-177.

Table 1: Variables used in the Chain Graph Model

Block 1 - Socioeconomic

1) Marital Status - *mar*
1. married/common-law/partner
2. single
3. widowed/separated/divorced (Ref.)

2) Education - *edu*
1. < secondary
2. secondary grad
3. some post-sec
4. coll/univ degree (Ref.)

3) Income adequacy - *inc*
1. lower than middle
2. middle
3. higher than middle (Ref.)

4) Social Involvement - *ssi*
1. no
2. low (score 1~4)
3. high (score 5+) (Ref.)

5) Recent life events - *rle*
1. no recent life event
2. have recent life events (Ref.)

6) Childhood & Adult Stressors - *cas*
1. no childhood/adult stressor
2. have childhood/adult stressor (Ref.)

Block 2 - Lifestyle

1) Body Mass Index - *bmi*
1. underweight
2. normal weight
3. overweight
4. obese (reference) (Ref.)

2) Smoker type - *smo*
1. daily smoker
2. occasional smoker
3. former smoker
4. never smoked (Ref.)

3) Alcohol drinker type - *dri*
1. regular drinker
2. occasion drinker
3. non-drinker now
4. never drank (Ref.)

Block 3 - Health

1) Health Utility Index - *hui*
1. <0.8
2. 0.8~0.8999
3. 0.9~0.9499
4. >=0.95 (Ref.)

2) Chronic Conditions - *chr*
1. no
2. yes (Ref.)

Table 2: Main and Interaction effects from Logistic Regression of variables in the Chain Graph Models

a) Males 35-49 in 1994

Blocks	Response	Wave 1- 1994	Wave 3 - 1998	Wave 5 - 2002
Lifestyle variables	1) *bmi*	*rle*	*mar, inc*	*Inc*
	2) *smo*	*edu, inc, ssi, cas, mar, edu*inc, edu*cas*	*mar, inc*	*mar, inc*
	3) *dri*	*edu,, ssi, cas*	*inc*	*mar, inc*
Health variables	1) *chr*	*edu, cas, rle, ssi*	*inc, bmi, smo*	*Inc*
	2) *hui*	*rle, cas, edu, smo*	*inc,bmi, dri, smo, mar*	*mar, inc, bmi*

b) Females 35-49 in 1994

Blocks	Response	Wave 1- 1994	Wave 3 - 1998	Wave 5 - 2002
Lifestyle variables	1) *bmi*	*ssi*	*Nil*	*Nil*
	2) *smo*	*edu, inc, ssi, mar*inc, edu*inc*	*mar*	*mar, inc*
	3) *dri*	*edu, inc, rle, cas*	*mar, inc*	*mar, inc*
Health variables	1) *chr*	*rle, bmi, edu*	*mar, dri, bmi, bmi*dri*	*mar, bmi*
	2) *hui*	*inc, rle, cas, ssi, edu, phy*	*inc, bmi, dri, smo, mar*	*Mar, inc, bmi, smo*

c) Males 50-64 in 1994

Blocks	Response	Wave 1- 1994	Wave 3 - 1998	Wave 5 - 2002
Lifestyle variables	1) *bmi*	*edu, ssi*	*inc*	*mar, inc*
	2) *smo*	*rle, ssi, edu, mar*	*mar, inc*	*inc*
	3) *dri*	*edu*	*inc*	-
Health variables	1) *chr*	*cas, rle*	*bmi, dri*	*mar, bmi, smo*
	2) *hui*	*inc, cas, ssi, smo*	*inc, smo*	*mar, inc*

d) Females 50-64 in 1994

Blocks	Response	Wave 1- 1994	Wave 3 - 1998	Wave 5 - 2002
Lifestyle variables	1) *bmi*	*cas, ssi, rle*	*inc*	*Nil*
	2) *smo*	*edu, cas, ssi, edu*ssi*	*mar*	*Inc*
	3) *dri*	*mar, inc, cas*	*mar, inc*	*mar, inc*
Health variables	1) *chr*	*cas, bmi, bmi*dri*	*dri, bmi*	*smo, bmi*
	2) *hui*	*rle, ssi, bmi, smo, dri*	*mar, dri, smo*	*inc, bmi, smo, dri*

Table 3: Logistic regression coefficients for directed edges in the Chain Graph Model from wave 1 variables to wave 3 and wave 5 health variables, classified by age groups 35-39 and 50-64 in 1994 and by sex

N.B. chr4, chr8, chr2 refer to chronic conditions 1994, 1998 and 2002 respectively; similarly hui4, hui8 and hui2 refer to health utility index in 1994, 1998 and 2002 respectively.

Age group 35-49

Females N = 1543

Wave 1 | Wave 3

1) Response: chr8 # variables: 7 - mar, cas, ssi, dri, bmi, chr4, hui4

GOF = .59 R^2 = .34 Sig: .007

	mar1	mar2	cas1	chr4
∃	-0.73	-1.25	-0.37	-2.23
$\Phi_{\exists}$	.21	.31	.13	.14

2) Response: hui8 # variables: 10 - mar, edu, inc, rle, cas, ssi, bmi, dri, chr4, hui4

GOF = .45 R^2 = .33 Sig: .005

<0.8:	inc1	rle1	ssi1	bmi1
∃	0.80	-0.56	0.77	-1.76
$\Phi_{\exists}$	.28	.19	.26	.41

	bmi2	chr	hui1	hui2
∃	-1.03	-0.82	2.67	0.99
$\Phi_{\exists}$	.25	.20	.25	.29

0.8 - 0.8999:	chr	hui1
∃	-1.12	1.20
$\Phi_{\exists}$	.24	.30

0.9-0.9499::	mar2	rle1	hui1	hui3
∃	1.24	-0.55	1.35	0.72

Males N = 1332

Wave 1 | Wave 3

1) Response: chr8 # variables: 5 - edu, inc, cas, hui4, chr4

GOF = .23 R^2 = .24 Sig: .01

	chr4
∃	-1.73
$\Phi_{\exists}$	.13

2) Response: hui8 # variables: 5 - edu, rle, bmi, dri, hui4

GOF = .35 R^2 = .28 Sig: .01

<0.8:	bmi1	bmi3	Hui41	Hui42
∃	2.03	-0.79	3.58	1.25
$\Phi_{\exists}$	.56	.30	.30	.31

0.8 - 0.8999:	dri1	dri2	dri3	hui41	hui42
∃	-1.84	-1.83	-3.12	1.51	1.16
$\Phi_{\exists}$	.42	.54	.75	.34	.31

0.9-0.9499::	hui41	hui42	hui43
∃	1.56	1.01	0.94
$\Phi_{\exists}$	.25	.24	.18

$\Phi_\exists$	.37	.16	.23	.19

Waves 1 & 3 |Wave 5

3) Response: chr2 # variables: 7 - ssi, dri, bmi, chr4, chr8, hui4, hui8

GOF = .11 R^2 = .42 Sig: .007

	ssi41	bmi41	bmi42	chr4	hui41	chr8
$\exists$	-0.74	-1.44	-0.87	-0.95	0.86	-2.0
$\Phi_\exists$	.22	.31	.26	.17	.26	.16

4) Response: hui2 # variables: 7- edu, inc, ssi, chr4, hui4, chr8, hui8

GOF = .99 R^2 = .37 Sig: .007

<0.8:	inc42	ssi41	hui41	hui42	chr8
$\exists$	0.87	0.77	2.06	1.12	-0.70
$\Phi_\exists$	.20	.27	.24	.27	.22

	hui81	hui82	hui3
$\exists$	2.60	2.37	1.34
$\Phi_\exists$	.25	.30	.24

0.8 - 0.8999:	hui41	hui42	hui81	hui82	hui83
$\exists$	1.41	1.0	1.69	2.20	1.53
$\Phi_\exists$	.29	.31	.33	.34	.26

0.9-0.9499::	hui41	hui42	hui43	hui83
$\exists$	1.40	1.21	0.84	0.75
$\Phi_\exists$	.24	.23	.18	.20

Waves 1 & 3 |Wave 5

3) Response: chr2 # variables: 6 - cas, smo, dri, chr4, chr8, hui8

GOF = .24 R^2 = .34 Sig: .008

	smo2	chr4	chr8
$\exists$	0.95	-1.31	-1.50
$\Phi_\exists$	.36	.16	.15

4) Response: hui2 # variables: 4- inc, dri, hui4, hui8

GOF = .03 R^2 = .22 Sig: .012

<0.8:	Hui41	hui81	hui82
$\exists$	1.28	2.30	1.17
$\Phi_\exists$	.28	.28	.34

0.8 - 0.8999:	hui41	hui42	hui81	hui82
$\exists$	1.54	1.05	1.72	0.95
$\Phi_\exists$	.34	.35	.35	.28

0.9-0.9499::	inc2	hui43	hui82	hui83
$\exists$	-0.51	0.76	0.90	0.63
$\Phi_\exists$	.19	.18	.27	.18

