BLOOD DONOR CORRELATES: CANADA AND TORONTO

AN EXAMINATION OF BLOOD DONOR CORRELATES: CANADA AND TORONTO

By

P.J. SABERTON, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2010) McMaster University

(School of Geography and Earth Sciences) Hamilton, Ontario

TITLE: An Examination of Blood Donor Correlates: Canada and Toronto

AUTHOR: P.J. Saberton, B.Sc. (Brock University)

SUPERVISOR: Professor K.B. Newbold

NUMBER OF PAGES: vii, 61

ABSTRACT

BACKGROUND: Canada's aging, heterogeneous population presents a challenge with respect to maintaining a sufficient national blood supply. Patterns of donation and correlates of donor data will be identified through analysis of the data.

DATA: Geo-coded blood donor and donor clinic data are provided by Canadian Blood Services. Blood donor data is provided for the fiscal year 2006-2007 indicating the total number of donors for each Canadian postal code, excluding the province of Québec. Potential correlates of blood donation are selected based on social and economic characteristics, as well as descriptors of city size and geographical location in the urban hierarchy measures of accessibility, and capacity of donor clinics.

METHODS: Data is aggregated to $n = 3{,}746$ census tracts in 40 Census Metropolitan Areas (CMA) across the country and then to $n = 992$ census tracts for the Toronto CMA. The number of donors per population in each Canadian census tract is regressed against the set of potential donation correlates. For the Toronto CMA model, the donor count in each census tract is regressed against similar potential correlates.

RESULTS: A number of factors are found to influence blood donation in Canada including the proportion of younger residents, English ability, proportion of people with immigrant status, higher education, and a population-based measure of accessibility.

These findings are confirmed when a model involving the city of Toronto is created. The Toronto model achieves similar correlates as the national model with the addition of variables that are unique to the city of Toronto. These unique attributes involve travel, employment, and gender.

CONCLUSION: While a number of correlates of blood donation are observed across Canada, important contextual effects across metropolitan areas are highlighted. These contextual effects are supported by the uniqueness of the Toronto model's secondary correlates. The thesis concludes by summarizing what these findings contribute to the field of blood donation in Canada. Further mention is also given regarding the role of spatial filters as a tool in regression analysis.

ACKNOWLEGEMENTS

I would like to firstly thank Bruce Newbold for his guidance, assistance, and insight into the world of health geography. Beyond his leadership with this project, Bruce provided me with far more opportunities and experiences than I could have ever imagined were possible as a graduate student. These opportunities included being able to attend and present my work at numerous events and conferences. As a part-time student, these experiences made me feel more involved in the education environment and helped me better understand the field of health geography and geography in general.

Secondly I would like to thank Antonio Páez for his guidance and instruction in the field of spatial analysis and econometrics. Antonio has always shown me extensive amounts of patience and understanding as I learned the concepts of these areas of study. The wisdom that Antonio has taught me is indescribable.

I would also like to thank Nancy Heddle for her instruction in the field of epidemiology. As a health professional, this field intrigues me and Nancy has helped me understand how it relates to health geography.

Lastly, I would also like to thank my family for their support throughout the graduate study process. How extensive your love and understanding has been in helping me achieve my goals, cannot be overstated.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS, CHARTS, DIAGRAMS

CHAPTER 1: Introduction

1.1 Research Outline

The factors and influences involved in the Canadian blood donation process are comprised of many issues and aspects. One of the leading issues consists of interpreting which attributes motivate people to donate. What makes this issue an important topic of discussion is the fact that the blood donor rate in Canada, while adequate on a daily or weekly basis, is consistently low. This can be partly attributed to the fact that Canada's national blood collection system is a volunteer based structure. Because of this, a better understanding of motivating factors involved in blood donation will likely lead to more efficient marketing along with improved donor turnout. The objective of this thesis is to discover these motivating factors through regressive models. The model correlates will hopefully assist in providing insight into ways of increasing and maintaining high donor turnout rates. By identifying the characteristics of current donors donor officials gain a better understanding of who to target when developing marketing and policy outlines. The correlates, or independent variables, in the model are represented by census characteristics that have shown in previous works to significantly relate to donor participation. Once these models are developed, further examination is necessary to discover whether patterns of blood donation follow regional or city-size patterns.

At the onset of this research, it was expected that a national blood donor model could easily be developed that would help identify which populations are more inclined to donate within Canada's urban centres. However, exploration of blood donor data reveals clustered, spatial patterns that are not captured by the model's independent variables. The problem of autocorrelation is addressed by introducing spatial filters into the model as independent variables. This removes the latent pattern from the residual part of the model and replaces it into the model itself by way of the synthetic filter variable.

The spatial filtering process proves to be a good complement to the initial variables as its inclusion in the regressive model helps improve the overall fit of the model. The completed model provides the answer to our initial goal which is to discover the characteristics that best describe Canadian donors. However, the model does highlight the unexpected issue of contextual effects between Canadian cities. This research question leads to the development of a second paper focusing on the donation correlates for the city of Toronto. The same modeling procedure is applied to the Toronto model where spatial filters are used to explain autocorrelation and improve the fit of the model. The Toronto model provides a platform to take our research one step further where the size of the spatial filter is reduced and replaced in part by new variables that correlate highly with the filter. The result is a regressive model with a reduced synthetic element, as well as the introduction of new variables that are unique to Toronto donors.

Both papers answer the main research objective, which is to identify the characteristics of blood donors in Canada. The issue of autocorrelation is addressed using spatial filters and the process of identifying the contextual effects within the data has begun. This research is immature in that models for other Canadian cities need to be developed before an effective policy can be designed. Once a better understanding of the contextual effects is achieved, concrete marketing and policy practices can begin. It is also hoped that the models completed here involving spatial filters help researchers in other areas of social science accept their use as a tool to eliminate or explain any autocorrelation present in their data.

1.2 Chapter Outline

The work presented here involves two research papers. The first paper, which is presented in the second chapter, is comprised of a national blood donor perspective where models are developed that represent an overview of Canadian blood donation. The results of this paper lead into the development of the second paper in Chapter 3, where an individual model is developed for the city of Toronto. Because the modeling techniques for both papers are similar and involve the same data there is considerable overlap in the two chapters. The fourth chapter ties the two papers together and summarizes the thesis work.

CHAPTER 2: Geographical Variations in the Correlates of Blood Donor Turnout Rates: An Investigation of Canadian Metropolitan Areas.[1]

2.1 Abstract

2.1.1 Background: Like other countries, Canada's population is aging, and the implications of this demographic change need to be better understood from the perspective of blood supply. Analysis of donor data will help to identify systematic patterns of donation and its correlates.

2.1.2 Data: Geo-coded blood donor and donor clinic data are provided by Canadian Blood Services. Blood donor data is provided for the fiscal year 2006-2007 indicating the total number of donors for each Canadian postal code, excluding the province of Québec.

---

[1] Published in *International Journal of Health Geographics* 2009, **8:**56.

Potential correlates of blood donation are selected based on social and economic characteristics, as well as descriptors of city size and geographical location in the urban hierarchy measures of accessibility, and capacity of donor clinics.

2.1.3 Methods: Data is aggregated to $n = 3{,}746$ census tracts in 40 Census Metropolitan Areas (CMA) across the country. The number of donors per population in a census tract is regressed against the set of potential donation correlates. Autocorrelation is tested for and results adjusted to provide parsimonious models.

2.1.4 Results: A number of factors are found to influence donation across the country, including the proportion of younger residents, English ability, proportion of people with immigrant status, higher education, and a population-based measure of accessibility.

2.1.5 Conclusion: While a number of correlates of blood donation are observed across Canada, important contextual effects across metropolitan areas are highlighted. The paper concludes by looking at policy options that are aimed toward further understanding donor behaviour.

2.2 Background

Blood products play an important role in modern medical procedures that can both save and extend life. It is therefore critical for any health system to ensure that the volume of blood supply is sufficient to satisfy the demand, and Canada is no exception. In Canada, with the exception of Québec, the organization responsible for collecting and distributing the national blood supply is Canadian Blood Services (CBS). This is a national, not for-profit charitable organization that operates 40 permanent collection sites and more than 20,000 donor clinics annually, and is charged with overseeing the safety of the blood supply, educating the public and recruiting blood donors. CBS operates under guidelines that establish that blood donation is a voluntary endeavour open to anyone in good health, 17 years or older, and weighing at least 50 kg [1]. To ensure the safety of the blood supply, a number of restrictions are implemented by CBS that essentially define the meaning of being in "good health". Thus, there are a number of disqualifying conditions that impede a potential donor from becoming one, including having tested positive for West Nile Virus (WNV) and infectious diseases [2]. Despite these safety restrictions, there is still a large pool of potential donors to support the system, estimated at about "12.5 million eligible donors in Canada" [3]. The reality, however, is that donor participation tends to be limited due to concerns for health effects and a general lack of education on the part of the public regarding the importance and lack of risk involved in donating blood. There is evidence, for instance, that up to twenty-five percent of Canadians mistakenly believe that donating blood is less than completely safe [4]. The negative effect becomes evident when considering the disappointingly low number of actual donors: of the potential pool, it is estimated that "only 3 per cent of adult Canadians donate blood while virtually all Canadians will need blood or blood products in their lifetime" [5]. While the system has so far proved sufficient (i.e. medical procedures are not routinely cancelled because of blood being unavailable), relying on such a small percentage of the Canadian population to provide the amounts of blood required to sustain the system in the long term is not advisable (A. Steed, personal communication, July 7, 2008). The limitations of placing the burden of sustaining the blood supply on a small number of donors becomes particularly salient within the context of an aging society, such as Canada's, for which projections indicate that one in five people will be at least 65 years old by 2021, and approximately 6% of the population will be older than 80 [6]. Given general population health trends, this strongly suggests that the actual pool of donors will shrink in the future, at the same time that the number of potential users of blood products will almost certainly grow.

To effectively achieve a reliable supply of blood, an initial large enrolment of first time donors and their subsequent retention as repeat donors is needed. Increasing the total number of Canadians donating blood for the first time is necessary to help meet national demands, while retaining these donors is important because of the higher cost of continuously recruiting new donors. Also, repeat donors provide a safer supply of blood, with a lower incidence of infectious diseases [7]. The key to attracting new donors (and subsequently repeat donors) requires a number of actions, which include: 1) Developing economically efficient and effective marketing campaigns; 2) Finding ways of increasing the convenience of donating, by optimizing operation hours, locations, etc.; and 3) Educating the public about the importance, ease, and lack of risks of giving blood. These are all required elements that could encourage Canadians to make blood donation part of their lifestyles.

The challenge of achieving these objectives is underlined by the difficulties of reaching and serving the large, heterogeneous, target market of all eligible Canadian donors. In order to be efficient and effective, recruitment plans aimed at increasing donor levels must be targeted at untapped populations, or those that are most likely to donate in each part of the country. In this regards, it is known that the patterns of donation vary between cities and regions. Figure 1 shows, for example, the proportion of the Canadian population (excluding Québec) living in Census Metropolitan Areas, or CMAs, defined by Statistics Canada as one or more contiguous municipalities, totaling at least 100,000 inhabitants, and situated around a major urban core with a population of at least 50,000. As seen in the figure, over 70% of the population live in CMAs and slightly above 40% in the 5 largest (population greater than 1 million) metropolitan areas. And, while the overall donor rate for the 40 CMAs is not very different from the population rate (calculated as number of donors over total population, not only eligible population), it can be seen that the largest metropolitan areas in the country tend to have substantially lower donor rates. There are a number of reasons why variations in donation patterns could arise, including varying economic, social, cultural, demographic, and historical factors of a given region, which may affect the motivating forces towards whether or not to donate [8-10]. More generally, there is a noticeable amount of heterogeneity with respect to patterns of volunteering inter-provincially, as well as between varying sized cities [11,12]. This suggests that residents in different areas may display, in addition to varying demographic attributes, different incentives and outlooks on volunteering and health. Therefore, there is a need to better understand the correlates of blood donation. This is a question of considerable interest since the answer may help to identify planning objectives, as well as to define whether outreach and service plans are developed nationally, regionally, or even locally.

