Calculating Distances in Alberta

Geographic Methodology Series No. 4

Prepared by
Health Surveillance Branch
Alberta Health and Wellness
Edmonton, Alberta

First Version: March 2004
Latest Revision: October 2004

Alberta
HEALTH AND WELLNESS

For more information contact:

Health Surveillance
Alberta Health and Wellness
P.O. Box 1360
10025 Jasper Avenue
Edmonton, Alberta
T5J 2P4

Phone: 780-422-4518
Fax: 780-427-1470

Series title: Alberta Health and Wellness Geographic Methodology Series
ISBN: 0-7785-2712-3 (online)
Geographic Methodology Series No. 4
Calculating Distances in Alberta
ISBN: 0-7785-3452-9 (pdf)

Note: Users of the information presented in this document in the analysis of health data must insure that products conform to the Alberta *Health Information Act*.

Executive Summary

This document is part of a series of reports that illustrates and documents the geographic methods required to properly analyze health data in Alberta. The descriptions and methods used are consistent across these reports. Together they provide all the information required to properly understand the spatial component of health data.

This report documents the methodologies currently used to calculate distances.

A number of health issues involve distance calculations. One example is in the determination of access standards for particular levels of health services. This reports provides essential information on alternative methods that can be used and highlights recommended methods relevant to health settings in Alberta. The report concludes with a warning about the use of crow-flies distances for health access calculations.

Acknowledgments

The following team prepared this report:

Alberta Health and Wellness

Erik Ellehoj	Geographic Information System Consultant
Dr. Donald Schopflocher	Biostatistician

The team acknowledges the contributions of the following reviewers in helping to insure completeness and accuracy in the report:

Corrine Truman	Alberta Health and Wellness
Tom de Beyer	Alberta Health and Wellness

Table of Contents

List of Figures

List of Tables

I. Introduction

A number of decisions within a health setting are associated with distance calculations. Distances are used to determine the availability of services to the population (i.e. what percentage of the population is within a certain distance or travel time of a basic care facility?). Distances may also be used to determine facility closures in areas where there may an overlap of services to a small population. Distances are also used in creating funding models. For example, more remote areas may receive additional funding to compensate for the higher transportation costs to facilities.

Most distance calculations are performed within a GIS environment since this is a very time-consuming task to perform manually. However, most GIS are poorly suited to provide a variety of distance estimates without specific add-on software. As a result, the default options are often used. This report seeks to document the methods available to calculate distances and the impact that they have in these estimates.

II. Crow-Fly Distances and Variants

The default option in all GIS is to calculate distances without any barriers in any direction. A travel distance of 50km is simply a circle centered on the starting point with a radius of 50km. There are problems with the assumptions made by this model, and these are discussed at the end of this section.