Table 3: Contd. **Age group 50-64**

Females N = 1103

Wave 1 |Wave 3

1) Response: chr8 # variables: 6 - edu, bmi, smo, dri,, chr4, hui4

GOF = .13 R^2 = .32 Sig: .008

	chr4
∃	-2.05
$\Phi_∃$	.19

2) Response: hui8 # variables: 5 -mar, inc, rle, dri, hui4

GOF = .04 R^2 = .24 Sig: .01

<0.8:	inc1	hui41	hui42
∃	0.87	2.09	1.25
$\Phi_∃$	.26	.23	.31

0.8 - 0.8999:	hui41	hui42
∃	1.67	1.38
$\Phi_∃$	.33	.42

0.9-0.9499::	mar2	rle1	dri2	hui42
∃	-1.27	-0.46	-1.02	1.24
$\Phi_∃$	.50	.18	.33	.27

Waves 1 & 3 |Wave 5

3) Response: chr2 # variables: 8 - mar, edu, inc, rle, cas, bmi, chr4, chr8

GOF = .05 R^2 = .50 Sig: .006

	cas1	chr4	chr8
∃	-0.96	-1.66	-2.57

Males N = 900

Wave 1 |Wave 3

1) Response: chr8 # variables: 3 - dri, hui4, chr4

GOF = .03 R^2 = .24 Sig: .016

	dri41	chr4	hui41
∃	-1.21	-1.80	0.63
$\Phi_∃$	.48	.17	.24

2) Response: hui8 # variables: 5 - edu4, inc4, smo4, chr4, hui4

GOF = .96 R^2 = .28 Sig: .01

<0.8:	inc1	smo1	smo3	chr4	hui41	hui42
∃	1.24	0.94	0.93	-0.62	2.43	1.44
$\Phi_∃$	.32	.34	.31	.24	.28	.37

0.8 - 0.8999:	hui41	hui42
∃	1.35	1.95
$\Phi_∃$	.42	.43

0.9-0.9499::	chr4	hui41	hui42
∃	-0.62	1.17	1.00
$\Phi_∃$	.23	.29	.35

Waves 1 & 3 |Wave 5

3) Response: chr2 # variables: 6 - edu4, rle4, bmi4, hui4, chr4, chr8

GOF = .26 R^2 = .40 Sig: .008

	bmi2	chr4	chr8
∃	-1.09	-1.36	-1.75

$\Phi_{\exists}$	.28	.29	.27

4) Response: hui2 # variables: 7- mar, edu, rle, smo, chr4, hui4, hui8

GOF = .001 R^2 = .33 Sig: .007

<0.8:	rle	smo1	chr4	hui41	hui42
∃	-0.74	0.81	-0.87	1.21	1.30
$\Phi_{\exists}$	.20	.27	.23	.28	.33

	hui81	hui83
∃	1.93	0.86
$\Phi_{\exists}$	.25	.27

0.8 - 0.8999:	smo1	hui42	hui83
∃	0.91	1.54	1.32
$\Phi_{\exists}$	.33	.36	.30

0.9-0.9499::	edu43	chr4	hui41	hui42	hui43
∃	-0.72	-0.60	0.87	1.18	0.69
$\Phi_{\exists}$	.26	.20	.27	.32	.23

$\Phi_{\exists}$	.37	.30	.24

4) Response: hui2 # variables: 5- mar, rle, dri, hui4, edu4

GOF = .01 R^2 = .32 Sig: .01

<0.8:	Hui41	hui43	hui81
∃	1.80	0.84	2.05
$\Phi_{\exists}$	.30	.31	.31

0.8 - 0.8999:	dri41	hui42	hui43	hui82	hui83
∃	-1.49	1.71	1.13	1.93	1.17
$\Phi_{\exists}$	.52	.41	.38	.49	.35

0.9-0.9499::	hui81
∃	1.02
$\Phi_{\exists}$	.32

Figure 1: Direct and Indirect Relationships involved in a simple hypothesis involving a ternary relationship

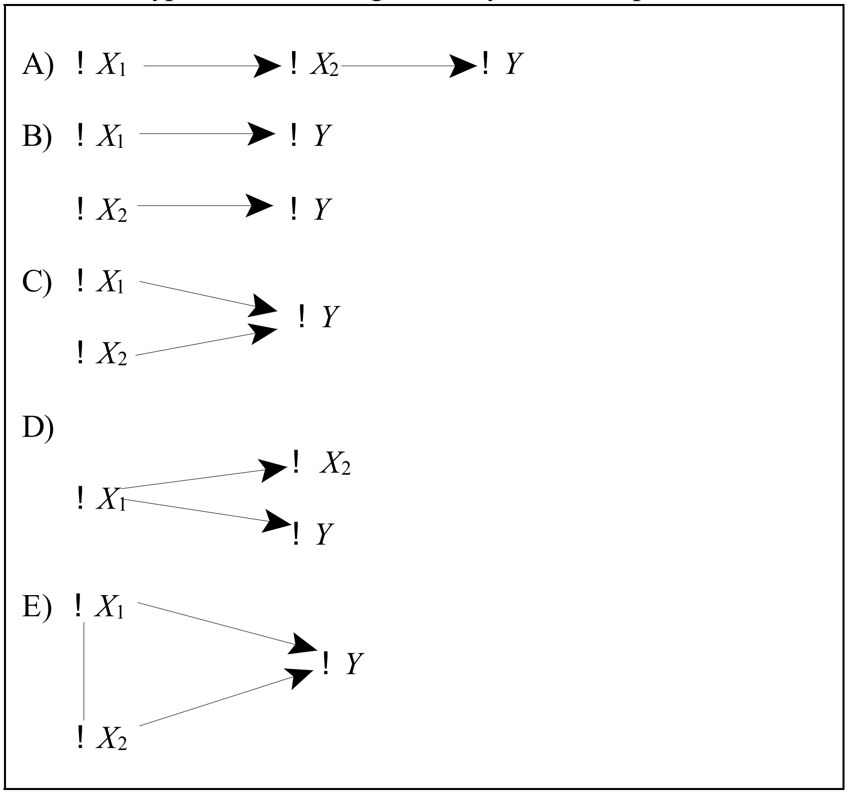
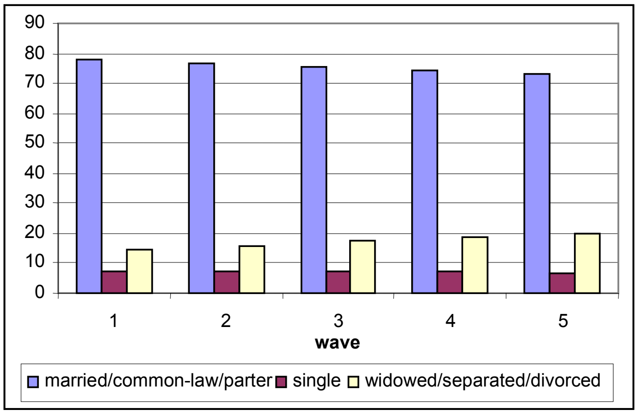
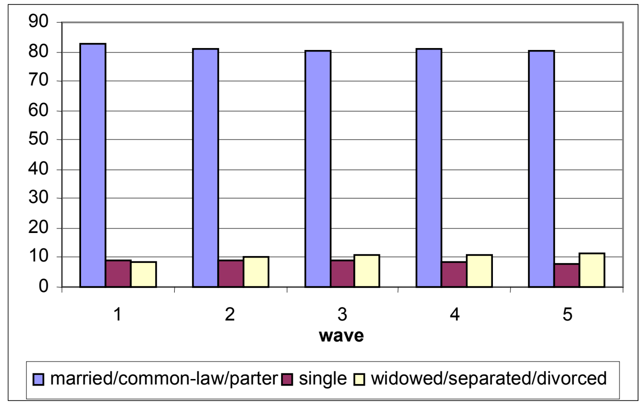


Figure 2: Changes in the variables over time from Wave 1 (1994) to Wave 5 (2002) – Age group 35-49