With the above considerations in mind, the objective of this paper is to investigate a wide array of correlates of donor turnout rates at the census tract level in order to determine the factors that influence the number of donors (a census tract is defined by Statistics Canada as a small, relatively stable area, usually with a population between 2,500 and 8,000 inhabitants). Furthermore, we also seek evidence of geographical segmentation of the Canadian population in terms of blood donation behaviour. More concretely, our objective is to determine whether the socio-demographic profile of various cities display patterns of similarity based on city size and region of the country, or whether on the contrary, there are geographical variations in the way donation is influenced by various factors. Besides the work of van der Pol et al. [13], there is scarcely any international research into this topic. In Canada, a recently released report explores the situation in Québec [14], but the present paper is the first effort to investigate the situation in the rest of the country. Statistical analysis of donor data helps to identify systematic patterns of donation and its correlates, which in turn provides insights into the potential of various incentives, levels of service, or marketing practices, including whether these need to be directed to a particular city based on its demographic profile, geographic location, or size.

2.3 Data

The research reported in this paper is based on geo-coded donor and donor clinic data provided by Canadian Blood Services for the purpose of this study. Blood donor data were provided for the fiscal year 2006-2007 in aggregate form, indicating the total number of donors for each Canadian postal code, by place of residence. Canadian postal codes range in size from approximately zero to more than 60,000 homes, averaging about 8,000 households for each postal code. Converting this data to the number of donors for each Canadian census tract allows all donor and donor clinic information to be linked to 2006 Census attribute information. The population consists of $n = 3{,}746$ census tracts in 40 Census Metropolitan Areas across the country containing a total of 310,767 unique donors. Number of donors per thousand population are shown in Figure 2 for each of the 40 CMAs in the analysis further illustrating the important differences between donor turnout rates in each location. Donor clinic data were provided in point-based form with longitude and latitude references. The clinic database is exhaustive, and includes a total of 19,671 clinics at approximately 1,600 unique locations.

Potential correlates of blood donation are selected based on social and economic characteristics that have been demonstrated by previous research to correspond to high or low donor turnout. Other variables describe the size of metropolitan areas or are introduced to account for fixed (city-specific) effects in the form of dummy variables. The objective of these dummy variables is to capture any contextual variation specific to a metropolitan area that is not attributable to other explicative factors. Finally, we also introduce measures of accessibility and capacity of donor clinics (see Table 1).

Socio-economic and demographic variables (age, income, employment, and education) are obtained from the Census on the basis of census tract aggregations. We collect variables describing the demographic profile of census tracts. Hollingsworth and Wildman [8] and Burnett [9] indicate that donation rates vary in relation to the number of individuals present for different age groups. In order to capture variations due to demographic structure, four age groups were defined as follows. First, the population of individuals between the ages of 15 and 24 is defined as the 'school aged' population, to include people who are exposed to CBS' school learning programs and on-site donor clinics. While the minimum age for donation is 17, Statistics Canada reports population counts in groups, with the closest matching groups beginning at age 15. The population of individuals aged 25 to 54 is characterized as 'working age donors', who may have less time to donate, and therefore would have lower donor turnout rates. The pre-retirement cohort including those aged 55 to 64, are considered senior donors. The fourth age group is comprised of the population of individuals who are 65 and over. These donors are more likely to receive healthrelated deferments and are potentially large users of the blood supply. A large population of this demographic is expected to have a negative effect on the donor rate.

A variable regarding language is also included. This variable is meant to test whether individuals whose first official spoken language is English are more responsive to CBS educational programs. Thus, the total number of English speaking people for each census tract is selected as an explanatory variable. Given the importance of immigration to the makeup of the Canadian population, the immigrant population is considered a relevant variable as well. The potential effect of this variable is ambiguous. Lacking other evidence, there are plausible explanations both in favour and against donor turnout, for example, if immigrants are more civically minded towards their host country, or on the contrary come from places where volunteering is less well regarded or blood donation is considered more risky.

Caro and Bass [15] found that adults who are employed are more likely to volunteer their time than those who are unemployed. Thus, it is expected that the variable for the total number of unemployed individuals will correlate negatively with donor turnout. Individuals employed in health related fields are exposed to the importance of blood donation in life-saving and life-prolonging procedures, either through training and education, or from work-related experiences. Thus, the number of people employed in health related professions in each CT is expected to have a positive impact on donor turnout. Since rates of donation are known to increase with level of education [16], the number of individuals with a college certificate or higher is included as an explanatory variable. The idea of wealthier individuals being more likely to donate also follows most models including Jirovec and Hyduk [17], and income is therefore expected to have a positive correlation with donation patterns.

In order to identify potential variations in blood donation patterns, we introduce the total population of each Census Metropolitan Area as a macro-level descriptor of each area. This variable is used to test the proposition that there are systematic variations in the correlates of donation following city size.

Finally, we also introduce variables that describe the levels of service provided by blood donor clinics around the country, and how accessible the services are. Accessibility variables are calculated using the two-step floating catchment area proposed by Radke and Mu [18] and applied by Wang and Luo to identify health professional shortage areas [19]. Since some donor clinic locations are permanent and others are temporary or ambulatory, and their levels of services vary widely, instead of simply recording presence we consider for each donor clinic location the number of beds available at the event (B_i) and the total service time in hours (T_i). With this, a measure of service is obtained as total number of beds-hours available at the location. The two-step catchment area procedure begins by centering a catchment area on donor clinic location $i = 1, 2, ..., n_i$ (the number of clinics), and searching all census tract centroids within a threshold (Euclidean) distance d_0 of that location. The level of service at the location is calculated as follows:

$$L_i = \frac{B_i T_i}{\sum_{d_{ij} \le d_0} P_j} \tag{1}$$

where P_j is the population in the census tract j in thousands. The level of service is thus measured in beds-hours per thousand people. The second step of the procedure then consists of "floating" the catchment areas to census tract centroid $j = 1, 2, ..., n_j$ (the number of census tracts) and calculating the level of service accessible to each census tract by adding the level of service of all donor clinics within a distance d_0 of census tract j:

$$ACC_j = \sum_{d_{ij} \le d_0} L_i \tag{2}$$

Accessibility is calculated based on residential population (ACCPOP). With regards to the selection of a critical distance, it has been noted that the use of travel behaviour information can yield valuable information to determine the catchment areas [20]. Unfortunately, very little is known about the travel behaviour of potential donors and no data are available to assess it. In particular, donors are coded by place of residence, but it is unknown whether they travelled from their home, workplace, or other location at the time of donation. Lacking other information, we decide to calculate the accessibility indicators using bands between 1 km and 10 km in 1 km increments. Band selection is based on the statistical fit and properties of the models, as described below.

2.4 Methods

To implement the analysis, the number of individual blood donors divided by the census tract population is taken as the dependent variable to estimate the coefficients of a log-linear regression model. The specification used in this analysis is the following regression model (in matrix form):

$$\log(\boldsymbol{D}) = \mathbf{X}\boldsymbol{\beta} + \varepsilon \quad (3)$$

where $\boldsymbol{D}$ is the vector of donor rates for each census tract; $\mathbf{X}$ is the matrix of explanatory variables; $\boldsymbol{\beta}$ is the vector of regression coefficients; and ε is a vector of independent and identically distributed error terms. Given the geocoded nature of the data, we also test for potential autocorrelation in the residuals of the model. It is well known that autocorrelation leads to statistical problems that may affect the quality of inference and policy prescriptions derived from the analysis [21,22]. We test for autocorrelation in the residuals of the model by means of Moran's I coefficient of spatial autocorrelation [23], implemented using a first order contiguity matrix $\mathbf{W}$ with element $w_{ij} = 1$ if census tracts i and j share a length of border, and 0 otherwise. The matrix is row-standardized so that the sum of elements in a row is always exactly 1. The interpretation of a vector multiplied by a row-standardized matrix is as a moving average, since it gives for each location the average of the values in contiguous areas.

In order to address the issue of residual spatial autocorrelation, we propose to introduce a spatial filter to absorb any detected residual pattern [24]. The filtering approach used in this study is based on the eigenvectors of the following matrix:

$$(\mathbf{I}\text{-}\mathbf{11}'/n)\mathbf{W}(\mathbf{I}\text{-}\mathbf{11}'/n) \tag{4}$$

where **I** is the identity matrix, **1** is a $n \times 1$ vector of ones, and **W** is the contiguity matrix of the system [25]. The eigenvectors represent patterns of latent autocorrelation, and combinations can usually be found that proxy omitted variables responsible for the residual pattern. The precise selection and number of eigenvectors for a filter depends on statistical criteria. Since eigenvectors are orthogonal, it is possible to follow a step-wise procedure whereby an eigenvector is introduced into the model and its significance tested. The first significant eigenvector (in our case the first one to attain a regression coefficient with a p-value of less than 0.10), is multiplied by its corresponding coefficient and introduced as an explanatory variable in the subsequent search for additional eigenvectors. Other significant eigenvectors are incorporated as part of the filter by summing them, after multiplying by their corresponding regression coefficient, to the previous version of the filter. The procedure continues until a desired level of spatial autocorrelation is reached (in our case, a non-significant z-score $\leq |0.05|$). The spatial filter is a synthetic variable, not necessarily meaningful by itself, but useful to remove residual autocorrelation, which improves the statistical quality of the model, all the while helping to ensure that other coefficients in the model are not afflicted by omitted variable bias [26].

2.5 Results

We estimate an initial model (Model 1) with a selection of variables from those shown in Table 2. The dependent variable is transformed using the natural logarithm operator. The dependent variables are not transformed. With few exceptions, the coefficients are significant at the $p < 0.05$ level, and have expected and/or reasonable signs. With respect to the demographic variables, the proportion of people in the 15-24 year range correlates positively with number of donors, as do the pre-retirement cohort of 55- 64 (at a marginal level of significance of $p = 0.0648$). In contrast, as the proportions of working age population and seniors increase, the donor yield tends to decrease. Other variables that correlate positively with donor rates are the proportion of English speakers, highly educated individuals, and the proportion of people employed in health-related occupations. Intriguingly, the effect of wealth is negative: neighbourhoods with higher average household income tend to correlate with a lower number of donors. Similarly, although not unexpectedly, the proportion of immigrants correlates negatively with number of donors.

Extensive analysis of different distance bands for the accessibility variables leads to the adoption of 4 km bands. This distance, although impossible to validate based on any empirical measure of travel behaviour, does not strike us as being unreasonable, since it represents a relatively short trip to the donation site from either the place of residence or the place of employment. Population-based accessibility, contrary to expectations, displays a negative sign, although the coefficient is not significant.

The fit of the model is fairly high, with the model explaining about 65% of the variance, although the large number of city-specific dummy variables (33 out of 39 candidates, not counting a reference city) indicates that there tend to be significant and substantial contextual effects. Spatial autocorrelation analysis of the residuals (which are in the same units as the dependent variable) indicates that unfortunately the assumption of independence cannot be sustained for this model. Calculation of variance inflation factors (vif) in contrast indicates that multicollinearity is not a problem.

In order to investigate the effect of metropolitan area effects and to address the issue of residual spatial pattern we estimate a second model. Autocorrelation is problematic because it can lead to bias and wrong inference. The second model is different from the first one in two important respects. First, we aim at capturing some of the contextual variation embedded in the city-specific dummy variables, and to this end we introduce CMA population as a macro-level descriptor of metropolitan areas. This variable is interacted with other explanatory variables in the model to produce expanded coefficients that consist of a direct and interaction effect that permits a mapping of the net effect as a function of the expansion variable. And secondly, we deal with residual pattern evident from autocorrelation analysis by means of the spatial filtering approach previously described. The results of the analysis appear in Table 3. Note that residual autocorrelation is effectively removed by the spatial filter. The goodness of fit has also improved somewhat, since now the model explains about 74% of the variance.

