ABSTRACT

Widespread recognition is given to the idea that an individual's health is influenced by the ways an individual works, rests and plays as well as genetic endowments. More recently, some researchers have suggested that the relationships between individual 'characteristics' and health are influenced by the context or community of the individual. Although the notion of the community as a determinant of health is not new, the incorporation of the community in empirical research on the determinants of health has been based on simple statistical models that fail to reflect the complex nature of the individual-community interface. In this paper we use the methods developed in educational research to show how separate statistical models for variations in health between communities and between individuals can be combined to provide a multi-level model for the determinants of health of populations.

ABRÉGÉ

L'idée que la santé d'un individu est influencée par son travail, ses loisirs et ses plaisirs, et par son bagage génétique, est généralement acceptée. Récemment, des chercheurs ont suggéré que la relation entre santé et caractéristiques des individus dépend du milieu de vie ou de la communauté de résidence. Depuis longtemps, on reconnaît en santé publique que la communauté est un déterminant de la santé. Dans les recherches empiriques, la communauté a été considérée comme toutes les autres variables individuelles dans des modèles statistiques qui ne traduisent pas la complexité de la relation individu-communauté. Les procédures statistiques développées dans les sciences de l'éducation proposent des modèles qui combinent les variations des états de santé dues aux caractéristiques des individus et à celles de la communauté. Un modèle multi-niveaux pour l'analyse des déterminants de la santé de population est décrit et proposé.

Modelling the Community as a Determinant of Health

Stephen Birch, DPhil,^{1,2} Greg Stoddart, PhD,^{1,3} François Béland, PhD⁴

Recent conceptual frameworks for the determinants of health in populations emphasize the potential importance of context for the relationships between various factors and individual health.¹⁻⁵ This notion of contextual or community-level influences, however, provides major challenges for the development of empirical models aimed at providing an 'evidence-base' for understanding and influencing policy on the determinants of health. In this paper we show how the statistical basis of traditional approaches to estimating community effects fails to reflect the complex nature of the conceptual frameworks. The statistical model is then developed to provide a more appropriate basis for estimating community effects which draws on the methods already used in other areas of research.

Communities, individuals and community effects on individuals

A major feature of the recent conceptual frameworks has been the explicit recognition given to the 'pathways' along which community effects operate (see Figure 1 based on Evans and Stoddart⁵). For example, the effect of an individual's income on that individual's health may be 'determined' by 'community' level factors, such as the level of unemployment in the community. Empirical estimation of these

effects depends on the development of statistical models that incorporate the role of community in accordance with these hypothesized pathways. Although consideration of community-level influences on health is by no means new in health research, careful consideration of the precise meaning of a community effect, and explicit consideration of the statistical specification of these effects have been absent from most empirical research on the determinants of health of populations.⁶

As a starting point for incorporating the role of community in empirical models, we define the *community* as *the social, physical, cultural and economic characteristics of the individual's environment*. A *community effect* is then represented by

a variation in an individual-level dependent variable embodied within communities that is explained through person-level processes that differ among communities.

So, although defined at the community level (i.e., all members of that community are exposed to that community variable), the community effect is concerned with influences on (or through) individuals according to the community to which they belong. As Diez-Roux⁷ notes, for context to affect health "it must get into the body" and hence "it ultimately operates through individual level processes". So the community provides a level at which to assess the impacts of context on the effects of individuals' characteristics on individuals' health.

Implications of community effects for empirical models

We can consider variations in health status of individuals in terms of variations in individual-level variables (e.g., individuals'

1. Centre for Health Economics and Policy Analysis and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario
2. McMaster Institute for Environment and Health
3. Population Health Program, Canadian Institute for Advanced Research
4. Groupe de Recherche Interdisciplinaire en Santé, Université de Montréal, Montréal, Québec

Correspondence: Dr. Stephen Birch, Centre for Health Economics and Policy Analysis, McMaster University, 1200 Main St. West, Hamilton, ON, L8N 3Z5, Tel: 905-525-9140, ext. 23244

incomes) and variations in community-level variables (e.g., prevailing levels of unemployment) in terms of

$$h_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (1)$$

where

h_i = health status for individual i ;

x_{0i} = base category, a set of 1s representing individual i with average income and living in a community with average unemployment;

x_{1i} = income (centred about the mean) for individual i ;

x_{2i} = level of unemployment in individual i 's community (centred about the mean);

ε_i = individual error; variations in individual health status not accounted for by income and level of unemployment;

β_0 = health status for individual with average income;

β_1 = change in health status of individual associated with the change in that individual's income;

β_2 = change in health status of individual associated with the change in unemployment in that individual's community.