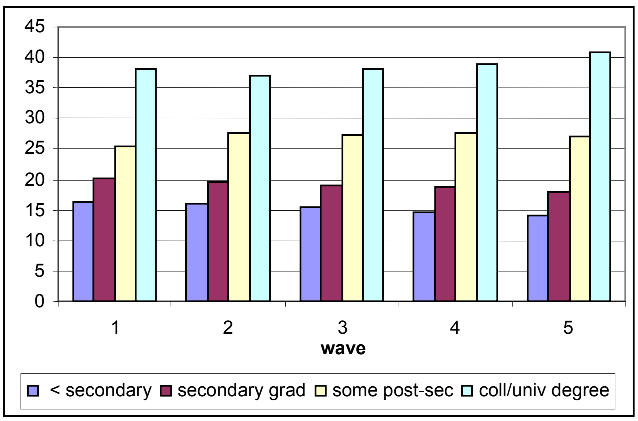
Female | **Male**

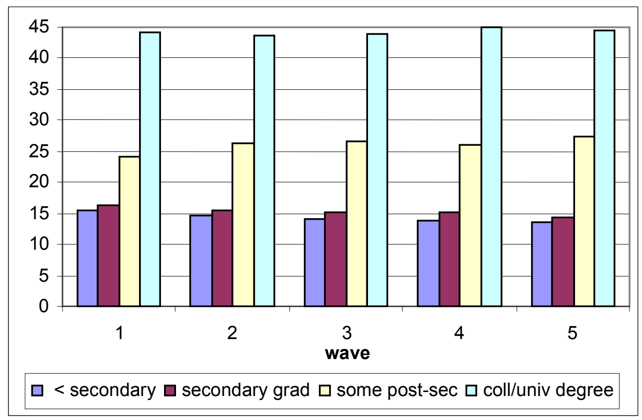
Marital status:



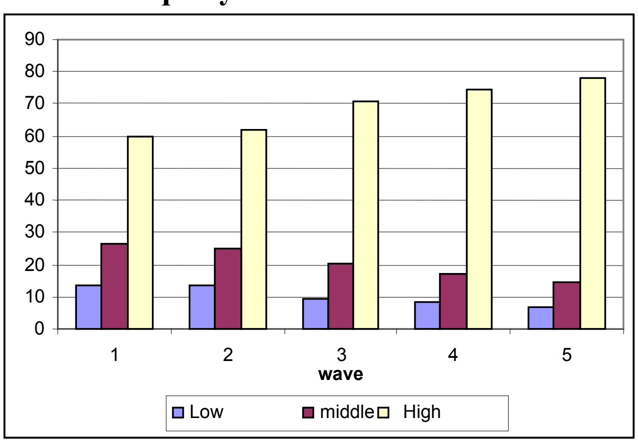


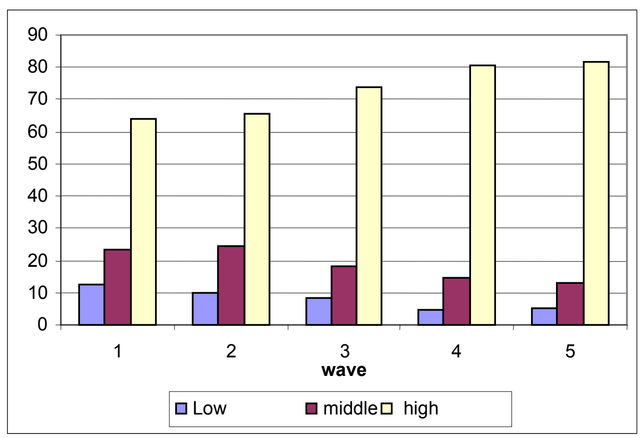
Educational attainment:



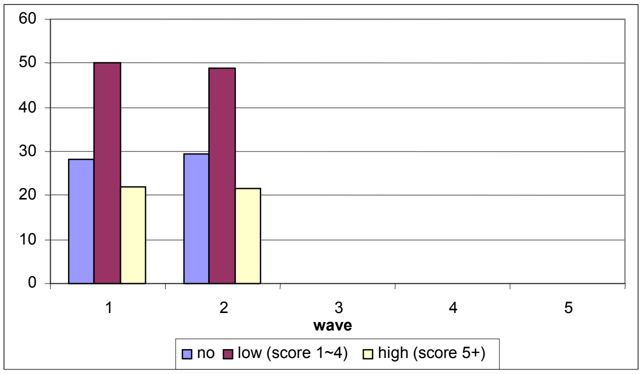


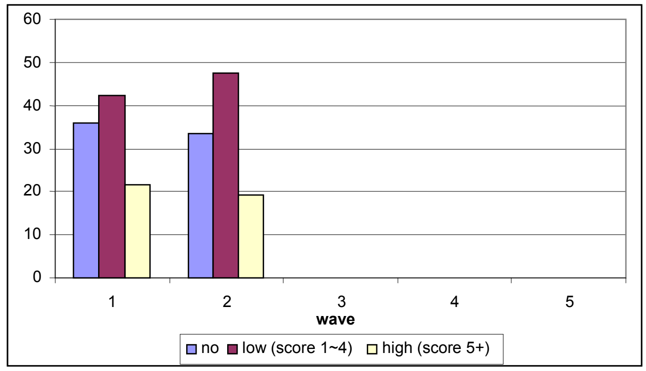
Income adequacy:



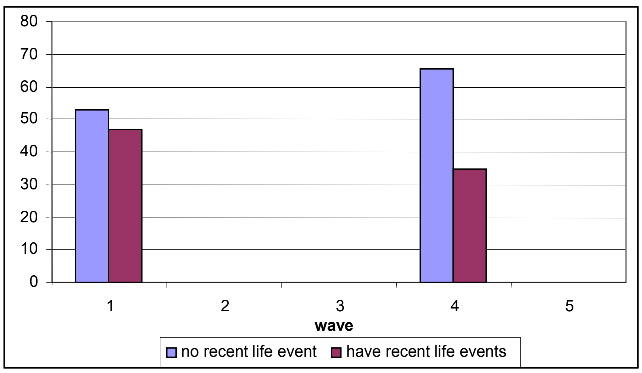


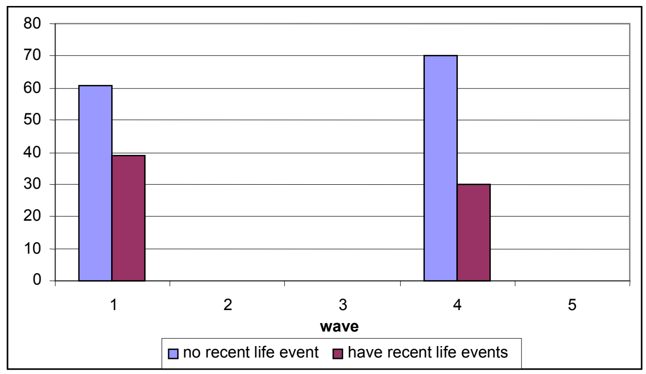
Social involvement:



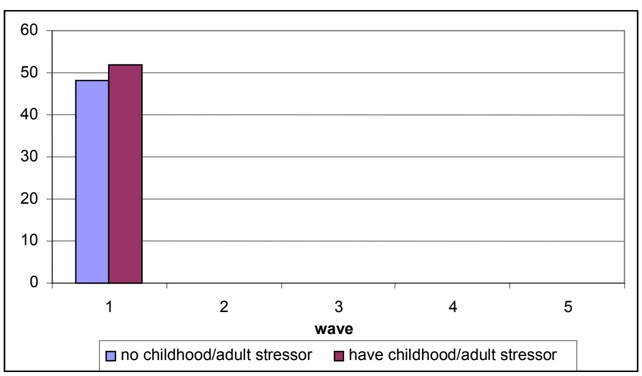


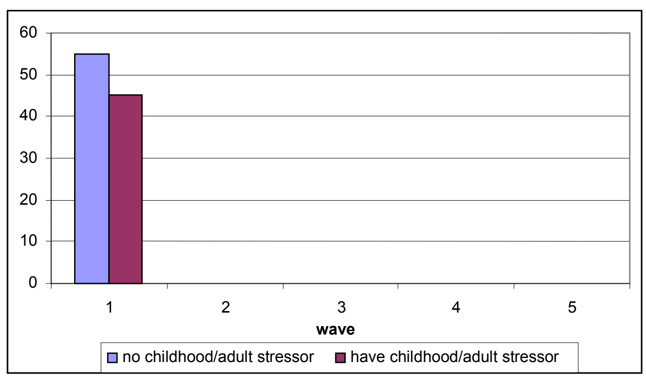
Recent life events:



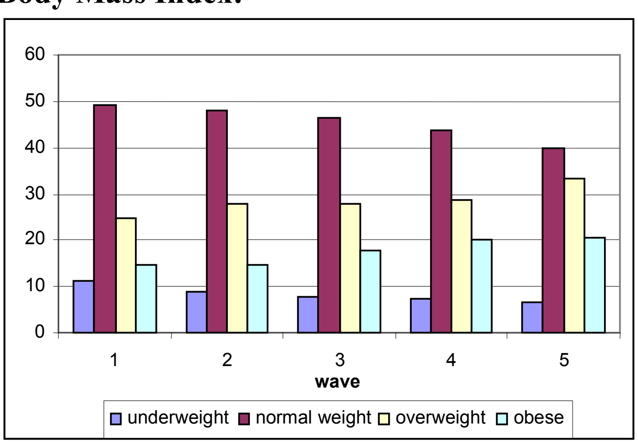


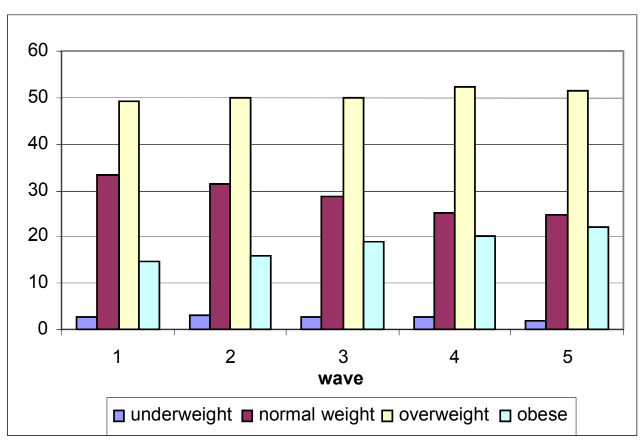
Childhood/adult stressor:



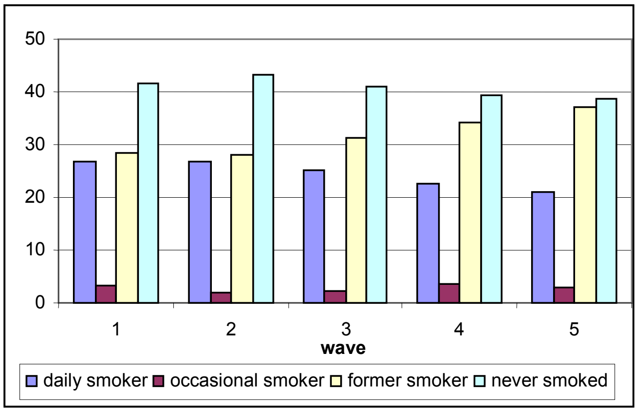


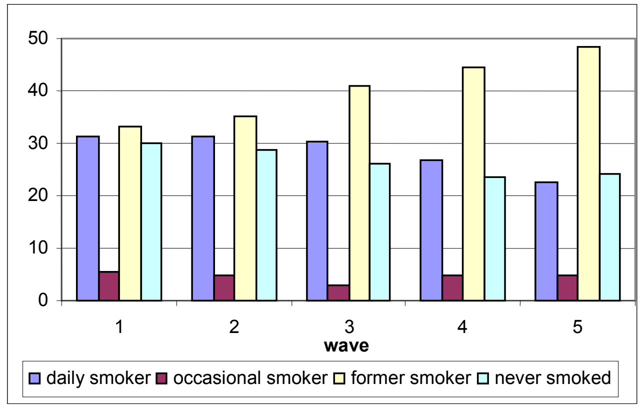
Body Mass Index:



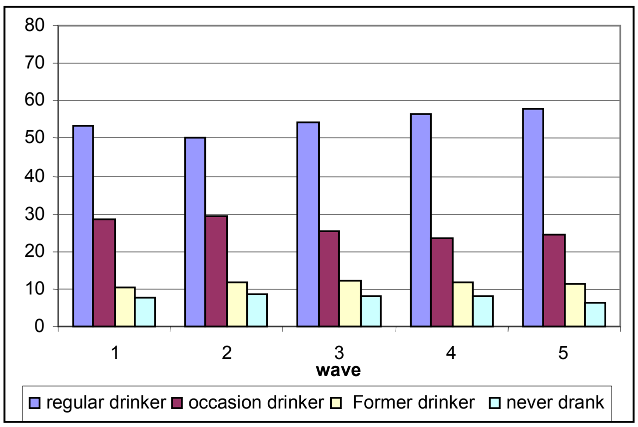


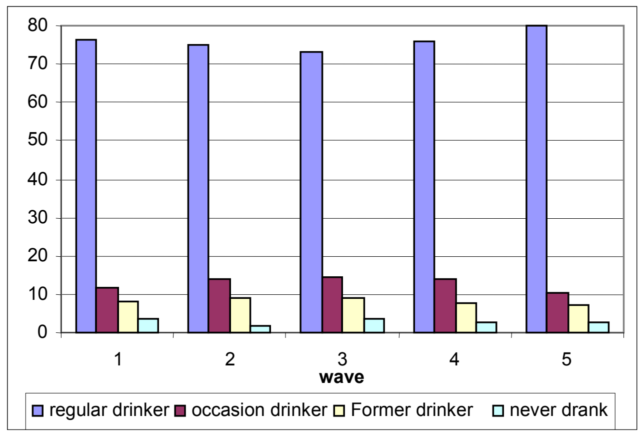
Smoker type:



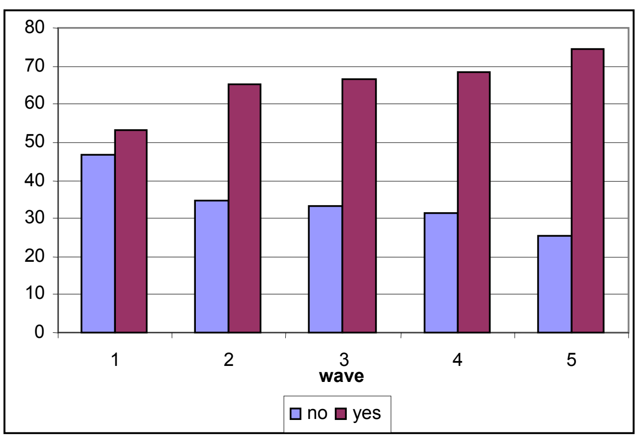


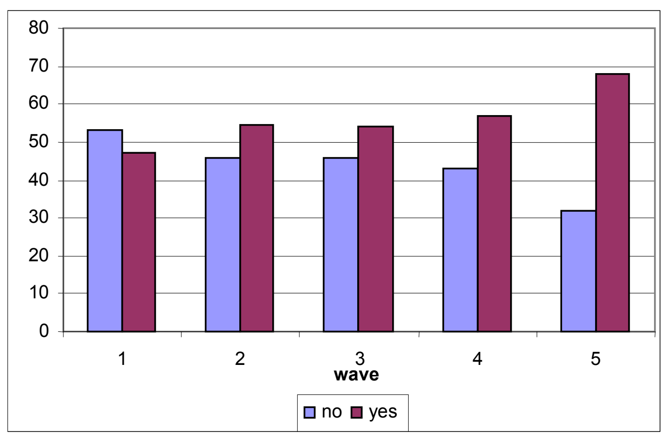
Drinker type:



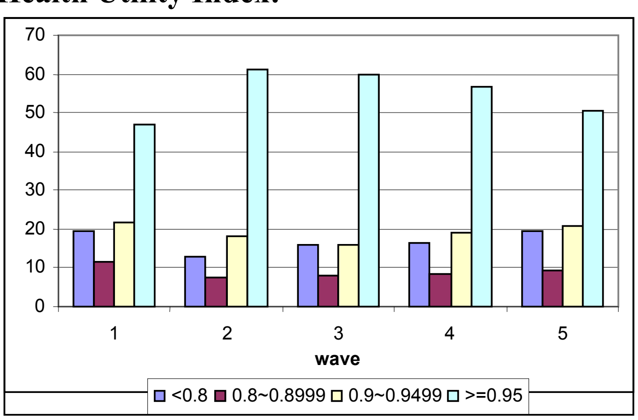


Chronic condition:





Health Utility Index:



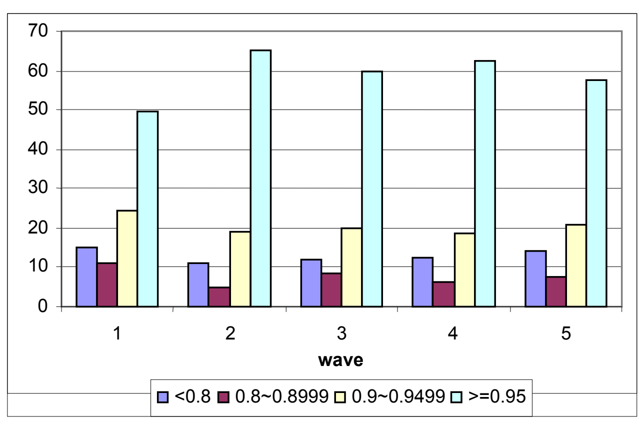
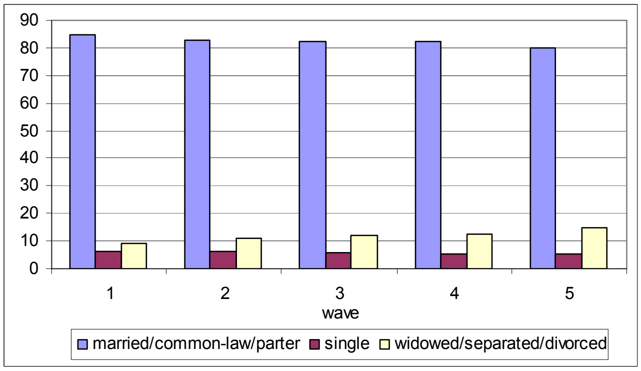
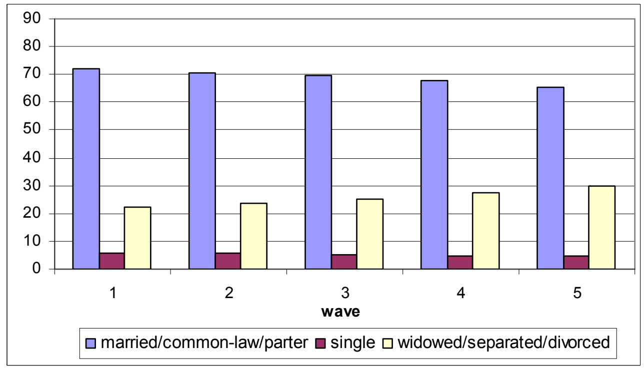


