University of Alberta

Spatial analysis to locate new clinics for diabetic kidney patients in the underserved communities in Alberta

by

Labib Imran Faruque

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Epidemiology

Department of Public Health Sciences

©Labib Imran Faruque
Fall 2012
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

ABSTRACT

Background: Canadians often live far from health care facilities, which may compromise their care. However, no objective method exists for selecting new facilities from potential locations. We used a new method for selecting optimum clinic locations and characterized remote-dwellers clinically.

Method: We used two methods for locating remote-dwelling Albertans with diabetes and chronic kidney disease (defined by estimated glomerular filtration rate of 15-60 ml/min/1.73m^2): plots of unadjusted density of patients per 100 km square; and SaTScan analysis which presents prevalent patient clusters with CKD rates (adjusted for population size).

Results: We studied 32,278 patients with concomitant CKD and diabetes. Density plots localized one large cluster. However, SaTScan technique and buffer analysis detected additional clusters in the northwest and southeast regions of Alberta. Identified clusters had higher hospitalization rates.

Conclusions: SaTScan objectively identifies clusters of underserved high-risk CKD patients and may be helpful for decision-makers in planning potential new facility locations.

ACKNOWLEDGEMENTS

Thanks to my supervisor, Dr. Marcello Tonelli. It is a great opportunity for me to work with him. It's difficult to express my gratitude with few words. Without his supervision and suggestions, this project cannot go far. I learned a good number of skills working with him. In future, it is my cherished desire to contribute in science in the way as he is doing right now.

Thanks to the committee members, Dr. Scott Klarenbach, Dr. Neesh Pannu and Dr. Dean Eurich who reviewed this thesis and provided critical feedback which helped me a lot to reshape the draft and present the arguments and thoughts in a coherent way to facilitate the understanding of the readership.

Thanks to Dr. Brenda Hemmelgarn for her constructive feedback on the manuscript. In addition, I would like to thank Ms. Natasha Wiebe for preparing the datasets and critical revision of the draft. The project would not be accomplished without kind support of Dr. Bharati Ayyalasomayajula who helped to create the maps in ArcGIS. Ms. Sophanny Tiv presented the maps in a wonderful way.

Thanks to the faculties of the School of Public Health. I learned the core concepts of epidemiology and biostatistics from the classroom orientations, which built my confidence to apply these principles in my thesis project.

Thanks to Alberta Kidney Disease Network (AKDN) and Interdisciplinary Chronic Disease Collaboration (ICDC) for giving me this excellent trainee position to undergo the masters program in the School of Public Health in the University of Alberta.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviation

CKD	Chronic Kidney Disease
DM	Diabetes Mellitus
SaTScan	Spatial, Temporal or Space-time Scan Statistics
eGFR	estimated Glomerular Filtration Rate
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
MDRD	Modification of Diet in Renal Disease
OGTT	Oral Glucose Tolerance Test
HbA1C	Hemoglobin A1C
CVD	Cerebrovascular Disease
CHF	Congestive Heart Failure
COPD	Chronic Obstructive Pulmonary Disease
HIV	Human Immunodeficiency Virus
PUD	Peptic Ulcer Disease
PVD	Peripheral Vascular Disease
AMI	Acute Myocardial Infarction
CVA/TIA	Cerebrovascular Accident/ Transient Ischaemic Attack
AKI	Acute Kidney Injury
CATH	Cardiac Catheterization
PCI	Percutaneous Coronary Intervention
CABG	Coronary Artery Bypass Grafting
ESRD	End Stage Renal Disease

Chapter 1: Introduction

1.1 Literature review

1.1.1 Diabetes

According to the Canadian Diabetes Association, diabetes mellitus is a metabolic disorder caused by defective insulin secretion and /or defective insulin action leading to the presence of hyperglycemia and associated with fatal consequences including damage, dysfunction and failure of kidneys, eyes, nerves, heart and blood vessels. Diabetes is classified into type 1, type 2, gestational diabetes and other specific types. Diagnostic criteria for diabetes include fasting plasma glucose of ≥7 mmol/L or casual plasma glucose of ≥11.1 mmol/L with classic symptoms of diabetes such as polyuria, polydipsia and unexplained weight loss, or plasma glucose of ≥11.1 mmol/L two hours after a 75gm OGTT [1].

1.1.2 Chronic kidney disease

Chronic kidney disease (CKD) is characterized by the presence of kidney damage or a decreased level of kidney function for a period of three months or more. It is characterized by a staging system of stage 1 to stage 5. In stage 1, eGFR is 90 ml/min with more than 90% of kidney function preserved; stage 3 with eGFR 30 to 59 ml/min (approximately 30 to 59% of normal kidney function preserved) and stage 5 with ≤15 ml/min (less than 15% of function preserved). For patients with stage 5 CKD, renal replacement therapy (dialysis or transplantation) may be recommended. Early symptoms appear at stage 4 along with anemia and other metabolic abnormalities [2]. As the kidney damage progresses, abnormal levels of protein can appear in the urine (proteinuria or albuminuria). Early diagnosis with annual monitoring of kidney function

with protein measurement in urines and effective management at earlier stages by medications such as ACE-inhibitors can slow the disease progression [2].

1.1.3 Chronic kidney disease in diabetes

A diabetic person is considered to have chronic kidney disease if s/he has persistent albuminuria or significantly reduced kidney function of eGFR≤60 mL/min. Proteinuria can be either microalbuminuria (urinary albumin 30 to 300 mg/day) or overt nephropathy as >300 mg/day [1].

1.1.4 CKD and diabetes are common in Canada and Alberta

1.1.4.1 Diabetes in Canada and Alberta

171 million people were estimated to have diabetes in the world in 2000 [3]. The International Diabetes Federation estimated the global age-standardized prevalence of diabetes among adults aged 20 to 79 years of 6.4% in 2010 as 285 million people worldwide [4]. In Canada, there were 1.3 million patients in 2000 and the direct cost attributed to diabetes is 3.5 percent of public healthcare expenditure [5]. Using 2008-2009 data from the Canadian Chronic Disease Surveillance system, around 2.4 million Canadians aged one year or older were diagnosed with either type 1 or type 2 diabetes, representing 6.8% of the total population [4]. About 90 percent of them were of type 2 diabetes [5].

According to the Alberta Diabetes Surveillance System (ADSS), there were 205,726 individuals with diabetes with a prevalence rate of 5.5% among the general population in Alberta in 2009 [6]. The trend of prevalence is increasing from 3.83 in 1995 to 5.81 in urban areas and 3.71 to 5.76 in the rural areas in 2006 [7]. In a study from ADSS in Alberta, the number of diabetic

patients was 147498 in 2007 (4.5%) and expected to be 513433 in 2035 (11.1%). Associated health care cost is expected to increase from 673 million dollar in 2007 to 2.27 billion dollars in 2035 [8].

1.1.4.2 Chronic kidney disease in Canada and Alberta

A systematic review of 26 studies on CKD prevalence in different populations in the world revealed that the median prevalence was 7% for people aged 30 years or more and 23% to 36% for those aged 64 years or more [9].

According to the Kidney Foundation of Canada, approximately 2.6 million Canadians had or were at risk for CKD in 2012 [10]. In Canada, the prevalence of CKD was 35.7% among participants from long term care facilities in the elderly aged ≥65 years in Ontario [11] and 35.4% among community dwelling elderly in the Calgary Health region aged ≥65 years using the Modification of Diet in Renal Disease (MDRD) equation [12]. In another study in Alberta, the age and sex adjusted prevalence was estimated to be 59.5 per 1000 populations among First Nations and 67.5 per 1000 population among non-First Nations [13].

1.1.4.3 Diabetic kidney disease in Canada and Alberta

According to Canadian Diabetes Association, fifty percent of diabetic patients present with CKD. In Canada, CKD associated with diabetes is the primary cause of kidney failure [1]. Individuals with diabetes are 5.9 times more likely to be hospitalized with kidney disease and 12 times more likely to develop ESRD than those without diabetes according to the 2008-2009 estimates [4]. Diabetic nephropathy contributed to the 35% of the incident cases of ESRD in 2010 [14].

Data from the Alberta Diabetes Surveillance System (ADSS) showed that the rate of developing ESRD was 12 times greater in patients with DM comparing without DM in 2009. In addition, the prevalence of ESRD patients with DM increased from 39% in 1997 to 56% in 2009 [15].

1.1.5 Patients with concomitant DM and CKD are at high risk of poor health outcomes

1.1.5.1 Cardiovascular risk

The prevalence of cardiovascular disease defined as myocardial infarction, heart failure, coronary heart disease death or stroke, is high among patients with diabetes, and is two to three times higher for type 2 DM patients than people without diabetes [16-20]. Recent studies presented that CKD is a potent predictor of cardiovascular disease risk (CVD) [21, 22]. The adjusted hazard ratio for cardiovascular events was 2.0 for estimated glomerular filtration rate (eGFR) of 30 to 44 ml/min/1.73 m^2 and 2.8 for eGFR of 15 to 29 ml/min/1.73 m^2 compared to eGFR $\geq$60 ml/min/1.73 m^2 [21]. CKD patients were two times more likely to have CVD and five times more likely to be hospitalized due to congestive heart failure (CHF) compared to patients without CKD [23]. Therefore, patients presenting with either diabetes or CKD are at higher risk of CV events. In a retrospective study with 1 million US Medicare patients of 65 years of age or older between 1998 and 1999, compared to patients without CKD or DM, patients with both conditions had higher incidence of atherosclerotic vascular disease (adjusted HR 1.41, 95% CI 1.37, 1.44) and congestive heart failure (adjusted HR 1.79, 95% CI 1.75, 1.83). The incidence of acute myocardial infarction (AMI) was 1.6 per 100 person years in non-diabetic non-CKD patients, 3.2 per 100 person years in diabetic non-CKD patients, 3.9 per 100 person years in non-diabetic CKD patients and 6.9 per 100 person years in diabetic CKD patients [24]. Hence the cardiovascular risk is additive for patients having both of these conditions [25-27], justifying

people with diabetes and CKD as a high risk condition warranting specific intervention.

1.1.5.2 Treating risk factors can improve outcomes in people with diabetes and CKD

Although this population with uncontrolled hypertension is at high risk, certain treatments can improve outcomes. Hypertension resulted in albuminuria and progressive kidney damage in diabetic nephropathy patients. Recent large scale clinical trials presented the clinical benefit of controlling hypertension on the risk of progressing to established diabetic nephropathy. According to the RENAAL study, reducing systolic BP of less than 140 mm Hg resulted in 50% lower risk of ESRD in patients with type 2 diabetes and nephropathy [28]. Management of dyslipidemia with statin had been recommended to have a beneficial effect for patients with diabetes and renal impairment [29]. In a pooled analysis with diabetic patients with CKD, anemia was significantly associated with both CVD and all cause mortality in concomitant presence of CKD. Anemia was found to be a significant risk factor in participants with CKD and diabetes for MI/fatal coronary heart disease (HR 1.64, 95%CI 1.03, 2.61) or all cause mortality (HR 1.88, 95%CI 1.33, 2.66) [30].

1.1.6 Remote-dwellers: health outcomes and health service

Due to the universal health care system in Canada, patients do not need to consider the cost of physician visits. However patients living in remote and rural regions need to take into account travel time and distance, which have certain implications for health outcomes. Specifically, residing at a distance from providers may impede access to care and compromise health outcomes. Therefore, optimizing the location of current and future health facilities is an important potential tool for improving health outcomes.

1.1.6.1 Remote-dwelling Canadians with kidney failure have poor health outcomes

In a prospective observational cohort of 20,994 hemodialysis patients, the respondents were asked to choose the travel time required to reach his/her dialysis unit from one of the provided four options of ≤15, 16 to 30, 31 to 60 and >60 minutes. They found that patients traveling longer than 60 minutes were more likely to die compared to those travelling 15 minutes or less (RR 1.20, P value=0.05). Although it was not statistically significant, there was a trend towards an increased risk (5%) of hospitalizations for those travelling longer than 60 minutes compared to those within 15 minutes [31].

Another study of peritoneal dialysis patients in Canada examined the impact of distance of patients' residences from nephrologists in selecting peritoneal dialysis and on mortality risks. The distance categories of patients' residences were of 50 km, 50 to 150, 150 to 300 and >300 km from nephrologists. Patients residing further away had significantly higher mortality compared to those living within 50 km (adjusted HR 1.17, 95% CI 1.07, 1.27 for 50 to 150 km; 1.07 (0.95, 1.21) for 150 to 300 km and 1.15(1.00, 1.32) for >300 km distance [32].

In a study with 31,452 chronic kidney disease patients with less than eGFR of <45 ml/min/1.73 m^2, the distance from the closest nephrologist was categorized into 0 to 50, 50 to 100, 100 to 200 and >200 km. Patients living >200 km away from the closest nephrologists were more likely to die [adjusted OR 1.23, 95%CI 1.12, 1.34] or be hospitalized [adjusted OR 1.5, 95% CI 1.3, 1.7] than those living within 50 km of a rural nephrologist [33].

1.1.6.2 Health service delivery in remote regions: current status

In Canada:

In 2000, approximately 6.7 million people lived in rural Canada (21.7 percent of the total population), but rural Canada had only 41,502 (17.9%) registered nurses, compared to urban Canada which had 187,819 (80.8%) registered nurses. The number of registered nurses was 62 in rural Canada and 78 in urban Canada per 10000 population. In rural Alberta, the number of registered nurses was 4301 compared to 17719 registered nurses in urban Alberta and the number of registered nurses was 56 in rural Alberta and 80 in urban Alberta per 10000 population [34].

According to the Canadian Institute for Health Information, only 12.6% of pharmacists were working in the rural and remote regions of Canada in 2009. In Alberta, 513 (14%) pharmacists were employed in the rural and remote communities compared to 3158 (86%) pharmacists for the urban communities [35]. In a study in Quebec, patients with three major chronic diseases, atherosclerosis, osteoporosis and diabetes were enrolled to compare rural and urban communities in terms of chronic disease management and health related issues. The study reported that material resources utilization (index revascularization or osteodensitometry) rates and the specialist consultation rates were statistically lower in small towns and rural areas compared to metropolitan areas. In addition, morbidity rates were statistically higher for atherosclerosis and diabetes in small towns and in rural areas [36].

In the United States:

In a US study with 973 adult diabetic patients, after adjusting for age, seasonality and insulin usage, the driving distance from home to the site of primary care was significantly associated with HbA1c (0.07% per 10 km, 95% CI 0.03 to 0.11%) in rural participants. The study summarized that there was an increase of 0.25 percent of HbA1c level for every 35 km of driving distance [37]. Since good glycemic control (lower HbA1c) is associated with better clinical outcomes, this suggests worse outcomes among remote-dwellers with diabetes.

Another US study with 781 diabetic patients found decreasing insulin usage among diabetic adults with higher driving distance to primary care facilities. After adjusting for social, demographic and clinical factors, the odds ratio for insulin use for those living within 10 km was 2.29 (95% CI 1.35, 3.88) compared to those living outside 10 km and OR of insulin use for each kilometer of driving distance was 0.97 (CI 0.95, 0.99) [38].

1.1.6.3 Physicians supply in remote Canadian communities

Centre for Rural and Northern Health Research (CRaNHR) study reported that physicians were not equally distributed in urban and rural remote communities across the provinces in Canada. In 1993, below 45^{o} north latitude (Halifax, Toronto and Southwestern Ontario), the population physician ratio was 476 and 91% population resided within 5 km of a physician and from the north latitude of 45^{o} to 49^{o} (Montreal, Vancouver, Ottawa, Winnipeg), the ratio was 448 and 87% resided within 5 km of a physician. However, in 65-69^{o} north latitude (northern part of Yukon and middle parts of the Northwest Territories) there are 3974 people served by a single physician and nearly two-thirds (64%) of the population resided at least 100 km away from the nearest

physician [39].

In 1996, around 9.8% of Canadian physicians worked in small town and rural areas and the disparity was more acute in regards to specialist physicians [40]. In 2000, around 4% of medical specialists served in rural communities [41]. In 2004, around 16% family physicians and 2.4% specialists were located in rural and small town in Canada and served 21.1% of national population resided there. Out of 6502 internal medicine specialists, only 123 worked in rural area in comparison to 6373 specialists served for the urban patients [42].

It had been shown that 95% of the family physician visits were held by older persons in their own communities in northern Ontario whereas two-thirds of the specialist visits occurred in the other cities requiring remote-dwellers to travel for long distances [43].

In summary, effective treatments are available for diabetic kidney disease, but remote-dwellers often receive suboptimal care. Therefore, geography can be considered as potentially reversible barrier for patient care in Canada. We could build new clinics to deliver effective treatments to remote-dwellers if we had better information on the optimal locations for such clinics. Studies are needed that show how to overcome this potential barrier - and make recommendations for policy makers about how to improve health services for remote regions. Accessible nephrology services in the remote communities might improve the quality of care for the rural Canadians and reduce disease burden by avoiding untoward health conditions including cardiovascular morbidity, hospitalizations and mortality. If possible, multidisciplinary clinics should be prioritized as remote residents can benefit most from having early diagnosis to coordinated care.

Sometimes establishment of a new facility is based on political influence or simply guesswork about the local burden of disease, which is potentially inefficient and inequitable. Using available advanced technology and existing health care information, a systematic approach can be adopted which can help policy makers to take efficient decisions. In this context, geographic information system (GIS) analysis may be a useful tool to study the access of care in distant locations. In the next section, we discussed GIS and its potential applications to health research.

1.1.7 Usage of Geographic Information Systems (GIS) in health science

A geographic information system (GIS) is a computer based system including hardware and software tools to capture, organize, analyze and display spatially referenced data, expressed as Cartesian coordinates, latitude and longitudes or other units [44]. It combines cartography and multivariate statistical analysis to investigate statistical relationship among variables influencing health outcomes varying place to place (spatial analysis) and illustrates the results in visual maps [45].

GIS applications are as follows: disease surveillance such as disease mapping and disease modeling; risk analysis related to ecological or environmental health; health access and planning using network analysis; community health profiling; and general or methodological applications such as spatial clusters, smoothing, kriging, autocorrelation and regression mapping [45].

1.1.7.1 Spatial analysis and GIS

Statistics based spatial analysis:

Space-time interaction tests are widely used in epidemiological studies [46]. Space time

clustering or interaction refers to a tendency for groups of cases to occur in clusters in the same sub area over the same sub period of time [47]. The Knox method was a popular statistical method to test space-time interaction [47-49].

The Knox method can be applied to information related to time and location of each case and without having any controls. Whether each possible pair of cases is classified as close or not close to each other depends upon some critical cutoff distance and time (specified a priori). The observed number of pairs locate close spatially and temporally are compared with an expected number based on the null hypothesis of no space-time clustering. If too many cases are classified as close in terms of both time and space, then space-time interaction is established [46, 47]. However it has been suggested that this method can be biased when population growth is not constant across all geographical units and this is referred to as population shift bias [48].

A Finnish Study used Bayesian ecological modeling (Markov chain Monte Carlo and Bayesian spatial conditional autoregressive model) and GIS to identify the risk of type 1 diabetes with the residential location in respect to level of urbanization in Finland and revealed association between risk of diabetes and rural residence [50]. Bayesian spatial conditional model is computationally intensive [51]; the need to assume appropriate prior distribution for the spatial random effects [52], subjective specification of such prior distribution is a long standing debate in statistics. The complex nature of the spatial and spatio-temporal models complicate the modeling process, therefore efficient computation algorithms and dimension reduction methods are required [53].

In a US study, the geographical distribution of 91,507 incident end stage renal disease (ESRD) cases were presented by using a statistical smoothing technique which smooths rates according to the variability of the local ESRD incidence rates, ignoring geographical information on clustering of events and reported areas of high and low incidence areas. This technique uses empirical Bayes approach requires gamma prior distribution specific to each county [54].

These are some examples of the statistics based spatial analysis approach using advanced statistical techniques. However, their potential drawbacks limit the applicability of this approach.

GIS based spatial analysis:

For GIS based approach to model and map disease clusters, two approaches (kernel estimation and cartogram) have been suggested to identify the spatially proximate incidences of a particular disease [55].

In the South East region of Sweden, kriging (a geostatistical technique) was used to estimate the spatial distribution of incident diabetes cases in children taking spatial autocorrelation into account [56]. Kriging, a popular spatial technique for interpolation and smoothing, estimates point values from the surrounding known point value, considered as a method of spatial prediction [44]. Disadvantages of kriging include marked smoothing effect which tend to underestimate high original value and overestimates low values [44]; the need to use empirical Bayes estimation or other methods to smooth data prior to kriging to deal with heterogeneous variances in the regional estimates and potentiality of negative interpolations [57]; and substantially more computing and modeling time as compared with available alternatives [58].

In a study in UK, high density areas (referred to as hotspots) of traffic accidents were the areas of interest to emphasize the spatial patterns of accidents in road safety campaign. The paper used GIS and kernel density estimation to identify high density zones and compared the findings with the results of K-means clustering algorithm [59]. Kernel density estimation transforms point data to continuous density surface map so that density can be estimated for any point of the surface. The entire study region is divided into grid cells of predetermined size and circles with predefined radius known as kernels are placed around the centre of each grid cell. Density values are calculated for each cell in the whole study region [60]. Kernel density estimation uses a decay function to assign smaller values for those points locating still in the neighbourhood but more distant from a case [61]. The limitations of this method include static bandwidth, which does not consider underlying population density. Although the adaptive bandwidth uses the background population to calculate different sizes of kernel for each individual case, it needs to limit the maximum distance of the bandwidth as well. Therefore, for both static and adaptive methods, bandwidth limits are arbitrarily selected, resulting in over or undersmoothing the original data and impact on subsequent analysis [61].

A Swedish study of 1871 children ≤16 years of age with Type 1 DM investigated how socioeconomic factors and population density contribute to the geographical variation of type 1 diabetes in Sweden. Rural-urban gradients were estimated using point in polygon search in GIS software. The result showed that children in areas of low average family income, large families and relatively lower education status had reduced risk and those living in affluent areas were at higher risk of developing type 1 diabetes in five counties in south-eastern Sweden. However, the background population and socio-economic data were aggregated at a predefined grid scale of

4000 square meter [62].

In all these studies, either statistics based or GIS based spatial analysis techniques were used for chronic diseases including diabetes and ESRD patients. However, each spatial technique has certain limitations. A review described the capabilities and limitations of GIS in spatial analysis, spatial statistics and modeling. This study revealed that other programs are needed to perform spatial analysis in addition to commercially available GIS softwares such as ArcGIS or MapInfo [63]. In spatial epidemiology, studies mostly focused on disease mapping, disease clustering or geographical correlation analysis [57]. For this project, we were interested to study disease clusters. Therefore, the following sections discuss various methods for assessing disease clusters, and their potential applications, advantages and disadvantages.

1.1.7.2 Disease cluster estimation and GIS mapping

The term ‘cluster’ is commonly used in epidemiology when the observed number of cases of a disease is higher than the expected number of disease cases in a defined community over a specific period of time for a particular disease. The cluster estimation method deals with the null hypothesis of spatial randomness. This technique can be used to determine whether the observed counts of diabetic kidney cases for a particular community are higher than expected for that community in Alberta.

To test the spatial randomness with the null hypothesis that disease risk is the same in all parts of the map, three types of tests are commonly used: global clustering test, cluster detection tests and focused cluster tests. Global clustering tests examine the clustering pattern throughout the study

area without specifying the location of clusters, e.g Pearson's method, Cuzick-Edwards test etc. Cluster detection tests identify local clusters along with their locations and test their significance at the same time, e.g; Besag-Newell test, spatial scan statistics, Turnbull's method, bivariate local indicator of spatial association (LISA) etc. Focused cluster tests are used when disease risk is presumed to be higher close to a specific geographical feature; available methods include Lawson and Waller score test and Bithell's linear risk score test [64].
In addition to these methods, there are other computational methods such as the hierarchical method.

Among these three types of tests, the research question should determine the type of test to be chosen. We were interested in location of clusters and their significance; therefore, we chose to perform cluster detection tests. As mentioned above, there are several potential choices for cluster detection tests, so we reviewed their applications to determine the best test statistic for our project.

1.1.7.2.1 Cluster detection methods

Hierarchical method:

The hierarchical cluster analysis (Ward's method) was used to identify healthy and unhealthy neighborhoods in terms of the clustering of adverse health outcomes (diabetes prevalence), eating behavior and physical activity in Nashville, Tennessee [65]. The applied Ward's method was used for defining clusters of diabetic patients, after accounting for case complexity or case-mix [66]. Hierarchical methods are not considered to be useful for these analyses, because they assume that the distance between any two cases is expected to be the same in the entire dataset

which may be incorrect [67].

Spatial autocorrelation:

In order to plan for community based interventions, spatial pattern of diabetes and high-risk socioeconomic areas were identified by using statistical spatial autocorrelation and GeoDa software [68]. This spatial technique (bivariate local indicator of spatial association [LISA]), clarified the relationship between socioeconomic determinants of health and diabetes prevalence in London, Ontario [68]. Moran's I statistic is used as a global index of spatial autocorrelation; however this statistic does not account for underlying population heterogeneity. For LISA technique, the Moran's I statistic is decomposed to a standard score to detect local clusters. The normal score is used to assess the significance of candidate clusters, and locations with unusually high (statistically significant) scores are designated as clusters [69].

Besag and Newell method:

A study was conducted in a Kenyan district to examine the geographical and temporal patterns of hospital admissions with severe malaria. In this study, the Besag and Newell test was used to assess clusters in space and time [70]. The Besag and Newell method reduces computation time by performing its search selectively centered on a case point without examining areas without any cases. However, this method may miss some clusters due to its selective nature, especially in datasets where clusters have varying case counts [67].

Turnbull's method:

To investigate the occurrence of breast and lung cancer incidences in relation to proximity from

point source pollution, major highways and floodplains of rivers in Michigan USA, the researchers performed smoothing techniques and spatial cluster statistics such as global (local Moran's test), local (Turnbull's method), two focused tests (Lawson and Waller score test and Bithell's linear risk score test) using GeoDa, ClusterSeer and TerraSeer's STIS softwares [71]. Turnbull's cluster evaluation permutation procedure (CEPP) searches for clusters in geographic window of varying size while maintaining a constant population at risk; due to the same population size the calculated rates can be represented as case counts [69].

