AN INVESTIGATION OF THE ASSOCIATION BETWEEN CONTEXTUAL VARIABLES AND THE RISK OF SMALL FOR GESTATIONAL AGE BIRTH IN QUEBEC

Nathalie Savard[1], Louis-Paul Rivest[2] and Patrick Levallois[3]

ABSTRACT

Until now, authors have rarely put emphasis on how small for gestational age (SGA) birth risk is associated with contextual factors. Contextual features investigated with SGA were almost exclusively of economic nature. In this cross-sectional study of 667 254 births to mothers living in Quebec from 2000 to 2008, the authors investigate whether community-defined neighbourhood social and economic features account for differences in risk of SGA. Census data of individuals as well as Canadian Community Health Survey data of 100 832 residents in 143 neighbourhoods of Quebec excluding known unwanted effects resulting from survey methodology are used to obtain tertiles of neighbourhood features that comprise material deprivation, social and racial isolation, cigarette use, food insecurity and access to services. Factor analysis highlights the key concepts measured by contextual variables. A compromise between stepwise and best subset selection is used to select variables to include in the multivariate logistic model on SGA. Survey variables on food insecurity and cigarette use are selected in the model with census data on material deprivation as reflected by education, racial isolation as measured by non-official language usage, social isolation as reflected by both inverse household size and by proportion not married. A one tertile increase of each variable but cigarette use is associated with different SGA risks when controlling for all other characteristics in the model.

KEY WORDS: Small for Gestational Age; Fetal Growth retardation; Social Environment; Health Behaviour; Health Surveys; Logistic Models.

RÉSUMÉ

Jusqu'à présent, peu d'emphase a été mise sur la façon dont le risque de faible poids pour l'âge gestationnel (*SGA*) est associé aux facteurs du contexte social. Les caractéristiques du contexte investiguées en lien avec le *SGA* sont presque uniquement de nature économique. Dans cette étude transversale de 667 254 naissances de mères qui résident au Québec entre 2000 et 2008, les auteurs investiguent si des variations de caractéristiques sociales et économiques de l'environnement sont associées à des différences de risque de *SGA*. Des données du recensement des individus ainsi que des données de l'Enquête de Santé dans les Collectivités Canadiennes à propos de 110 832 résidents de 143 voisinages du Québec qui excluent des effets indésirables résultant de la méthodologie d'enquête sont utilisées pour obtenir des tertiles de caractéristiques des voisinages incluant la défavorisation matérielle, l'isolement social et racial, l'utilisation de cigarette, l'insécurité alimentaire et l'accès aux services de santé. Une analyse factorielle met en valeur les concepts-clé mesurés par l'information contextuelle. Un compromis entre les méthodes de sélection de variables pas-à-pas et du meilleur sous-ensemble est utilisé pour choisir les variables à inclure dans le modèle logistique multivarié de *SGA*. Les variables d'enquête d'insécurité alimentaire et d'utilisation de cigarette sont retenues dans le modèle en plus de celles de recensement sur la défavorisation matérielle reflétée par l'éducation, l'isolement racial reflété par l'usage d'une langue non-officielle ainsi que l'isolement social reflété à la fois par l'inverse de la taille des ménages et la proportion non-mariée. Une augmentation d'un tertile de chacune des variables à l'exception de l'utilisation de cigarette est associée à un risque différent de *SGA* lorsque toutes les autres caractéristiques du modèle demeurent constantes.

MOTS CLÉS : Small for Gestational Age; Retard de croissance intra-utérin; Environnement social; Comportement de santé; Enquête de santé, Modèle logistique.

1. INTRODUCTION

In public health, the measure of small for gestational age (SGA) is frequently used to indicate whether or not the foetus has suffered from foetal growth retardation. Individuals born SGA are more prone to develop chronic illnesses such as type II diabetes, hypertension, metabolic syndrome and coronary diseases (World Health Organization, 2006). Until