The results clearly indicate that introducing the CMA population as a variable helps to express contextual variation in a more systematic way: the need for city-specific dummy variables is reduced from 33 to 19. The resulting model is more parsimonious and at the same time more informative. Interaction terms (linear and quadratic) with our CMA population variable were initially attempted for all explanatory variables; all non-significant coefficients were dropped from the analysis. The results for nonexpanded coefficients are in line with the previous model, positive for proportion of people aged 15-24 and 55-64 years (the latter with a marginally significant value of $p = 0.0577$), negative for proportions of seniors, unemployed, and immigrants, and positive for proportion of English speakers and highly educated population. Two variables further lose their significance in this analysis: proportion of people 25-54 years and those in health related occupations. In addition to reducing the number of dummy variables, using CMA population to estimate expanded coefficients also provides some valuable insights into the contextualizing effect of city size on two correlates of blood donation: Average Household Income and Population-based Accessibility. The first variable with a significant expanded coefficient is that corresponding to average household income. This variable was negative in Model 1. Now, the direct effect is seen to be positive and significant, which is more in line with our prior expectations regarding the effect of wealth. Furthermore, the net effect at a given level of the CMA population variable can be calculated as follows:

$$\beta_{\text{AVEHHDINC}} = 0.0127 - 0.0049\, POP_{\text{CMA}} \tag{5}$$

The net effect, shown in Figure 3, reveals an interesting pattern of variation according to city size. Whereas a positive and relatively large association is observed for census tracts in smaller cities, the effect tends to vanish with increasing city size, in fact becoming negative for the largest cities in the system. Thus, whereas wealthy census tracts in smaller cities tend to be more generous in their yield of donors, this is less, or not at all, the case in bigger cities. This is what could be termed a "stingy big-city" effect.

The accessibility variable based on population, which was negative but not significant in Model 1, is found to display significant city size variations, with a non-linear net effect that is calculated as follows:

$$B_{\text{ACCPOP3K}} = -0.00002 - 0.0005\, POP_{\text{CMA}} \tag{6}$$

The (negative) direct effect in this case is revealed to be non-significant, and there is evidence that the effect of accessibility increases with size of population: as seen in Figure 4, improvements in access to levels of service are likely to have a greater impact the bigger the city is.

2.6 Discussion

The analysis indicates that there are a number of common factors that influence donor turnout rates across Canadian metropolitan areas. Some of the factors, furthermore, are shown to display systematic variations according to one contextualizing factor, namely CMA population. The results highlight some of the challenges and opportunities facing the blood supply system in Canada. First, the demographics present a significant challenge, with the most generous age group being the 15-24 year cohort. This is a group that, due to long-term demographic trends will shrink both in absolute and relative terms, at the same time that the senior population (which correlates negatively with donor turnout rates) tends to increase. With respect to the younger population, it is unclear if cultural factors are in play to make a more active cohort remain so over their lifetime, and it is possible that newer younger generations will continue to produce more donors than the current working-age population (25-54 years). On the contrary, if limitations to donor behavior are due not primarily due to cultural factors but rather to other circumstances, for example more strict time constraints as this population integrates more fully as part of the workforce, then a different challenge arises in terms of finding ways to continue to support the altruistic behaviour of young donors as they enter the workforce, for example by increasing the convenience and levels of service at employment-rich locations.

A second result that seems to suggest a challenge is the negative relationship between proportion of immigrants and donor rates, given the importance of immigration in Canadian demographic processes. It could be argued that the effect of immigrants on number of donors is coloured by variations in income, education levels, or language proficiency (q.v. the positive effect of English). A cautious argument, mindful of the potential for committing an ecological fallacy, is that since these variables have already been controlled for in the model, the negative relationship could rather be attributed to cultural factors and attitudes. It is important to note that immigrants are not by any means a homogenous group, and in fact there could be important differences between various immigrant populations. Clearly, more research would be needed to determine the extent to which immigrant's donor behaviour is different from the host population due to socio-economic and cultural reasons.

Education appears to be an important factor that influences donor yield. The positive coefficient for this variable is second only to that associated with the proportion of unemployed population in terms of magnitude. This seems to confirm the importance of education in motivating donors, and it is possible that the objectives of marketing campaigns are easier to comprehend by highly educated individuals. A suggestion would be to target marketing efforts to this population, or to device marketing strategies that more directly speak to segments of the population with various educational attainment levels. Household income tends to have a smaller positive impact or to become negative for larger population centres.

Population-based accessibility to donor clinic services stands out as an important policy variable that needs to be considered, in particular in light of contextual effects that indicate a positive relationship between city size and accessibility.

Our attempt to contextualize the effect of various correlates of donor turnout rates resulted in a more parsimonious and informative model that produced a better fit while resorting to a smaller number of city-specific dummy variables. Out of 39 possible dummy variables, the first model used 33 dummy variables which included all of the Prairie and Atlantic CMAs. Thus, the initial model, while appearing to have significant strength in explaining the variation of donor behaviour across Canada, failed to capture more in full the donor patterns for these regions. Within Ontario, the cities represented are the North-Eastern CMAs such as Sudbury and North Bay as well as some of the cities that contain parts of the main 400 series and QEW highways around Lake Ontario, although beyond these observations there appears to be no definitive pattern within Ontario. In British Columbia the initial model includes only the suburban cities as the dummy variables represent Vancouver and the more rural CMAs in the Eastern parts of the province.

The second model in contrast reduced the need for dummy variables (19) and accounted for two Atlantic CMAs as well as seven Prairie cities. The model also better represents British Columbia by having dummy variables only for some of the Northern CMAs. While the number of Ontario cities included in the model decreases from 16 to 9, the regional representation in the second model is not much different from the initial model where there is no clear pattern in the distribution of dummy variable cities across the province. In this paper our objective has been to develop a big picture of the factors influencing donor turnover rates, and the analysis has successfully identified common factors and, where warranted, their variability according to city size. The continued existence of significant contextual effects, on the other hand, suggests that further analysis at the metropolitan level is necessary.

2.7 Conclusion

Blood products are an essential component of modern medicine and necessary to support many life-saving and life-prolonging procedures. CBS has successfully managed to build an active donor base of approximately 425,000 donors and 916,000 whole blood donations in 2008. However, the total number of donors and donations still needs to be increased in order to meet longer term projections of demand [3]. Considering the fact that turnout rates are but a small proportion of the potential donor pool, concerted action is called for in order to ensure the sustainability of a system that currently relies on a small number of Canadians to provide for the whole country. Actions will necessarily involve campaigns designed to encourage a greater number of Canadians to adopt blood donation as part of their lives, and to facilitate the practice.

The ideal answer to the dilemma of how to increase the nation's supply of blood is to increase the number of new donors and to retain them as repeat donors. This requires an understanding of the correlates of blood donation across the country. In order to support new efforts to encourage blood donation, in this paper we investigated the correlates of donation to determine the factors that may influence donor turnout rates. The results indicate that a number of significant correlates of blood donor turnout rates behave consistently across geographical regions and urban sizes. In some cases, in contrast, patterns of variation across cities of various sizes were detected. Understanding these patterns can assist in the economical efficiency of any marketing plan as well as maximize results by ensuring that both message and services target individuals in each city who are most likely to donate blood.

A number of suggestions and directions for further research are indicated as follows. As previously indicated, it is possible to incorporate travel behaviour information directly into the assessment of catchment areas. Lacking this, in our analysis we have opted for a purely statistical strategy for selecting accessibility bands, and while the results are reasonable, they may or may not reflect the actual distance that people who donate are willing to travel to reach a donor clinic. An important constraint is that at the moment donor's information is collected, CBS asks for place of residence but not place of employment, nor whether the trip to donate was home- or non-homebased. It is therefore currently not possible to determine, for instance, the typical trip length of a donor visiting a clinic. Relatively simple changes to the way data is collected should help to develop a better understanding of the conditions surrounding trips to donor clinics. Along this way lay more refined estimates of catchment areas that could in fact be different for home-based (i.e. residential population) and non-home-based (i.e. employment-based accessibility). This information is relevant for obvious operational reasons.

Along with the spatial characteristics of donor clinics (i.e. their location), another factor of interest is the possibility that younger cohorts are more likely to donate due to cultural or other factors. If time constraints are in fact responsible for the reduced participation of the working-age population, it would be important to understand the time use patterns of people in this age cohort, in order to fine tune, for instance, the hours of operation of donor clinics, in addition to their locations. Presently, to the best of our knowledge, there is no research available on the time use patterns of donors, and CBS does not collect time use information. Collection of time use data appears a promising way to better understand the context of donation. The analysis presented in this paper, being at the aggregate level, does not lend itself to the study of time use patterns, however, and other methods useful to investigate behaviour at the level of the individual would be indicated. Individual level analysis would have in addition, the benefit of circumventing the ecological fallacy, and would provide better tools to disentangle the effect, for instance, of immigration status, income, English proficiency, education, etc. [27] This is a matter of ongoing research.

Table 1: Variables and Definitions

Variable	Definition and units	Min	Max	Mean	SD
Dependent Variable					
DONORS	Donor rate in census tract	1	683	82.96	60.38
Socio-economic and demographic characteristics					
POP15TO24	Proportion of population of 15-24 years of age in census tract	0.04	0.66	0.14	0.03
POP25TO54	Proportion of population of 25-54 years of age in census tract	0.15	0.77	0.44	0.06
POP55TO64	Proportion of population of 55-64 years of age in census tract	0.02	0.26	0.11	0.03
POP65+	Proportion of population of 65+ years of age in census tract	0.00	0.68	0.13	0.07
ENGLISH	Proportion of English-speaking population in census tract	0.03	1.00	0.69	0.20
IMMIG	Proportion of immigrant population in census tract	0.00	0.82	0.26	0.18
UNEMP	Proportion of unemployed population in census tract	0.00	0.12	0.03	0.01
HEALTHOCC	Proportion of population in census tract in health-related occupations	0.00	0.13	0.03	0.01
HIGHED	Proportion of population in census tract with bachelors degree or higher	0.00	0.68	0.21	0.11
Ave Household Income	Average Household income ($10,000)	1.73	81.08	7.91	3.96
Geographic variables					
CMAPOP	Total CMA population (1,000,000s)	0.06	5.11	1.93	1.99
City Name	City-specific dummy variable				
Service variables					
Pop Based Access	Proportion of population-based accessibility within 4 km band (beds-hour/1000 people)	0.00	73.82	9.90	9.07