Ripley's K function and G statistics:

In addition to understand the spatial pattern of a single epidemic, epidemiologial data can be combined across time to understand the temporal stability of disease incidence in a community. One study used Ripley's K function and G_i statistics to explain the clustering pattern in space over time and reported localized and periodic sources of typhoid fever in Washington [72]. Due to the increasing size of the databases and computational complexity, methods like the K function (Ripley 1976) [73] require prolonged analysis time [67], which may be impractical for larger datasets.

DMAP spatial filtering:

By using DMAP spatial filtering software and GIS maps, statistically significant hotspots of diabetes related health problem were investigated for 3522 patients who visited emergency departments in Los Angeles. This hotspot analysis identified the need to develop targeted interventions for the communities with food insecurity, scarce activity spaces and access to health services [74]. Spatial filtering use predetermined circle sizes either by geographical or

constant or near constant population size rather than constant geographic size; therefore the cluster size depends on the filter size following the radius of the circles and does not adjust for multiple testing [75].

The limitations of the cluster detection methods are: inability to consider underlying background population, difficulty in handing increasing size of the database, often unrealistic assumptions that must be satisfied for appropriate use of the statistical methods, and over-reporting of "significant" clusters. Turnbull in 1990 [76], Kulldorff in 1997 [77], Openshaw in 1987 [78] and Besag and Newell in 1991 [79] developed methods to address inhomogeneous background population (see Appendix 1 for detailed comparison of these methods). For many of these methods, clusters must be assumed to follow a particular shape (such as a circle), even though there is no biological rationale for such a distribution. Spatial Scan statistics can minimize these limitations [67]. A more detailed explanation and examples of spatial scan statistics is provided in the next sections.

1.1.7.2.2 Spatial Scan statistics, SaTScan software and its application

Spatial scan statistics using SaTScan software is commonly used to determine whether an event of interest is randomly distributed over space and time in a defined geographical location for a specific period of time. Therefore this technique is useful to identify significant local spatial clusters of patients with particular clinical characteristics (see Appendix 2 for details). The method uses Poisson model, Bernoulli model, permutation model or exponential model. Unique or aggregated geographical locations can be used as a unit of interest [80].

The advantage of this technique as compared to other existing cluster detection methods is as follows: it considers inhomogeneous background population, can adjust for categorical variables, can account for temporal trends, is robust to missing data, and scans multiple datasets at a time to define space-time clusters [80]. SaTScan is particularly advantageous when the approximate cluster size is not known a priori. These multiple theoretical advantages make SaTScan an attractive potential method for GIS analyses.

In terms of power and precision, the spatial scan statistics has been considered to perform well [69, 81]. A US study evaluated power and precision of eight different tests to assess the performance and reported that SaTScanE (elliptic window) gained the best power, followed by a global cluster detection test, TangoPDM, then SaTScanO (oval window), Turnbull's CEPP, Besag-Newell's method and the last group with lower power was Moran's I statistic. Moran I statistic has been found to have the worst performance among the eight tested methods [69].

SaTScan technique has been gaining popularity for cluster detection in different patient populations [82-95] (see Appendix 3 for details). SaTScan can be used for adjustments of other covariates. A study in North Carolina performed both unadjusted and adjusted spatial analysis for age, race, parity, smoking and Medicaid status by adding one or two covariates at a time in Spatial scan and presented how the geographical distribution of gastroschisis cases (an abdominal wall birth defect) changed according to the effect of a specific covariate [90]. Another method of accounting for the relation between clusters and differences in the underlying population is to do SaTScan analysis followed by regression analysis. In a study in Winnipeg spatial and non-spatial linear regression analyses were performed to test the association between

prevalent diabetes clusters and socio-demographic, environmental and lifestyle factors [92]. SaTScan is used with other technique for refining detection of meaningful clusters using SaTScan analysis such as Cuzick-Edwards test, Besag-Newell test using Dcluster package [93], spatial filtering [94], global spatial autocorrelation using GoeDa [95].

Newer cluster detection methods have been developed in recent decades to adjust for inhomogeneous population distribution and intensive computational time, limitations with the earlier methods. Besag and Newell method and GAM method identify clusters by using overlapping circles, the former one performs selective searches and the later detect clusters regardless of the administrative boundaries. CEPP method uses overlapping circles and addresses multiple testing issues by Bonferroni correction for a cluster size known a priori. However, SaTScan technique is useful for the unknown cluster size and addresses multiple testing by likelihood ratio tests.

Therefore SaTScan seems to perform well for detection of spatial and spatio-temporal clusters. This spatial scan statistic is equally applicable for communicable and non-communicable disease populations, and results can potentially complement other statistical and spatial techniques, such as those previously used by our group to examine facility locations for Albertans with kidney disease. Hence, we selected SaTScan as the method we would use in the current study to investigate cluster effects among people with CKD and diabetes in Alberta.

1.2 Hypothesis and research objectives

The prevalence of diabetes and chronic kidney diseases is increasing worldwide. Patients suffering from both diabetes and chronic kidney disease are at high risk of cardiovascular mortality and morbidity including atherosclerotic heart disease and congestive heart failure. Proper management of diabetes, chronic kidney disease and their consequences (such as uncontrolled hypertension, anemia and dyslipidemia) can result in improvement of clinical consequences in this patient population.

Geographical distance from the specialized services is associated with sub-optimal process of care and results in adverse health outcomes. Provision of specialized nephrology services in the remote distant communities can mitigate this unmet care gaps and minimize adverse health consequences. Hence it is important to identify new facility locations to maximize clinical benefits for the remote-dwelling diabetic kidney patients.

However, the best way to choose potential new locations is yet to be determined. Such decisions are often based on perceptions about a given community's demand for services or political factors. In addition, comprehensive efforts, extensive planning, long term time commitment and huge economic resources are required to build up a new clinic. Selection of the best clinic locations is the key component to the optimum utilization of these limited resources, and therefore a more objective method would be helpful for the decision makers.

Among the spatial analytic techniques and cluster detection tests, spatial scan statistic (SaTScan) addresses underlying inhomogeneous population density without specifying cluster size or shape

a priori. It has been used for spatial cluster detection in many disease populations and has a better power and precision than alternatives. Spatial analysis with SaTScan software and generating maps using ArcGIS software would allow us a comprehensive understanding of the spatial pattern of disease rather than simply mapping the density plots using a predefined scale. Thus, use of SaTScan and ArcGIS together is superior to use of ArcGIS alone, as it allows us to identify the prevalent clusters of CKD and DM in underserved communities without any predefined criteria for cluster size or radius. A second advantage of SaTScan is that its superior resolution allows more precise location of potential new clinics – rather than simply identifying large geographical areas (100 km square grids, or 10,000 km^2) where new clinics might be needed.

We used data from a population based registry and performed spatial analysis using SaTScan software and buffer analysis using ArcGIS software to identify potential locations for the new clinics that might serve the maximum number of high risk patients.

Our primary goal was to investigate the spatial distribution of remote-dwellers with both diabetes and CKD in Alberta suggesting the optimal locations for new nephrology clinics that could serve this population at high risk of adverse health outcomes.

Our second objective was to estimate the proportion of high risk diabetic kidney patients (in terms of demographic and clinical factors) residing within the identified clusters in Alberta.

References

1. Clinical practice guidelines for the prevention and management of diabetes in Canada. In: *Can J Diabetes.* vol. 32: Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2008; 2008: S1-S201.
2. Living with kidney disease. 4 edn: The Kidney Foundation of Canada; 2006.
3. Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia: World Health Organization. International Diabetes Federation; 2006.
4. Diabetes in Canada: Facts and figures from a public health perspective. In. Ottawa: Public Health Agency of Canada; 2011.
5. An economic tsunami the cost of diabetes in Canada: Canadian Diabetes Association; 2009.
6. Balko SU, Hugel G, Johnson JA: Diabetes Trends in Alberta. vol. 5: Alberta Diabetes Surveillance System; 2011.
7. Johnson JA, Balko SU, Hugel G, Low C, Svenson LW: Increasing incidence and prevalence with limited survival gains among rural Albertans with diabetes: a retrospective cohort study, 1995-2006. *Diabetic medicine : a journal of the British Diabetic Association* 2009, 26(10):989-995.
8. Lau RS, Ohinmaa A, Johnson JA: Predicting the Future Burden of Diabetes in Alberta from 2008 to 2035. *Can J of Diabetes* 2011, 35(3):274-281.
9. Zhang QL, Rothenbacher D: Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC public health* 2008, 8:117.
10. Facing The Facts 2012: The Kidney Foundation of Canada.
11. Garg AX, Papaioannou A, Ferko N, Campbell G, Clarke JA, Ray JG: Estimating the

prevalence of renal insufficiency in seniors requiring long-term care. *Kidney international* 2004, 65(2):649-653.

12. Hemmelgarn BR, Zhang J, Manns BJ, Tonelli M, Larsen E, Ghali WA, Southern DA, McLaughlin K, Mortis G, Culleton BF: Progression of kidney dysfunction in the community-dwelling elderly. *Kidney international* 2006, 69(12):2155-2161.
13. Gao S, Manns BJ, Culleton BF, Tonelli M, Quan H, Crowshoe L, Ghali WA, Svenson LW, Hemmelgarn BR: Prevalence of chronic kidney disease and survival among aboriginal people. *Journal of the American Society of Nephrology : JASN* 2007, 18(11):2953-2959.
14. Annual Report: Treatment of End-Stage Organ Failure in Canada, 2001 to 2010 : Canadian Organ Replacement Register.
15. Klarenbach S, Hemmelgarn BR, Jindal KK, Tonelli M: Chapter 8. Diabetes and Kidney Disease in Alberta : *Alberta Diabetes Atlas.* 2011.
16. Kannel WB, McGee DL: Diabetes and cardiovascular disease. The Framingham study. *JAMA : the journal of the American Medical Association* 1979, 241(19):2035-2038.
17. Manson JE, Colditz GA, Stampfer MJ, Willett WC, Krolewski AS, Rosner B, Arky RA, Speizer FE, Hennekens CH: A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women. *Archives of internal medicine* 1991, 151(6):1141-1147.
18. Stamler J, Vaccaro O, Neaton JD, Wentworth D: Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. *Diabetes care* 1993, 16(2):434-444.
19. Fox CS, Coady S, Sorlie PD, Levy D, Meigs JB, D'Agostino RB, Sr., Wilson PW, Savage

PJ: Trends in cardiovascular complications of diabetes. *JAMA : the journal of the American Medical Association* 2004, 292(20):2495-2499.

20. Buse JB, Ginsberg HN, Bakris GL, Clark NG, Costa F, Eckel R, Fonseca V, Gerstein HC, Grundy S, Nesto RW *et al*: Primary prevention of cardiovascular diseases in people with diabetes mellitus: a scientific statement from the American Heart Association and the American Diabetes Association. *Diabetes care* 2007, 30(1):162-172.

21. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY: Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *The New England journal of medicine* 2004, 351(13):1296-1305.

22. Weiner DE, Tighiouart H, Amin MG, Stark PC, MacLeod B, Griffith JL, Salem DN, Levey AS, Sarnak MJ: Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies. *Journal of the American Society of Nephrology : JASN* 2004, 15(5):1307-1315.

23. Collins AJ, Li S, Gilbertson DT, Liu J, Chen SC, Herzog CA: Chronic kidney disease and cardiovascular disease in the Medicare population. *Kidney international Supplement* 2003(87):S24-31.

24. Foley RN, Murray AM, Li S, Herzog CA, McBean AM, Eggers PW, Collins AJ: Chronic kidney disease and the risk for cardiovascular disease, renal replacement, and death in the United States Medicare population, 1998 to 1999. *Journal of the American Society of Nephrology : JASN* 2005, 16(2):489-495.

25. Van der Meer IM, Ruggenenti P, Remuzzi G: The diabetic CKD patient--a major cardiovascular challenge. *Journal of renal care* 2010, 36 Suppl 1:34-46.

26. Patel T, Charytan DM: Cardiovascular complications in diabetic kidney disease.

Seminars in dialysis 2010, 23(2):169-177.

27. Radbill B, Murphy B, LeRoith D: Rationale and strategies for early detection and management of diabetic kidney disease. *Mayo Clinic proceedings Mayo Clinic* 2008, 83(12):1373-1381.

28. Toto RD: Reducing cardiovascular events in high-risk patients: the challenge of managing hypertension in patients with diabetic renal disease. *J Clin Hypertens (Greenwich)* 2007, 9(11 Suppl 4):16-25.

29. Molitch ME: Management of dyslipidemias in patients with diabetes and chronic kidney disease. *Clinical journal of the American Society of Nephrology : CJASN* 2006, 1(5):1090-1099.

30. Vlagopoulos PT, Tighiouart H, Weiner DE, Griffith J, Pettitt D, Salem DN, Levey AS, Sarnak MJ: Anemia as a risk factor for cardiovascular disease and all-cause mortality in diabetes: the impact of chronic kidney disease. *Journal of the American Society of Nephrology : JASN* 2005, 16(11):3403-3410.

31. Moist LM, Bragg-Gresham JL, Pisoni RL, Saran R, Akiba T, Jacobson SH, Fukuhara S, Mapes DL, Rayner HC, Saito A *et al*: Travel time to dialysis as a predictor of health-related quality of life, adherence, and mortality: the Dialysis Outcomes and Practice Patterns Study (DOPPS). *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2008, 51(4):641-650.

32. Tonelli M, Hemmelgarn B, Culleton B, Klarenbach S, Gill JS, Wiebe N, Manns B: Mortality of Canadians treated by peritoneal dialysis in remote locations. *Kidney international* 2007, 72(8):1023-1028.

33. Rucker D, Hemmelgarn BR, Lin M, Manns BJ, Klarenbach SW, Ayyalasomayajula B,

James MT, Bello A, Gordon D, Jindal KK *et al*: Quality of care and mortality are worse in chronic kidney disease patients living in remote areas. *Kidney international* 2011, 79(2):210-217.

34. Supply and Distribution of Registered Nurses in Rural and Small Town Canada, 2000: Canadian Institute for Health Information; 2002.

35. Pharmacists in Canada, 2009. In. Ottawa: Canadian Institute of Health Information; 2010.

36. Vanasse A, Courteau J, Cohen AA, Orzanco MG, Drouin C: Rural-urban disparities in the management and health issues of chronic diseases in Quebec (Canada) in the early 2000s. *Rural and remote health* 2010, 10(4):1548.

37. Strauss K, MacLean C, Troy A, Littenberg B: Driving distance as a barrier to glycemic control in diabetes. *Journal of general internal medicine* 2006, 21(4):378-380.

38. Littenberg B, Strauss K, MacLean CD, Troy AR: The use of insulin declines as patients live farther from their source of care: results of a survey of adults with type 2 diabetes. *BMC public health* 2006, 6:198.

39. Ng E, Wilkins R, Pole J, Adams OB: How far to the nearest physician? vol. 8; 1997: 19-31.

40. Pitblado JR, Pong RW: Are Physicians Distributed Equitably in Canada? From the report Geographic Distribution of Physicians in Canada. In.: Health Canada; 1999.

41. Rural Health in Rural Hands Strategic Directions for Rural, Remote, Northern and Aboriginal Communities. In.: Ministerial Advisory Council on Rural Health; 2002.

42. Geographic Distribution of Physicians in Canada: Beyond How Many and Where: Canadian Institute for Health Information; 2005.

43. Pong RW, Pitblado JR: Don't take 'geography' for granted! Some methodological issues

in measuring geographic distribution of physicians. *Canadian Journal of Rural Medicine* 2001, 6(2):103-112.

44. Moore DA, Carpenter TE: Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic reviews* 1999, 21(2):143-161.

45. Nykiforuk CIJ, Flaman LM: Exploring the Utilization of Geographic Information Systems in Health Promotion and Public Health : Centre for health promotion studies. The School of Public Health. The University of Alberta; 2008.

46. Kulldorff M, Hjalmars U: The Knox method and other tests for space-time interaction. *Biometrics* 1999, 55(2):544-552.

47. Siemiatycki J, Colle E, Aubert D, Campbell S, Belmonte MM: The distribution of type I (insulin-dependent) diabetes mellitus by age, sex, secular trend, seasonality, time clusters, and space-time clusters: evidence from Montreal, 1971-1983. *American journal of epidemiology* 1986, 124(4):545-560.

48. Samuelsson U, Johansson C, Carstensen J, Ludvigsson J: Space-time clustering in insulin-dependent diabetes mellitus (IDDM) in south-east Sweden. *International journal of epidemiology* 1994, 23(1):138-142.

49. Gustafsson B, Carstensen J: Evidence of space-time clustering of childhood acute lymphoblastic leukaemia in Sweden. *British journal of cancer* 1999, 79(3-4):655-657.

50. Rytkonen M, Moltchanova E, Ranta J, Taskinen O, Tuomilehto J, Karvonen M: The incidence of type 1 diabetes among children in Finland--rural-urban difference. *Health & place* 2003, 9(4):315-325.

51. Carroll C, Johnson DS, Dunk JR, Zielinski WJ: Hierarchical Bayesian spatial models for

multispecies conservation planning and monitoring. *Conservation biology : the journal of the Society for Conservation Biology* 2010, 24(6):1538-1548.

52. Osei PP: Statistical Methods for Disease Mapping. In: African Institute for Mathematical Sciences (AIMS); 2010.

53. Arab A, Hooten MB, Wikle CK: Hierarchical spatial models.

54. Foxman B, Moulton LH, Wolfe RA, Guire KE, Port FK, Hawthorne VM: Geographic variation in the incidence of treated end-stage renal disease. *Journal of the American Society of Nephrology : JASN* 1991, 2(6):1144-1152.

55. Koch T, Denike K: GIS approaches to the problem of disease clusters: a brief commentary. *Soc Sci Med* 2001, 52(11):1751-1754.

56. Samuelsson U, Lofman O: Geographical mapping of type 1 diabetes in children and adolescents in south east Sweden. *Journal of epidemiology and community health* 2004, 58(5):388-392.

57. Berke O: Exploratory disease mapping: kriging the spatial risk function from regional count data. *International journal of health geographics* 2004, 3(1):18.

58. Azpurua M, Ramos KD: A Comparison of Spatial Interpolation Methods For Estimation Of Average Electromagnetic Field Magnitude. *Progress In Electromagnetics Research M* 2010, 14:135-145.

59. Anderson TK: Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Anal Prev* 2009, 41(3):359-364.

60. Thornton LE, Pearce JR, Kavanagh AM: Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. *Int J Behav Nutr Phy* 2011, 8.

61. Carlos HA, Shi X, Sargent J, Tanski S, Berke EM: Density estimation and adaptive bandwidths: A primer for public health practitioners. *International journal of health geographics* 2010, 9.

62. Holmqvist BM, Lofman O, Samuelsson U: A low incidence of Type 1 diabetes between 1977 and 2001 in south-eastern Sweden in areas with high population density and which are more deprived. *Diabetic Med* 2008, 25(3):255-260.

63. Chung K, Yang DH, Bell R: Health and GIS: toward spatial statistical analyses. *Journal of medical systems* 2004, 28(4):349-360.

64. Kulldorff M: Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association* 2006, 101(475):1289-1305.

65. Schlundt DG, Hargreaves MK, McClellan L: Geographic clustering of obesity, diabetes, and hypertension in Nashville, Tennessee. *The Journal of ambulatory care management* 2006, 29(2):125-132.

66. Fischer CJ, Stiefel FC, De Jonge P, Guex P, Troendle A, Bulliard C, Huyse FJ, Gaillard R, Ruiz J: Case complexity and clinical outcome in diabetes mellitus. A prospective study using the INTERMED. *Diabetes & metabolism* 2000, 26(4):295-302.

67. Conley J, Gahegan M, Macgill J: A Genetic Approach to Detecting Clusters in Point Datasets. In. GeoVISTA Center, Department of Geography, The Pennsylvania State University.

68. Tompkins JW, Luginaah IN, Booth GL, Harris SB: The geography of diabetes in London, Canada: the need for local level policy for prevention and management. *International journal of environmental research and public health* 2010, 7(5):2407-2422.

69. Huang L, Pickle LW, Das B: Evaluating spatial methods for investigating global

clustering and cluster detection of cancer cases. *Statistics in medicine* 2008, 27(25):5111-5142.

70. Schellenberg JA, Newell JN, Snow RW, Mung'ala V, Marsh K, Smith PG, Hayes RJ: An analysis of the geographical distribution of severe malaria in children in Kilifi District, Kenya. *International journal of epidemiology* 1998, 27(2):323-329.

71. Guajardo OA, Oyana TJ: A critical assessment of geographic clusters of breast and lung cancer incidences among residents living near the Tittabawassee and Saginaw Rivers, Michigan, USA. *Journal of environmental and public health* 2009, 2009:316249.

72. Hinman SE, Blackburn JK, Curtis A: Spatial and temporal structure of typhoid outbreaks in Washington, D.C., 1906-1909: evaluating local clustering with the Gi* statistic. *International journal of health geographics* 2006, 5:13.

73. Ripley BD: 2nd-Order Analysis of Stationary Point Processes. *J Appl Probab* 1976, 13(2):255-266.

74. Curtis AJ, Lee WA: Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. *International journal of health geographics* 2010, 9:43.

75. Ozdenerol E, Williams BL, Kang SY, Magsumbol MS: Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *International journal of health geographics* 2005, 4:19.

76. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC: Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American journal of epidemiology* 1990, 132(1 Suppl):S136-143.

77. Kulldorff M: A spatial scan statistic. *Commun Stat-Theor M* 1997, 26(6):1481-1496.

78. Openshaw S, Charlton M, Wymer C, Craft A: A Mark 1 Geographical Analysis Machine

for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1987, 1:335-358.

79. Besag J, Newell J: The Detection of Clusters in Rare Diseases. *J Roy Stat Soc a Sta* 1991, 154:143-155.

80. Purpose. SatScan™ software for the spatial, temporal, and space-time scan statistics. [http://www.satscan.org/.]

81. Aamodt G, Samuelsen SO, Skrondal A: A simulation study of three methods for detecting disease clusters. *International journal of health geographics* 2006, 5:15.

82. Amin R, Bohnert A, Holmes L, Rajasekaran A, Assanasen C: Epidemiologic mapping of Florida childhood cancer clusters. *Pediatric blood & cancer* 2010, 54(4):511-518.

83. Reinhardt M, Elias J, Albert J, Frosch M, Harmsen D, Vogel U: EpiScanGIS: an online geographic surveillance system for meningococcal disease. *International journal of health geographics* 2008, 7:33.

84. Wen L, Li C, Lin M, Yuan Z, Huo D, Li S, Wang Y, Chu C, Jia R, Song H: Spatio-temporal analysis of malaria incidence at the village level in a malaria-endemic area in Hainan, China. *Malaria journal* 2011, 10:88.

85. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, Diallo M, Diawara S, Ouattara A, Diakite M *et al*: Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC public health* 2006, 6:286.

86. Bakker MI, Hatta M, Kwenang A, Faber WR, van Beers SM, Klatser PR, Oskam L: Population survey to determine risk factors for Mycobacterium leprae transmission and infection. *International journal of epidemiology* 2004, 33(6):1329-1336.

87. Tiwari N, Adhikari CM, Tewari A, Kandpal V: Investigation of geo-spatial hotspots for

the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. *International journal of health geographics* 2006, 5:33.

88. Qi X, Tong S, Hu W: Spatial distribution of suicide in Queensland, Australia. *BMC psychiatry* 2010, 10:106.

89. Nkhoma ET, Ed Hsu C, Hunt VI, Harris AM: Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980 - 2001. *International journal of health geographics* 2004, 3(1):25.

90. Root ED, Meyer RE, Emch ME: Evidence of localized clustering of gastroschisis births in North Carolina, 1999-2004. *Soc Sci Med* 2009, 68(8):1361-1367.

91. Chaturvedi K: Geographic Concentration of Diabetes Prevalence Clusters in Texas and Their relationship to Age and Obesity.

92. Green C, Hoppa RD, Young TK, Blanchard JF: Geographic analysis of diabetes prevalence in an urban area. *Soc Sci Med* 2003, 57(3):551-560.

93. Zhang Z, Carpenter TE, Chen Y, Clark AB, Lynn HS, Peng W, Zhou Y, Zhao G, Jiang Q: Identifying high-risk regions for schistosomiasis in Guichi, China: a spatial analysis. *Acta tropica* 2008, 107(3):217-223.

94. Trooskin SB, Hadler J, St Louis T, Navarro VJ: Geospatial analysis of hepatitis C in Connecticut: a novel application of a public health tool. *Public health* 2005, 119(11):1042-1047.

95. Fang L, Yan L, Liang S, de Vlas SJ, Feng D, Han X, Zhao W, Xu B, Bian L, Yang H *et al*: Spatial analysis of hemorrhagic fever with renal syndrome in China. *BMC infectious diseases* 2006, 6:77.

Chapter 2: Spatial analysis to locate new clinics for diabetic kidney patients in the underserved communities in Alberta[1]

2.1 Introduction

We identified prevalent patients with CKD and concomitant diabetes using data from the Alberta Kidney Disease Network [1] and the provincial health ministry (Alberta Health and Wellness; AHW), and located their residence relative to existing nephrology clinics using postal codes. Initially we used a commonly applied crude method (shaded maps to illustrate population density) to represent the geographical distribution of the target population. However, this method does not provide sufficient spatial resolution to localize new clinics. Therefore, we proceeded with SaTScan analysis -- a more complex method that provides more detailed results such as expected number of cases, annual rates, relative risk and log likelihood ratio--to detect clusters of patients who lived at greater distances from existing facilities, and graphically plotted the results.