[1] Nathalie Savard, Institut National de Santé Publique du Québec, 1000 route de l'Église, Québec (Québec) G1V 3V9, nathalie.savard.8@ulaval.ca
[2] Louis-Paul Rivest, Université Laval, 1045 rue de la médecine, Québec (Québec) G1V 0A6, Louis-Paul.Rivest@mat.ulaval.ca
[3] Patrick Levallois, Institut National de Santé Publique du Québec and Université Laval, 945 avenue Wolfe, Québec (Québec) G1V 5B3, Patrick.Levallois@msp.ulaval.ca

today, authors have put more emphasis on maternal determinants of SGA and less on contextual risk factors even though contextual features or residents' opinion about contextual features appear to be associated with health (Jarman et al., 1991; Pampalon et al., 2009a). Indeed, several individual risk factors including maternal age, parity, weight gain, and smoking have been identified, but contextual features have only recently been investigated and resulted in a set of risk factors including socioeconomic deprivation, race, immigration and social isolation (Elo et al., 2009; Nkansah-Amankra et al., 2009; Nkansah-Amankra et al., 2010). Those factors are almost exclusively of economic nature. They proxy the more complex underlying social processes rather than focus on social processes themselves and how those processes vary from neighbourhood to neighbourhood and contribute to the gaps in health (Braveman et al., 2005) or birth outcomes (Morenoff, 2003). A better understanding of how and why neighbourhoods affect birth outcomes may suggest new public health promotion, prevention and intervention strategies at the community level.

In this paper, we seek to distinguish important social and economic factors and evaluate the association between these factors and Quebec's neighbourhood risk of SGA. Section 2 presents the methodology used and section 3, the results. A discussion is provided in section 4.

2. METHODOLOGY

2.1 Study population and setting

The population of this cross-sectional study consists of live singleton births to mothers living in Quebec, Canada from 2000 to 2008, with the exception of the northern regions of *Nord-du-Québec*, *Terres-Cries-de-la-Baie-James* and *Nunavik*. Births with available information on sex, weight, gestational age, mother's age as well as postal code of residence on the birth certificate are used. Neonates have to be aged between 22 and 43 completed weeks so that the SGA status can be established and to have a plausible birth weight-gestational age combination according to a criteria detailed in Alexander et al. (Alexander et al., 1996).

Neighbourhood limits are defined by the borders of the Local Community Service Center territories (LCSC). They have an average size of 46 727 individuals. Also, LCSCs are basically the hierarchic level at which Quebec's perinatal services are organized.

2.2 Variables

Neonates with a weight, gestational age and sex combination located below the 10th percentile on a standardized scale are SGA. The Canadian reference used has been built by Kramer (Kramer et al., 2001).

Neighbourhood-level portraits are drawn from Statistics Canada's Canadian Community Health Survey (CCHS) (Béland, 2002) as well as from census profiles. Questions used were selected on the basis of a literature review and had to meet the following criteria: a link has been suspected with perinatal outcomes at an individual or contextual level, it was of public health interest and its answer was available at the neighbourhood level.

2.3 Source of data

CCHS is a cross-sectional survey that is conducted in four year-cycles to this day (2000-2001, 2003, 2005 and 2007-2008) with multiple data collection methods. We utilize survey questions that have not been modified and that focus on the same population from one cycle to the next. To increase statistical power, the four survey data cycles are pooled following the approach described by Thomas (Thomas et al., 2009). Answers to the survey are known to be potentially affected by data collection method and slight differences in methodology between cycles (Béland, 2002; Lapointe et al., 2007; Thomas et al., 2009). Hence, for every question, we estimate the baseline, LCSC, data collection, cycle and data collection-cycle effects with a logistic regression model that accounts for the complex sampling design. We also examine the variance of the LCSC mean proportion estimate (obtained from 500 replicate bootstrap-variances), and, for relatively stable estimates (coefficient of variation $\leq$ 33%), we rebuild proportion-like values from the baseline and LCSC effects alone using the formula below: $\frac{e^{\beta_0+\beta_{LCSC}x_{LCSC}}}{1+e^{\beta_0+\beta_{LCSC}x_{LCSC}}}$.

LSCSs census measurements are obtained by aggregating Census tract and Census subdivisions data. Census tract profiles are preferred when they are available (which occurs when the statistical area is classified as a census metropolitan area or

as a census agglomeration); otherwise, census subdivision profiles are selected. The resulting information about each LCSC consists of 2000 and 2006 population-weighted means for the tracts and subdivisions that are part of the LSCS territory in the 2006 geography reference files (Postal Code Conversion File and M-34).