Figure 2 contd: Changes in the variables over time from Wave 1 (1994) to Wave 5 (2002) – Age group 50-64

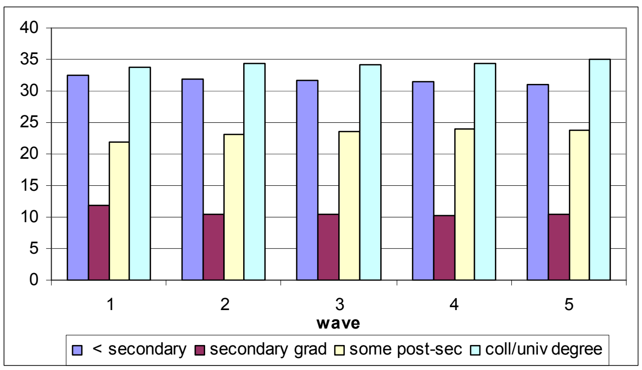
Female **Male**

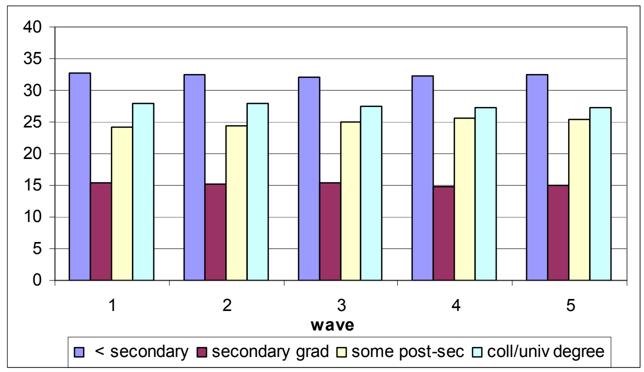
Marital status:



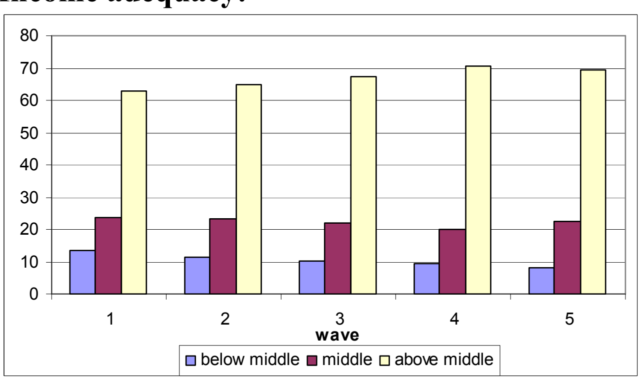


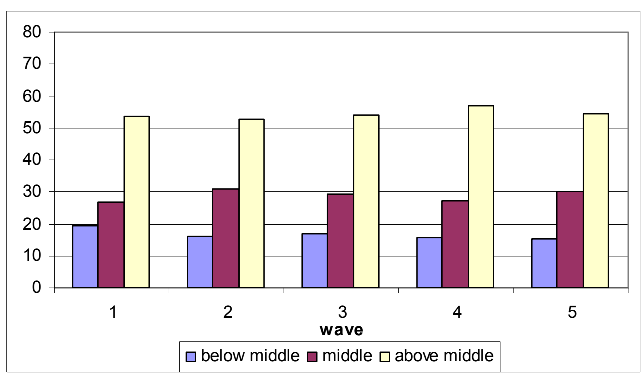
Education:



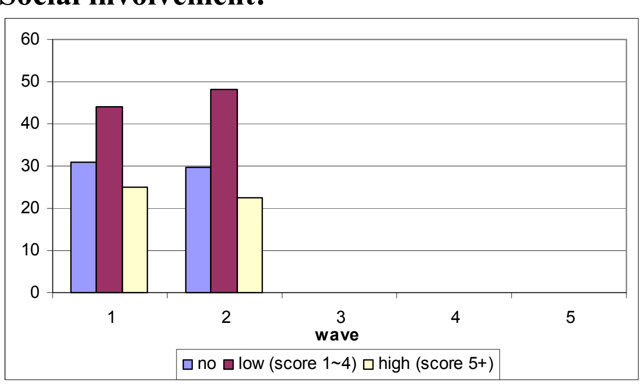


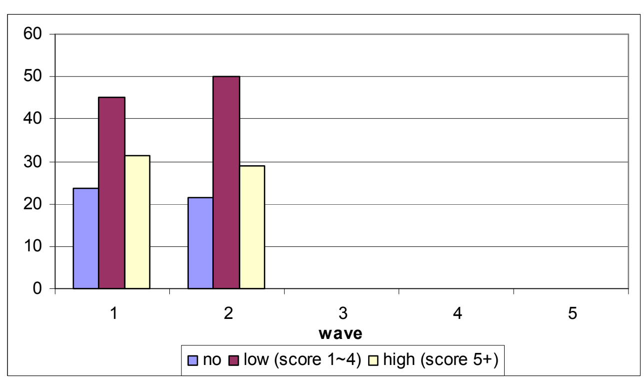
Income adequacy:



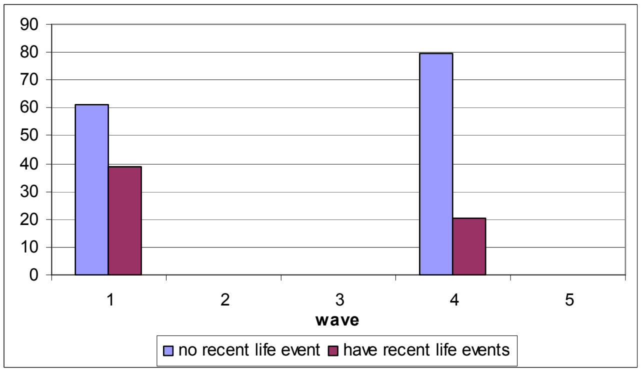


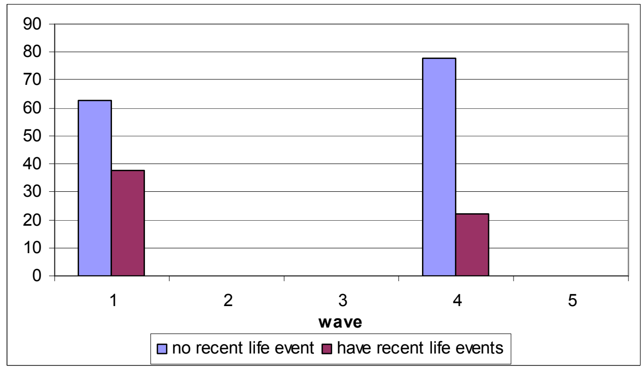
Social involvement:



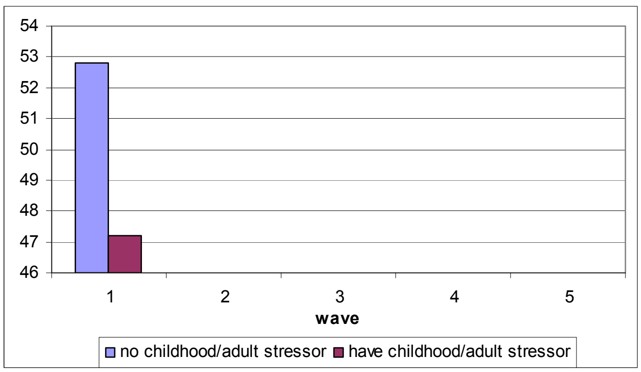


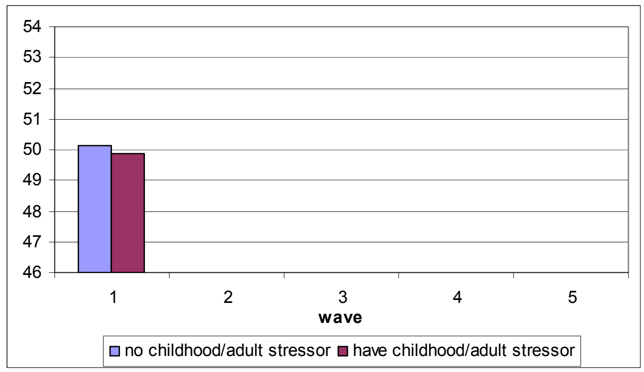
Recent life events:



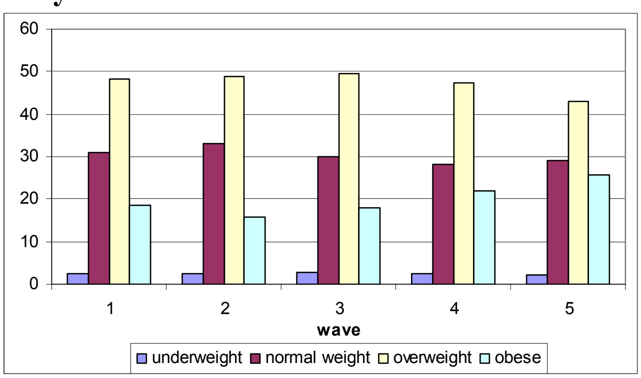


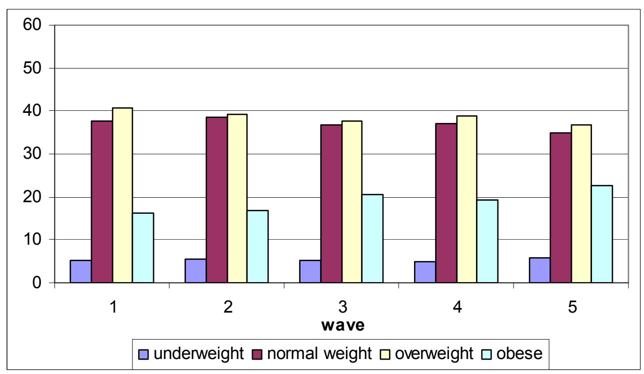
Childhood/adult stressor:



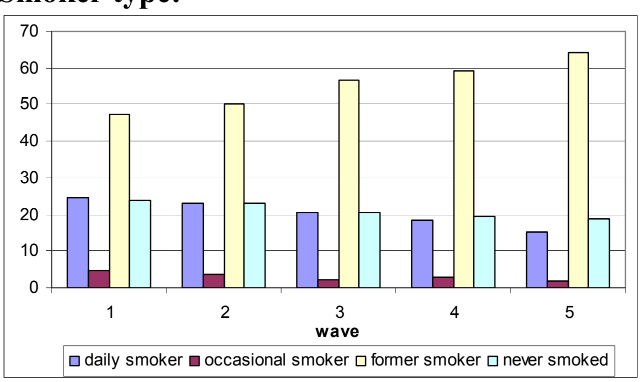


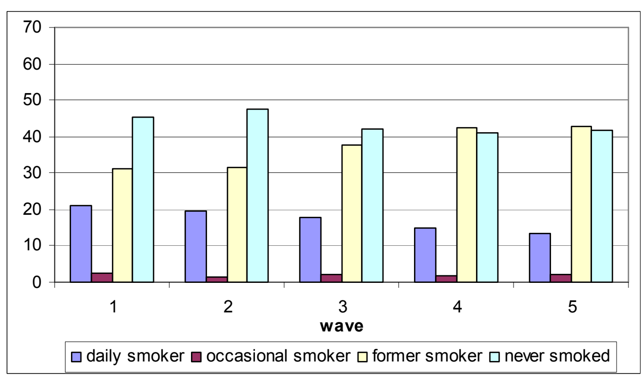
Body Mass Index:



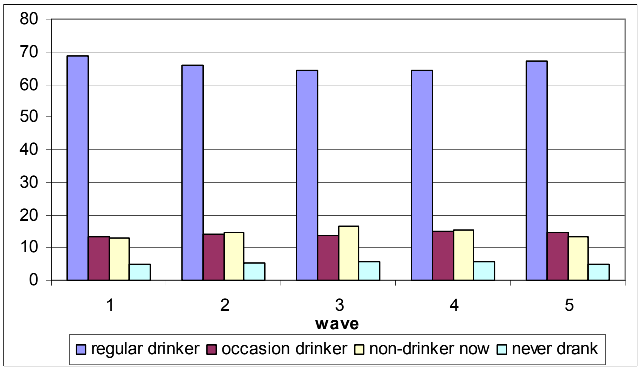


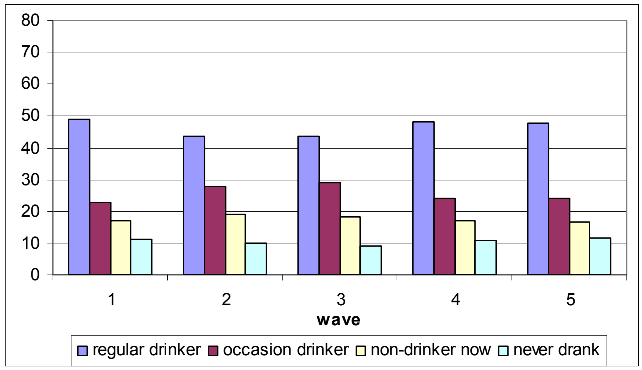
Smoker type:



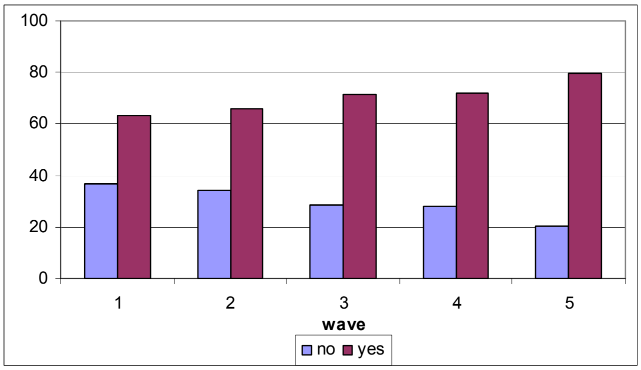


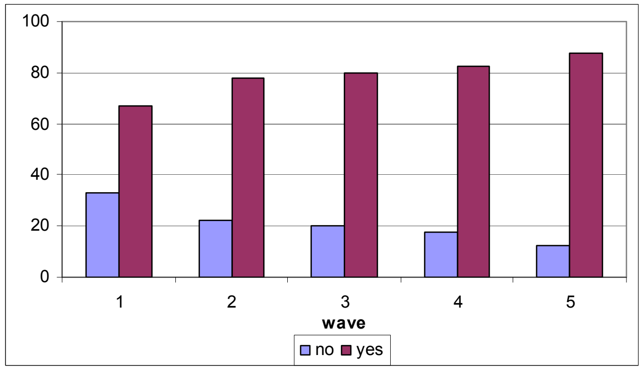
Drinker type:



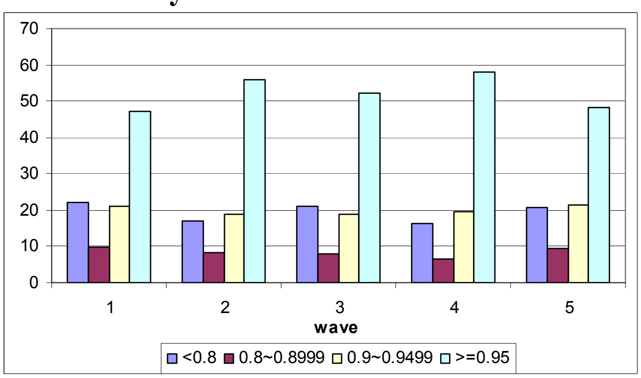


Chronic condition:





Health Utility Index:



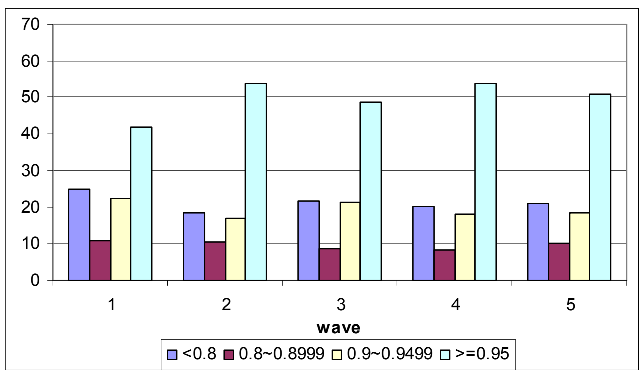


Figure 3a: Conditional Independencies or Undirected Edges, Wave 1 - for the age group 35-49 in 1994, Males and Females