2.2 Methods

2.2.1 Identification of patients with chronic kidney disease (CKD) and diabetes

We selected cases with stage 3-4 CKD (defined as estimated glomerular filtration rate [eGFR] 15-59.9 ml/min/1.73m^2) from all outpatients aged over 18 years who had serum creatinine measured in Alberta at least once between 1st May 2002, and 31st December 2008 and concomitant diabetes mellitus (DM) during the calendar years 1994-2008. We used the last serum creatinine available during the study period and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) study equation to estimate eGFR [2]. We used validated

[1] A version of this chapter including methods and results of spatial analysis has been accepted for publication in Nephrology Dialysis Transplantation on 15th June 2012

algorithms to identify diabetes cases (two physician billing claims in a two-year period or one hospital discharge ever with a diagnosis of diabetes, excluding gestational diabetes) [3].

We focused on identifying patients with earlier (stage 3-4) CKD in Alberta during the study period. Therefore, we excluded 1292 (3.84%) patients who died, out-migrated from Alberta (N=41, 0.12%) and 36 (0.11%) people who were treated with chronic dialysis or kidney transplantation during the study period, drawn from a total of 33,647 cases with diabetes and eGFR 15-59.9 ml/min/1.73m^2.

We assessed comorbidity using the Charlson score (Table 1) for the remaining 32,278 cases, ascertained using physician claims and hospitalization data together with validated algorithms [4]. The total number of Alberta residents (n=2,795,541) were retained in the dataset for use in analyses requiring the total population in each postal code.

2.2.2 Identification of residence locations

We identified the postal code associated with each participant's home address during each calendar year by using the AHW registry file. We used the last available postal code during the study period to represent the residence location for participants whose records were associated with multiple postal codes. All except 159 (<0.1%) of participants had valid postal code data. We obtained the latitude and longitude coordinates associated with the centroid of each postal code by matching these postal codes to the 2008 Postal Code Conversion File. The resulting dataset included residential postal code (with latitude and longitude) as well as demographic and clinical data for each person who was insured by AHW during the calendar years 2002-2008.

Some regions in Canada, including rural areas, are changing and growing – which can potentially require changes to the address system used for mail delivery. For example, the postal code identifiers for some areas have been updated accordingly – including changes from rural to urban status or changes in the specific area referred to by a given postal code. Despite these changes in rural areas postal codes tend to cover a larger geographical area than in urban counterparts, implying less geographical precision when mapping postal codes to residence locations. To overcome these limitations of the postal codes, we used the latest available postal codes, which are the smallest geographical units that are readily available for spatial analysis.

2.2.3 Density plots

We generated the shaded density plots of patients with diabetes and CKD in each geographical unit by using ArcMap (ArcGIS Desktop Release 10, Environmental Systems Research institutes, Redlands, CA). This image was created by a simple point in raster count operation using 100 km square grid cells in a 10 transverse Mercator projection commonly used with spatial datasets. We categorized patients according to the number of cases per 100 km square grid and plotted these categories in the map using different colour shades. The yellow shades represented the lower densities and the brown shades indicated the higher densities of case patients per square grid. We displayed the locations of existing nephrology clinics on the map.

This analysis only represented the case counts per 100 km square grid cells, without considering the total number of population per square grid in the denominator. This simple count-based technique is only efficient when the underlying population is constant across grids. However, there might have been rapid population growth or seasonal variation in population counts in

some communities which may contribute to errors in estimating higher than expected occurrence of disease burden for that community. Therefore, it has been recommended that comparisons should be based on rates rather than the case counts [5]. We calculated disease rates by using spatial analyses with SaTScan software at the postal code level in addition to density plots in the 100 km square predefined grid scale with ArcGIS software. We estimated disease rates in underserved communities by considering existing clinic locations in our analysis.

2.2.4 Clinic locations

A list of all 17 existing clinics providing specialist nephrology care to stage 3-4 CKD patients was obtained from the provincial renal programs. Because we were interested in patients residing at a distance from nephrology service, we established non-mutually exclusive categories of >50; >100; >150 and >200 km, each representing the distance between a patient's residence and the closest nephrology clinic. Buffer zone analysis was used to identify participants in the underserved communities who were living far from the existing nephrology clinics. Accordingly, when the 50 km buffer zone was created, we dropped postal codes of all people living within 50 km radius of existing clinics, retaining only postal codes of those who were residing more than 50 km away. A similar method was used to create non-mutually exclusive 100 km, 150 km and 200 km buffer zones. We used ArcGIS10.0 software (Environmental Systems Research Institute, Inc. (ESRI), California, USA) to create these buffer zone analyses. We geocoded postal codes of the patients' residences based on latitude and longitude. Four buffer zones were created following 'crow fly buffer' technique using the buffer tool in ArcGIS software. The crow fly buffer technique does not take into account road networks and thus travel distance.

The crow fly buffer, also known as circular buffer or airline buffer, uses the Euclidian straight line distance from a point of interest (in our case, existing nephrology clinics) and produces a circular buffer zone. This is a simple method requiring less time and expertise in GIS methods. The potential limitation with this technique is its inability to consider travel barriers; for example bridges, highways, waterways etc. On the other hand, road network buffer also uses a specific distance but extends from a point of interest by following existing road networks, and finally the endpoints are connected by drawing the lines which may result in irregular shaped network buffer [6]. Therefore, network buffer analysis includes detailed road network and would provide more accurate network based measures for actual distance addressing barriers for travel distance taken into analysis [7]. Again, circular buffers would capture overall land areas within a specific distance but road network buffer would capture the areas which would actually accessible to the people using road networks [6]. Therefore, it is possible that the network buffer will miss some pockets of patients within the regions which do not have well developed roadways or when the road network is not updated in the dataset.

In regions with higher connectivity of road networks, crow fly and network buffers provide approximate results but with lower connectivity especially in suburban areas, we may not get approximate results using both techniques [8]. However, in this study we presented four buffer scenarios as a sensitivity analysis to account for this limitation. For example; 100 km crow fly buffer might have dropped some postal codes which might not be dropped if we would perform our analysis with 100 km network buffer. But those postal codes were considered while doing analysis using 50 km crow fly buffer scenario. Thus, 150 km and 200 km buffer scenario might have accounted for 100 km and 150 km buffers.

2.2.5 Spatial analysis

For each of these four buffer zones we did spatial analysis using SaTScan software (Martin Kulldorff and Information Management Services Inc, USA) [9-11] as previously described for patients with other diseases [12-20] setting the maximum population size at 0.5%, 1%, 5% and 10%. Maximum population size 0.5% indicates 0.5% of the total population of the study region (for this study, 0.5% of population of Alberta). It implied that the cluster size in terms of the population counts within the cluster can be maximum 0.5% of the population of Alberta. This explanation is equally applicable for 1, 5 and 10% of population size as well. As we were only interested in clusters of patients who were prevalent at the end of 2008, we did not estimate the spatial variation in temporal trends. The spatial scan statistic using SaTScan software imposed a circular window on the map of Alberta. The window was centered one by one on each of the given grid points (in this analysis we used the latitudes and longitudes of the residents' postal codes) located throughout the study region. For each of the postal code, the radius of the window varied continuously in size from zero to the upper limit specified *a priori* (in this analysis, we specified the upper limit at 0.5%, 1%, 5% and 10% of the population size). For each postal code of interest, a circle radius of 0 implies that the window was defined solely by that single postal code. To test for the existence of clusters, the SaTScan software progressively increased the circle radius up to the maximum size (meaning that surrounding postal codes were progressively included until the maximum population size for the cluster was reached). In this way the software was able to examine a large number of distinct geographical windows to test for the presence of CKD clusters.

For each window, our analyses used Monte Carlo simulation to test the null hypothesis that there

was no statistically significant cluster of prevalent diabetic kidney disease cases within the window in question. The alternative hypothesis was that there was an elevated risk within the window as compared to outside for each of the scanning window locations and sizes. While gradually scanning a circular window across the entire map, the technique noted the number of observed and expected cases inside the circle at each location, by this way the clusters were detected (see Appendix 2 for details). Analyses used a Poisson probability model to estimate the rate of people with both diabetes and CKD within each potential cluster, and took the maximum likelihood function values for all window locations and sizes; the cluster with the greatest maximum likelihood ratio (reflecting the highest ratio of observed to expected cases) was considered as the primary cluster. Other statistically significant clusters that did not overlap with the primary cluster were identified as secondary clusters, and were ranked according to their likelihood ratio test statistic. We plotted polygon clusters in the map of Alberta by using the "point to polygon" tool [21] of ArcGIS software. In maps, we used yellow rectangles for the top three clusters with (ordered by the number of cases), black stars for other clusters, faint grey circles for clusters not visible in ArcGIS maps and red polygons to represent the amalgamated clusters.

The decision to consider different distance categories and population sizes were taken a priori. However, the selection of population sizes or distance categories required subjective assessment because we need to take into account pragmatic considerations, contextual evidences and lack of any gold standard cluster detection method.

The institutional review boards for the Universities of Alberta and Calgary approved the study.

Analyses were performed using STATA 11, SaTScan, ArcGIS 10 and Stat Transfer software.

2.2.6 Risk profile analysis

From the SaTScan results for four population sizes (0.5, 1, 5 and 10%), we identified the population size which best defined the cluster. Then we proceeded with presenting one cluster from each of the four different buffer scenarios (50, 100, 150 and 200 km) for that specific population size based on either primary clusters or maximum number of patients residing within that clusters to perform risk profile analysis. A cluster at 50 km buffer scenario indicates that the observed counts of diabetic kidney patients in a community 50 km apart from the existing clinic is higher than the expected number of cases for that specific community. Likewise, a cluster at 100 km buffer scenario denotes higher than expected number of diabetic kidney cases for an underserved community 100 km away from the existing clinic. Similarly, cluster at 150 and 200 km buffer scenario represent higher case counts in the underserviced communities far away from the existing clinics.

For cluster estimation we limited our dataset from 1st May 2002 to 31st December 2008; however, for this risk profile analysis we extended our follow up for the same 32278 patients from May 1st 2002 to March 31st 2009 to capture all relevant outcomes for these sets of patients. We evaluated the following baseline variables to characterize the patients residing in those clusters and for the total study participants: socio-demographic factors such as age, sex, rural status, socioeconomic status, quintile and clinical characteristics for example, prior hospitalizations, laboratory parameters and medication history. Socioeconomic status was classified into five categories, Aboriginal (all ages), normal, subsidy, welfare assistance and

pensioner (anyone 65 and older). Quintiles were derived from postal code conversion file (PCCF) based on dissemination areas from the census data of the year of 1996, 2001 and 2006. The 1st quintile represents the lowest income quintile and the 5th represents the highest income quintile. We presented the neighbourhood poverty index using a map that presented shaded 50 km square grids. Prior hospitalizations were due to all cause, acute myocardial infarction (AMI), cerebrovascular accident/ transient ischemic attack (CVA/TIA), congestive heart failure (CHF), acute kidney injury (AKI), cardiac catheterization (CATH), percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG) on or prior to the first date of serum creatinine measurement. The laboratory parameters included first serum creatinine (umol/L) measurement, CKD with proteinuria, CKD with heavy proteinuria, median dipstick urinalysis, median albumin: creatinine ratio (mg/mmol), hemoglobin A1C (HbA_1C) percentage and low density lipoprotein (LDL). Use of statins, angiotensin converting enzyme inhibitors (ACEi) and angiotensin II receptor blockers (ARBs) were also presented.

CKD with proteinuria was defined by documentation of A, B or C within 6 months of the first available serum creatinine. A) albumin: creatinine ratio ≥60 mg/mmol, B) protein: creatinine ratio ≥100 mg/mmol or C) protein ≥ 2+ dipstick urinalysis (regardless of eGFR). CKD with heavy proteinuria was defined by, within 6 months of the date of the first available serum creatinine, one or more documented occurrences of albumin: creatinine ratio ≥180 mg/mmol, protein:creatinine ratio ≥300mg/mmol or protein ≥ 3+ dipstick urinalysis (regardless of eGFR). Median dipstick urinalysis is defined as the median of all available measurements within 6 months of the date of the first available serum creatinine, continuous values ranging from 0 to 4 where 0 means negative, 1 for trace, 2 for one plus, three represents two plus and four represents

three plus. These values were recategorized as follows: 0 was assigned to 0, 1 was assigned to 1-2 and 2 was assigned to 3-4. Median ACR, HbA1C% and LDL were calculated as the median value of all available measurements within 6 months of the date of the first available serum creatinine. The medication histories for three drugs represent any use over the year prior to or on date of first available serum creatinine. However, these data were available only for patients 65 and older age group.

We tabulated the comorbid conditions for each of the participants using the following variables: CKD staging by using CKD-EPI equation, hypertension, Charlson (Deyo) index with co morbidities differences and mental health conditions. Charlson comorbidities included cancer, cerebrovascular disease (CVD), congestive heart failure (CHF), chronic obstructive pulmonary disease COPD, dementia, human immunodeficiency virus (HIV), metastatic cancer, myocardial infarction (MI), mild liver disease, moderate/severe liver disease, paraplegia, peptic ulcer disease (PUD), peripheral vascular disease (PVD) and rheumatic disease. These Charlson comorbidity variables were defined using claims and hospitalization date from within the three years or prior to the date of the first available serum creatinine. Mental health conditions were classified as affective disorder, problematic substance use and psychotic disorder. Presence of one instance of an ICD code associated with affective disorder or psychotic disorder were used to diagnose affective or psychotic disorder. Substance use was defined as presence of one instance of an ICD code associated with outpatient face-to-face clinical encounter with a mental health specialist [22].

The outcomes were categorized into a clinically relevant outcome, a process based outcome and

a composite renal outcome. The clinically relevant outcome was hospitalization due to all causes and cardiac causes. The process bases outcome was represented by HbA1c percentage. Doubling of serum creatinine (sustained) was used as a composite outcome – meaning that the value of last serum creatinine was doubled as compared to the first serum creatinine measurement.

Categorical variables were presented as proportions and continuous variables were presented as means and standard deviations. We presented the descriptive statistics in proportions or means to demonstrate the differences between the primary clusters; however, we did not perform any statistical hypothesis testing across the groups. We used STATA/MP 11.0 for all analysis.

2.3 Results

A map of Alberta presenting the location of major cities (population >50,000) and existing nephrology clinics is shown in Figure 1. Characteristics of the 32,278 included cases are presented in Table 1.

2.3.1 Cluster detection

The shaded map illustrates the density of these cases as the number of cases per 100 km square grid of Alberta in Figure 2. The density of patients (represented by the dark brown shades) was generally highest in the vicinity of existing nephrology clinics in the major cities of Alberta: Edmonton, Calgary, Lethbridge, Medicine Hat and Red Deer. The only exception was the northwestern region where we do not have any nephrology clinics at present. Intermediate densities of patients tended to outlie the more densely populated regions in the major cities and northeastern region. Areas with lower patient densities grids were diffusely distributed, and

mostly situated in the northern part of the province.

Table 2 presents the results of SaTScan analysis; for each scenario (>50, >100, >150, and >200 km from the closest nephrology clinic; not mutually exclusive), the number of postal codes, case counts and population counts are presented. Table 2 also shows the number of significant clusters identified in each scenario, and the radius and case counts of the primary cluster. Setting the maximum population size at 10% tended to identify clusters that were too large to be served by a single clinic being the largest primary cluster of 124.98 km. In contrast, setting the maximum cluster size at 0.5 or 1% of population tended to identify clusters that were too small (primary cluster of 110 m including only 17 cases) to justify placement of a new clinic – even when grouped together. A maximum population size of 5% seemed to perform well, and was selected as the primary measure of suitability for this analysis. As expected, using less stringent definitions of remote-dwelling patient (for example, the >50 km rather than the >200 km scenario) tended to identify larger numbers of patients.

Maps showing the locations of clusters under each scenario are shown in Figure 3 (see Appendix 6 for individual map). It is clearly evident that there is a tendency for smaller maximum population counts to identify impractically small clusters. When 5% was selected as the maximum population size, the primary cluster for the >50 km scenario was near the southern border of Alberta, and included the currently underserviced communities of Cardston, Pincher Creek, and Fort Macleod (Figure 3, Appendix 6 Map 3).

For the >100 km scenario, we identified one primary cluster and two secondary clusters with

higher case counts. The primary cluster was close to the underserved neighborhoods of Camrose and Wainwright with a radius of 67.42 km; secondary clusters were near Bonnyville; Peace River; and close to Grande Prairie. For the >150 km scenario, we identified a primary cluster in the currently underserviced community of Vermillion, and secondary clusters in Wainwright, Peace River and close to Grande Prairie. For the >200 km scenario, the primary cluster located in the vicinity of the currently underserviced communities of Grande Prairie, and secondary clusters were identified in Peace River, Wainwright and north of Grande Prairie (Figure 3, Appendix 6 Map 7,11 and 15).

Figure 4 shows the common themes in three scenarios (100, 150 and 200 km) using the maximum cluster size of 5% and 10%. When comparing with 10% as the maximum population size, although there was some slight variation in the cluster radii, the locations of the clusters were consistent throughout all the scenarios (Figure 4). Stratified analysis on CKD stage (stage 3; eGFR 30-59.9 ml/min/1.73m^2 and stage 4; eGFR 15-29.9 ml/min/1.73m^2) presents consistent results with the primary analysis (supplementary Table in Appendix 4 and Figure in Appendix 7).

2.3.2 Characteristics of identified clusters

According to the results presented from SaTScan analysis, setting 5% of population size in spatial scan statistics for four different buffer scenarios (50, 100, 150 and 200 km) provided better estimates to localize four underserved communities. We identified in total four primary clusters, each from one of the four buffer scenario: 50 km, 100 km, 150 km and 200 km. These primary clusters included the highest number of patients among the identified clusters from the respective analysis for each buffer scenario at 5% population size.

We presented socio-demographic characteristics of four clusters and the total study participants in Table 3. The primary clusters identified 664, 319, 262 and 186 patients at 50 km, 100 km, 150 km and 200 km respectively. Patients were mostly elderly, mean age varied from 74 to 77 years of old. Compared to the overall 15% prevalence of rural residence in the sample, all primary clusters had a higher proportion of rural residents (39 - 66%). However, the identified clusters were not accurately captured by the neighbourhood poverty index assessed at postal code level (see Appendix 5 for details).

Table 4 showed that the prevalence of all cause hospitalization was at least 59 percent for the study cohort. However, among residents of primary clusters, the prevalence of hospitalization was higher: 67% for cluster at 50 km buffer to 82% for clusters at 100 km buffer. AMI, CVA and CHF were the primary cardiac causes of hospitalizations prior to their entry into the study cohort.

The proportions of CKD cases with stage 3 and 4 were quite similar across the clusters (Table 5). Primary clusters had higher proportions of cancer, CVD, CHF and COPD and psychotic disorders compared to the total patient population, indicating a higher burden of overall illness and comorbidity.

Table 6 presents clinically relevant outcomes such as all cause hospitalization, average HbA1c level and incidence in doubling serum creatinine level. All cause hospitalization rates appeared higher for the clusters (69 to 76 %) compared to 62 percent for the overall cohort. Remote clusters located at 50 km and 150 km distance from the existing nephrology clinics had higher incidence of doubling serum creatinine and mean HbA1c compared to the overall study

population.

2.4 Discussion

2.4.1 Cluster locations

Most (71%) Albertans with diabetes and CKD live within 50 km of the nearest nephrology clinic – although a substantial proportion (8%) live more than 200 km away. We identified the residence locations of prevalent patients with diabetes and CKD who lived remote from the 17 established nephrology clinics in Alberta, and applied two methods for identifying underserved communities that would potentially benefit from new nephrology clinics. Simple density plots produced shaded maps with case counts per 100 km square grids and were useful for rough identification of underserved areas, although this method does not permit accurate or precise localization of the potential new clinic locations. The more computationally intensive SaTScan analysis was able to identify specific communities in the northwest and southeast region of the province with a higher-than-expected proportion of underserved patients. Our results were consistent across four different distance categories; the results were equally applicable for those living <50 km from the existing clinics, as compared to those who were residing >200 km from the clinics.

Our findings offer an objective way to locate new clinics that will serve the maximum number of remote-dwellers. The method described in the current paper is potentially complementary to an approach based on minimizing net (total) travel time for underserved patient populations that our group previously developed [23].

2.4.2 Cluster characteristics

Higher proportions of rural residents among the disease clusters reflect the underserved communities in the remote regions. Residents of remote clusters had higher Charlson scores at baseline reflecting greater comorbidity, compared to the non-remote dwellers. Remote-dwellers also had higher hospitalization rates, irrespective of the cause of hospital admissions.

Previous work from our group suggests that the quality of care delivered to Alberta CKD patients is associated with their residential locations; patients living further from the nephrologists were more likely to die or be hospitalized than those who were residing nearby and were less likely to receive markers of good quality care [24]. A recently published population based study in Alberta investigated the impact of remote locations on adverse clinical outcomes among 31337 patients with CKD and concomitant diabetes. The study reported that remote-dwellers had higher all cause hospitalization and all cause mortality [25].

Our present results were consistent with this previous paper. We excluded ESRD or death patients from our analysis to identity prevalent clusters of stage 3-4 CKD. However, analysis on hospitalization revealed similar conclusions that remote clusters were more likely to be hospitalized in comparison to the overall population.

2.4.3 Findings of the current study: placed in context

The literature review in chapter one demonstrates that limited accessibility and availability of health care facilities has significant potential consequences for health outcomes among chronic disease patients. However, we have limited evidence on how to choose the new facilities in an objective way for these underserved regions.

Previously published studies used different methods to improve health services or recommend new facility locations. One study in Alberta determined areas with geographic access within 90 minutes to a cardiac catheterization facility and concluded the preferred mode of transport on the basis of transfer time, by using ground and air ambulance using ArcGIS software. Alberta Chart of Call data was used for populated areas in which there was a community hospital. This study evaluated accessibility of patients to health services, three specialized cardiac facilities in Alberta, and recommended ways for faster travel time to improve access on time. Although the paper presented the proportions of populations by 90 minute transfer time distance in the province, it did not specify any particular underserved community based on population distribution [26].

In another study in Florida, GIS based model was used to propose new dental facilities based on accessibility of existing facilities from 5 to 15 miles of the nearest zip codes. Though they assumed that underserved communities were usually clustered in space, they did not perform any statistical analysis to confirm that hypothesis. In addition, they primarily focused on whether there was one or more dentists working in a given zip code, rather than considering population density into account [27].

These prior studies were therefore limited in terms of applicability and comparability of their methods and findings to the conduct and results of our study.

In the previous study conducted by researchers from our group, a cross section of 31452 CKD patients in the year of 2005 whose eGFRs were <45 ml/min/1.73m^2 was used to identify the ideal locations for up to four new nephrology clinics in Alberta by using GIS technique [23]. Buffer analysis and network analysis were performed to determine the clinic locations by minimizing patient traveling >120 minutes to see nephrologists. Among the four possible locations chosen by this different method, Grande Prairie was the first choice and Vermillion was the second which were consistent with our analysis. This increases confidence in our current results.

Out of presented 16 scenarios, it had been shown that without any clinic, 8.6% of remote-dwelling Albertans with CKD need to travel >120 minutes and setting four clinics would reduce to that proportion to 2.4 percent. However, placing two clinics, one in Grande Prairie and another one in Vermillion would reduce the patients living >120 minutes away by 57 percent and only 3.6 percent of patients need to travel >120 minutes. If a clinic would have been placed only in Grand Prairie, a city 400 km away from the nearest existing clinic with 109,000 total population, 876 patients would not need to travel >120 km. Likewise, creating two new clinics in two identified locations in Alberta would minimize travel distance for 1557 remote CKD patients [23].

The previous paper and current study are complementary to one another because they reach similar conclusions -- even though the earlier paper focused on net travel time without

considering the spatial distribution of patients served by the proposed clinics.

2.4.4 Study limitations

Our study has several limitations. Since we used postal code rather than street address to categorize the address of subjects, our analysis could be affected by ecological fallacy [28] and one can argue to extrapolate the findings to the individual level. The ecological fallacy refers to the fact that ecological associations can be different from the corresponding individual level associations within same population [28]. However, postal codes are the smallest geographical units and can be applicable to the individual level characteristics in most of the cases. In addition, we are not focusing on estimating disease burden or identifying risk factors, this fallacy should not have affected our inference much.

Secondly, we have some missing data for some postal codes, but the proportion was small (<0.1%) and would not be expected to affect our results.

Third, we failed to include patients with diabetes who could not access care in terms of physician visits or hospitalization. These patients might have CKD but were not diagnosed with concomitant diabetes and were thus excluded from our analysis. However, such patients are expected to be rare in Canada due to the universal health care system, and thus are unlikely to have affected our conclusions.

Moreover, travel distance or transportation infrastructure was not taken into account in this analysis, and instead we focused on the spatial distribution of the underserved communities.

In addition, we specified the maximum population size in SaTScan analysis and buffer zones in GIS analysis a priori, but for each of population sizes and buffer zones we presented sensitivity analysis in four ways and the results were more or less similar for all scenarios.

The Poisson model can underestimate the observed number with no case counts or overestimate the number with one or two case counts given that there are many zero case counts in the postal codes. This is explained by over dispersion or extra-Poisson variation with the mean number of case counts being less than the variance [29]. This occurs due to a single Poisson parameter, which is not enough to describe the population [29, 30]. However, we prioritized the maximum number of patients within the identified clusters for each scenario, which were less likely to bias our estimates. In addition, the traditional methods to handle over dispersion of Poisson data still results in lack of fit [29]. Besides, it has been suggested that we can use a model to account extra-Poisson variation when we are concerned with the inference related to the regression parameters and in situation where over dispersion routinely occurs [31]. On the other hand, it has been suggested that Poisson likelihood is robust to misspecification of distribution and have efficient properties even when the distribution does not follow Poisson [32]. In addition, many times Poisson models are valid and a method of choice for non-negative, discrete and random feature of events with its simplicity and applicability [31, 33].

There are certain limitations of SaTScan technique listed in the available literature. First, it does not have any cartographic features to explore the identified clusters, therefore the output needs to be visualized in the GIS (ArcGIS) software [34, 35]. Secondly, SaTScan depends on user-specified parameter choices and provides limited clue for appropriate maximum population size.