2.4 Statistical analysis

Since births that occurred outside of the province in 2008 might not be included in the file already, we carry out a t-test to insure there is no SGA difference before pooling 2008 data with 2000 to 2007 data. To highlight the underlying concepts measured by LCSC economic and behavioural features, an exploratory factor analysis is performed. First, all of the arcsine of square root proportions and the centered-reduced values of means were included. Then, the independent and inappropriately-represented variables (no communality exceeding |0.5|) are removed so that a final analysis can be run. The Heywood method is used when anomalies about communality estimates are encountered.

In the regression models for the proportion of SGAs, the LCSC and Census explanatory variables are coded in terms of tertiles, 1, 2 and 3, for the lower middle and upper part of the distribution. At this stage of the analysis this yielded results that are easier to interpret than those obtained by using the factors resulting from the factor analysis as explanatory variables. Tertiles are treated as continuous. A series of univariate logistic regressions is first adjusted to data. Then, a multivariate logistic regression model including economic and social features is fitted. For variable selection, we use a compromise between stepwise and best subset selection (Shtatland et al., 2001) since objections relating to the bias-variance tradeoff have been brought forth both in the case of the first (Hastie et al., 2009; Shtatland et al., 2001) and the second method (Hastie et al., 2009). We rely on the Schwarz information criterion for the comparison of non-nested models (Hastie et al., 2009; Shtatland et al., 2001) in the complete stepwise sequence to select the optimal number of variablesk_{SC}. Then, we identify and fit three of each $k_{SC} \pm 1$ term best-subset models. From there, results of three models presenting an acceptable goodness-of-fit Hosmer and Lemeshow statistic are compared and enable us to lower model uncertainty. The model that presents the best fit and number of term combination is finally adjusted to SGA. Regression models compensate for the presence of data overdispersion.

Quebec's Access to Information Commission as well as *Université Laval*'s Ethics Committee approved this research. Analysis are carried out using SAS 9.2. Results are interpreted at the 1% level.

3. RESULTS

3.1 Participants

Of the 676 165 singleton births recorded in all of Quebec's regions, 7 379 are of mothers from *Nord-du-Québec*, *Terres-Cries-de-la-Baie-James* and *Nunavik*, 850 cannot be assigned a LCSC number and 642 have no SGA status. The t value for the comparison of LCSC proportion of SGA according to data temporary status (proportion on LCSCs with 2000-2007 birth data versus proportion on LCSCs with 2008 data) is 1.61 with 267.74 degrees of freedom (p-value = 0.109). Therefore, temporary and permanent birth data are analyzed together. The remaining 667 254 births distributed from the 143 neighbourhoods show an average neighbourhood SGA risk of 8.22%. The CCHS survey was not designed for LCSC-level estimation and high coefficient of variations of some neighbourhood information was observed. Measurements with a CV larger than 33% were considered as missing. It led to missing value of the contextual variables in 25 LCSC territories out of 143.

3.2 Contextual factors

From the factor analysis, variables of social support as well as high alcohol consumption, low fruit and vegetable consumption, nutritional insecurity and access to health services are removed because they are independent or badly represented. Three factors account for 97.3% of the common variability of the remaining variables. These factors mainly represent material or educational deprivation, racial isolation and social isolation.

Table 1 – Relative risks of SGA according to an increase of one tertile of neighbourhood feature

Census	N	RR	[99 % C.I.]	Survey	N	RR	[99 % C.I.]
Academic degree – without	143	**1.05**	[1.01; 1.10]	Frequent high alcohol consumption	122	1.03	[0.99; 1.08]
Age *	143	1.03	[0.99; 1.08]	Food insecurity	118	**1.09**	[1.05; 1.13]
Household size **	143	**1.07**	[1.03; 1.11]	Less than five fruit/ vegetable	140	**1.04**	[1.00; 1.09]
Household size of one	143	**1.10**	[1.06; 1.13]				
Immigrants	142	**1.06**	[1.02; 1.11]	Low emotional support	126	1.00	[0.96; 1.05]
Income **	143	**1.08**	[1.04; 1.12]	Low tangible social	125	0.99	[0.94; 1.03]
Income – low	143	**1.10**	[1.07; 1.14]	No physical activity	134	**1.08**	[1.04; 1.12]
Non-official language	143	**1.07**	[1.03; 1.12]	Poor access to services	129	1.00	[0.96; 1.04]
Not married	143	1.02	[0.98; 1.06]	Smoking cigarette daily	141	**1.05**	[1.01; 1.09]
On foot/biking to work	143	**1.07**	[1.03; 1.11]	Someone smoking in home	141	1.01	[0.97; 1.06]
Territory †	143	1.00	[0.93; 1.08]				

* Mean tertile; ** Inverse mean tertile; † Rural vs urban; N: number of LCSC; RR: Relative risk; C.I.: Confidence interval.

