

## Research on social determinants and health: what sorts of data do we need?

Siegfried Geyer

Published online: 3 September 2009  
© Birkhäuser Verlag, Basel/Switzerland 2009

Studies dealing with the relationship between social determinants and health are usually conducted by means of surveys. Data on outcomes are assessed on an individual basis, e.g. subjective health or subjective well-being, and they are mostly cross-sectional. Although there are significant exceptions, this holds for the majority of studies. Depending on the research question it may be appropriate to choose cross-sectional designs, it may also make sense to do so if it is only second best, because having approximate solutions are better than having none at all. There are, however, good reasons to deviate from this tradition. So what sorts of data are needed?

Research on social determinants should no longer build on one sort of data only. Instead, data from different sources should be used in order to reduce sources of bias and to improve validity of the conclusions. Recent technical developments are increasingly permitting to integrate different types of information. This refers to micro-level data collected from individuals by means of interviews (e.g. attitudes, behavioural measures or personality characteristics) and to micro-data drawn from registries (e.g. individual data on sickness absence, data on the length of inpatient treatment). Technical advances also relate to types of data that are not individual based, such as the

quality of housing areas, the degree of environmental pollution or on the security of districts within towns or defined areas. The appropriate information may be provided by community authorities, by local and national weather services or other institutions collecting data on a regular basis. In this editorial the author outline some of the problems related to the use of data from different sources and highlight possible solutions.

**Subjective health is an incomplete and biased measure of impaired health. It needs to be complemented by clearly defined outcomes in order to examine the full range of effects of social determinants**

Subjective health is an indicator for a number of hidden phenomena such as distress, undiagnosed diseases or the decline of resilience due to age or to the occurrence of chronic difficulties. It is a general and unspecific measure that may be a good starting point for further exploration but it is difficult to draw precise conclusions based on this measure. In addition, assessing diseases, symptoms or impairments on a subjective basis runs the risk of under- or over-reporting. Underreporting may occur if symptom awareness is low, if individuals try to keep a healthy self-image in spite of a severe diagnosis, and if adaptation to a disease without permanent impairment is successful (e.g. diabetes). On the other hand increased symptom awareness may lead to an overestimation of disease rates.

Although questionnaire data are an important source of information for research, such measures are flawed by the well-known difference between attitudes, intentions and behaviour. In case of measuring behaviour directly it should be possible to monitor the adherence to therapeutic regimen or to what extent individuals follow

---

Siegfried Geyer is director of the Medical Sociology Unit at Hannover Medical School. He was president of the European Society of Health and Medical Sociology. His research interests are focussed on the role of social factors in the aetiology and course of diseases, in particular in the field of health inequalities.

---

S. Geyer (✉)  
Medical Sociology Unit, Hannover Medical School,  
30625 Hannover, Germany  
e-mail: siegfried.geyer@t-online.de  
URL: [http://www.mh-hannover.de/med\\_soziologie.html](http://www.mh-hannover.de/med_soziologie.html)

health-promoting lifestyle patterns. In a recent study (Schiel et al. 2009) behaviour registration was performed via mobile phones in order to assess the amount of exercise and for documenting the type of nutrition consumed by patients. Mobile phone data were used in combination with questionnaires and medical data.

### **Data from registries may be more complete than survey data and depict morbidity better and more accurately**

The concentration on survey methodology in studies on social determinants and health has shortcomings, and there are good reasons to consider using also data from other sources. Objections against registry data or other routinely used information refer to the variety of individuals and institutions in charge of data collection, with limited validity due to varying standards in registration and coding and also due to idiosyncrasies of coders who may bring in undesirable variability (Geyer and Wedegärtner 2007). These points cannot be brushed aside, but surveys are also susceptible to several sources of bias. This starts with large interindividual differences in the understanding of the same question (Schoon et al. 2003; Tourangeau et al. 2000). In surveys, the health status of individuals leads to differential willingness to participate. It has consistently been shown that in population surveys respondents are more likely to refrain from participation if they are ill or if they consider their health as bad (Hoffmann et al. 2004; Pirzada et al. 2004). Thus, the population under study appears healthier as it actually is and the prevalence of disease may be underestimated. In contrast, in patient studies the absence of symptoms may lead to non-participation, but a higher degree of severity may lead to increased willingness, because patients may expect more and better information and treatment. The latter condition may then lead to the conclusion that the population under study is more impaired as it actually is.

A possible way out is the use of micro-level data drawn from registries, e.g. information from health insurances, hospital records, cancer registries or from other registries that are currently being established for monitoring the health status of populations. Other routinely collected data can also be helpful because they are complete with respect to defined groups. A prominent example is school entry examinations. They are covering entire age cohorts of school children and provide information on their general health status and not only a limited number of defined health impairments.

Using these sorts of data also makes it possible to study the effects of social determinants on the most frequently occurring diseases such as myocardial infarctions, depression or diabetes. This also refers to diseases with low

prevalence but considerable significance for the health care system, because treatment is long lasting and expensive. Examples are diabetes type 1, congenital heart disease or cancers of low incidence. These data may have been collected for defined purposes other than research, e.g. accounting, book keeping or for purposes of internal statistics. Thus, they may have limits due to inclusion criteria and lack of completeness, nevertheless they are complete within certain limits and the shortcomings are known, and this makes it possible to explore the ins and outs of their validity.

The use of registry data can avoid such inaccuracies, and the knowledge of shortcomings and limits may improve the validity of substantive conclusions.