Setting maximum population size too large (such as 50% by default) may lead to cluster with unusually large geographical area and miss small clusters. On the other hand, small population size may miss clusters with larger size [34, 36]. Therefore, it has been suggested to choose population size based on contexts and run multiple scans for different maximum size parameters [34]. We have followed our analysis in this way and thus our results would less likely to be affected by these limitations of SaTScan technique.

2.5 Conclusion

The optimal method for selecting locations for establishment of new health care facilities in remote places is unknown. We used ArcGIS software to draw density plots and to perform buffer analysis and SaTScan software to conduct spatial analysis. We identified prevalent clusters of diabetes and chronic kidney disease, an important health condition. In addition, the remote clusters had higher hospitalization rates reflecting high risk for adverse clinical outcomes. Results were similar to prior studies using different methods. However, SaTScan technique has potential advantage over other available methods -- including buffer and network analysis -- because the latter requires expertise in GIS knowledge and has limitations related to unavailability of road network due to dirt or unpaved roads and arbitrary selection of starting point for the algorithm. Our findings indicate that we can use this analytic technique to determine potential locations of new health care facilities objectively for diabetic kidney patients in the underserved communities in Alberta.

References:

1. Hemmelgarn BR, Clement F, Manns BJ, Klarenbach S, James MT, Ravani P, Pannu N, Ahmed SB, MacRae J, Scott-Douglas N *et al*: Overview of the Alberta Kidney Disease Network. *BMC nephrology* 2009, 10.
2. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T *et al*: A new equation to estimate glomerular filtration rate. *Annals of internal medicine* 2009, 150(9):604-612.
3. Hux JE, Ivis F, Flintoft V, Bica A: Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes care* 2002, 25(3):512-516.
4. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA: Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care* 2005, 43(11):1130-1139.
5. Book chapter 3 and 19. In: *Epidemiologic Methods: Studying the Occurrence of Illness.* edn. Edited by Koepsell TD, Weiss NS. New York, USA: Oxford University Press; 2003: pp. 37-61 and 464-489.
6. Seliske LM: The Built Environment and Obesity-Related Behaviours in Canadian Youth. Kingston, Ontario, Canada: Queen's University; 2012.
7. Achuthan K, Titheridge H, Mackett R: Measuring Pedestrian Accessibility. In: *Geographical Information Science Research Conference: 2007; NUI Maynooth, Ireland*; 2007.
8. Oliver LN, Schuurman N, Hall AW: Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International journal of*

health geographics 2007, 6:41.

9. Kulldorff M: A spatial scan statistic. *Commun Stat-Theor M* 1997, 26(6):1481-1496.

10. Software: Kulldorff M. and Information Management Services, Inc. SaTScanTM v8.0: Software for the spatial and space-time scan statistics. http://www.satscan.org/, 2009.

11. Kulldorff M, Nagarwalla N: Spatial Disease Clusters - Detection and Inference. *Statistics in medicine* 1995, 14(8):799-810.

12. Kulldorff M, Feuer EJ, Miller BA, Freedman LS: Breast cancer clusters in the northeast United States: A geographic analysis. *American journal of epidemiology* 1997, 146(2):161-170.

13. Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N: Childhood leukaemia in Sweden: Using GIS and a spatial scan-statistic for cluster detection. *Statistics in medicine* 1996, 15(7-9):707-715.

14. Trooskin SB, Hadler J, Louis TS, Navarro VJ: Geospatial analysis of hepatitis C in Connecticut: a novel application of a public health too. *Public health* 2005, 119(11):1042-1047.

15. Green C, Hoppa RD, Young TK, Blanchard JF: Geographic analysis of diabetes prevalence in an urban area. *Soc Sci Med* 2003, 57(3):551-560.

16. Aamodt G, Samuelsen SO, Skrondal A: A simulation study of three methods for detecting disease clusters. *International journal of health geographics* 2006, 5:15.

17. Lin H, Liu Q, Guo J, Zhang J, Wang J, Chen H: Analysis of the geographic distribution of HFRS in Liaoning Province between 2000 and 2005. *BMC public health* 2007, 7:207.

18. Nkhoma ET, Ed Hsu C, Hunt VI, Harris AM: Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980 - 2001. *International*

journal of health geographics 2004, 3(1):25.

19. Root ED, Meyer RE, Emch ME: Evidence of localized clustering of gastroschisis births in North Carolina, 1999-2004. *Soc Sci Med* 2009, 68(8):1361-1367.

20. Qi X, Tong S, Hu W: Spatial distribution of suicide in Queensland, Australia. *BMC psychiatry* 2010, 10:106.

21. Wynne D: PointToPolygon.zip, generic tool for the software ArcGIS Desktop, ESRI. In. http://arcscripts.esri.com/details.asp?dbid=15974.

22. Frayne SM, Miller DR, Sharkansky EJ, Jackson VW, Wang F, Halanych JH, Berlowitz DR, Kader B, Rosen CS, Keane TM: Using administrative data to identify mental illness: what approach is best? *American journal of medical quality : the official journal of the American College of Medical Quality* 2010, 25(1):42-50.

23. Ayyalasomayajula B, Wiebe N, Hemmelgarn BR, Bello A, Manns B, Klarenbach S, Tonelli M: A novel technique to optimize facility locations of new nephrology services for remote areas. *Clinical journal of the American Society of Nephrology : CJASN* 2011, 6(9):2157-2164.

24. Rucker D, Hemmelgarn BR, Lin M, Manns BJ, Klarenbach SW, Ayyalasomayajula B, James MT, Bello A, Gordon D, Jindal KK *et al*: Quality of care and mortality are worse in chronic kidney disease patients living in remote areas. *Kidney international* 2011, 79(2):210-217.

25. Bello AK, Hemmelgarn B, Lin M, Manns B, Klarenbach S, Thompson S, James M, Tonelli M: Impact of remote location on quality care delivery and relationships to adverse health outcomes in patients with diabetes and chronic kidney disease. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant*

Association - European Renal Association 2012.

26. Patel AB, Waters NM, Ghali WA: Determining geographic areas and populations with timely access to cardiac catheterization facilities for acute myocardial infarction care in Alberta, Canada. *International journal of health geographics* 2007, 6:47.

27. Horner MW, Mascarenhas AK: Analyzing location-based accessibility to dental services: an Ohio case study. *Journal of public health dentistry* 2007, 67(2):113-118.

28. Morgenstern H: Ecologic studies in epidemiology: concepts, principles, and methods. *Annual review of public health* 1995, 16:61-81.

29. Bulsara MK, Holman CD, Davis EA, Jones TW: Evaluating risk factors associated with severe hypoglycaemia in epidemiology studies-what method should we use? *Diabetic medicine : a journal of the British Diabetic Association* 2004, 21(8):914-919.

30. Dossou-Gbété S, Mizère D: An Overview of Probability Models for Statistical Modelling of Count Data. *Monografías del Seminario Matemático García de Galdeano* 2006, 33:237-244.

31. Dean CB: Testing for Overdispersion in Poisson and Binomial Regression-Models. *Journal of the American Statistical Association* 1992, 87(418):451-457.

32. Berkhout P, Plug E: A bivariate Poisson count data model using conditional probabilities. *Stat Neerl* 2004, 58(3):349-364.

33. Wang YH, Nihan NL: Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. *Accident Anal Prev* 2004, 36(3):313-321.

34. Chen J, Roth RE, Naito AT, Lengerich EJ, MacEachren AM: Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. *International journal of health geographics* 2008, 7.

35. Robertson C, Nelson TA: Review of software for space-time disease surveillance. *International journal of health geographics* 2010, 9:16.

36. Chen J, MacEachren A, Lengerich E: Visual Analytics of Spatial Scan Statistic Results. In: *Geospatial Visual Analytics Workshop, GIScience conference.* Utah; 2008.

Chapter 3: Relevance to public health and future research

3.1 Overview

Patients with concomitant diabetes and chronic kidney disease have higher cardiovascular mortality and morbidity. Recently a population-based cohort study in Alberta demonstrated that patients with CKD (without diabetes) had higher rates of myocardial infarction (MI), compared to patients having diabetes (but without CKD). However patients having both conditions had higher first MI rate compared to those having either of the conditions [1]. In addition, remote-dwelling Albertans having both disease conditions had a lower likelihood of receiving recommended care (specialist consultations, medications and HbA1c or proteinuria assessment) and also had eventually higher hospitalization and mortality risk compared to their urban counterparts [2]. Receiving appropriate care may improve outcomes for these patients. However, we do not have any existing tools to select the optimal location of new health care facilities in these remote communities.

In this paper, we used spatial scan statistics and GIS maps to show the clusters of these high risk patients in underserved communities in Alberta, to inform the selection of candidate locations.

3.2 Implications in health service delivery

The Canada Health Act (1984) mandates insured persons to have reasonable and uniform rights of access to insured health care services without any financial or other barriers or any discrimination on the basis of income, age and health status [3]. Therefore, establishment of a hospital facility in a distant location can provide equity of care for the remote-dwellers in

addition to improving their health outcomes. In addition, their accessibility to care may be improved through their ability to secure a regular physician and obtain a referral to the specialist in the absence of direct financial barrier due to the universal health care system in Canada [4].

SaTScan technique has been widely used in studies aimed at cluster detection [5-8], outbreak investigation [9-11], disease surveillance and monitoring [12-14], evaluation of interventions [15, 16], risk factor assessment [17-20], network analysis [21], and spatio-temporal analysis [22-24]. In this study, SaTScan objectively identifies underserved communities, based on comparing the expected and observed distribution of patients who reside far from existing health facilities. This is the first study to use SaTScan to identify new clinic locations aimed at improving access to health care, to our knowledge. Although there are approaches based on estimated travel time to identify potential new clinic locations [25], such approaches optimize net travel time (for all patients within the province) rather than focusing on underserved patients specifically. Potential advantages of SaTScan include its ability to adjust for non-homogeneous population density across different study regions, to reduce pre-selection bias without specifying spatial size and locations of a cluster *a priori* and to address bias related to multiple comparisons by likelihood ratio-based estimates [5].

General practitioners (GP) providing primary care are the initial point of contact in the health care system in Canada [26]. There are conditions when patients are referred to a specialist for example to receive advice on diagnosis or management, to undergo specialized care due to lack of availability of further management options in primary care setting or to seek a second opinion [27]. Referral to specialist physician depends not only on general practitioner (GP)'s decision but

also how patients can negotiate with the GP to see a specialist or geographic variation in access to a specialist [28]. It has been shown that having chronic conditions and self-perceived health status are important determinants for the Canadians to consult with the physicians while adjusting for other factors such as age, sex, race, language, household income, urban rural residence or having regular family doctor [26]. GP or specialist consultation may vary according to the health condition of the patients; specialist care or hospital services may be important for heart disease, cancer patients or diabetes patients who need frequent follow up. In contrast, patients may seek care from GP for ambulatory conditions such as allergy. However, aged population with increased chronic disease burden is leading to increasing demand of coordinated care in the community with primary care physician and specialists along with other health care professionals [29].

Evidence-based decisions can improve the allocation and utilization of health care resources. Therefore, the availability of these findings to policymakers, may allow more informed decision making for the underserved communities, rather than having perceptions on patients demand or responding to extraneous factors (such as lobbying, media coverage or political influence) which may lead to inefficient decisions.

Providing services too much in an area or in an inappropriate area will result in wasted resources; on the other hand, providing too little care in another area which actually deserves services may lead to untoward clinical consequences and inequity. An educated guess can result in similar findings as compared to systematic approaches like SaTScan, as we reported in this study. However given the higher cost of a clinic establishment, this approach can be used to confirm

this intuition or in contexts where guesswork does not provide any obvious clue where to locate a new facility.

Some practical considerations should also be taken into consideration when considering the location for a new facility in a peri-urban or rural remote community; such as whether specialists, nurses or other health care professionals will be willing to work and live in these communities and serve the remote-dwellers.

3.3 Significance and future research

This is the first study using SaTScan analysis to objectively identify underserved communities that might be the most suitable locations for new nephrology clinics. The recommendations from this study will be useful to clinicians and decision-makers responsible for the care of remote-dwelling patients. Future studies should identify other barriers to optimum care delivery in remote areas and develop intervention strategies to improve the quality of care for underserved communities. This method should be validated in CKD populations in other provinces and territories in Canada. In addition, long-term studies should be performed to evaluate whether use of the findings from the SaTScan analysis leads to clinically meaningful benefit for patients and improve their quality of life. Although this study is focused on people with CKD, these methods are potentially applicable to other populations.

References:

1. Tonelli M, Muntner P, Lloyd A, Manns BJ, Klarenbach S, Pannu N, James MT, Hemmelgarn BR: Risk of coronary events in people with chronic kidney disease compared with those with diabetes: a population-level cohort study. *Lancet* 2012.
2. Bello AK, Hemmelgarn B, Lin M, Manns B, Klarenbach S, Thompson S, James M, Tonelli M: Impact of remote location on quality care delivery and relationships to adverse health outcomes in patients with diabetes and chronic kidney disease. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2012.
3. Madore O: The Canada Health Act: Overview and Options. In. Edited by Economics Division PRB, Library of Parliament. Ottawa; 2005.
4. Allin S: Does Equity in Healthcare Use Vary across Canadian Provinces? *Healthcare policy = Politiques de sante* 2008, 3(4):83-99.
5. Kulldorff M, Feuer EJ, Miller BA, Freedman LS: Breast cancer clusters in the northeast United States: A geographic analysis. *American journal of epidemiology* 1997, 146(2):161-170.
6. Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N: Childhood leukaemia in Sweden: Using GIS and a spatial scan-statistic for cluster detection. *Statistics in medicine* 1996, 15(7-9):707-715.
7. Green C, Hoppa RD, Young TK, Blanchard JF: Geographic analysis of diabetes prevalence in an urban area. *Soc Sci Med* 2003, 57(3):551-560.
8. Qi X, Tong S, Hu W: Spatial distribution of suicide in Queensland, Australia. *BMC psychiatry* 2010, 10:106.

9. Nisha V, Gad SS, Selvapandian D, Suganya V, Rajagopal V, Suganti P, Balraj V, Devasundaram J: Geographical information system (GIS) in investigation of an outbreak. *The Journal of communicable diseases* 2005, 37(1):39-43.

10. Ekong PS, Ducheyne E, Carpenter TE, Owolodun OA, Oladokun AT, Lombin LH, Berkvens D: Spatio-temporal epidemiology of highly pathogenic avian influenza (H5N1) outbreaks in Nigeria, 2006-2008. *Preventive veterinary medicine* 2012, 103(2-3):170-177.

11. Huang SS, Yokoe DS, Stelling J, Placzek H, Kulldorff M, Kleinman K, O'Brien TF, Calderwood MS, Vostok J, Dunn J *et al*: Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study. *PLoS medicine* 2010, 7(2):e1000238.

12. Waller LA, Turnbull BW, Clark LC, Nasca P: Chronic Disease Surveillance and Testing of Clustering of Disease and Exposure - Application to Leukemia Incidence and Tce-Contaminated Dumpsites in Upstate New-York. *Environmetrics* 1992, 3(3):281-300.

13. Andrade AL, Silva SA, Martelli CM, Oliveira RM, Morais Neto OL, Siqueira Junior JB, Melo LK, Di Fabio JL: Population-based surveillance of pediatric pneumonia: use of spatial analysis in an urban area of Central Brazil. *Cadernos de saude publica / Ministerio da Saude, Fundacao Oswaldo Cruz, Escola Nacional de Saude Publica* 2004, 20(2):411-421.

14. Gosselin P, Lebel G, Rivest S, Douville-Fradet M: The Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV): a real-time GIS for surveillance and decision-making. *International journal of health geographics* 2005, 4:21.

15. Ali M, Asefaw T, Byass P, Beyene H, Pedersen FK: Helping northern Ethiopian communities reduce childhood mortality: population-based intervention trial. *Bulletin of*

the World Health Organization 2005, 83(1):27-33.

16. Oviedo M, Munoz P, Dominguez A, Carmona G, Batalla J, Borras E, Jansa JM: [Evaluation of mass vaccination programmes: the experience of Hepatitis A in Catalonia]. *Revista espanola de salud publica* 2009, 83(5):697-709.
17. Stanca CM, Babar J, Singal V, Ozdenerol E, Odin JA: Pathogenic role of environmental toxins in immune-mediated liver diseases. *Journal of immunotoxicology* 2008, 5(1):59-68.
18. McNally RJ, Ducker S, James OF: Are transient environmental agents involved in the cause of primary biliary cirrhosis? Evidence from space-time clustering analysis. *Hepatology* 2009, 50(4):1169-1174.
19. Hayran M: Analyzing factors associated with cancer occurrence: A geographical systems approach. . *Turkish Journal of Cancer* 2004, 34:67-70.
20. Bakker MI, Hatta M, Kwenang A, Faber WR, van Beers SM, Klatser PR, Oskam L: Population survey to determine risk factors for Mycobacterium leprae transmission and infection. *International journal of epidemiology* 2004, 33(6):1329-1336.
21. Wylie JL, Cabral T, Jolly AM: Identification of networks of sexually transmitted infection: A molecular, geographic, and social network analysis. *Journal of Infectious Diseases* 2005, 191(6):899-906.
22. Nkhoma ET, Ed Hsu C, Hunt VI, Harris AM: Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, U.S., 1980 - 2001. *International journal of health geographics* 2004, 3(1):25.
23. Elias J, Harmsen D, Claus H, Hellenbrand W, Frosch M, Vogel U: Spatiotemporal analysis of invasive meningococcal disease, Germany. *Emerging infectious diseases*

2006, 12(11):1689-1695.

24. Jones RC, Liberatore M, Fernandez JR, Gerber SI: Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Rep* 2006, 121(2):133-139.

25. Ayyalasomayajula B, Wiebe N, Hemmelgarn BR, Bello A, Manns B, Klarenbach S, Tonelli M: A Novel Technique to Optimize Facility Locations of New Nephrology Services for Remote Areas. *Clin J Am Soc Nephro* 2011, 6(9):2157-2164.

26. Nabalamba A, Millar WJ: Going to the doctor. *Health reports / Statistics Canada, Canadian Centre for Health Information = Rapports sur la sante / Statistique Canada, Centre canadien d'information sur la sante* 2007, 18(1):23-35.

27. Grimshaw JM, Winkens RAG, Shirran L, Cunningham C, Mayhew A, Thomas R, Fraser C: Interventions to improve outpatient referrals from primary care to secondary care. *Cochrane Db Syst Rev* 2005(3).

28. Asada Y, Kephart G: Equity in health services use and intensity of use in Canada. *BMC health services research* 2007, 7.

29. Why Health Care Renewal Matters: Learning from Canadians with Chronic Health Conditions. In. Toronto: Health Council of Canada; 2007.

Table 1: Characteristics of participants with chronic kidney disease and diabetes

Characteristics (N=32,278)		**Proportion or Mean***
Age, mean (SD), y		74.9 ± 10.6
Female		17,261 (53.48)
Charlson score**		1.7 ± 1.6
eGFR(ml/min/1.73m^2)***		
	15-29.9	3224 (9.99)
	30-59.9	29054 (90.01)
Income****		
	Aboriginal	672 (2.08)
	Normal	6,394 (19.81)
	Subsidy	1,662 (5.15)
	Assistance	1,029 (3.19)
	Pensioner	22,521 (69.77)
Distance from closest nephrology clinic*****		
	>50 km and <100 km	4,234 (13.12)
	>100 km and <150 km	1,571 (4.87)
	>150 km and <200 km	1,080 (3.35)
	>200 km	2,504 (7.76)

* Values are n (%) or mean (SD) as appropriate.

** Charlson comorbidity [4]

*** eGFR is glomerular filtration rate estimated by the CKD-EPI equation [2]

**** Income as an indication of socioeconomic status: "Assistance" refers to participants with health insurance premium paid under a program sponsored by Alberta Employment, Immigration and Industry. "Subsidy" refers to participants who pay less than the full premium or no premium to Alberta Health and Wellness or in the premium is subsidized though a Government Sponsored Program. "Normal" refers to all other participants.

, **, ***** estimated based on the date of first serum creatinine available.

Table 2: Findings of SaTScan analysis, by buffer zone and maximum population size

Buffer	Postal codes*	Cases**	Total population***	Maximum population size****	Number of Significant clusters*****	Radius of primary cluster	Observed cases in the primary cluster
50 km	7971	8036	621,984	10 %	16	100.41 km	1189
				5 %	19	121.66 km	664
				1 %	30	0.50 km	69
				0.5%	26	0.50 km	69
100 km	5264	4235	387,911	10%	13	124.98 km	655
				5%	14	67.42 km	319
				1%	18	0.68 km	76
				0.5%	15	0.29 km	22
150 km	4638	3120	308,293	10%	10	75.75 km	427
				5%	14	53.12 km	262
				1%	14	0.29 km	22
				0.5%	14	0.29 km	22
200 km	3750	2214	237,201	10%	9	59.48 km	379
				5%	12	49.36 km	186
				1%	9	0.11 km	17
				0.5%	10	0.11 km	17

* 'Postal Codes' is the total number of postal codes outside the corresponding buffer scenario

** 'Cases' is the total number of cases residing within these postal codes located outside the correspoding buffer scenario

*** 'Total population' is the total number of population living within these postal codes located outside the corresponding buffer scenario

**** 'Maximum population size' means SaTScan maximizes the cluster size at certain percentages of the total study population.

***** 'Number of significant clusters' means the total number of primary and secondary clusters which were statistically significant at p value of <0.05 for a maximum population size outside the corresponding buffer scenario.

Table 3: Socio-demographic characteristics of patients with chronic kidney disease and diabetes in clusters detected through SaTScan analysis*

Characteristics**	**Primary cluster at 50 km buffer (N=664)**	**Primary cluster at 100 km buffer (N=319)**	**Primary cluster at 150 km buffer (N=262)**	**Primary cluster at 200 km buffer (N=186)**	**Total (N=32278)**
Mean age (± SD)	73.7 ± 11.1	76.1 ± 11.3	76.7 ± 10.2	76.3 ± 11.2	74.9 ± 10.6
Female	355 (53)	175 (55)	141 (54)	93 (50)	17,261 (53)
Rural	280 (42)	210 (66)	101 (39)	82 (44)	4993 (15)
Social economic status					
Aboriginal	80 (12)	0	10 (4)	2 (1)	672 (2)
Normal	111 (17)	48 (15)	41 (16)	26 (14)	6394 (20)
Pensioner	407 (61)	239 (75)	196 (75)	141 (76)	22521 (70)
Subsidy	48 (7)	22 (7)	13 (5)	11 (6)	1662 (5)
Assistance	18 (3)	10 (3)	2 (1)	6 (3)	1029 (3)
Quintile***	(n=614)	(n=313)	(n=208)	(n=113)	(n=30751)
1st quintile	137 (22)	78 (25)	49 (24)	2 (2)	6828 (22)
2nd quintile	73 (12)	92 (29)	36 (17)	30 (27)	6584 (21)
3rd quintile	44 (7)	45 (14)	71 (34)	1 (1)	6123 (20)
4th quintile	111 (18)	38 (12)	18 (9)	50 (44)	5911 (19)
5th quintile	249 (41)	60 (19)	34 (16)	30 (27)	5305 (17)

*Four primary clusters are presented based on 5% of population size for four buffer scenario: 50 km, 100 km, 150 km and 200 km

** N with percentages, unless otherwise indicated

*** 1 Lowest income quintile and 5 Highest income quintile

Table 4: Baseline clinical characteristics of patients with chronic kidney disease and diabetes in clusters detected through SaTScan analysis*

Characteristics**	**Primary cluster at 50 km buffer (N=664)**	**Primary cluster at 100 km buffer (N=319)**	**Primary cluster at 150 km buffer (N=262)**	**Primary cluster at 200 km buffer (N=186)**	**Total (N=32278)**
Prior hospitalizations***					
All cause	443 (67)	260 (82)	182 (69)	139 (75)	19144 (59)
AMI	33 (5)	23 (7)	12 (5)	25 (13)	1869 (6)
CVA/TIA	21 (3)	19 (6)	15 (6)	10 (5)	1394 (4)
CHF	25 (4)	21 (7)	14 (5)	3 (2)	863 (3)
AKI	5 (1)	9 (3)	6 (2)	1 (1)	571 (2)
CATH	19 (3)	9 (3)	7 (3)	6 (3)	1516 (5)
PCI	28 (4)	9 (3)	7 (3)	5 (3)	1259 (4)
CABG	23 (3)	6 (2)	9 (3)	7 (4)	1293 (4)
Laboratory parameters					
First serum creatinine umol/L (mean ± SD)	97.4 ± 39.3 (n=663)	106.6 ± 31.4 (n=311)	99.8 ± 33.3 (n=261)	98.8 ± 26.6 (n=181)	101.9 ± 32.4 (n=32106)
CKD with proteinuria	68/320 (21)	15/163 (9)	6/62 (10)	3/96 (3)	2122/21672 (10)
CKD with heavy proteinuria	35/320 (11)	6/163 (4)	2/62 (3)	2/96 (2)	836/21672 (4)
Median dipstick urinalysis	(n=205)	(n=151)	(n=49)	(n=83)	(n=17419)
0	143/205 (70)	107/151 (71)	33/49 (67)	67/83 (81)	12967/17419 (74)
1	39/205 (19)	33/151 (22)	10/49 (20)	14/83 (17)	3106/17419 (18)
2	23/205 (11)	11/151 (7)	6/49 (12)	2/83 (2)	1346/17419 (8)
ACR median (mg/mmol) (mean ± SD)	51.7 ± 130.5 (n=201)	9.8 ± 19.4 (n=34)	2.9 ± 3.4 (n=20)	8.7 ± 24.5 (n=33)	20.1 ± 68.3 (n=11753)
HbA1C% (mean ± SD)	7.5 ± 1.7 (n=426)	7.3 ± 1.4 (n=186)	7.2 ± 1.4 (n=96)	7.3 ± 1.6 (n=107)	7.3 ± 1.5 (n=20977)
LDL (mmol/L) (mean ± SD)	2.8 ± 0.9 (n=348)	2.9 ± 0.9 (n=177)	2.7 ± 0.9 (n=156)	2.9 ± 0.9 (n=103)	2.8 ± 0.9 (n=20656)
Medications					
Statin	145 (22)	79 (25)	81 (31)	43 (23)	9368 (29)