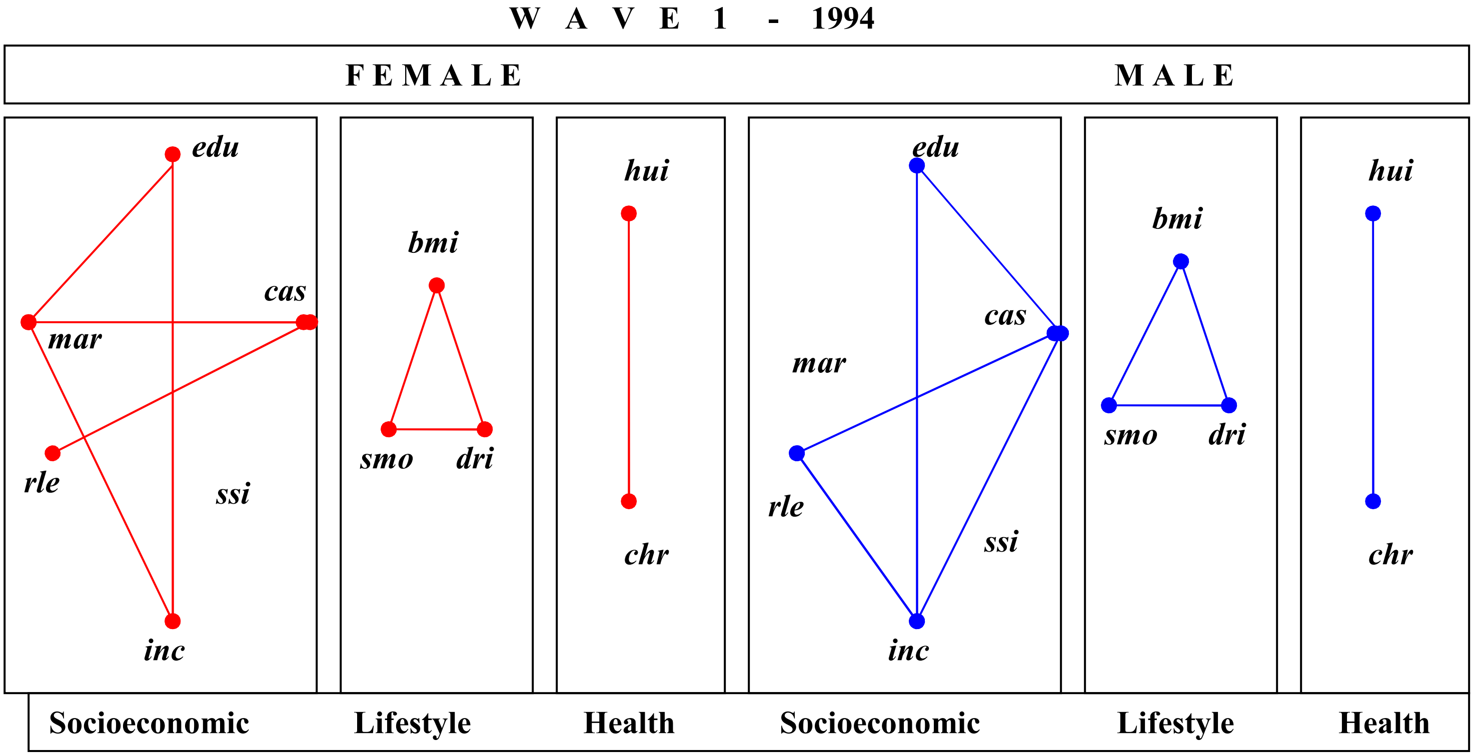


Figure 3b: Conditional Independencies or Undirected Edges, Wave 1 - for the age group 50-64 in 1994, Males and Females