Table 2: Regression Model 1			
Variable	ESTIMATE	p-value	vif
CONST	-5.3594	0.0000	-
POP15TO24	2.1200	0.0000	1.19
POP25TO54	-0.3102	0.1060	0.83
POP55TO64	0.4833	0.0648	1.42
POP65+	-0.7670	0.0001	1.56
ENGLISH	1.3164	0.0000	3.67
IMMIG	-0.3025	0.0005	4.47
HIGHED	1.2928	0.0000	2.16
Ave Household Income ($10,000)	-0.0042	0.0424	1.73
Unemployment	-8.5297	0.0000	1.61
Health Occupations	2.2787	0.0004	1.64
Pop Based Access (4 km)	-0.0001	0.1871	1.48
CALGARY (AB)	0.4020	0.0000	1.31
EDMONTON (AB)	0.3484	0.0000	1.43
LETHBRIDGE (AB)	0.7668	0.0000	1.08
MEDICINE HAT (AB)	0.2778	0.0030	1.05
RED DEER (AB)	0.5038	0.0000	1.07
KAMLOOPS (BC)	-0.3115	0.0001	1.07
KELOWNA (BC)	0.2961	0.0000	1.08
PRINCE GEORGE (BC)	0.5968	0.0000	1.08
VANCOUVER (BC)	0.0869	0.0001	1.29
WINNIPEG (MB)	0.5394	0.0000	1.33
FREDERICTON (NB)	0.6939	0.0000	1.07
MONCTON (NB)	0.7471	0.0000	1.17
SAINT JOHN (NB)	0.3886	0.0000	1.30
ST. JOHNS (NL)	0.5771	0.0000	1.26
HALIFAX (NS)	0.2439	0.0000	1.30
BARRIE (ON)	0.1815	0.0061	1.07
GUELPH (ON)	0.4340	0.0000	1.04
HAMILTON (ON)	0.1851	0.0000	1.23
KINGSTON (ON)	0.1578	0.0094	1.12
KITCHENER (ON)	0.4216	0.0000	1.13
LONDON (ON)	0.4961	0.0000	1.24
NORTH BAY (ON)	0.5870	0.0000	1.08
OSHAWA (ON)	0.1968	0.0000	1.14
OTTAWA (ON)	0.2935	0.0000	1.83
PETERBOROUGH (ON)	0.3333	0.0000	1.09
SARNIA (ON)	0.3200	0.0001	1.06
SAULT ST. MARIE (ON)	0.2152	0.0080	1.07
ST. CATHARINES (ON)	0.2572	0.0000	1.17
SUDBURY (ON)	1.1393	0.0000	1.20

THUNDER BAY (ON)	-2.9853	0.0000	1.14
WINDSOR (ON)	0.7685	0.0000	1.11
REGINA (SK)	0.6623	0.0000	1.15
SASKATOON (SK)	0.5078	0.0000	1.18
n = 3749, R^2 = 0.647, σ^2 = 0.153, σ = 0.394, *I* (z-score) = 0.2031 (22.0590)			

Table 3: Regression Model 2

Variable	ESTIMATE	p-value	vif
CONST	-4.8351	0.0000	-
POP15TO24	1.5889	0.0000	1.12
POP55TO64	0.3920	0.0577	1.36
POP65+	-0.3936	0.0001	1.38
ENGLISH	0.7055	0.0000	2.03
IMMIG	-1.1326	0.0000	2.66
HIGHED	1.5693	0.0000	1.76
Ave Household Income ($10,000)	0.0127	0.0000	3.62
*CMA Population	-0.0049	0.0000	3.60
Unemployment	-5.1708	0.0000	1.40
Pop Based Access (4 km)	-0.0002	0.0827	2.45
*CMA Population	0.0005	0.0036	2.34
EDMONTON (AB)	0.1866	0.0000	1.16
LETHBRIDGE (AB)	0.4725	0.0000	1.04
RED DEER (AB)	0.4772	0.0000	1.04
KELOWNA (BC)	0.2515	0.0000	1.04
PRINCE GEORGE (BC)	0.3723	0.0000	1.04
WINNIPEG (MB)	0.1264	0.0000	1.10
FREDERICTON (NB)	0.2218	0.0011	1.04
MONCTON (NB)	0.8303	0.0000	1.08
HALIFAX (NS)	0.0742	0.0281	1.12
KITCHENER (ON)	0.7040	0.0000	1.06
LONDON (ON)	0.2216	0.0000	1.07
NORTH BAY (ON)	0.3275	0.0000	1.04
OSHAWA (ON)	0.6779	0.0000	1.09
OTTAWA (ON)	-0.1509	0.0001	1.42
PETERBOROUGH (ON)	0.0481	0.0563	1.03
SUDBURY (ON)	0.2668	0.0000	1.09
THUNDER BAY (ON)	-3.3342	0.0000	1.11
WINDSOR (ON)	0.3924	0.0000	1.07
REGINA (SK)	0.3950	0.0000	1.06
n = 3749, R^2 = 0.739, σ^2 = 0.113, σ = 0.338, *I* (z-score) = -0.0024 (0.4872)			

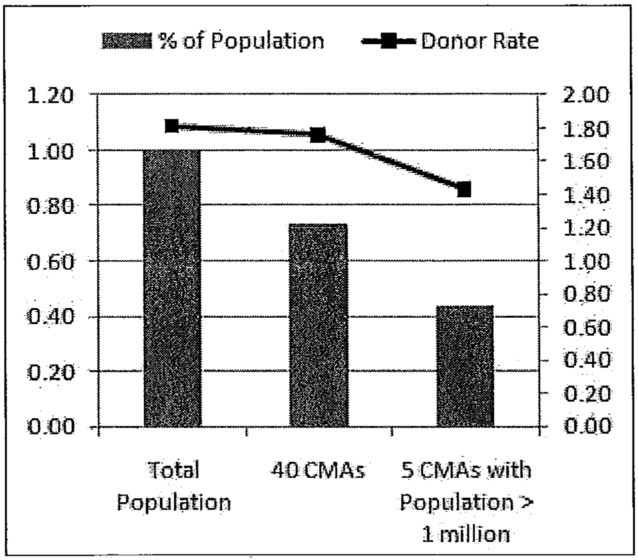

Figure 1 Donor rates by geographical context.

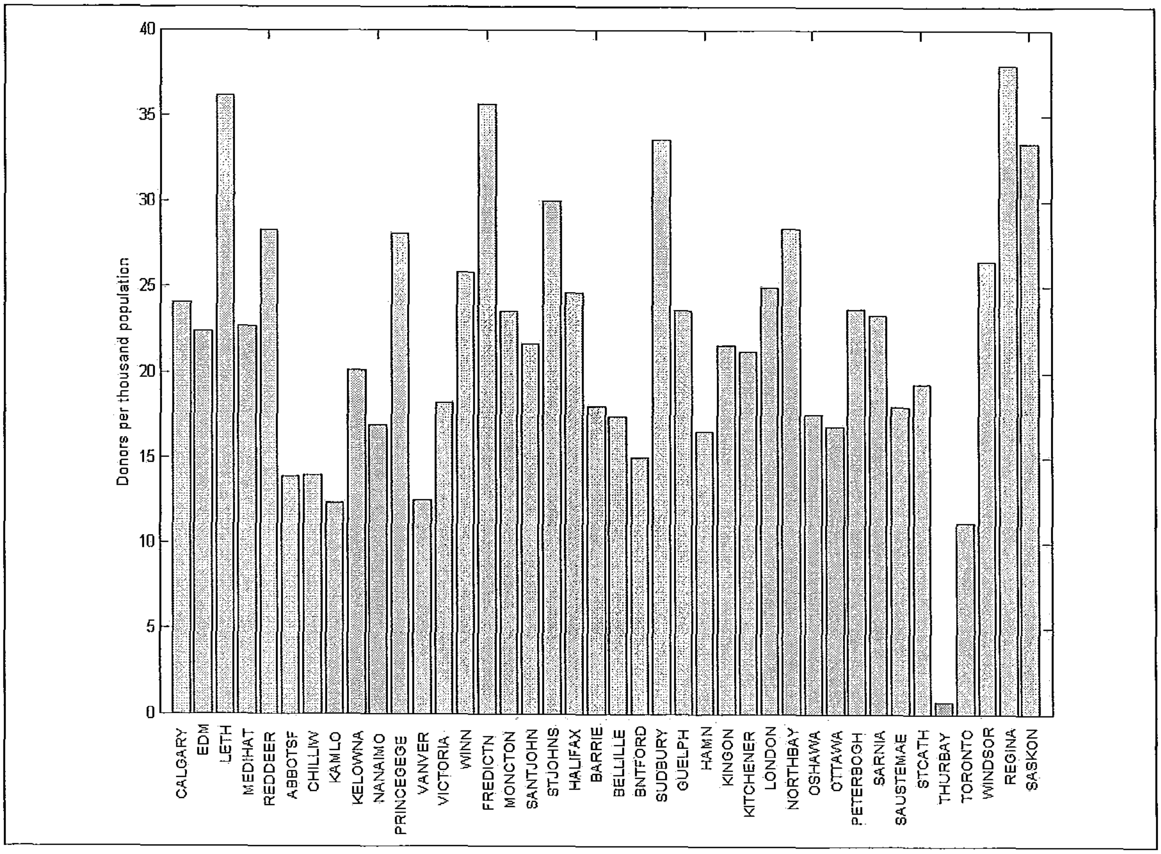

Figure 2 Donors per thousand population in target CMAs.

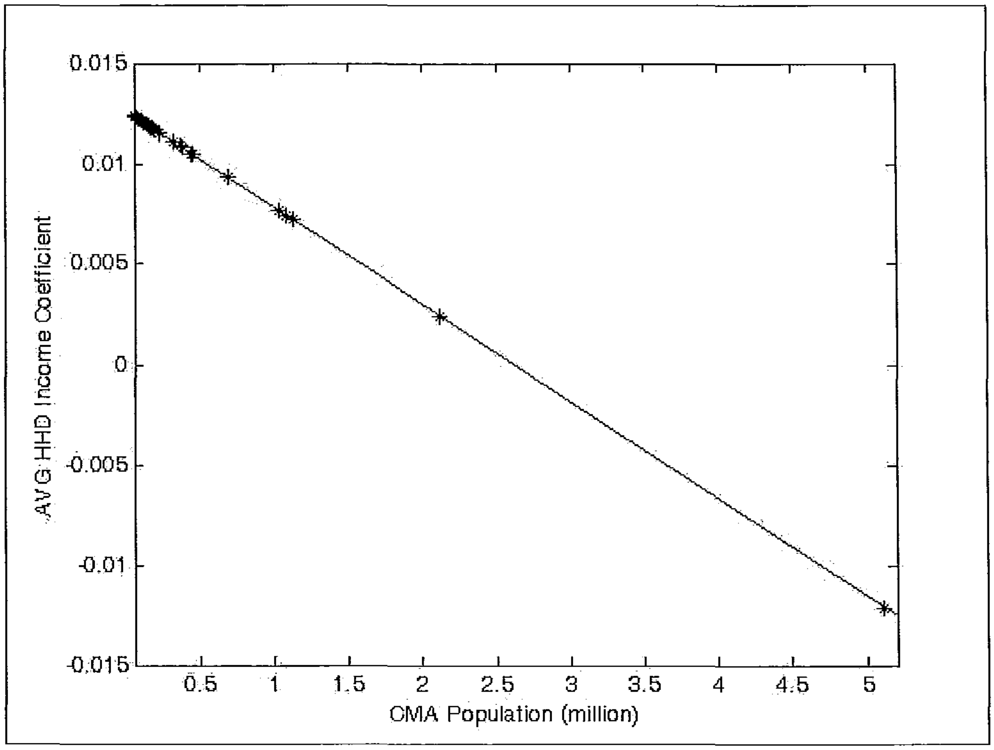

Figure 3 Average Household Income coefficient as a function of CMA population.