ARB	110 (17)	66 (21)	71 (27)	32 (17)	6216 (19)
ACEi	208 (31)	127 (40)	109 (42)	82 (44)	11583 (36)

*Four primary clusters are presented based on 5% of population size for four buffer scenario: 50 km, 100 km, 150 km and 200 km

** N with percentages, unless otherwise indicated

AMI acute myocardial infarction, CVA/TIA cerebrovascular accident/ transient ischaemic attack, CHF congestive heart failure, AKI acute kidney injury, CATH cardiac catheterization, PCI percutaneous coronary intervention, CABG coronary artery bypass grafting ARB angiotensin II receptor blockers; ACEi angiotensin converting enzyme inhibitors

First serum creatinine, umol/L First serum creatinine measurement; CKD with proteinuria is defined by, within 6 months of the date of the first available serum creatinine one or more documented occurrences of albumin:creatinine ratio ≥60 mg/mmol, protein:creatinine ratio ≥100 mg/mmol or protein ≥ 2+ dipstick urinalysis (regardless of eGFR); CKD with heavy proteinuria is defined by, within 6 months of the date of the first available serum creatinine one or more documented occurrences of albumin:creatinine ratio ≥180 mg/mmol, protein:creatinine ratio ≥300mg/mmol or protein ≥ 3+ dipstick urinalysis (regardless of eGFR); Median dipstick urianalysis category: 0: 0, 1: 1-2, 2: 3-4 whereas the value ranges 0=negative, 1=trace, 2=1+, 3=2+ and 4=3+; Median Albumin: creatinine ratio, mg/mmol is the Albumin:creatinine ratio, Median value of all available measurements within 6 months of the date of the first available serum creatinine; Hemoglobin A1C%, Median value of all available measurements within 6 months of the date of the first available serum creatinine; Low density lipid LDL, mmol/L, Median value of all available measurements within 6 months of the date of the first available serum creatinine

Table 5: Co morbidities of patients with chronic kidney disease and diabetes in clusters detected through SaTScan analysis*

Comorbidities**	**Primary cluster at 50 km buffer (N=664)**	**Primary cluster at 100 km buffer (N=319)**	**Primary cluster at 150 km buffer (N=262)**	**Primary cluster at 200 km buffer (N=186)**	**Total(N=32278)**
CKD-EPI Stage of CKD***					
15-29 ml/min/1.73 m^2	92 (14)	34 (11)	24 (9)	11 (6)	3224 (10)
30-59 ml/min/1.73 m^2	572 (86)	285 (89)	238 (91)	175 (94)	29054 (90)
Hypertension	483 (73)	266 (83)	207 (79)	146 (78)	24710 (77)
Charlson (Deyo)****(mean ± SD)	1.7±1.5	1.9±1.6	2.1±2.0	1.7±1.4	1.7±1.6
Cancer	57 (9)	34 (11)	27 (10)	17 (9)	2615 (8)
CVD	52 (8)	32 (10)	26 (10)	11 (6)	2537 (8)
CHF	84 (13)	66 (21)	48 (18)	23 (12)	3638 (11)
COPD	156 (23)	81 (25)	74 (28)	52 (28)	6768 (21)
Dementia	8 (1)	15 (5)	10 (4)	4 (2)	731 (2)
HIV	0	0	0	0	16 (0.1)
Metastatic cancer	4 (1)	2 (1)	3 (1)	0	193 (1)
MI	43 (6)	30 (9)	18 (7)	18 (10)	2922 (9)
Mild liver disease	5 (1)	2 (1)	2 (1)	1 (1)	405 (1)
Moderate/severe liver disease	0	0	1 (0.4)	0	78 (0.2)
Paraplegia	3 (0.5)	2 (1)	2 (1)	1 (1)	306 (1)
PUD	20 (3)	10 (3)	12 (5)	11 (6)	1221 (4)
PVD	29 (4)	14 (4)	19 (7)	9 (5)	1966 (6)
Rheumatic disease	31 (5)	4 (1)	6 (2)	4 (2)	774 (2)
Mental condition					
Affective disorder	154 (23)	81 (25)	57 (22)	30 (16)	7662 (24)
Substance abuse	10 (2)	2 (1)	2 (1)	1 (1)	188 (1)
Psychotic disorder	47 (7)	14 (4)	9 (3)	13 (7)	989 (3)

*Four primary clusters are presented based on 5% of population size for four buffer scenario: 50 km, 100 km, 150 km and 200 km
** N with percentages, unless otherwise indicated
*** eGFR is glomerular filtration rate estimated by the CKD-EPI equation
**** Charlson comorbidity
CVD cerebrovascular disease, CHF congestive heart failure, COPD chronic obstructive pulmonary disease, HIV human immunodeficiency virus, MI myocardial infarction, PUD peptic ulcer disease, PVD peripheral vascular disease

Table 6: Outcomes of patients with chronic kidney disease and diabetes in clusters detected through SaTScan analysis*

Outcomes**	Primary cluster at 50 km buffer (N=664)	Primary cluster at 100 km buffer (N=319)	Primary cluster at 150 km buffer (N=262)	Primary cluster at 200 km buffer (N=186)	Total (N=32278)
Hospitalization					
All cause	463 (70)	228 (71)	180 (69)	142 (76)	20060 (62)
Acute kidney injury	56 (8)	20(6)	10 (4)	7 (4)	2830 (9)
Cardiac cause					
AMI	25 (4)	12 (4)	13 (5)	11 (6)	1695 (5)
CABG	21 (3)	19 (6)	6 (2)	7 (4)	1008 (3)
CATH	22 (3)	9 (3)	6 (2)	8 (4)	1242 (4)
CHF	56 (8)	16 (5)	18 (7)	14 (8)	1958 (6)
CVA/TIA	28 (4)	16 (5)	8 (3)	9 (5)	1148 (4)
PCI	20 (3)	5 (2)	3 (1)	8 (4)	1051 (3)
Process based outcome					
HbA1C (mean ± SD)	7.2 ± 1.6 (n=641)	7.0 ± 1.3 (n=277)	7.2 ± 1.4 (n=227)	7.1 ± 1.3 (n=164)	7.1 ± 1.4 (n=30467)
Composite renal outcome					
Doubling S Creatinine	68/663 (10)	11/311 (4)	18/261 (7)	6/181 (3)	1843/32106 (6)

*Four primary clusters are presented based on 5% of population size for four buffer scenario: 50 km, 100 km, 150 km and 200 km
** N with percentages, unless otherwise indicated
*** These are the outcomes incurred by 32278 patients in the cohort began in May 1 2002 and ended in March 31 2009

Serum creatinine doubling, umol/L is defined by the last serum creatinine is 2 times the first serum creatinine measurement

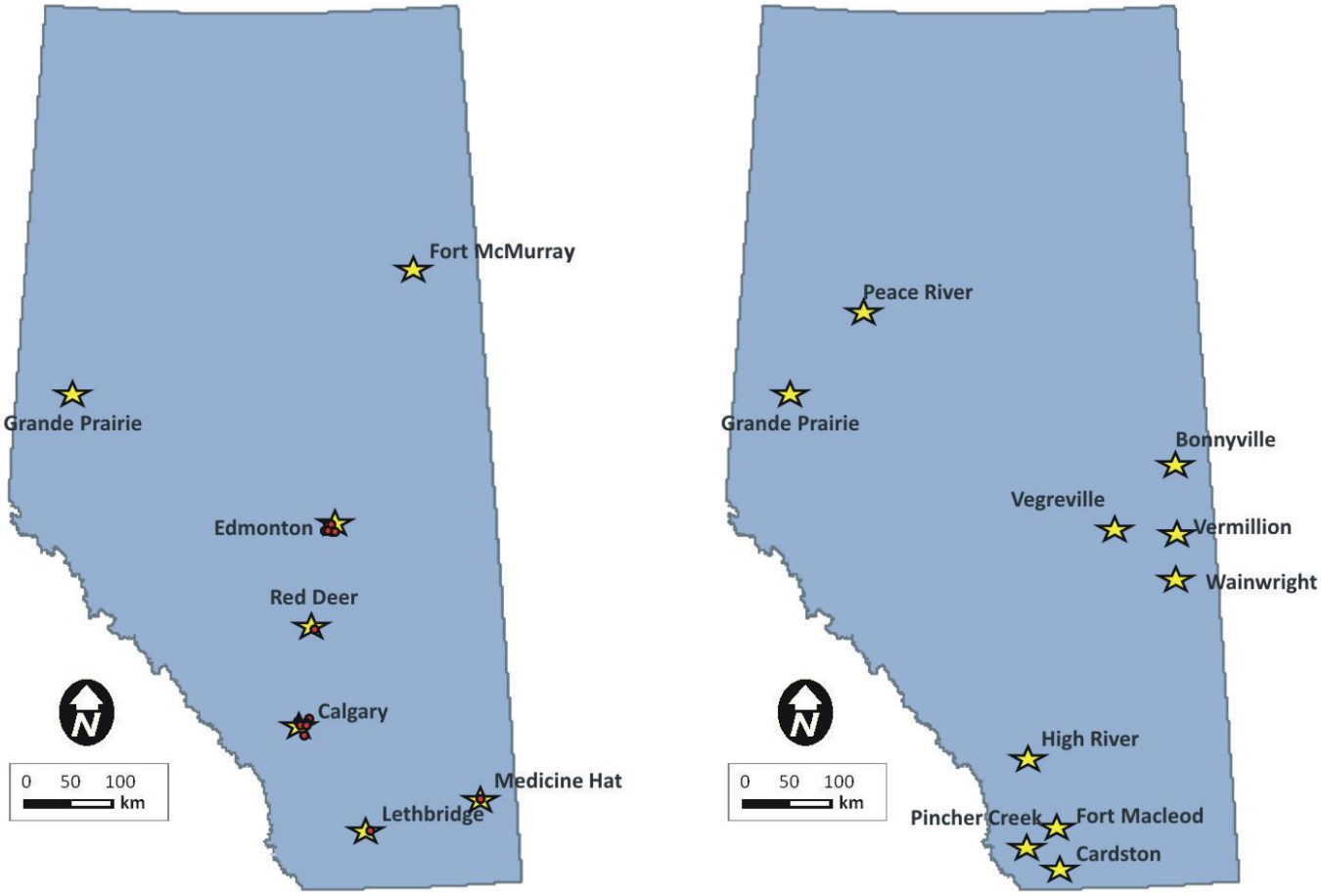


Figure 1: Major cities represented by stars and nephrology clinics by red dots (left panel) and other nearby cities (right panel) in Alberta
The left panel illustrates that the existing nephrology clinics are located in the major cities of Alberta; the right panel illustrates other medium-sized communities in the province.

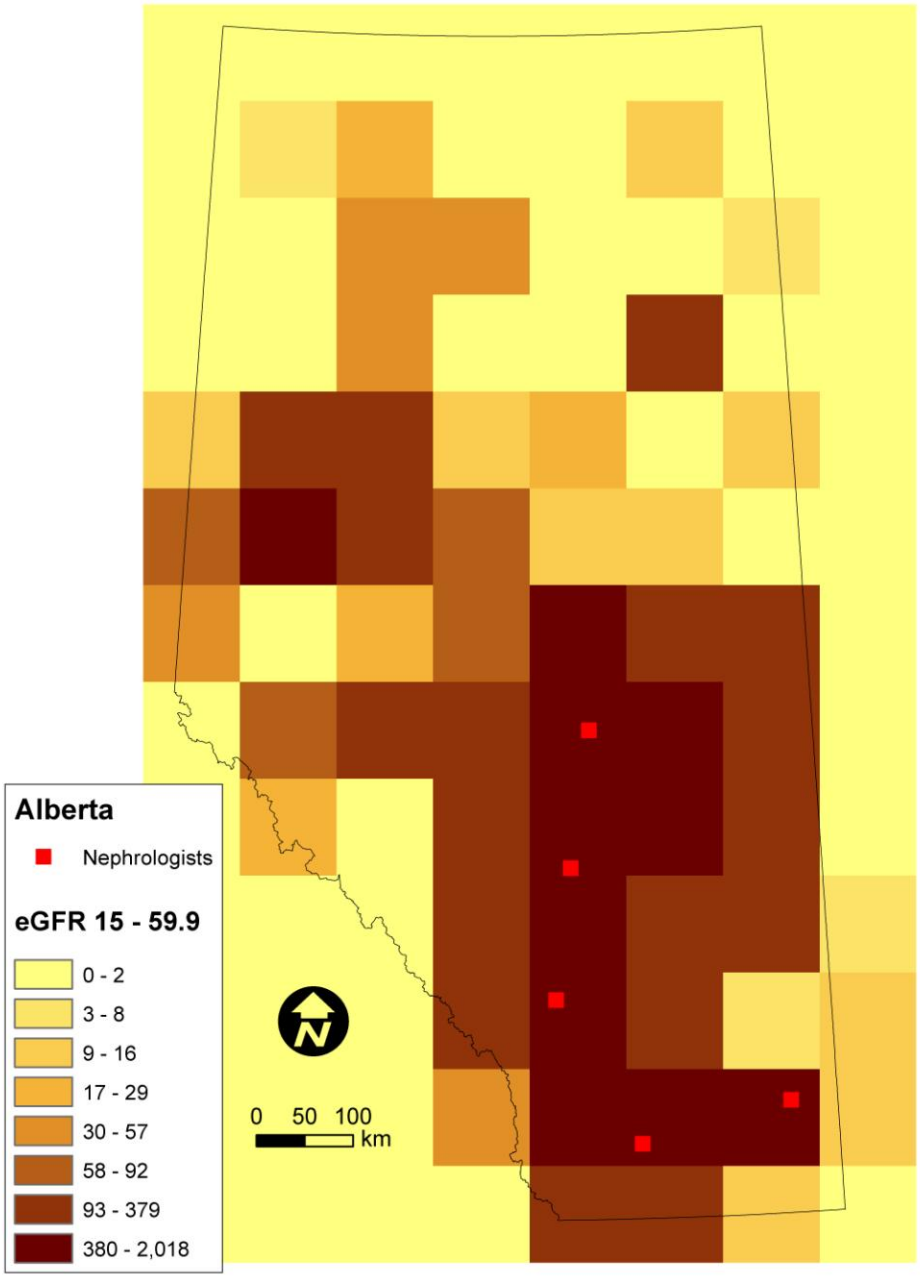


Figure 2: Shaded map presenting the distribution of people with diabetes and chronic kidney disease in Alberta
The figure shows that a substantial number of patients reside in areas located far from existing nephrology practice locations.

Figure 3: Location of clusters for four buffer scenario at 0.5%, 1%, 5% and 10% of population size

The figure is showing 16 maps of Alberta in four rows. Each row represents different cluster distributions in four maps varying only the maximum population size for a single buffer scenario. The purpose of this figure is to identify areas that were consistently shown to include clusters of underserved patients.

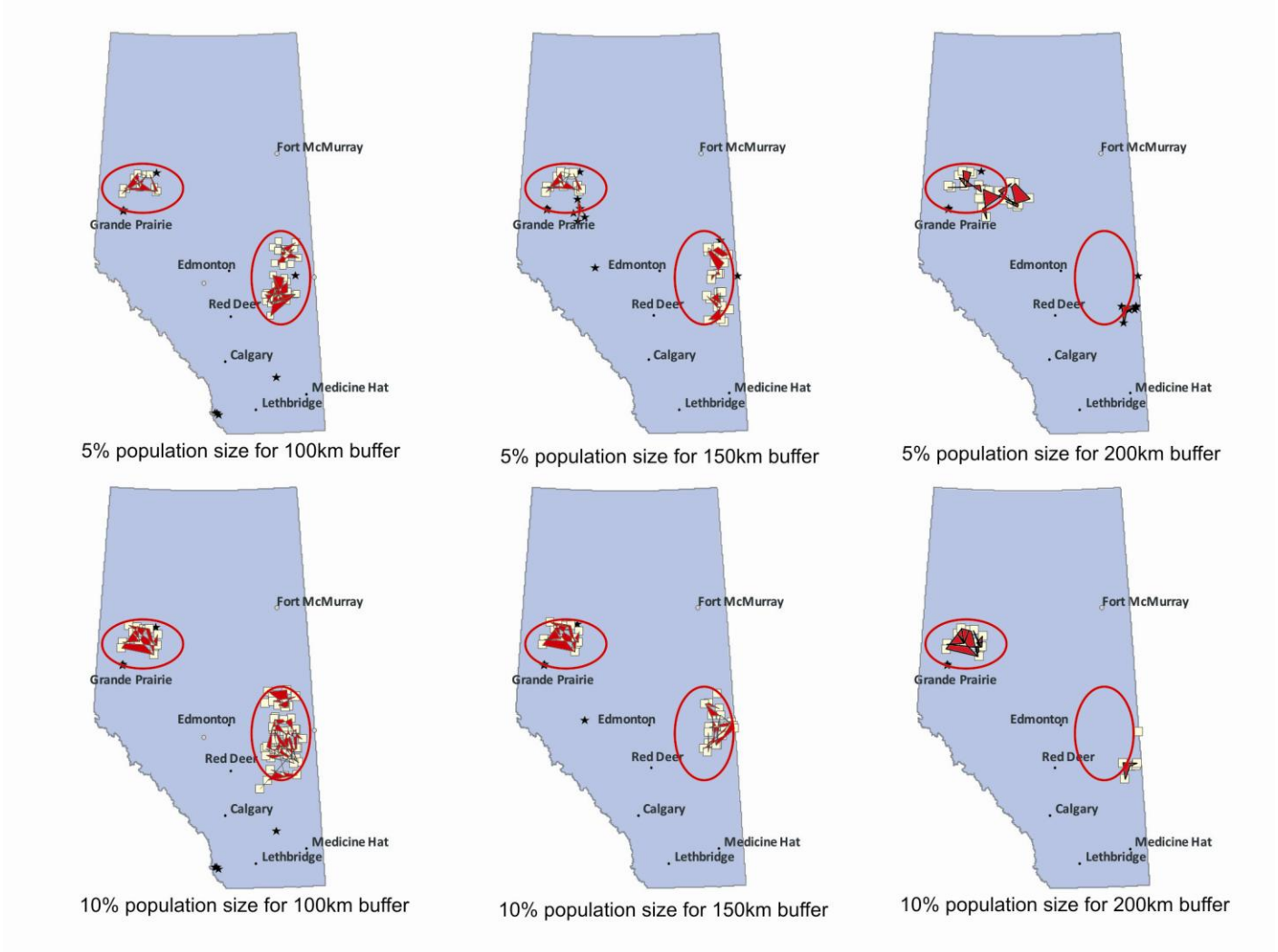


Figure 4: Consistently observed clusters at 5% and 10% of population size across three buffer scenarios (100 km, 150 km and 200 km)

This figure summarizes key findings from Figure 3, and demonstrates that the region in the horizontal oval (Grande Prairie) and the vertical oval (encompassing the communities of Bonnyville, Vegreville, Vermillion and Wainwright) were consistently identified as clusters of underserved patients, and are thus potential locations for new clinics.

Appendices

Appendix 1: Comparison between commonly used cluster detection methods

Name	Advantage	Disadvantage
Besag and Newell Method	It uses a multitude of overlapping circles to graphically identify possible clusters. Less running time is required for large datasets as it performed with selective search technique only centre upon case points.	If the centre of cluster is not located on case point, it may miss that cluster. Algorithm stops after meeting minimum case counts therefore chance of missing clusters of different case counts, when setting minimum at high counts may fail to capture smaller clusters, conversely, setting low counts may result in sub-clusters of a large cluster.
Geographical analysis machine (GAM) by Openshaw	It is useful for descriptive purpose. Graphically identify possible clusters using a multitude of overlapping circles of variable size. Ability to detect clusters regardless of their boundaries coincides with the administrative boundaries.	The circles do not vary continuously. It needs to perform separate significance test for each of the circle. Therefore multiple hypotheses testing with Bonferroni correction is required. There is chance of false positive results. Intensive running time is required for large datasets.
Cluster evaluation permutation procedure (CEPP) by Turnbull	If we know about the cluster size a priori, then we can use this technique. It constructs a test by using overlapping circles to detect clusters and address multiple testing problems. Ability to detect clusters regardless of their boundaries coincides with the administrative boundaries.	Cluster size must be specified a priori. Multiple hypothesis testing with Bonferroni correction still required as it address multiple testing problem for a pre-determined population size but while using different population size as recommended, multiple hypothesis need to be tested.
SaTScan by Kulldorff	If we do not know about the cluster size a priori, then we can use this technique. It uses likelihood ratio test to identify clusters of different size and adjust for multiple hypothesis testing.	If we do not specify any maximum population size, this technique by default chooses 50% of the population size and evaluates very small and very large clusters.

Reference:

Kulldorff M, Nagarwalla N: Spatial Disease Clusters - Detection and Inference. *Statistics in medicine* 1995, 14(8):799-810.

Kulldorff M: Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association* 2006, 101(475):1289-1305.

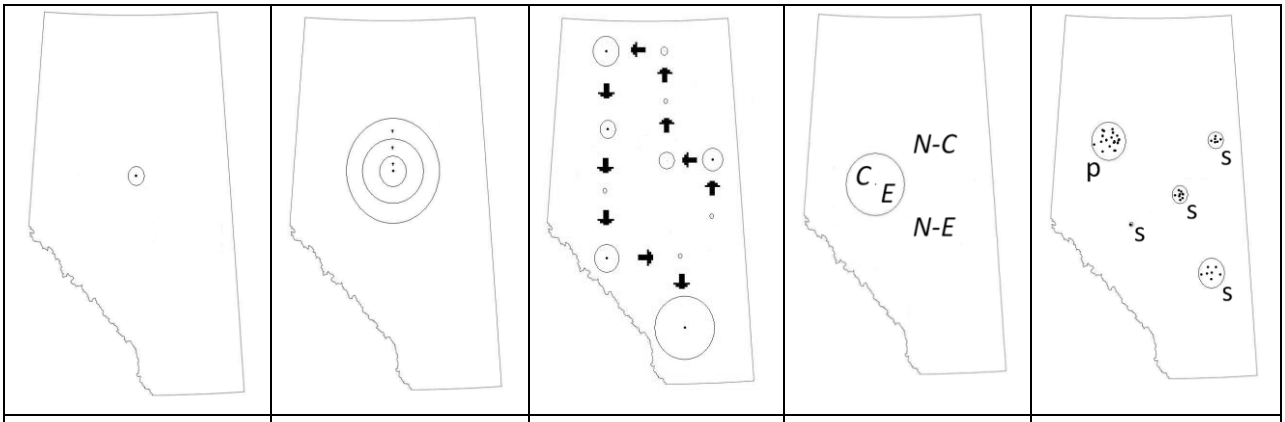
Kulldorff M: SaTScan™ User guide. For version 9.0 July 2010. In. http://www.satscan.org/.

Conley J, Gahegan M, Macgill J: A Genetic Approach to Detecting Clusters in Point Datasets. In.: GeoVISTA Center, Department of Geography, The Pennsylvania State University.

Appendix 2: A step by step visual example of one iteration of SaTScan analysis

Step 1: The spatial scan statistic imposed a circular window on the map of Alberta

Step 2: The radius of the window varied continuously in size from zero to the upper limit specified *a priori* (in this analysis, we specified the upper limit at 0.5%, 1%, 5% and 10% of the population size)



Step 1	Step 2	Step 3	Step 4	Step 5

Step 3: The window was centered one by one on each of the given grid points (latitudes and longitudes of the postal codes) located throughout the study region

In summary, for each of the postal code, the radius of the window varied continuously in size, finally we had an infinite number of distinct geographical windows with different sets of adjacent neighborhoods within them

Step 4: While gradually scanning a circular window across the entire map, the technique noted the number of observed and expected cases inside the circle at each location, by this way the clusters were detected.

Under the Poisson assumption, the likelihood function for a specific window is proportional to:

$$\left(\frac{C}{E}\right)^{C}\left(\frac{N-C}{N-E}\right)^{N-C} I\,(C > E)$$

Where N is the total number of cases over the study area,
C is the number of cases within the window, and
E is the expected number of cases within the window under the null-hypothesis.
Therefore, N-E is the expected number of cases outside the window.
I is an indicator function that is equal to 1 when the window has more cases than expected under the null hypothesis, and 0 otherwise.

By maximizing the likelihood over all windows, most likely disease cluster had been identified based on the maximum likelihood ratio. However, the distribution of this maximum likelihood ratio test statistic under null hypothesis and corresponding simulated P value is calculated by

Monte Carlo simulation.

Finally, the analysis identified primary cluster (p) and secondary clusters (s)

The non-overlapping additional clusters had been reported as secondary clusters if the likelihood ratio is larger than the likelihood ratio for primary cluster for at least one data set simulated under the null hypothesis.

Reference:

Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N: Childhood leukaemia in Sweden: Using GIS and a spatial scan-statistic for cluster detection. *Statistics in medicine* 1996, 15(7-9):707-715.

Kulldorff M, Feuer EJ, Miller BA, Freedman LS: Breast cancer clusters in the northeast United States: A geographic analysis. *American journal of epidemiology* 1997, 146(2):161-170.

Kulldorff M: SaTScan™ User guide. For version 9.0 July 2010. In. http://www.satscan.org/.