The influence of the community level of unemployment is introduced in the same way as individual-level characteristics. There is no recognition of the hierarchical nature of the data with individuals and their characteristics being 'grouped' into communities (e.g., families, households, neighbourhoods, schools, etc.). The effects of the level of unemployment in the different communities (i.e., the community context) are therefore confounded with the effects of the composition of the different communities (i.e., the distribution of individual characteristics in the different communities). Duncan et al.⁸ note that analyses of variations in health-related behaviours require methods that acknowledge "... it is individuals, not groups, that behave, thereby avoiding the ecological fallacy [and]... individuals behave in context, thus avoiding the atomistic fallacy." (pg. 727)

As an alternative approach, let us start with a model that attempts to explain variations in health in terms of variations in only individual-level variables, e.g.,

$$h_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \varepsilon_i \quad (1a)$$

Note that parameters β_0 and β_1 are restricted to be the same across all communities. Let us remove this restriction so that these parameters are allowed to vary across communities, i.e.,

$$\beta_{0j} = \beta_0 + \mu_{0j} \quad (2)$$

$$\beta_{1j} = \beta_1 + \mu_{1j} \quad (3)$$

where

μ_{0j} represents the variation among communities in health status of individuals of average income (i.e., the variation in the intercept);

μ_{1j} is the variation among communities in the increase in health status with increase in income (i.e., the variation in the slope).

Substituting equations (2) and (3) into equation (1a) gives

$$h_{ij} = (\beta_0 + \mu_{0j})x_{0ij} + (\beta_1 + \mu_{1j})x_{1ij} + \varepsilon_{ij} \quad (4)$$

Rearranging

$$h_{ij} = \beta_0 x_{0ij} + \mu_{0j} x_{0ij} + \beta_1 x_{1ij} + \mu_{1j} x_{1ij} + \varepsilon_{ij} \quad (4a)$$

Suppose $\mu_{0j} \neq 0$, $\mu_{1j} = 0$, i.e., the change in health associated with the change in income is independent of the community of residence (see for example Newachek⁹). The only effect of community is to shift the individual-level relationship (Figure 2a). Where $\mu_{0j} = 0$, $\mu_{1j} \neq 0$, the health status of individuals with average income is the same in all communities but the health-income relationship is associated with the community of residence (Figure 2b) (see for example Turner¹⁰). Finally where $\mu_{0j} \neq 0$, $\mu_{1j} \neq 0$, the community is associated with (or a determinant of) both the health of people with average income and the health-income relationship (Figure 2c).

The variability of community-level differences in health consists of the real differences between communities (i.e., true parameter variance) and the variance arising from the communities under consideration being only a sample of the population of communities (i.e., the sample variance). The parameter variance is accounted for in the empirical model by inclusion of the community level (equations 2 and 3),

but the community sample variance terms μ_{0j} and μ_{1j} are unexplainable in terms of any independent community-level variables. In this way, the model in equation 4 may not be well specified and hence risks generating incorrect parameter estimates.¹¹

Exploration of community characteristics in explaining variations in health status among individuals has traditionally taken the form of community-level variables and interactions between these variables and individual-level variables being introduced into an extended form of equation 1 as:

$$h_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i \quad (5)$$

Where x_{2i} is the value of the community-level variable for individual i (and all other individuals in that community), and β_2 and β_3 are the regression coefficients for the community-level variable and the interaction between community- and individual-level variables respectively. In practical terms this means that the predicted values of the dependent variable, health, based on the estimated relationship between health and say, the prevailing level of unemployment, are *restricted* to lie on the community-specific 'health - income' estimated relationships in Figure 2c (i.e., there is no sample variance). Attempts to estimate the relationship between community-level variables, individual-level variables and health therefore need to allow for the possibility that the community-level variable does not fully explain between-community variations in individual-level relationships (i.e., β_{0j} and β_{1j}).