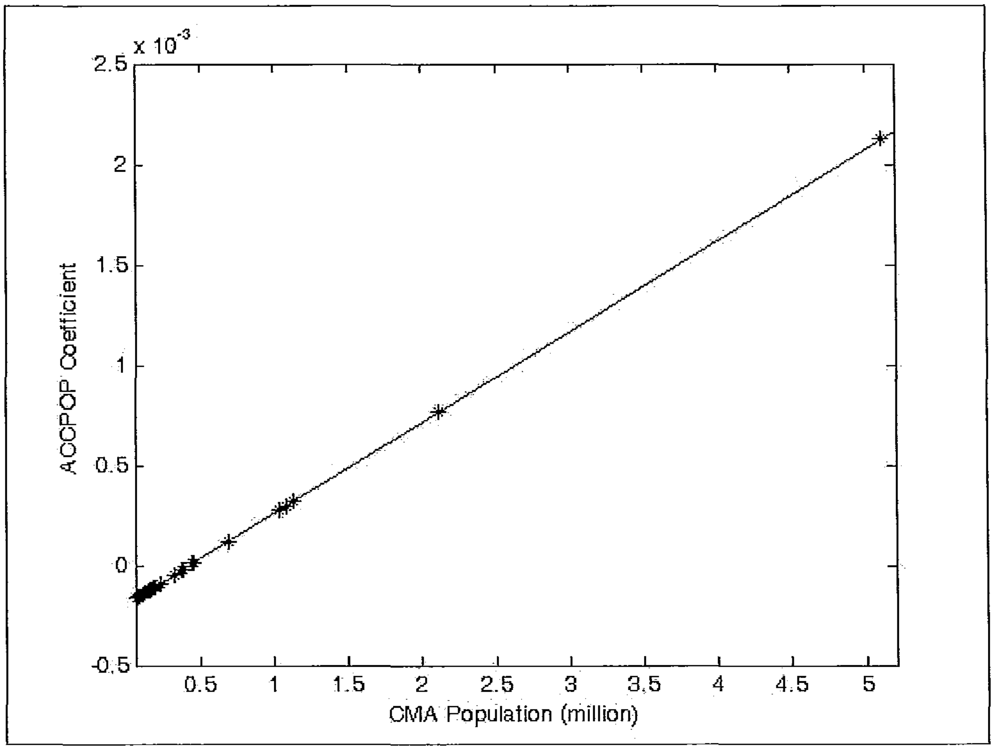

Figure 4 Population based (4 km) Accessibility coefficient as a function of CMA population.

2.8 References

1. Canadian Blood Services: **Basic Eligibility.** 2009 [http://www.bloodservices.ca/centreapps/internet/uw_v502_mainengine.nsf/page/E_Can_I_Donate]. 1-6-2009.

2. Canadian Blood Services: **Testing.** 2009 [http://www.bloodservices.ca/CentreApps/Internet/UW_V502_MainEngine.nsf/page/Testing?OpenDocument]. 1-6-2009.

3. Canadian Press: **Less than 4 percent of Canadians donate blood.** 2006 [http://www.bloodservices.ca/Centreapps/InterneUW_V502_MainEngine.nsf/page/Questions%20and%20Answers?OpenDocument#QA12].

4. Canadian Blood Services: **A Report to Canadians 2007/2008.** 2008 [http://www.blood.ca/CentreApps/Internet/UW_V502_MainEngine.nsf/resources/Annual+Reports/$file/07-08-CBS-Annual-Report-en.pdf]. 1-6-2009.

5. Talbot D, Mumford I: **Per Capita Blood Donations In Canada Trail Other Countries.** 1999 [http://www.blood.ca/CentreApps/Internet/UW_V502_MainEngine.nsf/resources/1999/$file/1999-10-07-per-capita-blood-donations-trail_eng.pdf]. 2-6-2009.

6. Statistics Canada: **Projected population by age group and sex according to medium growth scenario for 2006, 2011, 2021, 2026 and 2031, at July 1 (2006, 2011).** 2009 [http://www.statcan.ca/english/Pgdb/demo23a.htm]. 2-6-2009.

7. Goncalez T, Sabino EC, Chamone DF: **Trends in the profile of blood donors at a large blood center in the city of São Paulo, Brazil.** *Pan American Journal of Public Health* 2003, **13:**144-148.

8. Hollingsworth B, Wildman J: **What population factors influence the decision to donate blood?** *Transfusion Medicine* 2004, **14:**9-12.

9. Burnett J: **Examining the profiles of the donor and non donor through a multiple discriminant approach.** *Transfusion* 1982, **22:**138-142.

10. Hupfer MF, Taylor DW, Letwin JA: **Understanding Canadian student motivations and beliefs about giving blood.** *Transfusion* 2005, **45:**149-161.

11. Reed PB, Selbee LK: **Distinguishing Characteristics of Active Volunteers in Canada.** *Nonprofit and Voluntary Sector Quarterly*2000, **29:**571-592.

12. Reed PB, Selbee LK: **Patterns of Citizen Participation and the Civic Core in Canada.** 2000.

13. Pol M van der, Carins J, Galea G: **The efficient organization of blood donation: what determines the number of donors and donations?** *Transfusion Medicine* 2000, **10:**5-11.

14. Apparicio P, Charbonneau J, Dussault G: **La géographie du don de sang au Québec: une analyse exploratoire.** 2009 [http://www.ucs.inrs.ca/pdf/GeographieDuDonDeSang.pdf]. Institut national de la recherche scientifique - Urbanisation, Culture et Société.

15. Caro FG, Bass SA: **Receptivity to volunteering in the immediate postretirement period.** *Journal of Applied Gerontology* 1997, **16:**427-441.

16. Ownby HE, Kong F, Watanabe K, Tu Y, Nass CC: **Analysis of donor return behavior.** *Transfusion* 1999, **39:**1128-1135.

17. Jirovec RL, Hyduk CA: **Type of volunteer experience and health among older adult volunteers.** *Journal of Gerontological SocialWork* 1998, **30:**29-42.

18. Radke J, Mu L: **Spatial decomposition, modeling and mapping service regions to predict access to social programs.** *GeographicInformation Sciences* 2000, **6:**105-112.

19. Wang FH, Luo W: **Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas.** *Health & Place* 2005, **11:**131-146.

20. Páez A, Mercado RG, Farber S, Morency C, Roorda M: **Relative Accessibility Deprivation Indicators for Urban Settings: Definitions and Application to Food Deserts in Montreal.** *Urban Studies* 2010 in press.

21. Anselin L, Griffith DA: **Do Spatial Effects Really Matter in Regression-Analysis.** *Papers of the Regional Science Association* 1988, **65:**11-34.

22. Griffith DA, Layne LJ: *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets* New York: Oxford University Press; 1999.

23. Cliff AD, Ord JK: *Spatial Processes: Models and Applications* London: Pion; 1981.

24. Getis A, Griffith DA: **Comparative spatial filtering in regression analysis.** *Geographical Analysis* 2002, **34:**130-140.

25. Griffith DA: *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization* Berlin: Springer-Verlag; 2003.

26. Greene WH: *Econometric Analysis* 5th edition. Upper Saddle River: Prentice Hall; 2002.

27. Germain M, Glynn SA, Schreiber GB, Gelinas S, King M, Jones M, *et al.*: **Determinants of return behavior: a comparison of current and lapsed donors.** *Transfusion* 2007, **47:**1862-1870.

CHAPTER 3: Using visualization and filters to improve regression analysis of spatial data: An analysis of blood donation in metropolitan Toronto

3.1 Abstract

3.1.1 Background: Contextual effects and autocorrelation measures have shown to influence blood donor turnout in cities across Canada. Spatial filters are introduced to blood donor data for the city of Toronto to explore the effectiveness of filters at improving the contextual meaning of the model while improving the overall fit.

3.1.2 Data: Canadian Blood Services (CBS) provides geo-coded, blood donor, and donor clinic data for the fiscal 2006-2007 year. The data indicates the total number of donors for each Canadian postal code, excluding Québec. Explanatory characteristics are chosen based on measures of accessibility, donor clinic capacity, as well as previous findings regarding high correlates between the socio-economic variables with donor turnout.

3.1.3 Methods: The CBS data is aggregated to 992 census tracts within the Toronto Census Metropolitan Area (CMA). The count of donors within each census tract is regressed against known correlates. Once autocorrelation has been explained within the model via the introduction of a spatial filter, the magnitude of the synthetic filter is decreased by introducing variables that produce high correlation statistics with the spatial filter.

3.1.4 Results: The introduction of the spatial filter improves the fit of the model while also explaining the amount of autocorrelation present in the data. Replacement of the filter variable with true characteristics upholds the models high R^2 while improving the

validity of the model by reducing the size of the synthetic portion of the model from 68 to 54 eigenvectors.

3.1.5 Conclusion: The use of spatial filtering in regressive modeling goes beyond its use as a synthetic variable to represent autocorrelation. The filter can be further broken down into more detailed, true variables. In the case of Toronto blood donation, the filter helps reveal that a portion of donor behaviour in Toronto strays from the national model and relates directly to the work and travel habits of Torontonians, specifically involving commuting and employment differences related to gender.

3.2 Background

"[S]patial correlation … involves the correlation between values of the same variable at different spatial locations."[1] This autocorrelation across space is a common trait for most social variables at all spatial scales. However, while understanding the concept of spatial correlation helps researchers relate to ecological phenomena, it also hinders model representation because of the difficulty it presents in modeling spatial relationships. This difficulty is highlighted when rare event, count data is modeled because auto-Poisson models are restricted to circumstances involving only negative autocorrelation [2]. Count data for social interactions typically involve significant amounts of positive autocorrelation and the issue is usually addressed by using normalizing modeling techniques. An alternative to working with transformed data involves the development of a spatial filter variable that represents the latent autocorrelation in the model. This filter variable, first introduced by Getis and Griffith [3], provides a tool for explaining any present autocorrelation as well as improving the overall fit of the model. Griffith [2] has shown that extracting eigenvectors from geographical connectivity matrices such as the Moran's I based matrix can be used to index the amount of spatial autocorrelation within a dataset. The index, or spatial filter, consists of n orthogonal column vectors that describe Moran's I coefficient. Including this spatial filter as an independent variable will account for the autocorrelation that is present within the data enabling the use of auto-Poisson regression techniques. The spatial filter modeling process has demonstrated to be easier than MCMC and maximum likelihood methods while allowing for improved visualization of latent patterns [4].

With respect to socio-economic models, the spatial filter process has shown to be an effective tool in previous research involving blood donation in Canada [5]. This same research identifies that there is a strong contextual effect present in Canada with respect to blood donation. Using these findings, this paper sets out to construct an individual city model for Toronto that involves the same filtering process to identify the donor factors that are unique to the metropolitan area of Toronto.

Although the introduction of the spatial filter into the model will help explain the amount of autocorrelation present, the addition of synthetic variables is a slippery slope. Even if the added variable increases the fit of the model, the introduction of artificial variables to do so is questionable. Thus, once the filter is calculated and added to the model, attempts are made to substitute parts of the synthetic filter variable with true variables that correlate highly with the filter. Once these secondary variables are introduced into the model as a replacement to parts of the filter, a more realistic model is revealed which maintains a higher R^2 while decreasing the size of the synthetic part of the model by over 20% (From 68 eigenvectors to 54).

3.3 Data

Canadian Blood Services (CBS) has provided donor as well as blood clinic data for the research of this paper. The donor data was exhaustive and was provided in aggregate form as the number of donors in each Canadian postal code (excluding Québec). The data represents donor turnout for CBS' fiscal 2006-2007 year with each donor geo-coded to their place of residence. The postal code data was converted to the number of donors for each Canadian census tract (CT) for the purpose of modeling the data against 2006 Census attribute data. The data used for this paper involves the CBS data that is geo-coded within Toronto's Census Metropolitan Area (CMA). The 2006 Census lists 1,003 CT's within Toronto's CMA. Ten of these are removed for this study because of a lack of census information due to privacy concerns. The ten removed CTs do produce various voids within the CMA landscape, however all ten are contiguous with more than one other CT; this indicates that the continuity of the data is preserved even though there are ten "holes" in the data. Beyond the ten CTs removed because of censorship, one other CT is omitted because its inclusion produces significantly weaker regression results. The problematic CT is the tract representing the Toronto Island and its removal is understood by us to be trivial because of the island's unique location within the CMA. The island can only be reached by ferry and there are currently no donor clinics located on the island. Furthermore, much of the population residing on Toronto Island is known to be seasonal because of the reliance on the ferry transit and as such can be considered a region outside of the Toronto CMA.