Appendix 3: SaTScan used in different patient population with a variety of objectives

Field of study	Place	Software	Analysis	Objectives
Cancer investigations	Florida	SaTScan ArcGIS	Purely spatial and space-time analysis Create maps with cancer clusters	To assess childhood clusters of cancers including leukemia, lymphoma and brain cancer [82]
Infectious diseases	Germany	SaTScan EpiScanGIS	Spatio-temporal clusters Generate animated map	Online geographic surveillance system (EpiScanGIS) for meningococcal disease [83]
	Malaria–endemic county in Hainan province, China	SaTScan ArcGIS	Temporal and spatial variation Digital map of GIS based Analysis	To evaluate the variation of malaria clusters [84]
	Mali	SPSS SaTScan GPS GeoExplorerII ArcGIS	Classical ARIMA time series analysis model to perform global temporal analysis Space time clusters at household level Georeferencing	To identify high risk zone of childhood malaria to facilitate understanding of local pattern of malaria transmission and infection and adopt interventions [85]
	Five highly endemic Indonesian Islands	SaTScan ArcView	Spatial clustering Preparation of maps	To define clustering of leprosy patients [86]
	Almora district in Uttaranchal state in India	SaTScan	Purely spatial and space-time clusters	To identify statistically significant clusters of tuberculosis cases [87]
Suicide	Queensland, Australia	SaTScan MapInfo	Spatial cluster	To examine spatial distribution of suicide [88]
Poisoning	Texas counties	SaTScan ArcGIS	Spatial temporal clusters Produce maps	To identify clusters of accidental poisoning mortality, geographical variation of

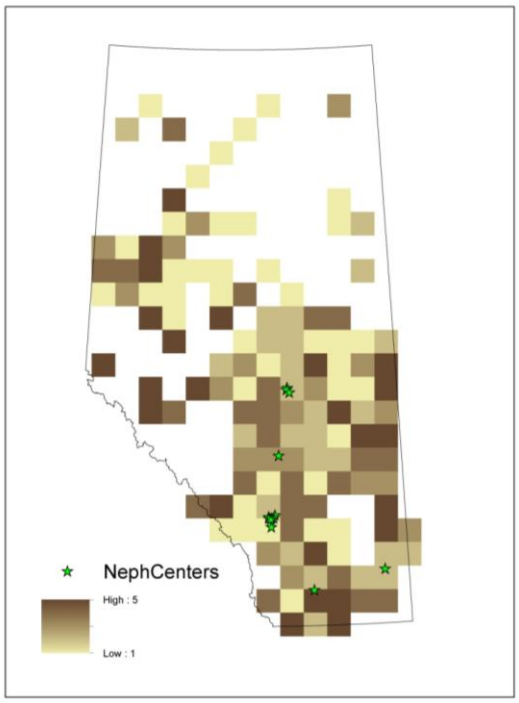
				clusters in terms of gender and ethnicity [89]
Adjustments for other covariates	North Carolina	SaTScan	Spatial clusters	To determine the location and extent of clusters of gastroschisis (an abdominal wall birth defect) [90]
Spatial scan statistics followed by regression analysis	Texas	SaTScan SPSS	Spatial cluster at the county level Logistic regression analysis	To identify spatially significant clusters of diabetes and the association between the prevalent clusters and age/obesity [91]
	Winnipeg, Canada	SaTScan ArcView	Spatial cluster Mapping of clusters	To detect prevalent clusters of diabetes [92]
With other techniques	Guichi in Anhui province, Eastern China	Dcluster SaTScan ArcGIS	Global clusters (Cuzick-Edwards test) Spatial clusters (Besag-Newell test and spatial scan statistics) Produce map	To locate high risk regions of acute schistosomiasis [93]
	Connecticut, USA	 SaTScan	Spatial filtering Spatial clusters	To eliminate random variation caused by small population and smaller cases per town To determine the significant clusters of hepatitis C virus infection [94]
	Mainland China	GoeDa SaTScan ArcGIS	Global spatial autocorrelation Clusters at the county level	To prioritize areas of Hemorrhagic fever with renal syndrome (HFRS) for public health planning and resource allocation [95]

Appendix 4: Supplementary Table: Results stratified by CKD stage

Buffer	Population	Maximum population size	Stage 3 CKD				Stage 4 CKD			
			Cases	Number of significant clusters	Radius of primary cluster	Observed cases in the primary cluster	Cases	Number of significant clusters	Radius of primary cluster	Observed cases in the primary cluster
50 km	621,984	0.5%	7157	22	0.50 km	59	879	3	0.14 km	8
		1 %		24	0.50 km	59		3	0.14 km	8
		5%		17	86.02 km	584		4	0.14 km	8
		10%		15	100.41 km	1052		4	133.20 km	115
100 km	387,911	0.5%	3799	15	0.56 km	58	436	2	0.090 km	6
		1 %		16	0.68 km	72		2	0.090 km	6
		5%		15	67.42 km	285		3	0.090 km	6
		10%		13	124.98 km	583		3	0.090 km	6
150 km	308,293	0.5%	2797	13	0.29 km	18	323	2	0.090 km	6
		1 %		14	0.29 km	18		2	0.090 km	6
		5%		13	53.12 km	238		3	0.090 km	6
		10%		8	75.75 km	379		3	0.090 km	6
200 km	237,201	0.5%	1990	9	0 km	12	224	2	0.090 km	6
		1 %		9	0 km	12		2	0.090 km	6
		5%		11	49.36 km	175		3	0.090 km	6
		10%		9	59.48 km	336		3	0.090 km	6

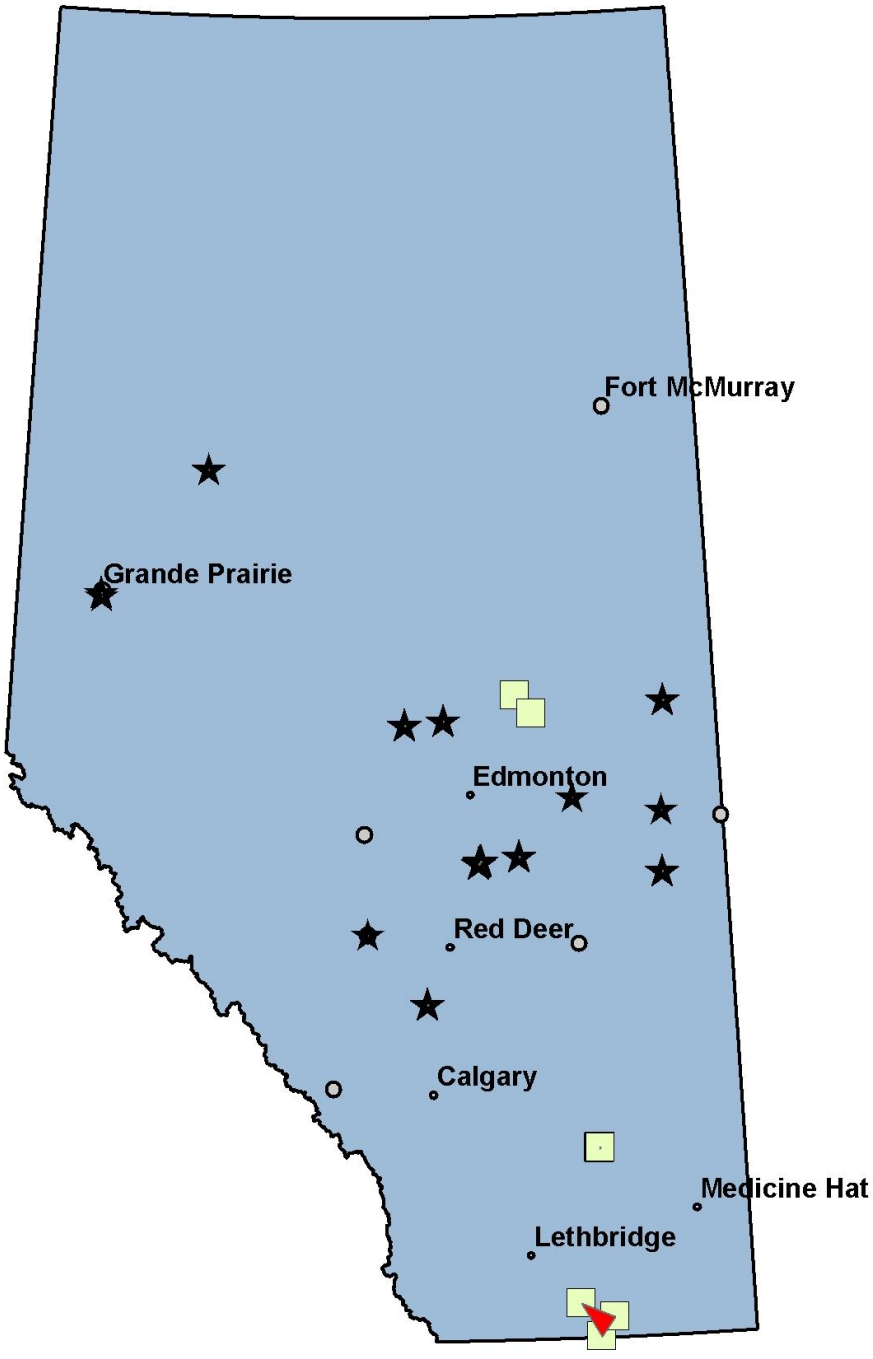
Appendix 5: Neighborhood poverty index assessment

In Alberta, poverty is correlated with distance from the closest facility – but the correlation is weak. Therefore, neighbourhood poverty index is not useful for identifying new facilities. We demonstrate this in the attached Figure:

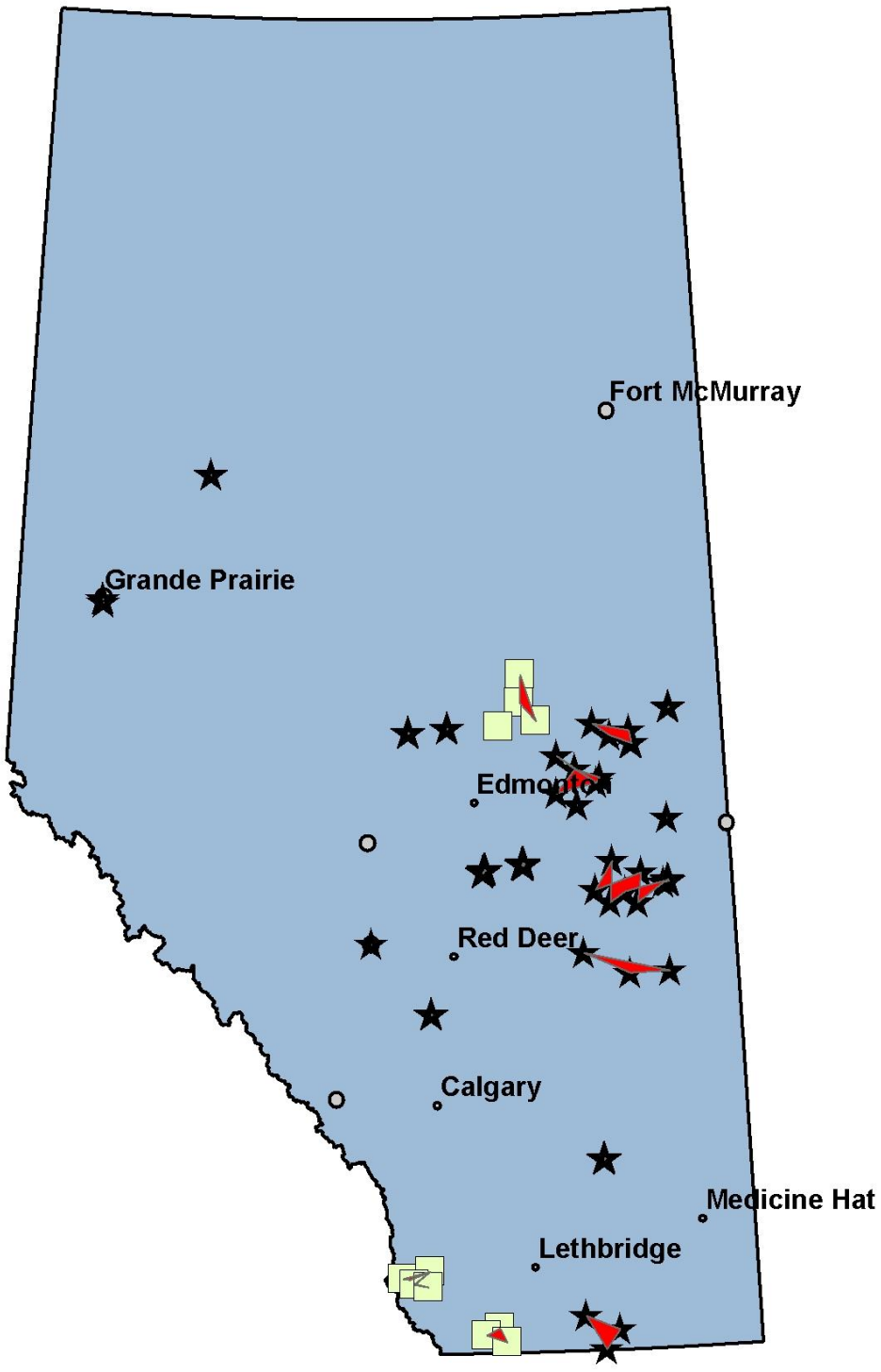


Using census data, we assigned a quintile value for neighbourhood income from 1 to 5 (1 for lowest and 5 for highest income) for each geographic unit (dissemination area). The Figure presents median neighbourhood income using shading for each 50 km square grid. The existing 17 nephrology clinics are also shown on the map (green stars). The Figure shows that patients with incomes in the lower quintile (yellowish shades) were scattered throughout the province. Grande Prairie, Vermillion and southern Alberta (where we identified the significant clusters of underserved patients using SaTScan) were surrounded by patients of both lower and higher income. Given these findings, we concluded that neighbourhood income cannot be used as the basis for new clinic locations.

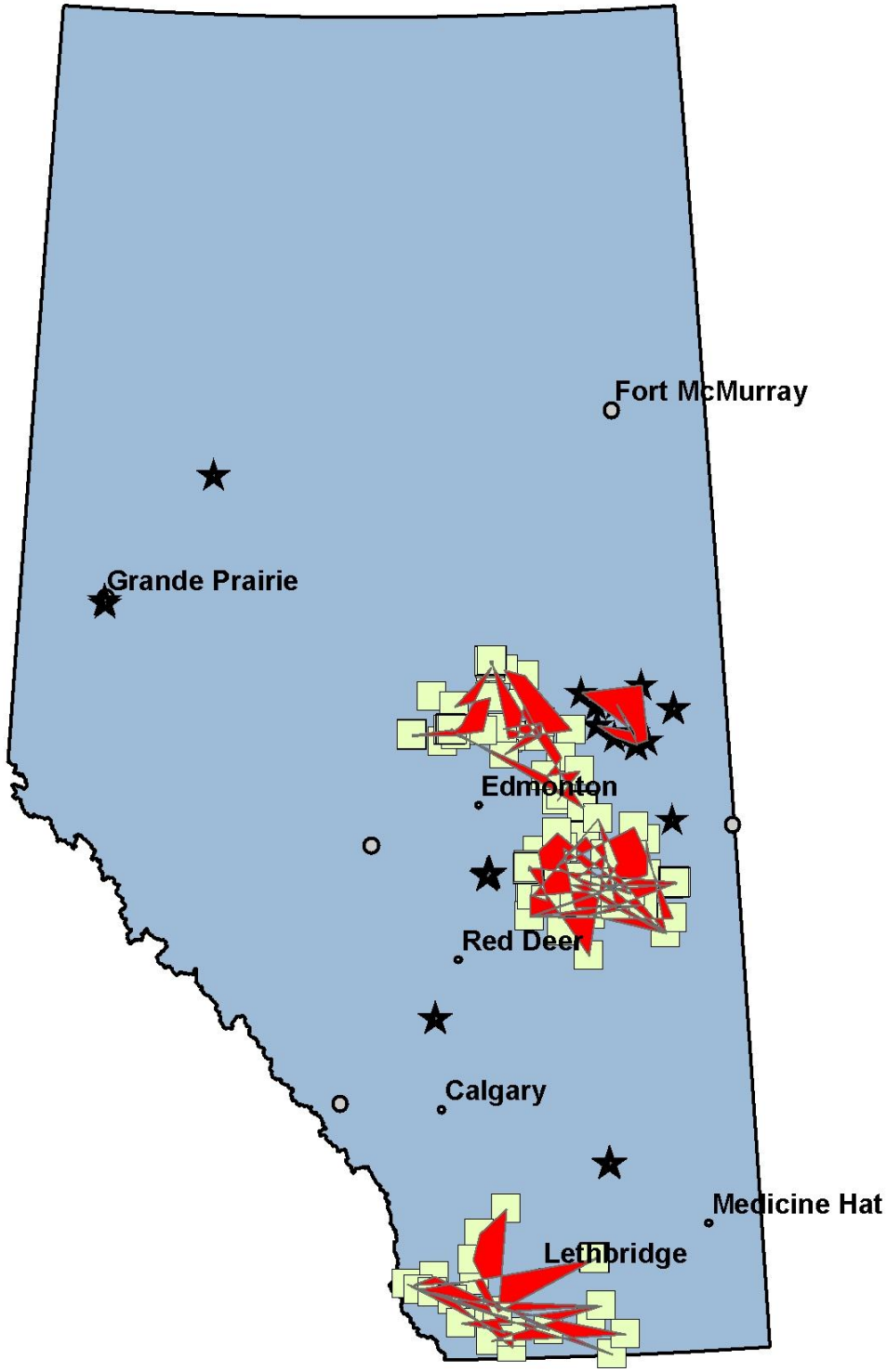
Appendix 6: Individual map presented for 16 scenarios (4 buffer X 4 population size) on Table 2



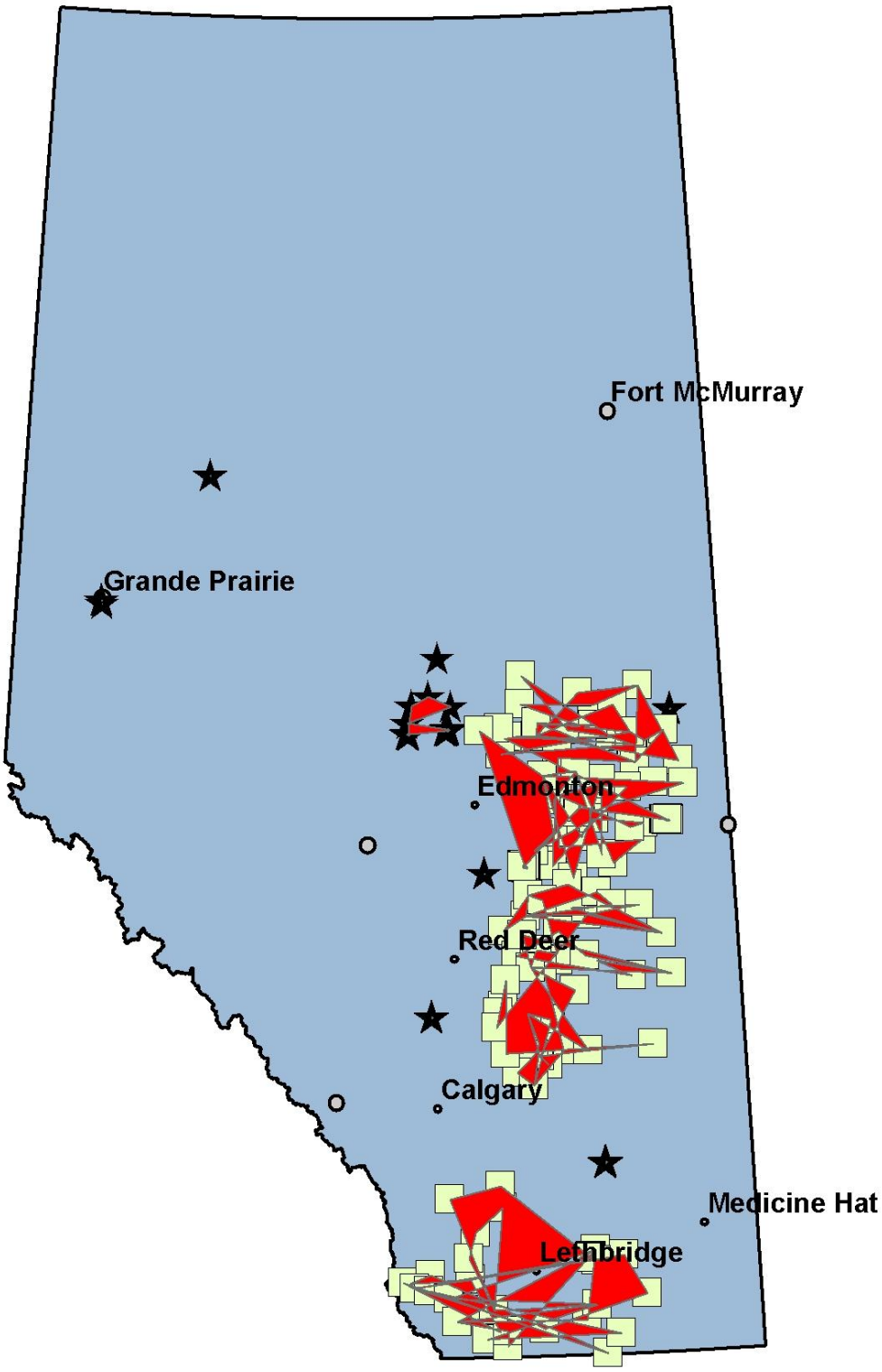
Map 1: 50 km buffer 0.5% population size