In principle a separate slope and a separate intercept term could be included for each community in the study. However this would involve the estimation of a large number of parameters and would require a very large sample size. Moreover, because communities represent a clustering of individuals the independence of residuals assumption of the OLS model would be violated and would lead to underestimates of standard errors.¹²

In education research, multi-level models that reflect the hierarchical structure of individual- and community-level data have been developed to explore variation at the community level and covariation between community- and person-level variables.¹³⁻¹⁵

$$h_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + \gamma_1 z_{1j} x_{0ij} + \gamma_2 z_{1j} x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \epsilon_{ij}) \quad (8)$$

Only in the special case of $\mu_{0j} = \mu_{1j} = 0$ can community-level variables be modeled as individual-level effects in a single level equation. Where either μ_{0j} , μ_{1j} or both are non-zero, single level equations are inadequate. Estimates of the variances of μ_{0j} and μ_{1j} across communities are therefore used to assess the importance of community. As described by Jones and Duncan,¹⁶ this involves

“...separate specification of statistical models at both the individual and community level and their combination into an overall model. The micro-model is a within-community equation while the macro-models are between-community equations in which the parameters of the within-community model are the responses in the between-community models.”

Predicted values for health are not constrained to lie on community-specific health-income relationships. Instead, the sample of communities is used to estimate distributions of community-level slopes and intercepts.

DISCUSSION AND SUMMARY

Much attention has been given to the nature and role of community influences on health in the development of conceptual frameworks on the determinants of health. Corresponding attention needs to be given to the empirical specification of these relationships in order to generate a valid ‘evidence-base’ on the determinants of health. We have illustrated how individual- and community-level models can be combined to explore the conditional nature of health determinants. Although the illustrations use simple functional forms at both the individual and community levels, more complex models can be used in the same way. Moreover software packages are available for estimating the relevant equations.¹²

The significance of these approaches is illustrated by reference to recent research on community influences on health and lifestyles. Blaxter¹⁷ considered the role of location in analyses of variations in lifestyle in the UK while Turner¹⁰ considered the

Following this approach, variance terms for the community-level relationships can be defined as:

$\sigma^2_{\mu 0}$ estimate of between-community differences in health of individuals with average income (the variance of the intercept);

$\sigma^2_{\mu 1}$ estimate of between-community differences in the health-income relationship (the variance of the slope);

$\sigma_{\mu 0 \mu 1}$ covariance of between-community differences in health of individuals with average income and between-community differences in the health-income relationship (the intercept-slope covariance).

Between-community differences in health are modeled by including the community-level variable (e.g., unemployment level) in the parameter equations (2) and (3) so that

$$\beta_{0j} = \beta_0 + \gamma_1 z_{1j} + \mu_{0j} \quad (6)$$

$$\beta_{1j} = \beta_1 + \gamma_2 z_{1j} + \mu_{1j} \quad (7)$$

where z_{1j} is the level of unemployment in community j . The γ parameters represent the relationship between community differences in health, after controlling for the income composition of the communities, and the level of unemployment in communities. Equation 6 represents the relationship between the level of unemployment of the community and differences between communities in the level of health of individuals with average income. Equation 7 represents the relationship between community differences in the health-income relationship and the level of unemployment of the community. Substituting community-level equations (6) and (7) into the individual-level model in equation (1a) gives

Figure 2. Community effects on the health-income relationship among individuals.

direct and indirect effects of community unemployment on the health of individuals in the United States. In both cases, relationships were estimated using models that failed to reflect either the hierarchical structure of the conceptual models or the hierarchical nature of the data sets used. Duncan et al.⁸ used a hierarchical (or multi-level) model to challenge Blaxter's findings concerning the importance of location in the distribution of lifestyles. The strong community-level differences in lifestyle observed by Blaxter were much smaller once account was taken of the variance/covariance structure of the data.