Using the remaining 992 CMA census tracts, the Toronto donor data reveals 55,965 unique donors. The data also takes into consideration the 2,119 Canadian Blood Donor clinics that are situated within a 3 kilometre buffer around the Toronto study area. These 2,119 clinics represent 335 unique donation sites. Figure 1 displays the study area including the unique clinic locations across the city. The map gives a clear outline of the higher rates of donation in the outer parts of the city complimented by the donor clinics that seem to cluster within the core of the metropolitan area. Toronto's large number of commuters and downtown employment lead us to believe that this pattern promotes the notion that Torontonians are more likely to donate during working hours.

Donor turnout has shown to correlate with numerous Census characteristics. Previous research [5] has confirmed that characteristics relating to age, employment, nationality, wealth, and accessibility to donor clinics all correlate significantly with Canadian donor rates. Thus, the independent variables that have shown to explain variability in the national donor rate are implemented into our model to help explain donor participation within Toronto. Table 1 outlines the explanatory variables used in the study.

The socio-economic and demographic variables are retrieved from the 2006 Canadian Census, measured as the proportion of individuals residing in each Toronto census tract who pertain to each statistic. The division of age into four age groups follows the significant findings in the national correlate model and is unchanged as Toronto's donor participation for each age demographic is not expected to vary from the national perspective. The national findings indicate a positive relationship between most adult donors and a negative association between individuals who are over 65.

Immigration has shown to be a negative influence towards donor levels in Canada [5]. Toronto has a large immigrant population, with the 2006 Census identifying over 2 million (or 43%) foreign born individuals in the Toronto CMA. While the national trend shows that immigration has a negative impact on donor turnout, it is unclear whether Toronto and its extensive immigrant population base will follow this pattern as the majority of Toronto immigrants are from Asia and the Middle East, whereas immigrant origins for smaller Canadian cities are mostly from the United Kingdom [6].

Donor turnout is also expected to correlate negatively with unemployment as related studies [7] as well as findings on the national level have supported the notion that Toronto will likely follow a similar understanding.

The number of individuals employed in health related fields is included as a socio-economic variable with the belief that areas housing higher numbers of healthcare workers will demonstrate to have higher donor rates. It is expected that there will be a positive relationship between donor turnout and healthcare employment because of the relationship between healthcare and blood donation. It is assumed that many of the workers involved in health are somewhat aware or educated regarding the need to donate blood.

Variables for education and income are also included in the model. These have both shown to correlate positively at the national level. While education is expected to continue this trend within Toronto, the influence of income on donor rates in Toronto might differ from the Canadian trend. National findings indicate that income acts as a positive factor in smaller metropolitan areas and decreases as the CMA population increases. Thus, it is expected that income has a negative association with donor rates within the Toronto CMA.

Distance to the central part of the city is used as an independent variable to capture any possible effects that the infrastructure of the city and travel behaviour of individuals has on donor turnout. Designating which part of a city is "central" is, however debatable. Thus, two focal points of measure are examined to help represent how the geography of Toronto might influence donor levels. The financial central business district (CBD) is used to embody the central part of the city's downtown commercial area that involves a large number of public transit commuters. The geographic mean, calculated as the mean of all the latitude and longitude for each of Toronto's CT centroids is conveniently located in an area that connects the city's major arterial 400 and 401 highways. This mean focal point is an ideal representation of the central part of the city employing automotive based commuter traffic; it represents the point that most outer area populations pass through when commuting to the city by car. Models using the geographic mean proved to be more robust and are kept as a better representative of the central part of the city with respect to donor turnout.

The accessibility of CBS' blood donor clinics are included as a variable to capture how the level of service at each clinic influences donor turnout. The two-step floating catchment area concept introduced by Radke and Mu [8] and implemented in the national model is used to describe the amount of accessibility at each clinic. This two-step process involves calculating the level of service at each clinic; the number of beds-hours at each clinic divided by the population (CT centroid) lying within a 3 kilometre band from each clinic location. Accessibility is then calculated as the sum of all levels of service that lie within a 3 kilometre Euclidean distance from each CT centroid. Accessibility is calculated using both residential as well as workforce populations in each CT.

A secondary group of census variables is added at a later stage of modeling to improve the validity of the model. These variables are retrieved from the 2006 census as variables that correlate highly with the filter variable. The variables include: the average number of bedrooms, employment rate for parents, females employed in transportation and warehousing, as well as male and female commuters.

Whether the average number of bedrooms in a home will correlate positively or negatively with donor turnout is uncertain. If the variable is more of an indicator of wealth, then its sign could follow the findings for donor rates regarding average household income in Toronto, where it is expected that regions of higher household income will donate less frequently. However, with Toronto's diverse population, the average number of bedrooms could also represent homes that involve a large number of family members living in one dwelling. If this is the case, then the number of bedrooms might have a positive coefficient.

The secondary variables representing employment are expected to follow the same correlate pattern as previous employment research [7] as well as previous findings involving donor turnout [5]. This means that higher rates of employment for individuals with children at home will likely result in higher donor participation. Similarly, census tracts that are home to an increased number of women working in transportation and warehousing are likely to be locations of higher donor turnout.

The expected correlate sign for male and female commuters is uncertain. The additional time that commuters spend each day in traffic could negatively influence their participation in such activities as blood donation. The transient nature of their daily lives however, makes commuters perhaps more likely to visit donation sites during their daily travels, such as places of employment.

3.4 Methods

Matrix form of generalized linear regression modeling is used to develop coefficient estimates for the count of donors in each Toronto census tract. Following a poisson distribution, the model takes the form:

$$\log(\boldsymbol{D}) = \mathbf{X}\boldsymbol{\beta} + \varepsilon \qquad (1)$$

where $\boldsymbol{D}$ is the vector of distribution parameters, $\boldsymbol{\beta}$ is the vector of coefficients, $\mathbf{X}$ is the matrix of explanatory variables, and ε is the vector of independent residuals. The residual errors are examined for possible autocorrelation before developing any conclusions or inference from the model. Spatial autocorrelation is tested for by use of a first order contiguity matrix with the Moran's statistic [9]. Positive Moran values reveal positive spatial autocorrelation while negative Moran values indicate negative autocorrelation. Moran values that are zero or insignificant indicate an absence of spatial pattern.

Any observed spatial pattern is explained by a spatial filter, developed to represent the concern of spatial autocorrelation. The method of using filtered variables to separate spatial effects from other non-spatial variables follows Griffith filtering technique [3]. This method involves using the following matrix based on the Moran's I statistic to obtain up to n orthogonal eigenvectors:

$$(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)\mathbf{W}(\mathbf{I}-\mathbf{1}\mathbf{1}'/n) \qquad (2)$$

where **1** is an nx1 column vector of ones, **I** is the identity matrix, and **W** is the nxn spatial proximity matix. Each element of the matrix **W** is assigned a value of 1 if the neighbouring entry (census tract) is contiguous [1]. The matrix is standardized such that the sum of each row equals one. The eigenvectors represent the latent autocorrelation present in the residual geography. A step-wise procedure is followed where each of the n eigenvectors are introduced into the model with the other explanatory variables, tested, and then either kept or eliminated based on their significance ($p<0.10$). As eigenvectors are accepted as significant, they are then multiplied by their coefficient and introduced into the model as an explanatory variable for the remaining eigenvector significance tests. This step-wise procedure is continued until an acceptable level of autocorrelation significance is reached. The final matrix of column eigenvectors is treated as a synthetic explanatory variable that improves or maintains the statistical integrity of the model while eliminating the residual, spatial autocorrelation.

3.5 Results

Our initial poisson regression estimation representing blood donor counts in Toronto can be seen in Table 2 (model 1). The model has a relatively high R^2 where approximately 74% of the blood donor variation in Toronto can be explained by the distribution of the independent variables across the city. Most of the independent variables are significant at the 95% level and have a positive or negative sign similar to what we would expect. All four age group categories correlate positively with donor counts as does the effect of education. These findings are easily explained except for the coefficient relating to populations over 65 years of age. While it has been shown at the national level that senior populations are negatively associated with donor turnout, the positive coefficient for this group in Toronto might be related to Toronto's efficient transit system and infrastructure that enables seniors to have better access to donor sites. This idea is supported by examination of the significance for the variable representing clinic access. While the coefficient for population based access to CBS clinics is insignificant in model 1, further modeling shows a positive relationship with home-based access in Toronto. Like the variable for those over 65 years of age, population based access is found to have a negative relationship with blood donation at the national level. A negative correlation for immigrant populations tends to support the argument that immigrants avoid getting involved in events that are unfamiliar to them. Residents that reside in census tracts that are places of higher workforces tend to also correlate positively with blood donation. In contrast, proportion variables for immigration, unemployment, and health occupations all correlate negatively with donor turnout. The negative association that unemployment has with donor levels follows previous findings (see [7]). The negative relationship found between those employed in healthcare occupations and blood donor turnout is opposite of what is expected. Those employed in health-related fields are thought to have an increased awareness regarding the importance of donating blood. The negative finding might indicate that healthcare workers in Toronto have less time to donate regardless as to whether they understand that there is a need for them to donate blood. The negative healthcare coefficient coupled with the significant findings regarding workforce populations and accessibility highlights the uniqueness of travel and employment behaviour in Toronto with respect to donor participation. This understanding is expanded upon in later models once new variables are added. Overall, the initial model is a useful exploratory tool but fails to address the spatial autocorrelation that is present in the data. This is demonstrated by the significantly positive Moran's coefficient (z-value of 14.304).

Blood donor research conducted at the national level [5] has shown that the inclusion of a spatial filter into a regressive, donor model removes the residual autocorrelation from the error while also improving the fit of the model. A second model that includes a spatial filter as a variable is thus developed to address the presence of spatial pattern within the data. It is clear from the results of the second model highlighted in table 2, that the filter eliminates the residual autocorrelation by allowing the autocorrelation to become part of the model rather than part of the residual error. The fit is shown to improve with the filters inclusion as the R^2 improved from 0.738 in model 1 to 0.827 in model 2. The elimination of residual autocorrelation also improves the significance of one of the other independent variables, changing the variable representing access to CBS donor sites to significant and positive. In model 1, when no filter is present the three kilometre, residential access measurement is negative and insignificant. This finding, along with the significant 'distance to geographic mean' variable, indicates a strong relationship between travel behaviour and blood donor participation in Toronto. A map of the synthetic filter variable for model 2 is displayed in Figure 2.

While the introduction of a synthetic, filter variable produces a better fitting model with no residual spatial pattern, the inclusion of a variable that has no real-world understanding is debatable. Simply adding unknown parts to a model might improve the strength of the model but clearly the findings become much more easily contested. Thus, a third model is created where new variables are introduced that correlate highly with the spatial filter depicted in Figure 2. This is done so that the size of the spatial filter and therefore the magnitude of the synthetic portion of the model are reduced. The newly added census variables are more detailed than the variables introduced in the initial models and constraints are made so that a correlation of less than 90% with all other variables is maintained before any new variable is considered. Table 2 (model 3), reveals how the substitution of variables for parts of the filter influences the structure of the model. The secondary variables all have their expected signs and not only help reduce the size of the filter (from 68 to 54), but their inclusion renders some of the initial variables representing unemployment, healthcare jobs, and residential access to CBS clinics as insignificant. The elimination of these initial variables indicates that the more detailed variables introduced into model 3 are doing a better job of explaining the pattern of blood donation. Although the R^2 in models 2 and 3 are the same, the reduction in the number of spatial filter eigenvectors indicates that the model is more robust with the more detailed variables. Furthermore, not only do the number of eigenvectors used in the model decrease when the new variables are introduced, but the size of the eigenvectors used also decreases. The eigenvectors in model 2 range from -0.5779 to 0.6551, or 1.233 units, while the eigenvectors in model 3 range from -0.4550 to 0.6380, or 1.093 units. This means that the overall influence of the spatial filter is reduced. While the reduction of the filter variable helps make our third model more believable, there is still a significant amount of autocorrelation unexplained in the model. This can be visualized by examining the extent of the spatial filter variable from model 3 (Figure 3).