Results of the univariate logistic regression models of SGA on LCSC contextual variables built from the census and the 100 832 CCHS responses are presented in Table 1. An increase of one tertile of the proportion without academic degree is associated with a 1.05 times higher SGA risk. Also, increases of one tertile in inverse mean household size, inverse mean income and in proportion: of household sizes of one, of immigrants, with low income, using a non-official language, going on foot or biking to work, with food insecurity, eating less than five fruit and vegetables a day, without physical activity and smoking cigarette daily are associated with significantly higher risks of SGA.

The optimal number of variables for the multivariate models of SGA is six. All of the three best five to seven variable models include terms of proportion without degree, proportion using a non-official language, inverse mean household size as well as proportion not married. One model also includes a term for tertile of proportion smoking cigarette daily, another model includes a term for tertile of proportion with food insecurity and the last one includes both. The latter best fits the data and has the optimal number of variables.

Results for the final model fit on all SGA births are shown in Table 2. At the 1% significance level, with all other variables held constant, a one tertile increase in material deprivation as reflected by the proportion of residents without academic degree is associated with a 1.07-fold increase risk of SGA. A similar conclusion of a 1.08-fold increase can be drawn with racial isolation as reflected by the tertile of the proportion using a non-official language. A one-tertile increase in social isolation reflected by the inverse average household size is associated with a significant 1.07-fold increased SGA risk, whilst a one-tertile increase of proportion not married is associated with a 0.945-fold lower risk. An increase of one tertile of a feature that was not well represented in the factor analysis, food insecurity, is also associated with a higher SGA risk when all others variables are held constant. The contextual feature of proportion smoking cigarette daily is not significantly associated with SGA at the 1% level of significance.

Table 2 – Relative risks of SGA according to an increase of one tertile of a neighbourhood feature accounting for the presence of other features in the model

Factor	Variable	D.F.	RR	Estimate	Std. Err.	Wald X^2	*p*-val.
Material	Without degree	1	1.07	0.07	0.02	18.24	**0.00**
Racial	Non-off. language	1	1.08	0.08	0.01	29.07	**0.00**
Social	Household size *	1	1.07	0.06	0.02	18.44	**0.00**
Social	Not married	1	0.95	-0.06	0.02	13.43	**0.00**
No specific factor	Food insecurity	1	1.04	0.04	0.01	7.46	**0.01**
No specific factor	Smoking cigarette	1	1.03	0.03	0.01	3.36	0.07

* Inverse mean tertile; D.F.: Degrees of freedom; RR: Relative risk; Std.Err.: Standard Error; *p*-val.: *p*-value.

4. DISCUSSION

In this study, we find significant associations of contextual material deprivation, contextual racial isolation, contextual social isolation and contextual food insecurity with SGA when all other variables are held constant.

Globally, our measure of SGA is calculated independently of contextual variables for the entire population and the census provides accurate measures of the independent variables. Our results are nevertheless subject to small bias toward the null value for the three following reasons. First, severely exposed foetuses might not have survived until birth and not have been included in the study. Second, LCSC information does not describe the relocated mothers' context. Last, explanatory variables are built using self-reported data. Our results might also be subject to confounding of unmeasured factors. Since prematurity is potentially associated with SGA, stratified analysis by prematurity will eventually be included as part of this project.

Two of the factors underlying our contextual data, material deprivation and social isolation, are part of a well known population-specific material and social deprivation index (Pampalon et al., 2009b). Information about racial isolation, when added to material and social deprivation, has also been incorporated into a deprivation index that was more correlated to infant health than to general health (Jarman et al., 1991). This work has identified birth outcomes risk factors of contextual socioeconomic measures, contextual racial and social isolation that have also been pointed out in previous studies (Elo et al., 2009; Nkansah-Amankra et al., 2009; Nkansah-Amankra et al., 2010). We find a novel association of social context of neighbourhood food insecurity with SGA.