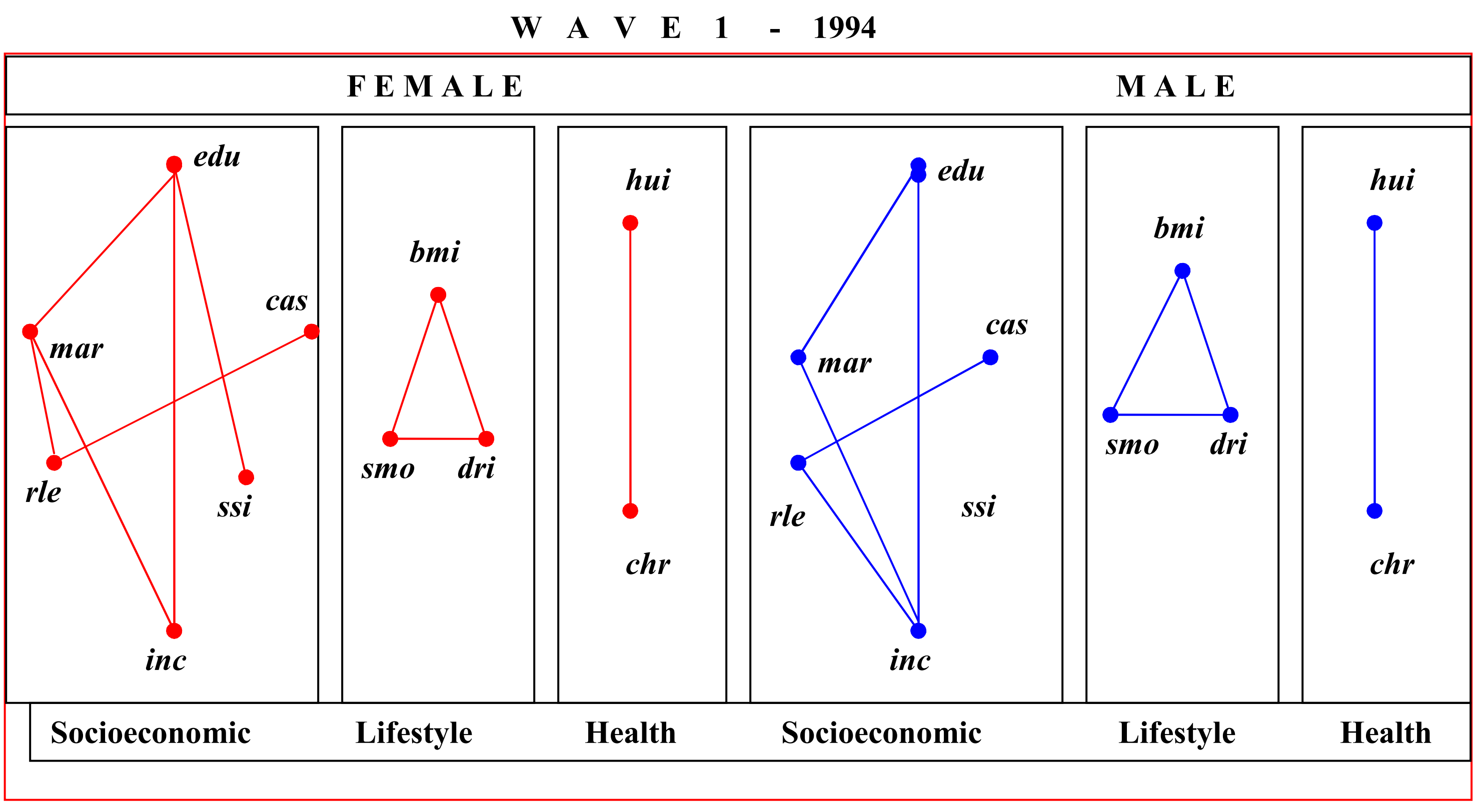


Figure 4a: Undirected and Directed Edges – for Males in age group 35-49 in 1994

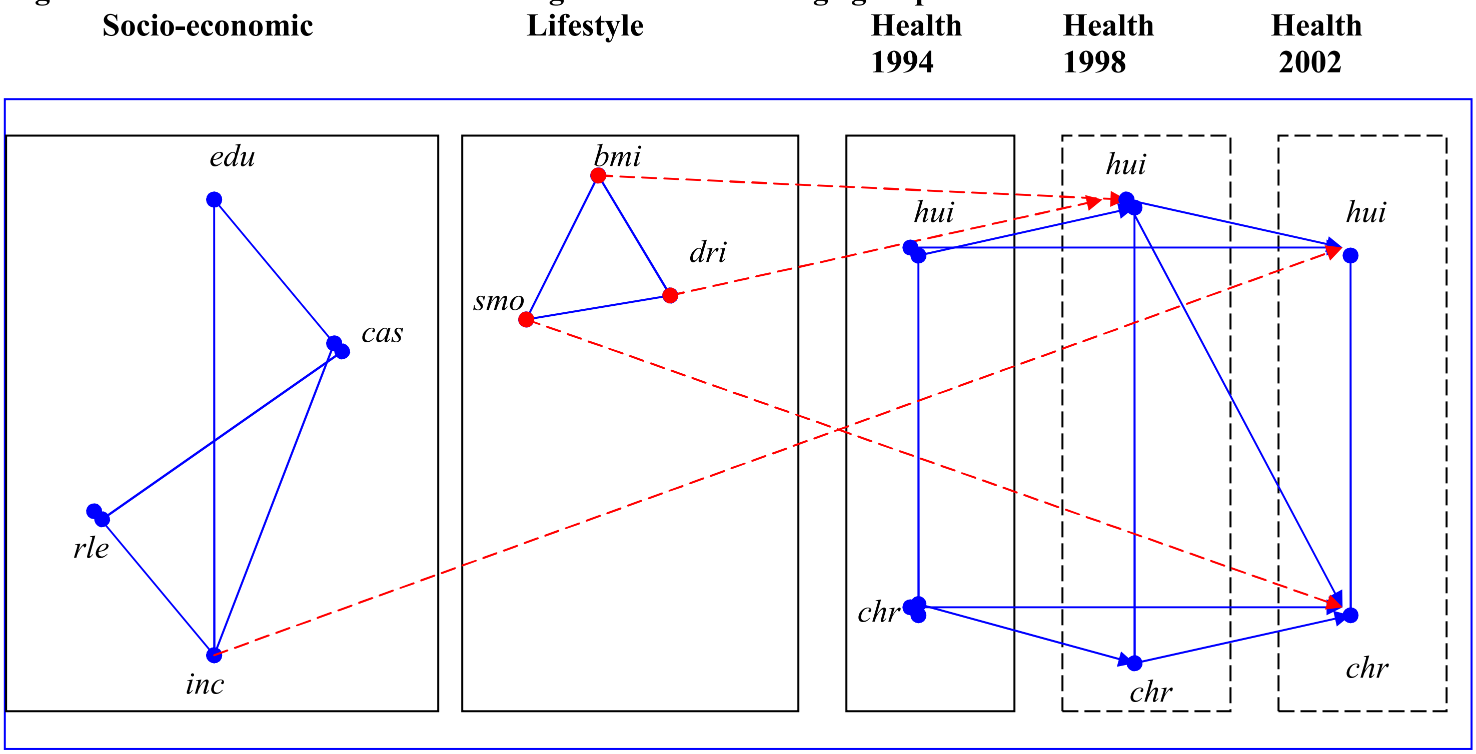


Figure 4b: Undirected and Directed Edges – for Males in age group 50-64 in 1994

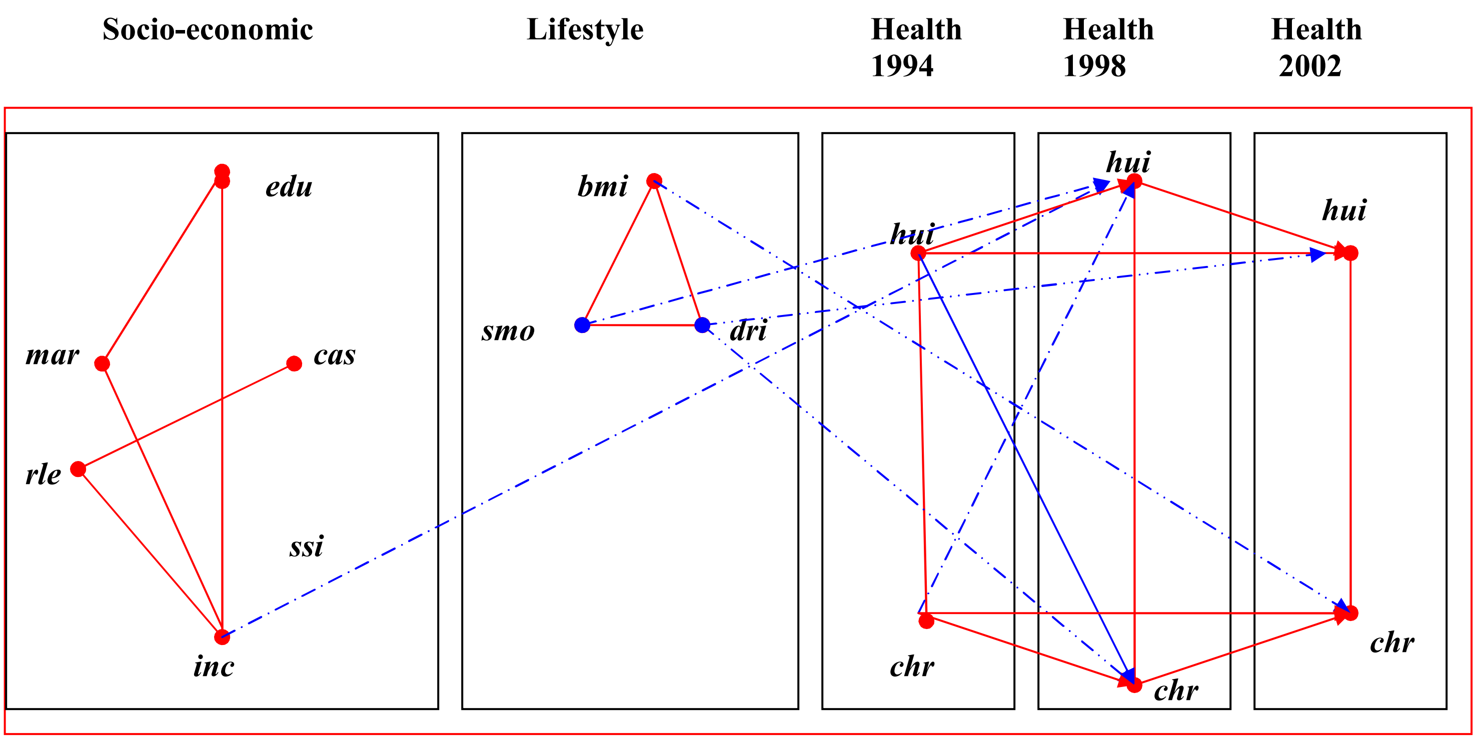


Figure 5a: Undirected and Directed Edges – for Females in age group 35-49 in 1994

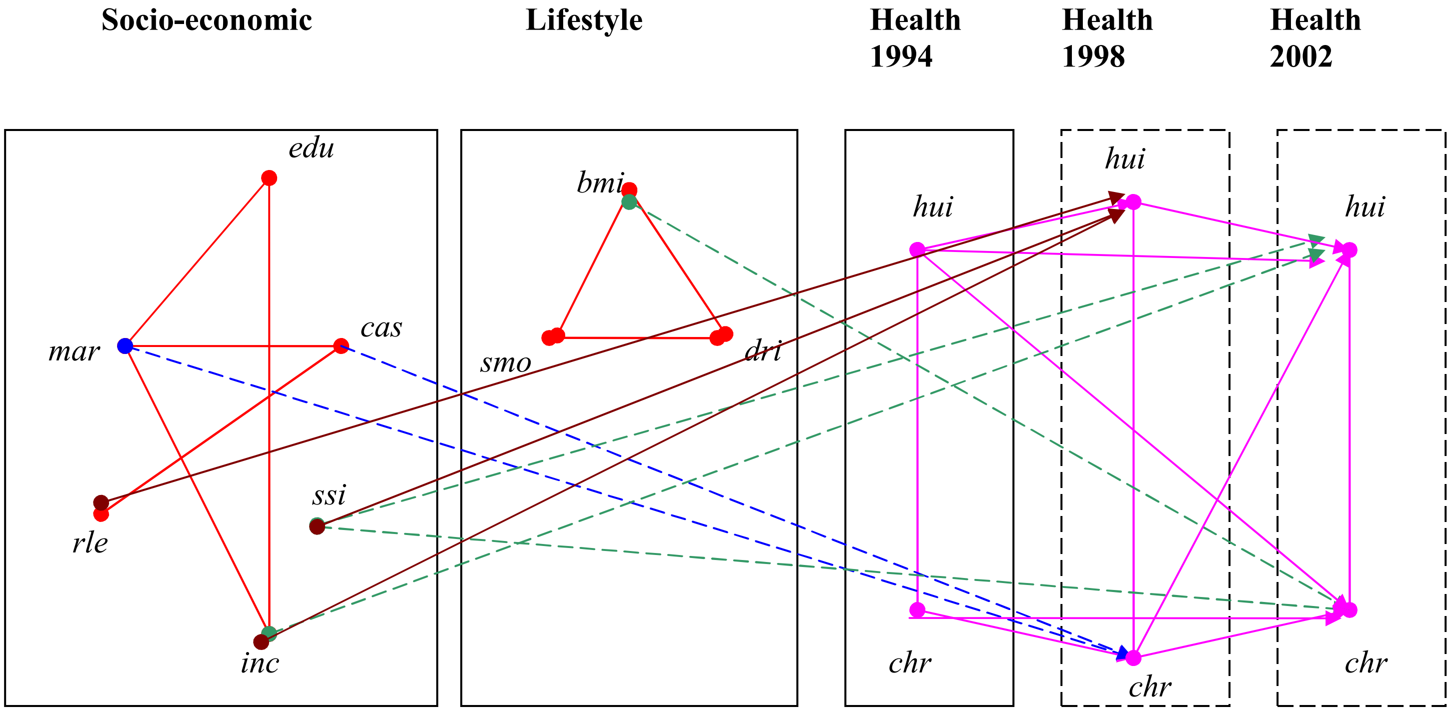


Figure 5b: Undirected and Directed Edges – for Females in age group 50-64 in 1994