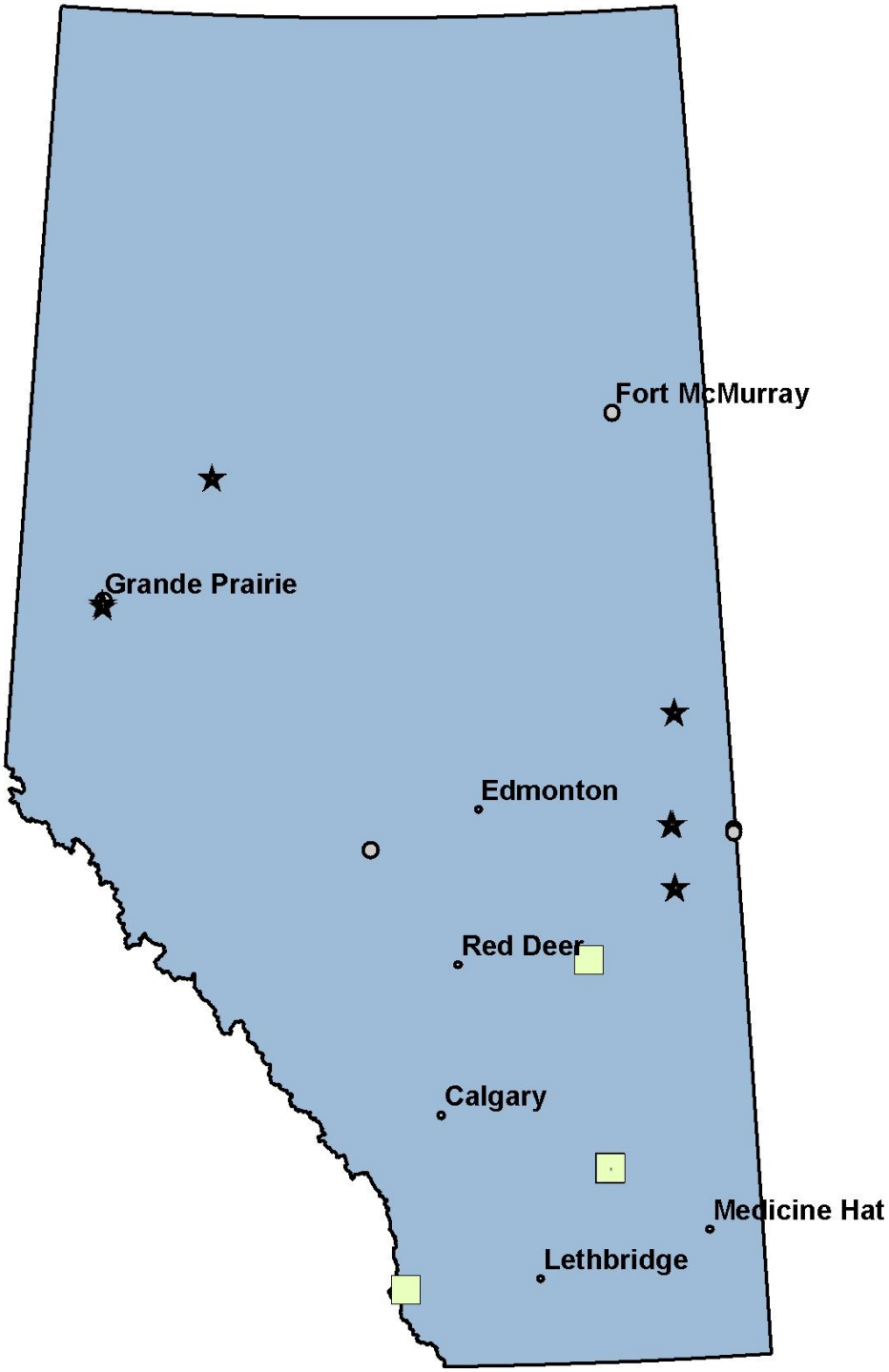
Map 2: 50 km buffer 1% population size



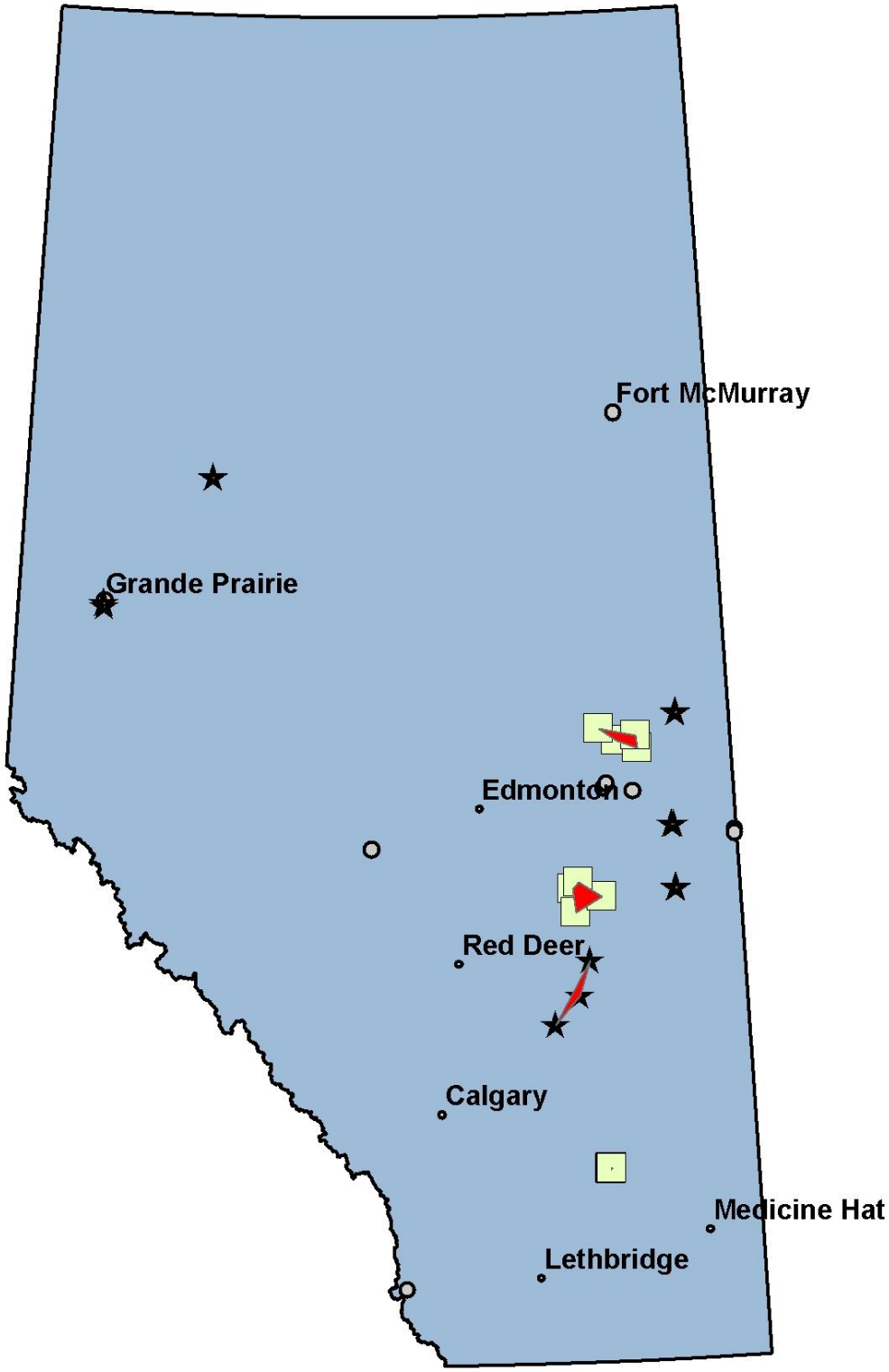
Map 3: 50 km buffer 5% population size



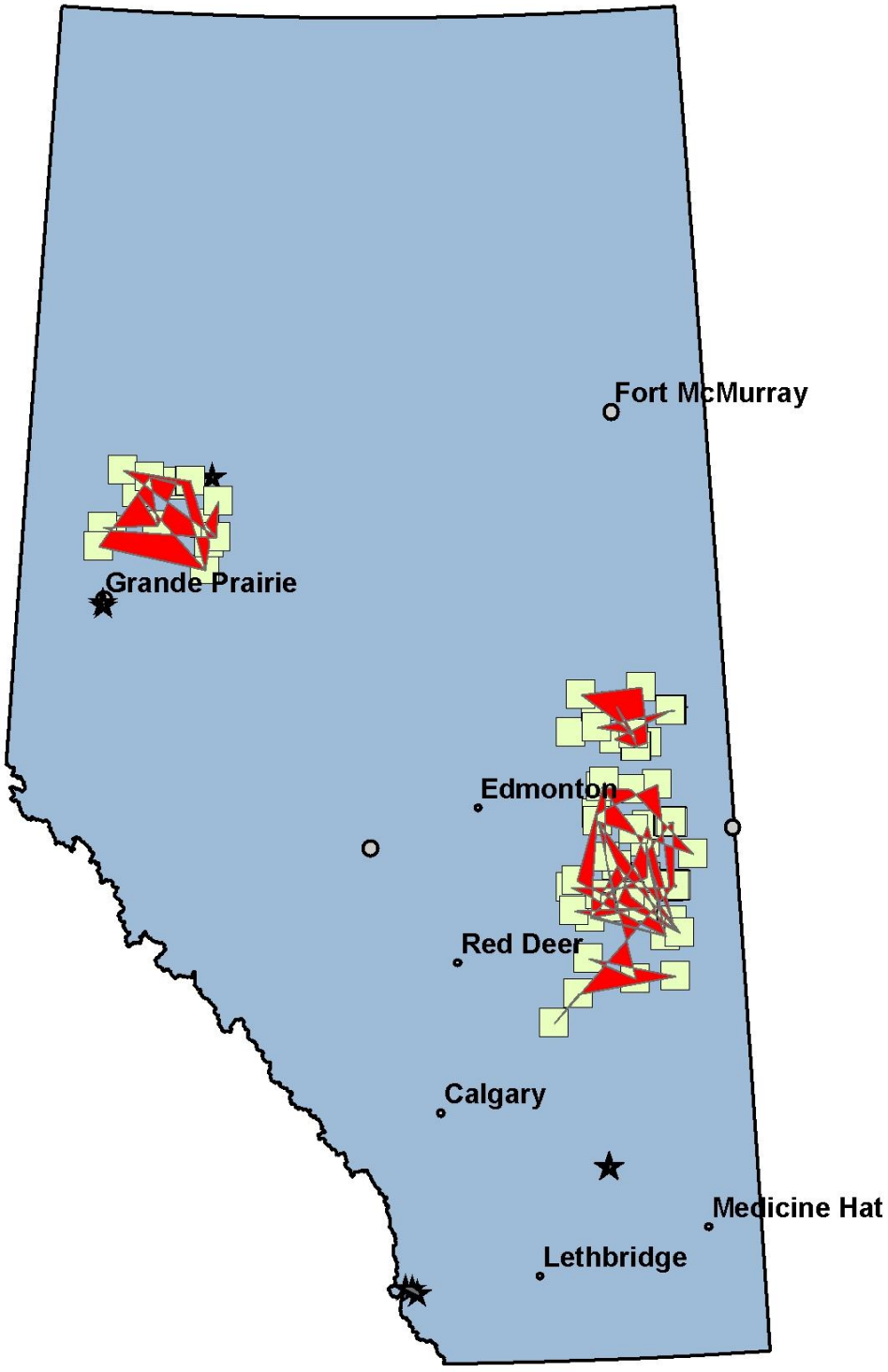
Map 4: 50 km buffer 10% population size



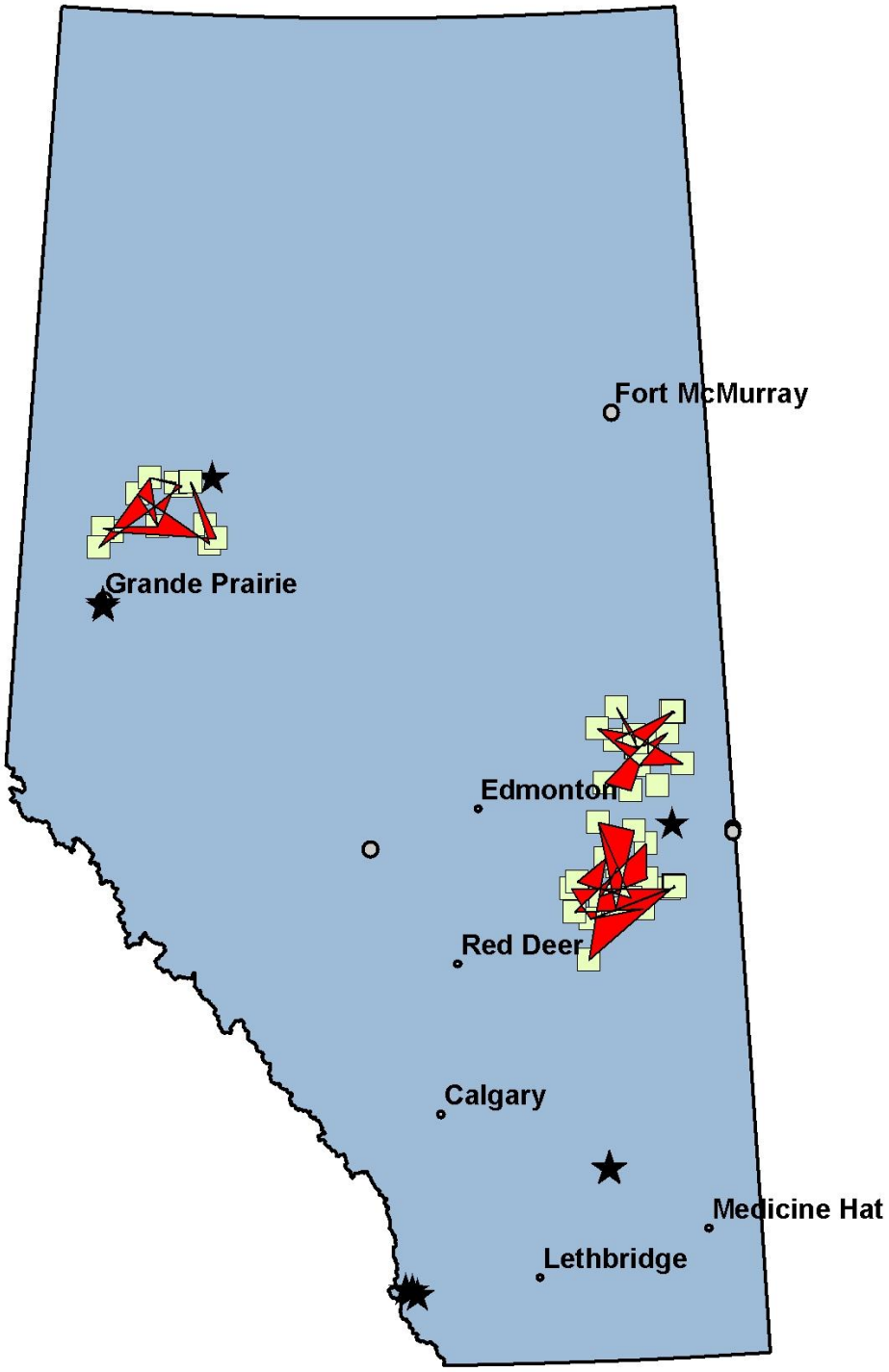
Map 5: 100 km buffer 0.5% population size



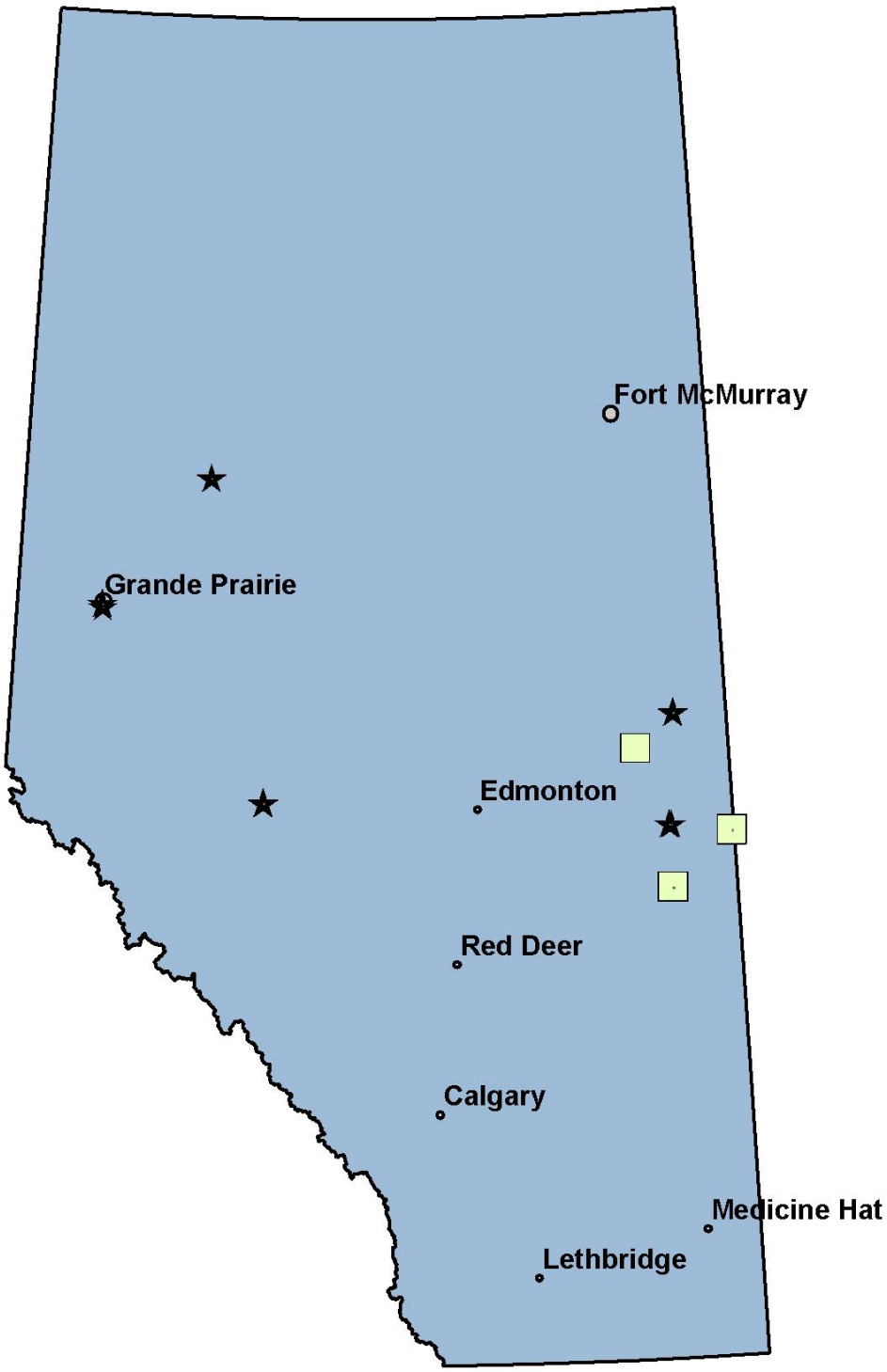
Map 6: 100 km buffer 1% population size



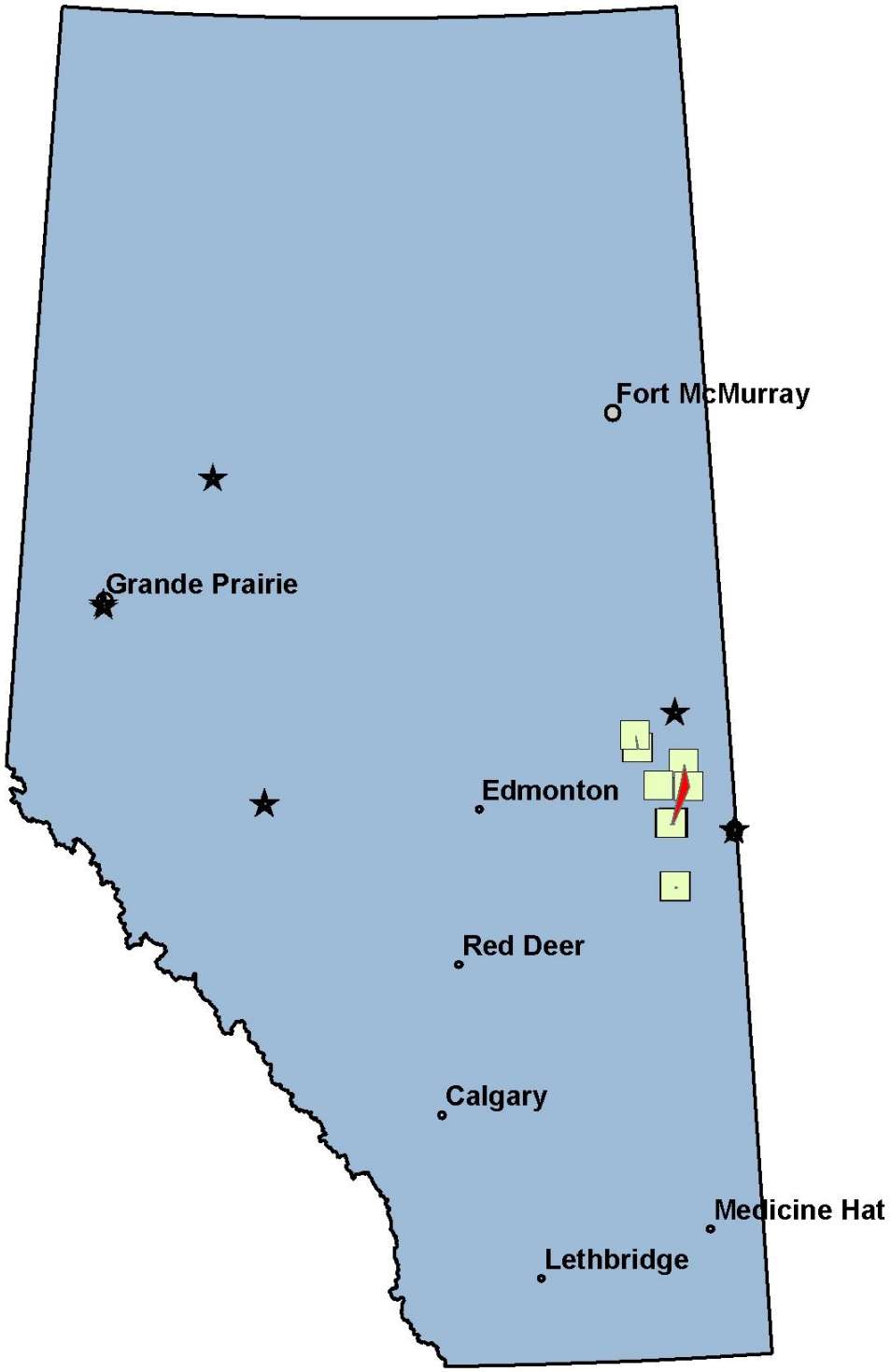
Map 7: 100 km buffer 5% population size



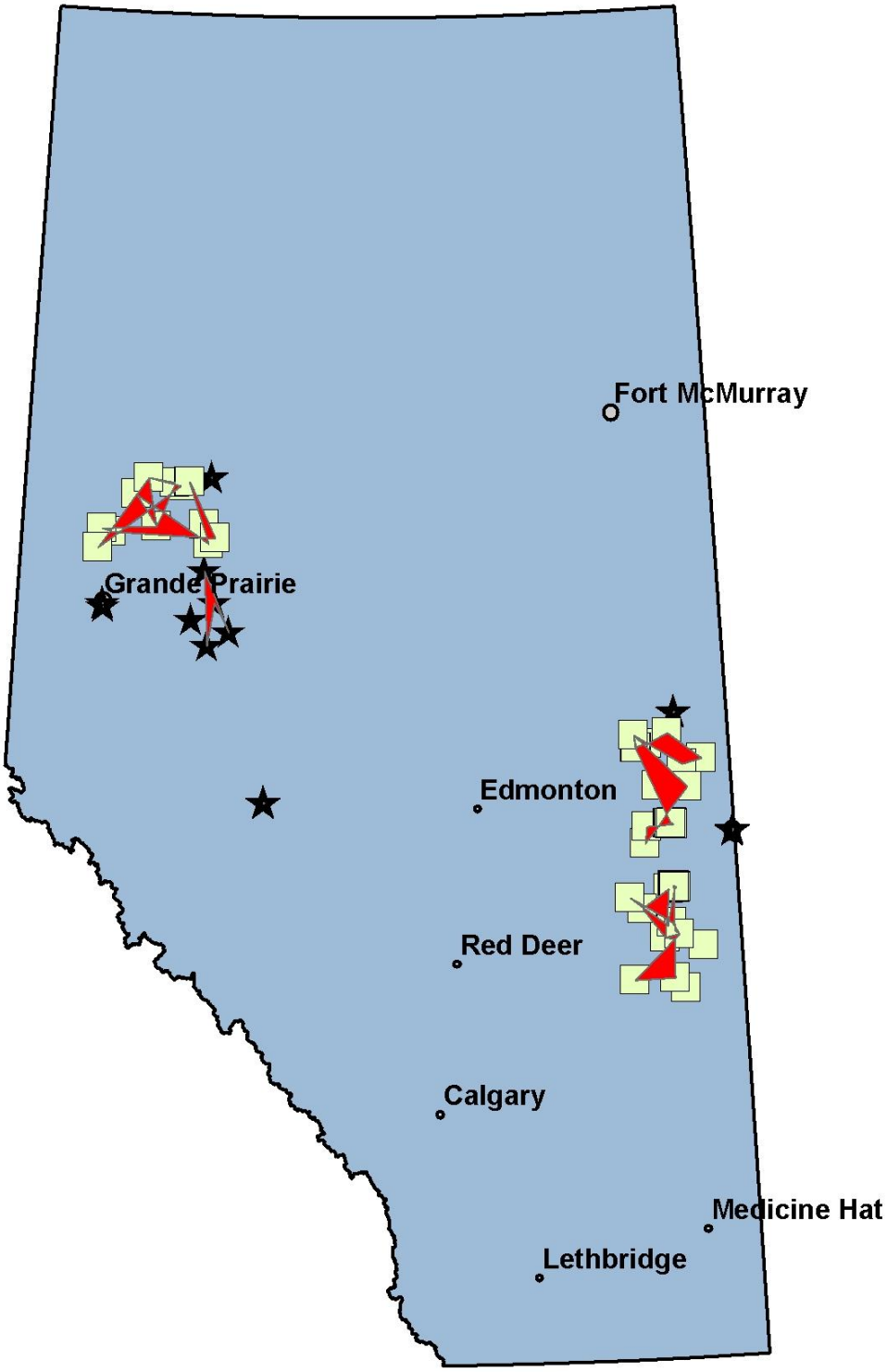
Map 8: 100 km buffer 10% population size



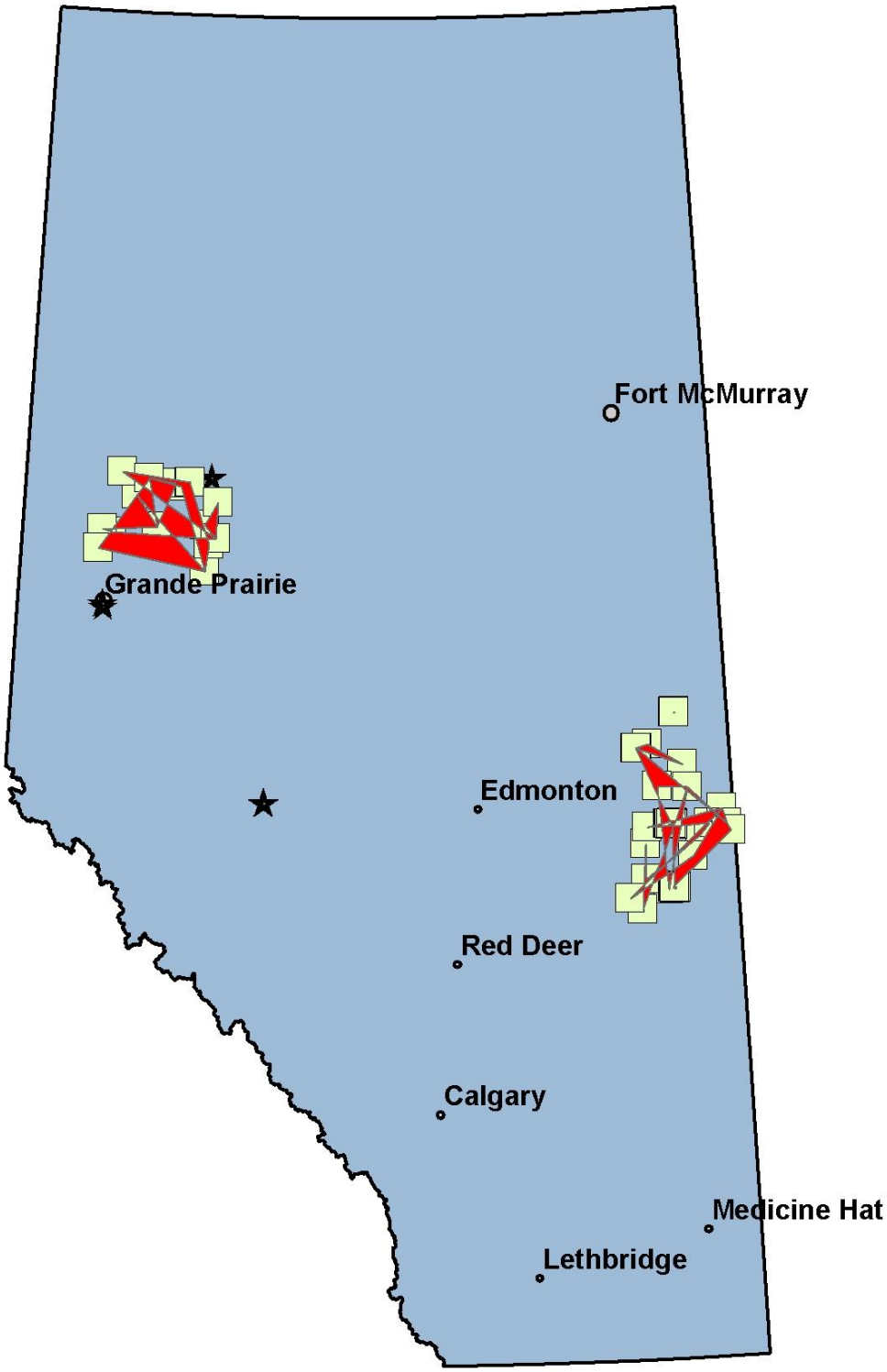
Map 9: 150 km buffer 0.5% population size