Finally, we note that observed contextual associations might represent real effects of the context on the mother, effects of the mother's individual characteristics or both. To help dissociate contextual effects from individual effects, individual characteristics should be included in further analysis.

ACKNOWLEDGEMENTS

This project was partly supported by the *Institut national de santé publique du Québec*, the *Centre interuniversitaire Québécois de statistiques sociales*, the *Centre de recherche en aménagement et développement* of *Université Laval* and the Faculty of medicine of *Université Laval*. The authors would like to thank Suzanne Gingras, Robert Pampalon as well as members of the sixth CIHR-NICHD-IHDCYH Summer Institute in Reproductive and Perinatal Epidemiology for helpful discussion.

REFERENCES

Alexander,G.R., Himes,J.H., Kaufman,R.B., Mor,J., & Kogan,M. (1996). "A United States national reference for fetal growth". *Obstetrics and gynecology*, **87(2)**, 163-168.

Béland,Y. (2002). "Canadian community health survey - methodological overview". *Health reports*, **13(3)**, 9-14.

Braveman,P.A., Cubbin,C., Egerter,S., Chideya,S., Marchi,K.S., Metzler,M., & Posner,S. (2005). "Socioeconomic status in health research: one size does not fit all". *The journal of the American Medical Association*, **294(22)**, 2879-2888.

Elo,I.T., Culhane,J.F., Kohler,I.V., O'Campo,P., Burke,J.G., Messer,L.C., Kaufman,J.S., Laraia,B.A., Eyster,J., & Holzman,C. (2009). "Neighbourhood deprivation and small-for-gestational-age term births in the United States". *Paediatric and perinatal epidemiology*, **23(1)**, 87-96.

Hastie,T., Tibshirani,R., & Friedman,J. (2009). *The elements of statistical learning: data mining inference and prediction* (2nd ed). New York: Springer Science.

Jarman,B., Townsend,P., & Carstairs,V. (1991). "Deprivation indices". *BMJ*, **303(6801)**, 523.

Kramer,M.S., Platt,R.W., Wen,S.W., Joseph,K.S., Allen,A., Abrahamowicz,M., Blondel,B., & Breart,G. (2001). "A new and improved population-based Canadian reference for birth weight for gestational age". *Pediatrics*, **108(2)**, E35.

Lapointe,F., & Courtemanche,R. (2007). "Avis méthodologique pour la combinaison de plusieurs cycles de l'Enquête sur la santé dans les collectivités canadiennes". (pp.1-12). Quebec Statistics Institute.

Morenoff,J.D. (2003). "Neighborhood mechanisms and the spatial dynamics of birth weight". *American journal of sociology*, **108(5)**, 976-1017.

Nkansah-Amankra,S., Dhawain,A., Hussey,J.R., & Luchok,K.J. (2009). "Maternal social support and neighborhood income inequality as predictors of low birth weight and preterm birth outcome disparities: analysis of South Carolina pregnancy risk assessment and monitoring system survey, 2000-2003". *Maternal and child health journal*.

Nkansah-Amankra,S., Luchok,K.J., Hussey,J.R., Watkins,K., & Liu,X. (2010). "Effects of maternal stress on low birth weight and preterm birth outcomes across neighborhoods of South Carolina, 2000-2003". *Maternal and child health journal*, **14(2)**, 215-226.

Pampalon,R., Hamel,D., & Gamache,P. (2009a). "A comparison of individual and area-based socio-economic data for monitoring social inequalities in health". *Health reports*, **20(4)**, 85-94.

Pampalon,R., Hamel,D., Gamache,P., & Raymond,G. (2009b). "A deprivation index for health planning in Canada". *Chronic diseases in Canada*, **29(4)**, 178-191.

Shtatland,E.S., Cain,E. & Barton,M.B. (2001). "The perils of stepwise logistic regression and how to escape them using information criteria and the output delivery system".

Thomas,S., & Wannell,B. (2009). "Combining cycles of the Canadian Community Health Survey". *Health reports*, **20(1)**, 53-58.

World Health Organization (2006). "Promoting optimal fetal development. Report of a technical consultation". (pp.1-61).