3.6 Discussion

Previous work has shown that many of the factors influencing blood donation in Canada are unique to each city. The models created here provide support to this idea. While the Toronto blood donor model shares many of the main correlates as the national model, dissection of the spatial filter reveals attributes relating to employment and travel that are likely unique to this area. This information may be useful to CBS when faced with decisions regarding clinic locations, or hours of operation.

The final model, with the reduced spatial filter, introduces five variables with three of the original variables becoming insignificant. The variables representing unemployment, health occupations, and clinic access by evening populations are replaced by more detailed variables involving mobility, employment, house size, and gender. The indication that gender plays a role in blood donation is not new ([11, 12]) but is a key development. It was understood at the onset of modeling that gender might play a role in blood donation; however the variable representing gender was omitted because the census tract level of aggregation indicates very little male/female variation. Now that it is clear that donor rates in Toronto are specifically related to gender roles with respect to employment and commuting, gender related statistics can be addressed in any further Toronto donor research.

The number of bedrooms in a home is likely related to whether the donor has children rather being an indicator of wealth or immigrant status. This is because the income and immigrant status variables both have negative coefficients while the coefficient representing the number of bedrooms is positive. The variable representing the number of individuals employed with kids also indicates that the presence of children positively influences blood donation in Toronto. Initial models for the city of Toronto included a variable representing the average number of children at home but proved to be insignificant. Thus, like the role of gender, perhaps the presence of children strongly affects donor turnout only when related to the employment status of the parents or house size.

The finding that females employed in transportation or warehousing are more likely to donate follows the findings that lower income earners are more likely to be blood donors in bigger cities like Toronto. Beyond this, further research needs to be done to better understand the role of gender with respect to donation and employment based in warehousing.

The variables representing female and male commuters are found to have negative coefficients. This can be explained by the extreme commuting times within Toronto which directly translate to a decreased amount of free time to volunteer or donate blood.

While the variables that are introduced in model 3 reduce the size of the spatial filter, much of the latent autocorrelation pattern still remains unknown. A comparison of Figures 2 and 3 gives evidence that while part of the pattern is explained by the variables introduced in model 3, a clear pattern still remains. Individuals who are more familiar with Toronto's infrastructure might be able to uncover other attributes that closely follow the patterns seen in Figure 3 to reduce the filter size further. Any additional variables that reduce the size of the filter will make the model even more parsimonious. This process can of course continue until the amount of autocorrelation present is insignificant or the range of the filter essentially becomes negligible.

3.7 Conclusion

The role that spatial autocorrelation plays in modeling Poisson distributed data has until recently been one of hindrance. This is related to the idea that "[t]he auto-log-Gaussian approximation… and the auto-logistic approximation circumvent the auto-Poisson's restriction to only situations involving negative autocorrelation…" [2]. The introduction of a spatial filter has shown to avoid this issue by creating a robust model that resolves the issue of autocorrelation without having to manage the intractable normalizing constant. This paper explores the use of spatial filters not only as a way to explain spatial pattern present in the model, but as a tool to help introduce new variables into the model that help replace part of the synthetic filter variable.

Beyond implementing the concept of using spatial filters to uncover new variables, the true purpose of this paper has been to address the contextual concept of Canadian donor turnout. This is done by developing a single-city model for Toronto that implements the use of the spatial filtering process to discover new variables that are unique to Toronto blood donors. This objective is considered met based on the nature of the variables that are introduced into the Toronto model. The variables, which relate to travel, employment, and house size and are not likely contributing factors in other Canadian cities based on Toronto's unique commuting culture. Although there have yet to be other city models developed, the national model does not indicate that these variables might be significant elsewhere. Although Vancouver is similar to Toronto with its large workforce, the commuting times in Vancouver are not known to be as severe as they are in Toronto.

While new variables have been researched that relate well to Toronto donor participation, the modeling process for Toronto donors can still be expanded upon. The process of introducing other variables to reduce the size of the filter can continue until the filter is no longer a significant part of the model. It is unclear whether the best way to reduce the filter size is to develop new variables similar in nature to the ones that we have uncovered, or whether there are other statistical areas that need exploring in order to find ways to represent any autocorrelation. What is known is that there is still a significant amount of pattern within the spatial filter variable that can be represented by some sort of attribute. While the procedure for developing new variables described here involves correlation analysis, it is certainly not the only method that can be used to discover variables to replace parts of the filter. Other methods might involve consulting individuals more familiar with the city of Toronto or a detailed analysis of cultural characteristics of the Toronto population that might reveal variables that are not represented within the census data.

Although the donor situation in large cities like Toronto might not be as bright as we would like it to be, the findings here give hope that the identification of donor behavior will help CBS improve donor turnout. While the information researched here should prove helpful to CBS, improved collection statistics will also assist in a better model. One of the constraints of this research has been the unknown of whether each donor's trip to donate is home or work based. The finding that employment statistics are a significant factor for donation indicates a need for more detailed donor data. Although it is clear from figure 1 that CBS has placed many of their clinics within reach of the working population, our accessibility modeling is purely statistical and any effort by CBS to identify trip-based information from each donor will help produce more robust models.

Toronto's weak donor rate (just over 1% by our calculations) indicates a strong need to understand donor behavior in this city. The variables developed here provide a substantial first step towards understanding how to address the low turnout, however further research is required before any sound policy can be implemented. Once the Toronto model has been expanded upon and other city models have been developed, the contextual effects of blood donors in Canada will be much more clear and more importantly, attempts to increase donor turnout can be made.

Table 1: Variables and Definitions

Variable	Definition and units	Min	Max	Mean	SD
Dependent Variable					
DONORS	Donor count in census tract	0	240	56.42	36.11
Primary Socio-economic and demographic characteristics					
POP15TO24	Population 15-24 years of age in census tract/1,000	0.03	3.48	0.69	0.32
POP25TO54	Population 25-54 years of age in census tract/1,000	0.18	10.67	2.36	1.12
POP55TO64	Population 55-64 years of age in census tract/1,000	0.02	1.90	0.53	0.20
POP65+	Population 65+ years of age in census tract/1,000	0.01	2.06	0.61	0.30
IMMIG	Population immigrant population in census tract/1,000	0.07	14.61	2.34	1.47
UNEMP	Unemployed population in census tract/1,000	0.00	1.08	0.19	0.10
HEALTHOCC	Population in census tract in health-related occupations/1,000	0.00	0.71	0.12	0.07
UNIVERSITY	Population in census tract with bachelors degree or higher/1,000	0.03	6.75	1.34	0.82
AVEHHDINC	Average Household income ($10,000)	2.66	81.08	9.18	5.62
WORKFORCE	Population working in census tract/1,000	0.06	108.6	2.46	6.64
Geographic variables					
DISTCTR	Distance from census tract centroid to geographic centre of study area (kms)	0.06	7.43	1.85	1.16
Service variables					
ACCPOP3K	Population-based accessibility within 3 km band (beds-hour/1000 people)	0.00	73.82	9.90	9.07
ACCJOB3K	Employment-based accessibility within 3km band (beds-hour/1000 people)	0.00	410.12	21.93	23.76
Secondary Socio-economic and demographic characteristics					
AVEBDR	Average number of bedrooms per dwelling	1.00	4.50	2.78	0.68
EMPKDS	Employment rate, population 15 years and over with children at home	25.00	93.90	76.68	7.03
FEMLABR	Females employed in Transportation and warehousing/1,000	0.00	0.25	0.04	0.04
MALCMUTR	Males who commute to work/1,000	0.00	1.27	0.10	0.17
FEMCMUTR	Females who commute to work/1,000	0.00	3.31	0.41	0.41

Table 2: Regression results.

Model 1			Model 2			Model 3		
VARIABLE	ESTIMATE	p-value	VARIABLE	ESTIMATE	p-value	VARIABLE	ESTIMATE	p-value
CONSTANT	2.8664	0.0000	CONSTANT	3.0482	0.0000	CONSTANT	2.1597	0.0000
POP15TO24	0.8713	0.0000	POP15TO24	0.6891	0.0000	POP15TO24	0.6013	0.0000
POP25TO54	0.2018	0.0000	POP25TO54	0.1787	0.0000	POP25TO54	0.1721	0.0000
POP55TO64	0.6093	0.0000	POP55TO64	0.4625	0.0000	POP55TO64	0.4354	0.0000
POP65+	0.0747	0.0001	POP65+	0.2034	0.0000	POP65+	0.2692	0.0000
IMMIG	-0.2604	0.0000	IMMIG	-0.3088	0.0000	IMMIG	-0.3080	0.0000
UNEMP	-1.2153	0.0000	UNEMP	-0.3047	0.0005	UNIVERSITY	0.2676	0.0000
HEALTHOCC	-0.6739	0.0000	HEALTHOCC	-0.2179	0.0325	AVEHHDINC	-0.0142	0.0000
UNIVERSITY	0.2189	0.0000	UNIVERSITY	0.2641	0.0000	WORKFORCE	0.0034	0.0000
AVEHHDINC	-0.0032	0.0017	AVEHHDINC	-0.0080	0.0000	DISTCTR	0.0170	0.0011
WORKFORCE	0.0014	0.0588	WORKFORCE	0.0028	0.0001	ACCJOB3K	0.0009	0.0000
DISTCTR	0.0001	0.0000	DISTCTR	0.0330	0.0000	AVEBDR	0.0936	0.0000
ACCPOP3K	-0.0006	0.3633	ACCPOP3K	0.0020	0.0020	EMPKDS	0.0091	0.0000
ACCJOB3K	0.0019	0.0000	ACCJOB3K	0.0004	0.0360	MALCMUTR	-0.1162	0.0004
			Filter	1.0000	0.0000	FEMCMUTR	-0.0590	0.0050
						FEMLABR	0.4142	0.0083
						Filter2	1.0000	0.0000
MoransI(Z)	14.304		MoransI(Z)	0.141		MoransI(Z)	0.290	
R^2=	0.738		R^2=	0.827		R^2=	0.827	
n=	992		n=	992		n=	992	
			Filter Size	68		Filter Size	54	
			Range of Filter	1.233		Range of Filter	1.093	

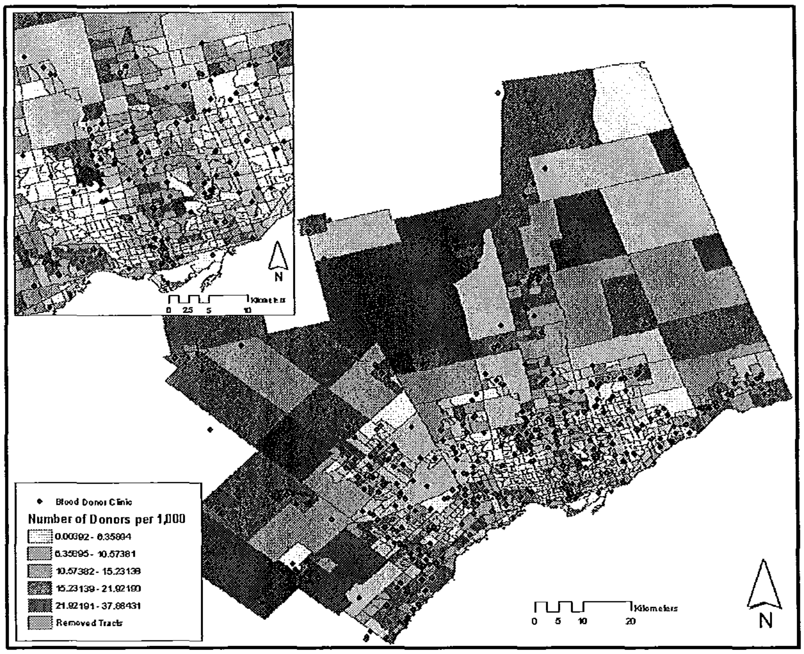

Figure 1 Distribution of donors and location of CBS clinics.