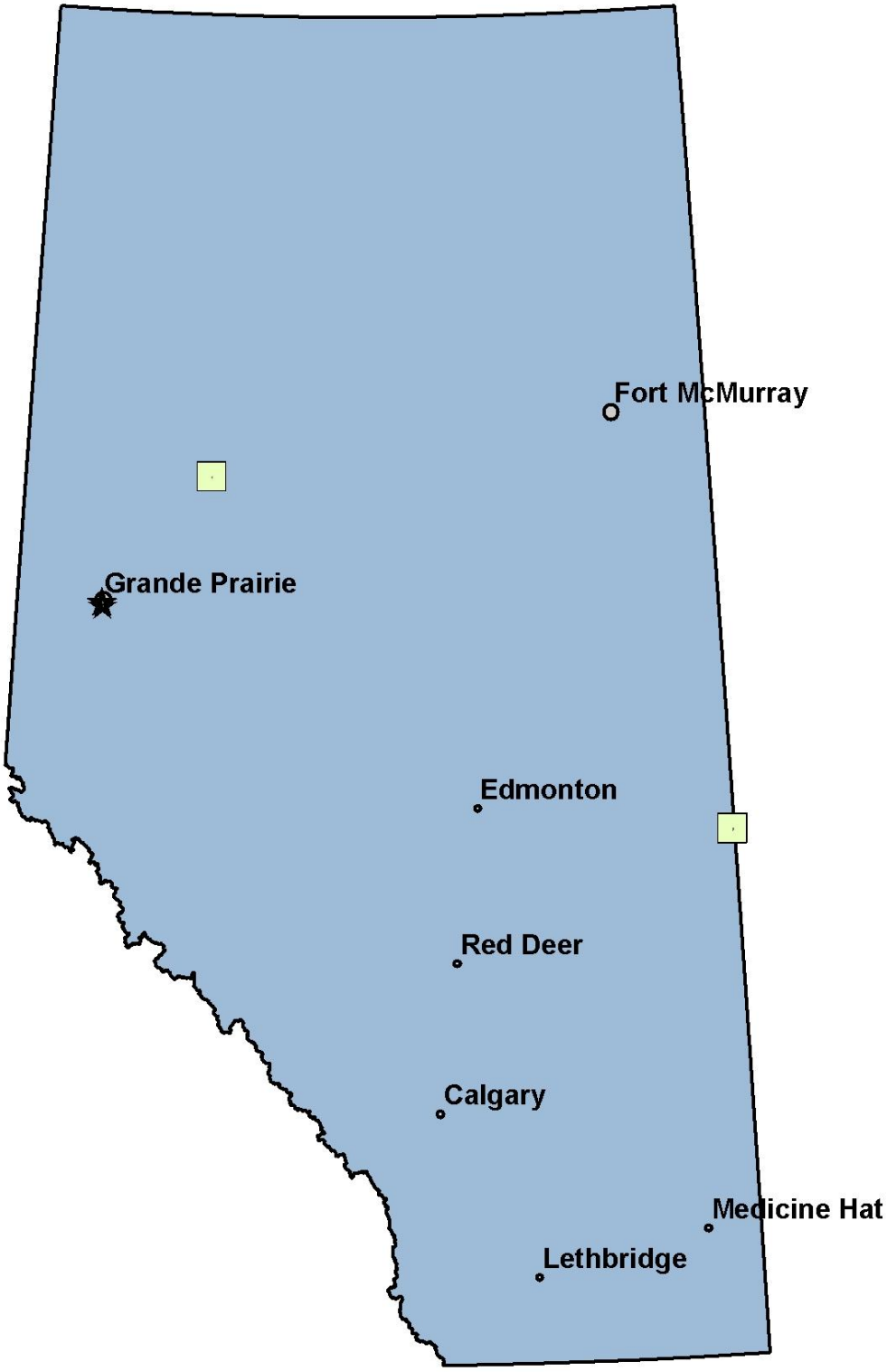
Map 10: 150 km buffer 1% population size



Map 11: 150 km buffer 5% population size

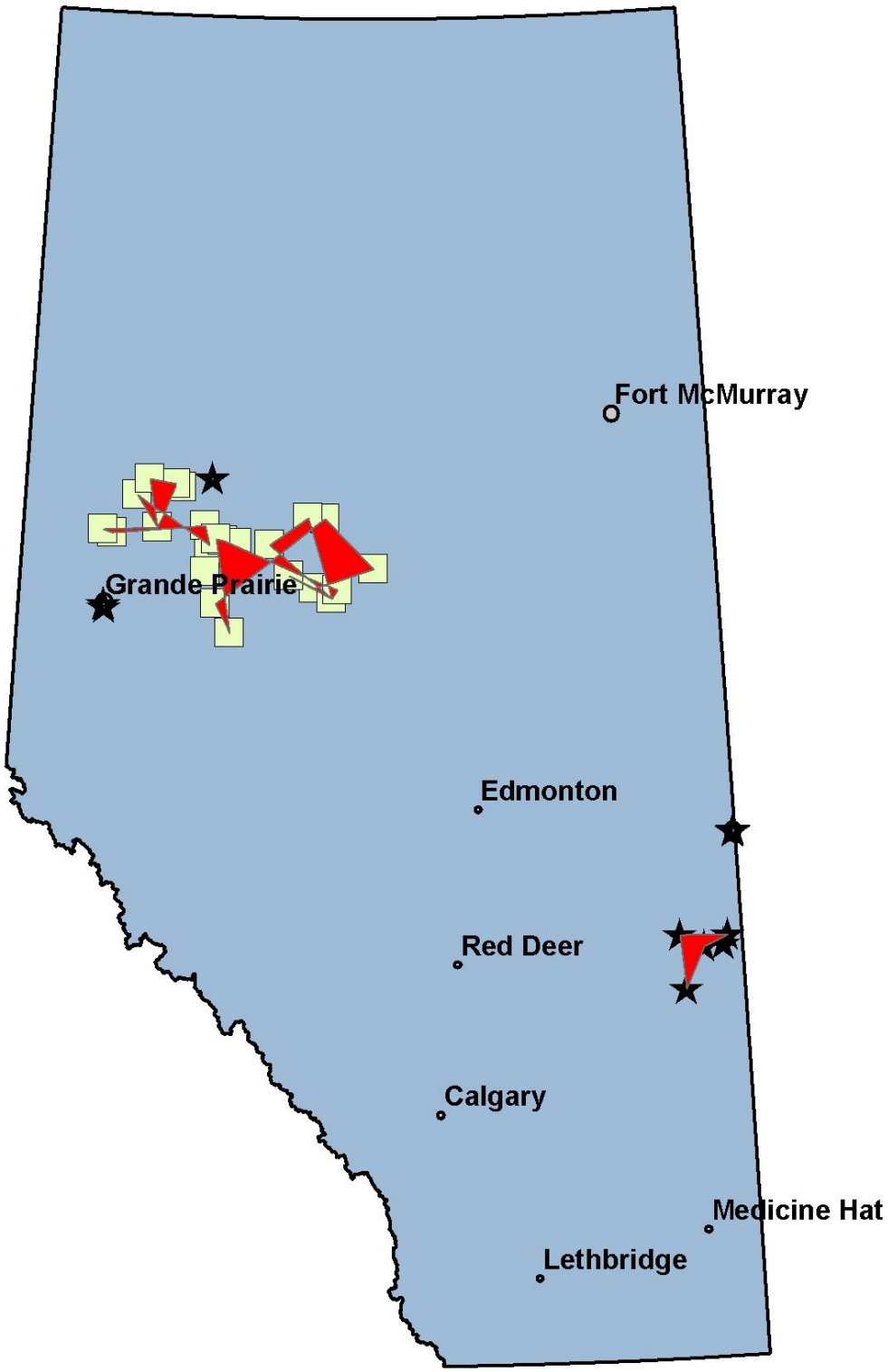


Map 12: 150 km buffer 10% population size

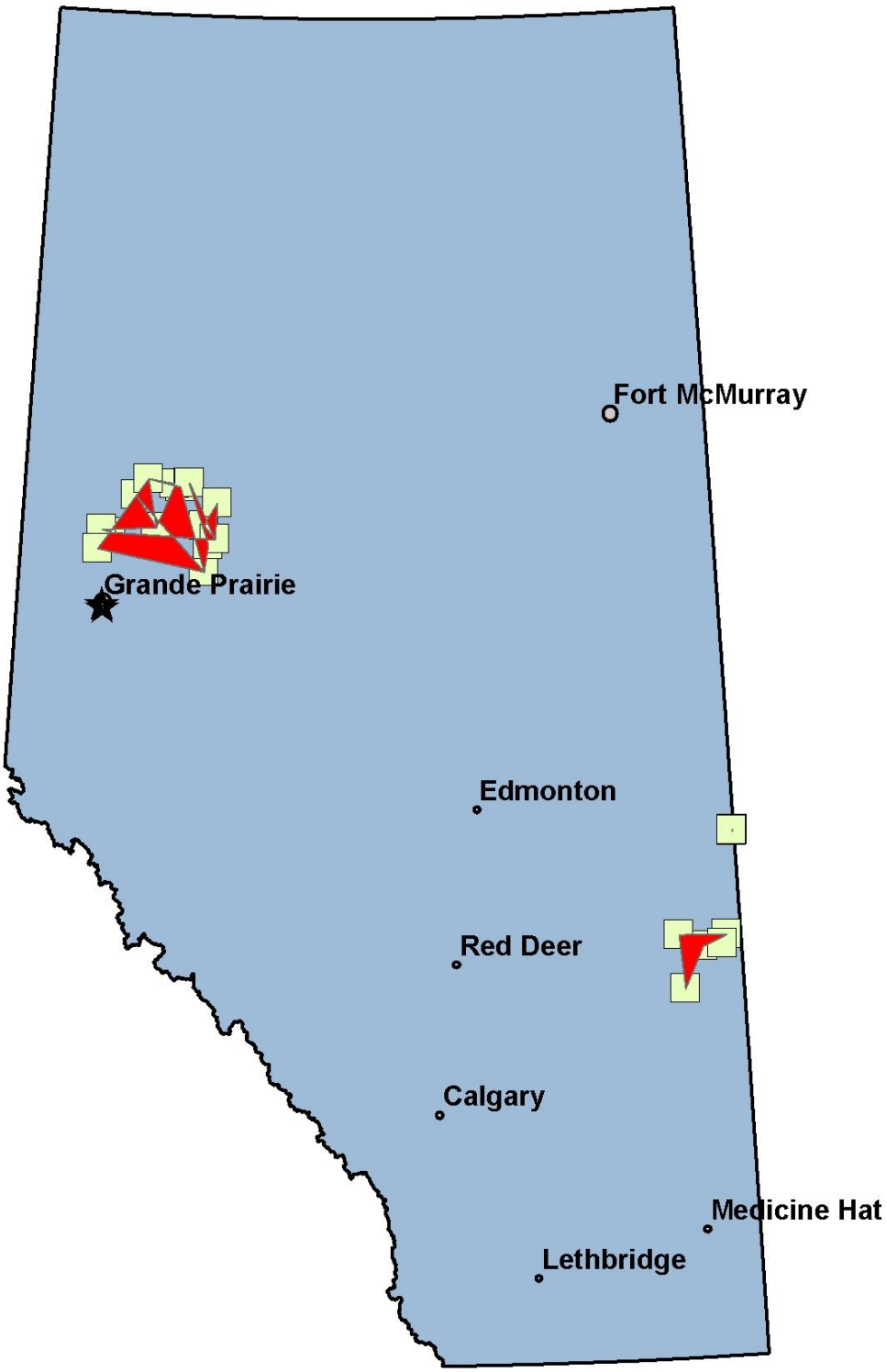


Map 13: 200 km buffer 0.5% population size

Map 14: 200 km buffer 1% population size

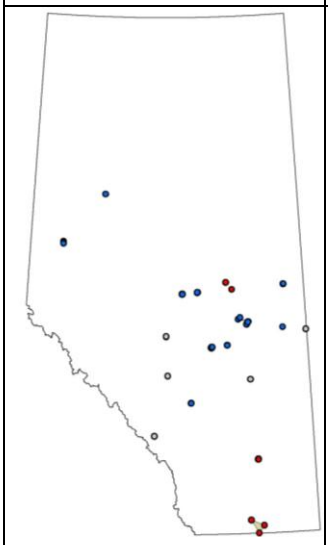
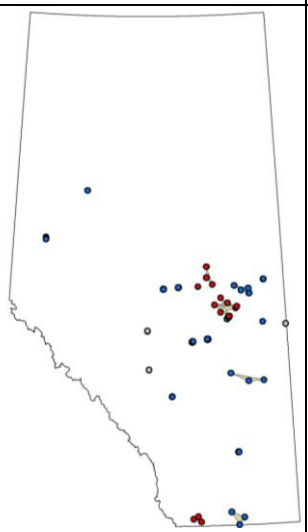
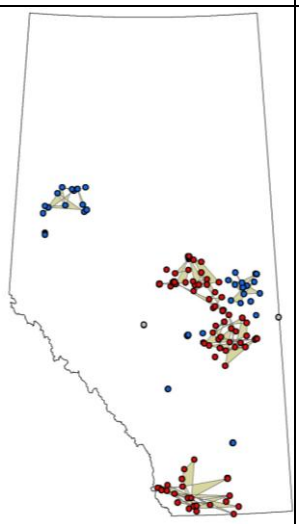
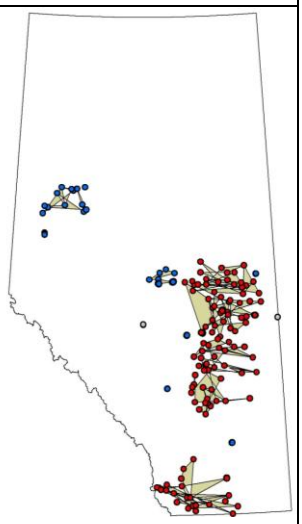
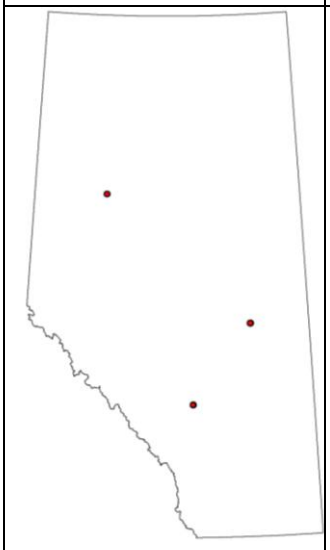
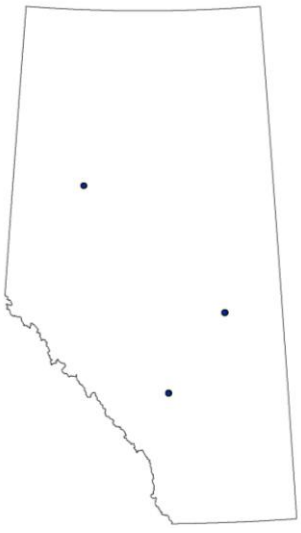
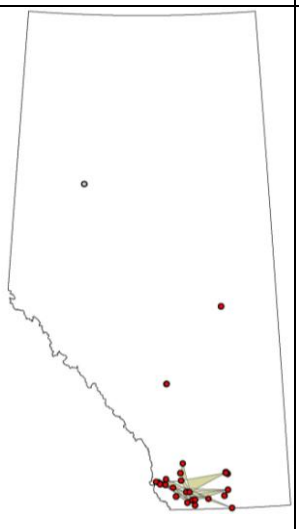
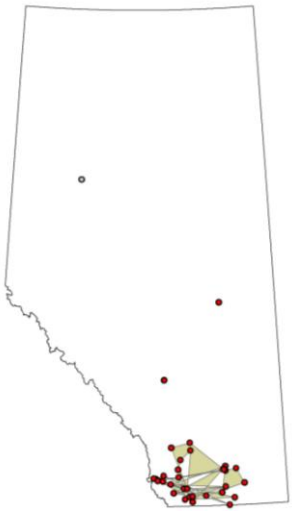


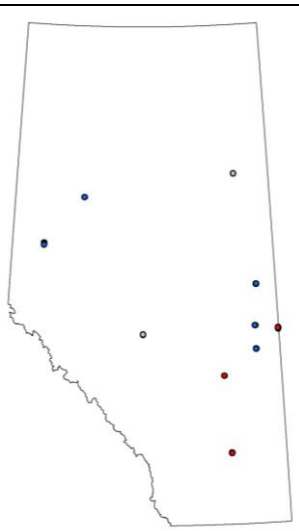
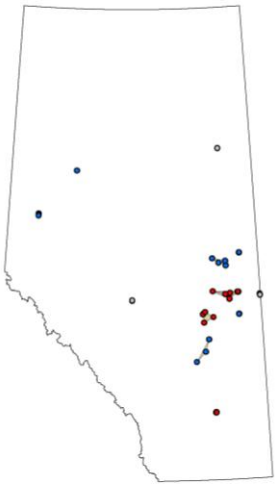
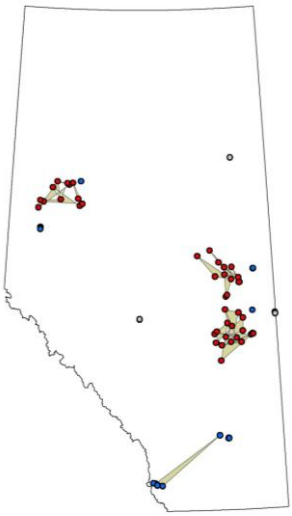
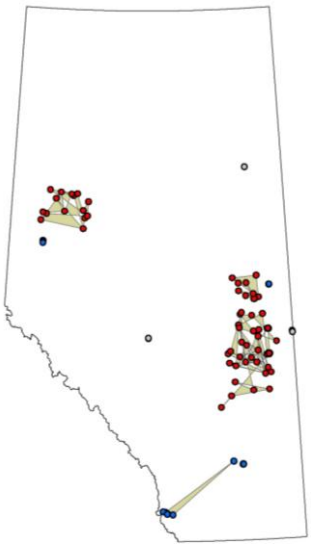
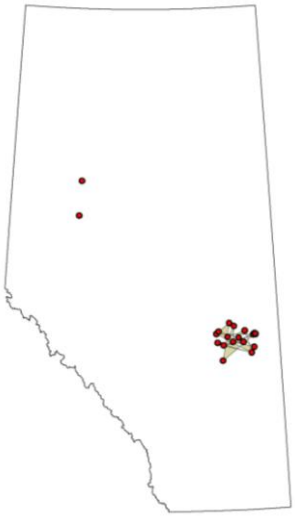
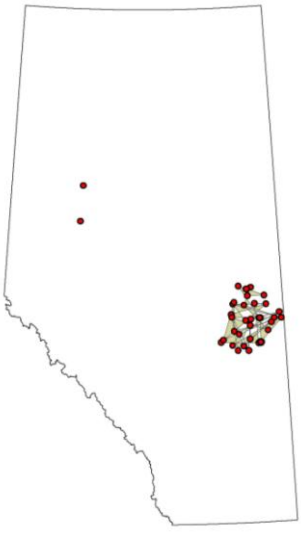
Map 15: 200 km buffer 5% population size

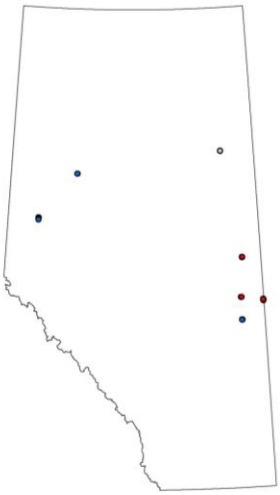
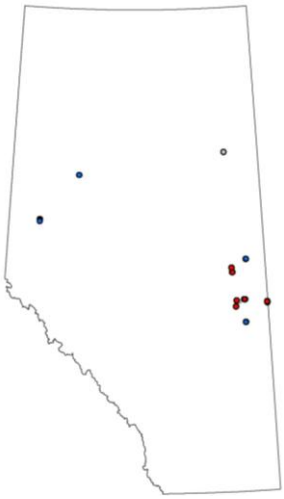
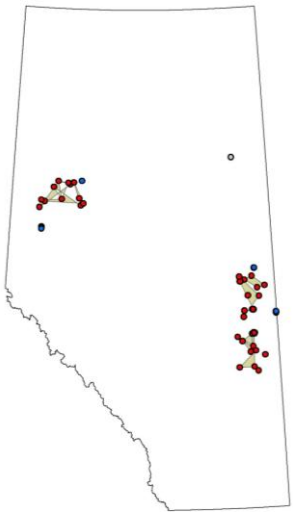
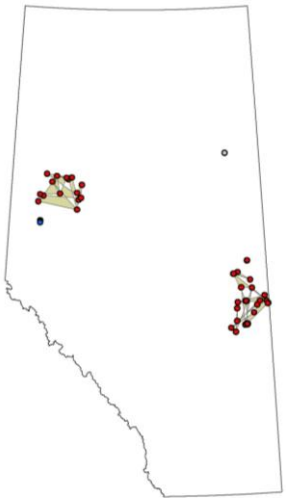
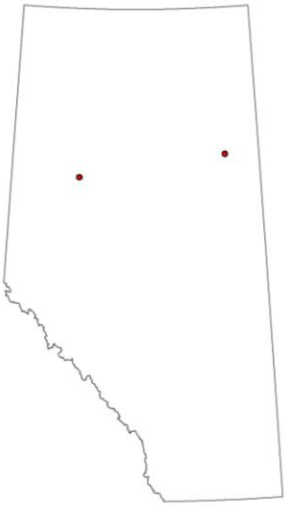
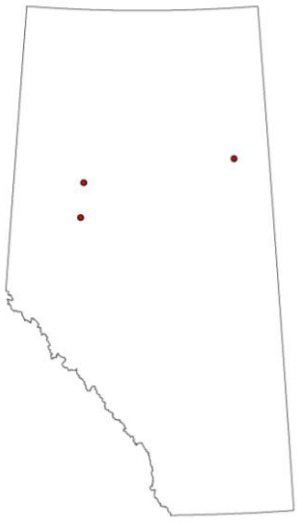
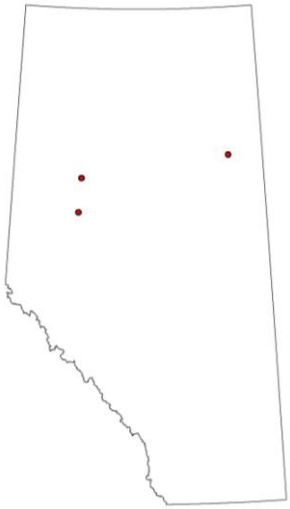


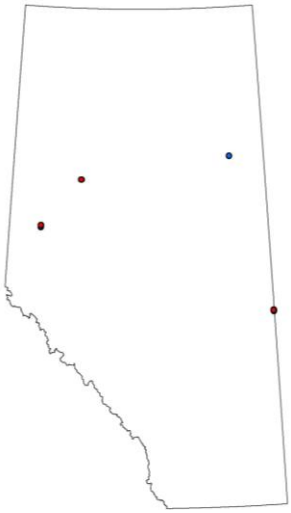
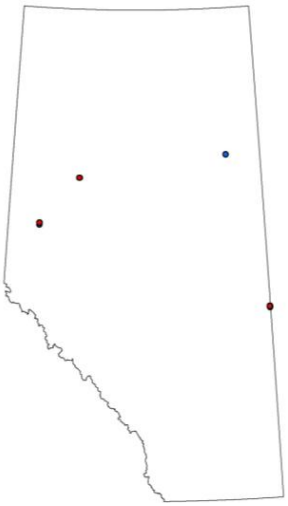
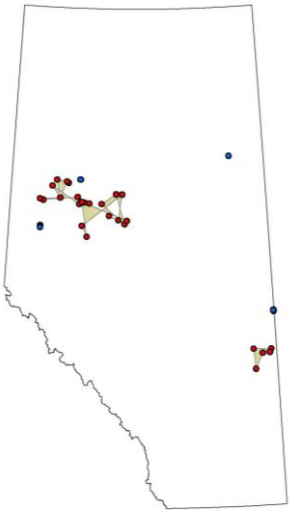
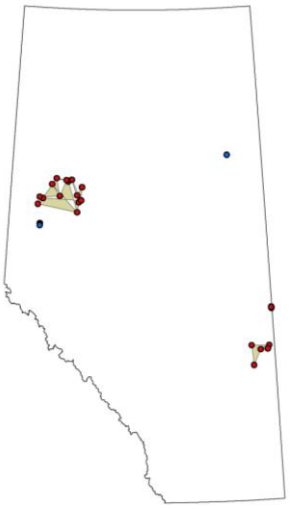
Map 16: 200 km buffer 10% population size

Appendix 7: Figure of 32 maps from stratified analysis on CKD stage from Supplementary Table

<table>
<tr><td colspan="4">50 km buffer scenario</td></tr>
<tr><td colspan="4">CKD Stage 3</td></tr>
<tr><td>0.5% population size</td><td>1% population size</td><td>5% population size</td><td>10% population size</td></tr>
<tr><td>  </td><td>  </td><td>  </td><td>  </td></tr>
<tr><td colspan="4">CKD Stage 4</td></tr>
<tr><td>0.5% population size</td><td>1% population size</td><td>5% population size</td><td>10% population size</td></tr>
<tr><td>  </td><td>  </td><td>  </td><td>  </td></tr>
</table>

100 km buffer scenario			
CKD Stage 3			
0.5% population size	1% population size	5% population size	10% population size
			
CKD Stage 4			
0.5% population size	1% population size	5% population size	10% population size
			

150 km buffer scenario			
CKD Stage 3			
0.5% population size	1% population size	5% population size	10% population size
			
CKD Stage 4			
0.5% population size	1% population size	5% population size	10% population size
			

200 km buffer scenario			
CKD Stage 3			
0.5% population size	1% population size	5% population size	10% population size
			
CKD Stage 4			
0.5% population size	1% population size	5% population size	10% population size