Significant community effects were found relating deprivation and income variation to the prevalence of chronic illness in the UK population using the same approach.¹⁸ These effects remained after including a wide range of individual-level determinants of health. Multi-level models have also been used to estimate the effects of individual- and community-level factors on the behaviour of primary-care physicians¹⁹⁻²² and the utilization of health-care services within communities.²³

It seems particularly pertinent to use these approaches in attempting to understand the complex pathways to health implied by the broader frameworks of the determinants of health. It may be that community-level variation is unimportant, or that community-level variables that explain this variation cannot be found. But this can only be established, as distinct from assumed, by using specifications of empirical models that enhance rather than

constrain the capacity of the model to reflect the underlying conceptual frameworks.

ACKNOWLEDGEMENTS

François Béland is funded in part by a research scholarship awarded by the National Health Research and Development Programme. We are particularly grateful to Jerry Hurley for helpful comments on several draft versions of the paper and to two anonymous reviewers. We thank Joanne Bannister and Nancy Bishop for help in preparation of the paper.

REFERENCES

- Hancock T. Lalonde and beyond: Looking back at a new perspective on the health of Canadians. *Health Prom* 1986;1:93-100.
- Gunning-Schepers L, Hagen J. Avoidable burden of illness: How much can prevention contribute to health? *Soc Sci Med* 1987;24:945-51.
- World Health Organization, Health and Welfare Canada, Canadian Public Health Association. *Ottawa Charter for Health Promotion*. Ottawa: Health and Welfare Canada, 1986.
- Sagan L. *The Health of Nations*. New York: Basic Books, 1987.
- Evans R, Stoddart G. Producing health, consuming health care. *Soc Sci Med* 1990;31:1347-63.
- Birch S, Stoddart G, Béland F. Modelling the community as a determinant of health. Centre for Health Economics and Policy Analysis Working Paper 97-9, McMaster University, Hamilton, Ontario, 1997.
- Diez-Roux A. Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216-22.
- Duncan C, Jones K, Moon G. Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med* 1993;37:725-33.
- Newacheck P. Access to ambulatory care for poor persons. *Health Serv Res* 1988;23:401-19.
- Turner B. Economic context and the health effects of unemployment. *J Health Soc Behav* 1995;36:213-29.
- Aitken M, Longford N. Statistical modelling in school effectiveness studies. *J Royal Statistical Society Series A*, 1986;149:1-43.
- Rice N, Leyland A. Multi-level models: Applications to health data. *J Health Serv Res Pol* 1996;1:154-64.
- Goldstein H. *Multilevel Models in Educational and Social Research*. London: Charles Griffin, 1987.
- Goldstein H. Non-linear multilevel models, with an application to discrete response data. *Biometrika* 1991;78:45-51.
- Goldstein H. Multilevel cross-classified models. *Sociological Methods and Research* 1994;22:364-75.
- Jones K, Duncan C. People and places: The multi-level model as a general framework for the quantitative analysis of geographical data. Paper presented at the Workshop on Spatial Modelling and GIS, Bristol, UK, June 1994.
- Blaxter M. *Health and Lifestyles*. London: Tavistock/Routledge, 1990.
- Jones K, Duncan C. Individuals and their ecologies: Analysing the geography of chronic illness within a multi level modelling framework. *Health and Place* 1995;1:27-40.
- Kristiansen I, Høltedahl K. Effect of the remuneration system on the general practitioner's choice between surgery consultations and home visits. *J Epidemiol Commun Health* 1993;47:481-84.
- Kristiansen I, Mooney G. The general practitioner's use of time: Is it influenced by the remuneration system? *Soc Sci Med* 1993;37:393-99.
- Scott A, Shiell A. Do fee descriptors influence treatment choices in general practice? A multi-level discrete choice model. University of Aberdeen, Health Economics Research Unit, Aberdeen, 1995.
- Scott A, Shiell A. A multi-level logit model of the effect of competition on general practitioners' behaviour. Paper presented at the Third European Conference on Health Economics, Stockholm, Sweden, August 20-22, 1995.
- Carr-Hill R, Hardman G, Martin S, et al. A formula for distributing NHS revenues based on small area use of hospital beds. York, UK: University of York, Centre for Health Economics, 1994.

Received: July 7, 1997

Accepted: September 25, 1998