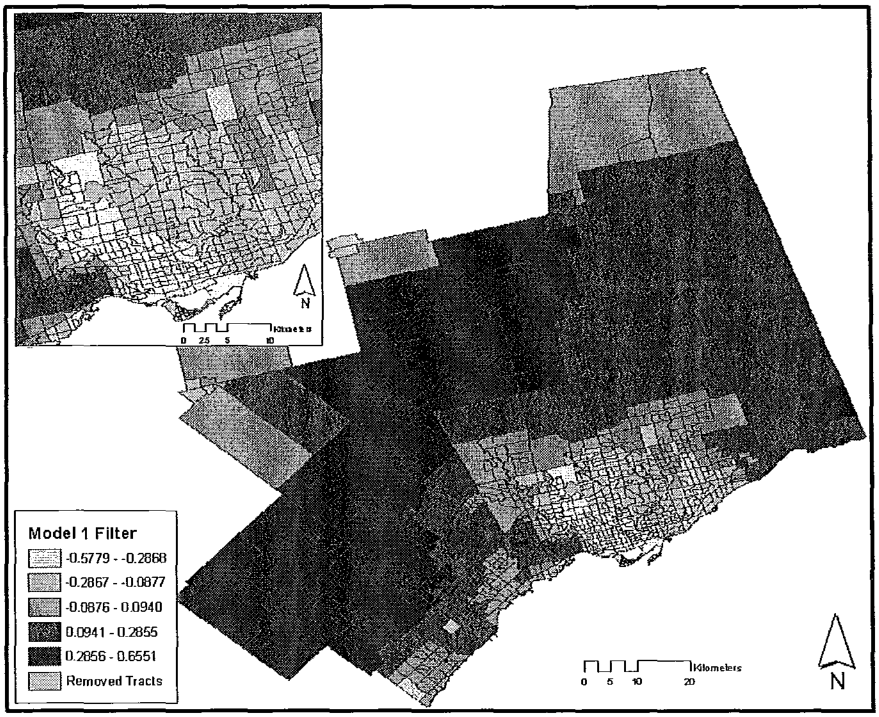

Figure 2 Distribution of Model 1 Spatial Filter.

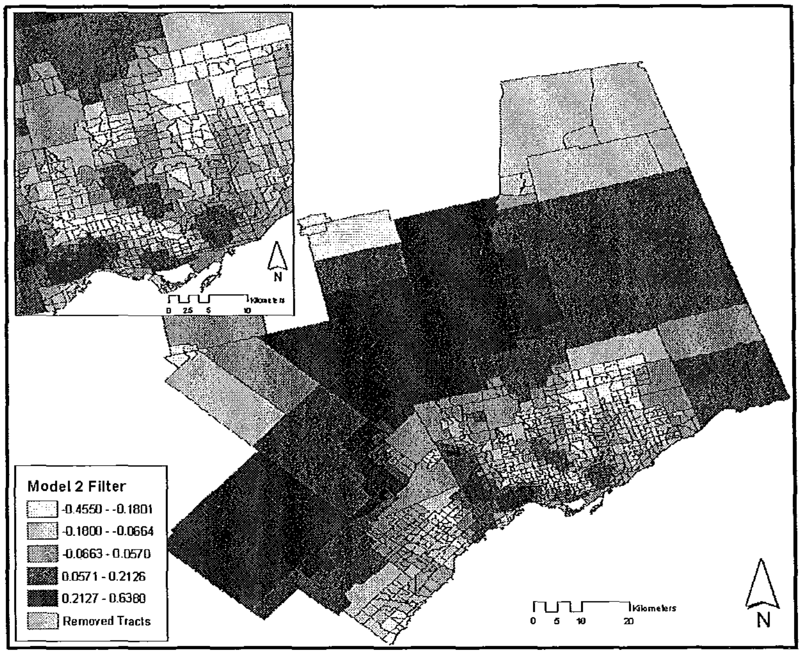

Figure 3 Distribution of Model 2 Spatial Filter.

3.8 References

1. Bailey TC, Gatrell AC: *Interactive Spatial Data Analysis* Harlow, England: Prentice Hall; 1995

2. Griffith D: **A Spatial Filtering Specification For The Auto-Poisson. Model.** *Statistics &Probability Letters* 2002, **58:**245-251.

3. Getis A, Griffith DA: **Comparative Spatial Filtering in Regression Analysis.** *Geographical Analysis* 2002, **34(2):**130-140.

4. Griffith DA: **A Spatial Filtering Specification For The Autologistic Model.** *Environment and Planning A* 2004, **36:**1791-1811.

5. Saberton PJ, Paez A, Newbold KB, Heddle NM: **Geographical Variations In The Correlates Of Blood Donor Turnout Rates: An Investigation Of Canadian Metropolitan Areas.** *International Journal of Health Geographies* 2009, **8:**56.

6. Statistics Canada: **Immigrant Population by Place of Birth, by Census Metropolitan Area** [http://www40.statcan.ca/l01/cst01/demo35c-eng.htm]. 6-24-2010.

7. Caro FG, Bass SA: **Receptivity to Volunteering in the Immediate Postretirement Period.** *Journal of Gerontological Social Work* 1997, **30:**427-441.

8. Radke J, Mu L: **Spatial Decomposition, Modeling and Mapping Service Regions to Predict Access to Social Programs.** *Geographic Information Sciences* 2000, **6:**105-112.

9. Cliff AD, Ord JK: *Spatial Processes: Models and Applications.* London. Pion; 1981.

11. Gillespie TW, Hillyer CD: **Blood Donors and Factors Impacting the Blood Donation Decision.** *Transfusion Medicine Reviews* 2002, **16(2):**115-130.

12. Wu Y, Glynn SA, Schreiber GB, Wright DJ, Lo A, Murphy EL, Kleinman SH, Garratty G: **First-time blood donors: demographic trends.** *Transfusion* 2001, **41:**360-364.

13. Griffith D: **A linear regression solution to the spatial autocorrelation problem.** *Journal of Geographical Systems* 2000, **2:**141-156.

14. Tiefelsdorf M, Griffith DA: **Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach.** *Environment and Planning A* 2007, **39:**1193-1221.

CHAPTER 4: Conclusion

4.1 Thesis Summary

The objective of the research presented here has been to better understand the motivating factors involved in Canadian blood donation. Donor participation in Canada is weak and not well understood. Because blood plays such an important role in the health of so many people, a better understanding of the factors involved in blood donation is necessary to help increase levels of repeat donation.

Initial regressive models representing the 40 CMAs in Canada are developed in an attempt to discover which factors contribute most to the donor process. These models reveal that unemployment and immigration status are the strongest indicators that a population will not donate, while populations that contain a large number of educated teens and young adults are the biggest indicators that a population will donate blood. While finding that these factors relate to donor participation is a step forward in understanding who is donating blood, further sociological exploration of the variables' significance needs to be explored before their impact can be fully understood. This is because the populations involved in these models are not static; as populations age, their propensity to donate will change as indicated by the differing age category coefficients in the models. Immigrant populations are also inclined to change. Not only does the makeup of the immigrant community change but the attitudes within the community tend to acclimatize to their Canadian environment. This settling effect is bound to have an influence on whether immigrants donate. Thus, while the number of immigrants may remain the same in certain CTs, through time the generational acceptance of Canadian ways might alter their blood donor habits. So, while these models have shown to produce variables that significantly relate to blood donor turnout, in depth understanding of the impact that each factor has on the future of blood donation should be explored before any policy objectives can be set forth.

The spatial filter concept has demonstrated that its inclusion in blood donor models assists in improving the overall fit and validity of the model. Its introduction however, still results in the large number of city-specific dummy variables even when the effects of population size are included. The strong contextual variation among cities indicates that further exploration is needed into which factors motivate donors in each individual city. A model for metropolitan Toronto area is developed to explore this research question. The Toronto model follows the same outline as the Canadian model where the initial variables are modeled with the spatial filter. However, once the filter has been established, alternative variables are introduced that correlate highly with the filter in an attempt to reduce the magnitude and range of the filter. This results in the removal of the initial variables representing access to evening populations, unemployment, and healthcare workers. These are replaced by more detailed variables representing similar attributes. While the process of subtracting three variables to introduce five new variables might seem redundant because the R^2 remains the same throughout the procedure, the introduction of the new variables reduces the size and magnitude of the filter while maintaining a high R^2. The unchanged R^2 is not an indication that the adjusted variable grouping has no affect on the strength of the model. It simply indicates that the filter's inclusion in the model helps explain any pattern present in the donor data; the remaining 17% can be attributed to random effects.

Many of the other Toronto correlates follow the Canadian model outline where the strongest indicators for blood donor turnout relate to teenage and young adult ages. One surprising difference is the positive coefficient for populations aged 65 and older. The coefficient for this attribute is negative in the national model and positive for Toronto's model. Because the variables introduced into the Toronto model all relate to travel and employment, it is thought that the senior populations' mobility is also influenced by Toronto's efficient transit system.

While both the Canadian and Toronto donor models address the task of finding characteristics of blood donation, further modeling involving other Canadian cities needs to be developed before a more detailed donor picture is complete. The uniqueness of the Toronto variables demonstrates the need for individual city models as their development will encourage comparisons to the Toronto model as well as assist in better understanding of the Canadian model. It is hoped that the research conducted here will help encourage further research into the field of blood donor factors, a field that has far too little understanding considering its importance in all of our lives.

4.2 Study Limitations

The research discussed here is one of the first attempts to describe the traits of Canadian blood donors. Because of this, the data provided to us by CBS was used to develop a broad, preliminary understanding of donor traits in Canada. While we have argued that this understanding will likely lead to better policy design based on CBS having a better grasp on who their donors are, this research neglects to examine the correlates of who is not donating. Although we feel that researching the characteristics of current donors is a good beginning for the end goal of understanding the motivating factors involved in Canadian blood donation, it does not necessarily help identify ways of getting non-donors to become regular donor participants.

It is discussed throughout the thesis that there is resistance within the statistical community towards introducing synthetic variables into a model for the sole purpose of increasing the model's fit. Many statisticians feel that the filter is an unnecessary tool if the proper variables are chosen from the modeling onset. The argument against this is that the goal of all modeling is the same: to create a simplification of a real-world phenomenon that will help us better understand the process in question. We have shown here that the filter is a useful tool for helping to understand the makeup of blood donors in Canada and Toronto. The Toronto modeling process exemplifies how adding "better" variables does not necessarily improve a models function. While the detailed correlates identified in the third Toronto model help reduce the amount of unexplained autocorrelation, the introduced variables are much too specific for the model to be used as a projecting tool. The second model's variables are more understandable and thus more useful in any predictive modeling endeavors.

4.3 Future Research

With blood donation having such an important role in every Canadian life, the need to build on the research started here is clear. Although we have developed correlates that relate to blood donation in Canada and Toronto, these models can be built upon by developing donor models for other Canadian metropolitan areas. The models described here can also be improved upon by developing better donor trip-based statistics as well as possibly examining the potential correlates of non-donors. A likely assumption is that the coefficients that correlate negatively with donors in our models will be the driving factors for non-donors. However, this assumption is unproven and may or may not be true. The development of models based on the motivating factors regarding non-donors will provide an even clearer picture of the donor habits of Canadians.

Beyond the development of models based on current donor data, further research involving qualitative models could also help better understand donor turnout in Canada. While the data used in our research was exhaustive, the models are broad. More detailed models, possibly involving the analysis of individual donors within a single donor clinic, or group of clinics could help identify more detailed factors that relate to blood donation.