### **Cross-sectional data are insufficient for examining the dynamics between social determinants and health as they are providing only a snapshot of the effects**

It has already been mentioned that there is no reason for a general rejection of cross-sectional designs. Nevertheless research examining the effects of social determinants on the development and course of diseases should make more use of longitudinal approaches. Doing so, it is possible to examine the length of time between exposure to stressors and changes in health which is a crucial question if causal associations shall be subject to study. This can be demonstrated with the case of unemployment where it is necessary to know the interval between the loss of a job and elevated coronary risks. Likewise in a registry-based Japanese study it was examined whether an earthquake was followed by increased mortality rates that emerged within a period of 6 weeks (Ogawa et al. 2000). The longitudinal observation revealed that in the subsequent period CHD mortality was lower than before the earthquake, and after another period it rose to the level before. It can be concluded that the event led to premature mortality primarily in vulnerable subjects, but this conclusion could not have been drawn if only a single measurement immediately after the event had taken place.

Besides circumscribed events or chronic conditions the health-related consequences of social change needs a longitudinal perspective, and whole populations cannot be studied longitudinally with survey data only. This can be shown with the social changes following the collapse of the Soviet Union. Within a few years the life expectancy dropped considerably (Men et al. 2003). It could be demonstrated that the most dramatic changes in health occurred in the period between 1991 and 1994 where a political and an economic crisis were intertwined, and 1998 mortality rates rose again during another economic downturn, but less pronounced than in the first period. In this case

mortality statistics and long-term socio-economic macro-data on economic decline were combined in order to demonstrate the impact of social and economic change on the health of populations.

### Integration of data from different sources: an example

The usefulness of integrating material from different sources may be further illustrated with an ongoing prospective study on the role of social determinants in the course of breast cancer (Geyer et al. 2008). Data on stressful circumstances, chronic difficulties and other social determinants that may have effects on the course of disease are repeatedly collected as qualitative information by means of personal and tape-recorded interviews. The information obtained during the interview is rated and converted into quantitative data using standardized classification procedures. Socio-demographic information and data on disease-related knowledge is assessed using standardized and self-administered questionnaires. Information on the severity of the disease was drawn from hospital records, and outcomes were collected using data on recurrences as drawn from cancer care centres. Finally, information on deceased patients was obtained from cancer registries. Taken as a whole, this combination yields a more comprehensive understanding and a smaller amount of missing data as compared to relying only on one source of data.

### Conclusion

Studies on social inequalities in health or those considering relationships between specific social determinants such as unemployment are requiring large databases in order to obtain interpretable results. This information may be obtained from registries, health care providers or other institutions collecting data as part of their everyday business. Large data bases also permit to study associations of social factors with a wide variety of outcomes, including those not belonging to the rather small group of endemic diseases. Some countries even permit to combine registry data with survey information which opens new opportunities for analysing the interplay between structural variables and subjective judgements.

Research dealing with relationships of social determinants and health at the behavioural level does not necessarily require large case numbers, but rather detailed information. Data on medical treatments, the utilization of medical and social services collected by various

institutions can be combined with interview data or with macro-level information such as the quality of housing areas, the availability of infrastructure or the amount of criminality.

Collecting data from different sources is time consuming and requires planning ahead. Very early in the course of a project informed consent will have to be obtained from patients as well as permissions from participating institutions. However, all this might be well worth the effort as appropriate linking of different data sources will lead to a better understanding of the relationship between social determinants and disease and it will reduce sources of bias.

Time is ready for using data from different sources and to combine them in a meaningful way. The research questions are there, and the technologies are waiting to be used (Hammarstrom and Janlert 1994).

### References

- Geyer S, Wedegärtner F (2007) Variabilität von Arzturteilen in Schuleingangsuntersuchungen. *Das Gesundheitswesen* 69:621–627
- Geyer S, Noeres D, Sassmann H, Mollova M, Prochnow A, Neises M (2008) Does the occurrence of adverse life events in patients with breast cancer lead to a change in illness behaviour? *Support Care Cancer* 16:1407–1414
- Hammarstrom A, Janlert U (1994) Unemployment and change of tobacco habits: a study of young people from 16 to 21 years of age. *Addiction* 89:1691–1696
- Hoffmann W, Terschüren C, Holle R, Kamtsiuris P, Bergmann M, Kroke S, Sauer S, Stang A, Latza U (2004) Zum Problem der Response in epidemiologischen Studien in Deutschland (Teil II). *Das Gesundheitswesen* 66:482–491
- Men T, Brennan P, Boffetta P, Zaridze D (2003) Russian mortality trends for 1991–2001: analysis by cause and region. *Br Med J* 327:964
- Ogawa K, Tsuji I, Shiono K, Hisamichi S (2000) Increased acute myocardial infarction mortality following the 1995 Great Hanshin-Awaji earthquake in Japan. *Int J Epidemiol* 29:449–455
- Pirzada A, Yan LL, Garside DB, Schiffer L, Dyer AR, Daviglius ML (2004) Response rates to a questionnaire 26 years after baseline examination with minimal interim participant contact and baseline differences between respondents and nonrespondents. *Am J Epidemiol* 159:94–101
- Schiel R, Kaps A, Bieber G (2009) Telemedizinische Erfassung körperlicher Aktivität, Ruhe und der Ernährung von Kindern und Jugendlichen mit Typ-1-Diabetes mellitus. In: 18. Rehabilitationswissenschaftliches Kolloquium. Innovation in der Rehabilitation- Kommunikation und Vernetzung vom 9.-11. März in Münster, Deutsche Rentenversicherung Bund, Berlin, pp 84–85
- Schoon I, Sacker A, Bartley M (2003) Socio-economic adversity and psychosocial adjustment: a developmental-contextual perspective. *Soc Sci Med* 57:1001–1015
- Tourangeau R, Rips LJ, Rasinski K (2000) *The psychology of survey response*. Cambridge University Press, Cambridge