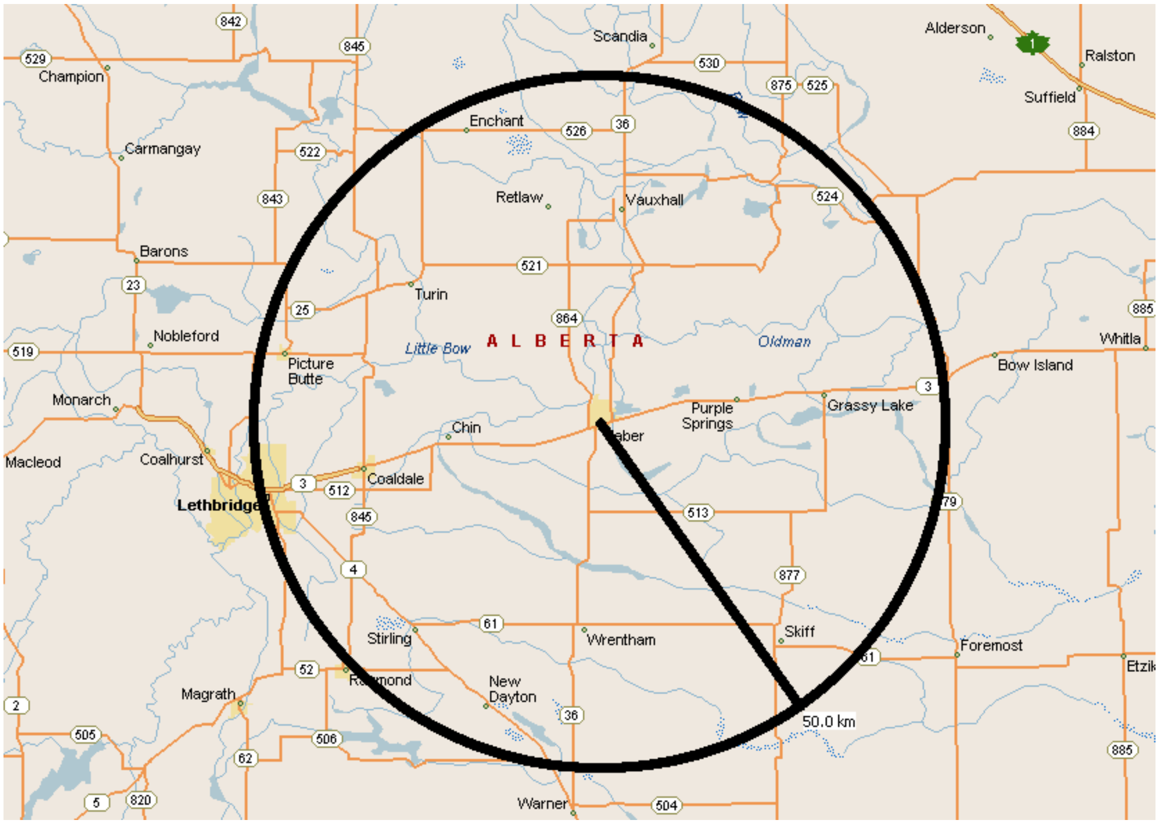

Figure 1: 1:50 km radius from Taber, AB

The manner in which the distance is calculated can vary significantly based on a number of factors. Several GIS systems offer the option to calculate distances based on projected coordinates or on great distance routes. The projected (or cartesian) option assumes that the portion of the world being examined has been projected onto a flat x-y plane. The characteristics of this plane are controlled by the selection of projection. In Alberta, a common projection is the 10 degree Transverse Mercator Projection, which has a maximum error of 0.9992 in the centre of the province (115^{O}). This projection should not be used outside the Province of Alberta. Each province has a commonly used projection and the error associated with distance calculations is proportional to the size of the province. If calculations over longer distances are required, such as across all of Canada, then this option is a very poor choice. A characteristic of the Alberta projection is that errors calculating distance in a North-South direction is minimized. Since the province is long and thin, this projection is relatively well suited to distance calculations. The distances are calculated using the Pythagoras theorem, which is appropriate only when the geographic information has been projected to a flat x-y plane.

The theorem can be stated as:

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

Where X_2-X_1 is the difference in the X axis and Y_2-Y_1 is the difference in the Y axis.

An alternative is the use of spherical methods to calculate distances. These calculations are based on the shortest travel route and are thus better estimates of distance over long distances. However, some of the methods can create difficulty with points that are very near to each other. A full review of available methods appears in Appendix 1: Calculation Methods for Crow-Fly Distances.

A third method is to account for road patterns while still keeping the methodology as simple as possible. In this scenario, the projected coordinates (Alberta modified 10 degree Transverse Mercator projection) for the starting and end points are used using the following formula.

$$abs(X_2 - X_1) + abs(Y_2 - Y_1)$$

Where X_2-X_1 is the difference in the X axis and Y_2-Y_1 is the difference in the Y axis.

The distances calculated correspond to the sum of the two shorter sides of the triangle. This is to account for the fact that many roads in Alberta follow the Township-Range grid and thus are either horizontal or vertical. Only a few roads follow the hypotenuse.

A summary matrix comparing the distances of 40 origins against 40 destinations was created in order to examine the differences between these techniques. The tables below form a subset of this spreadsheet to illustrate the patterns observed in the larger matrix.

Municipality	BANFF	BROOKS	CALGARY	CARDSTON
BANFF	0	266	119	272
BROOKS	266	0	150	183
CALGARY	119	150	0	214
CARDSTON	272	183	214	0
COLD LAKE	506	445	443	619
DRUMHELLER	201	114	91	255
EDMONTON	298	348	274	483
FORT MCMURRAY	674	685	647	846
GRANDE PRAIRIE	491	688	558	759
HIGH LEVEL	820	944	850	1063
JASPER	255	497	350	526
LAC LA BICHE	465	467	427	625
LETHBRIDGE	254	116	171	67
LLOYDMINSTER	444	327	357	507
MEDICINE HAT	366	103	252	211
RED DEER	170	229	129	340
SLAVE LAKE	468	579	485	699
STETTLER	235	204	159	350
WAINWRIGHT	372	263	284	440
WHITECOURT	328	470	358	570

Table 1: Distances calculated using the spherical formula (great circle route)

Municipality	BANFF	BROOKS	CALGARY	CARDSTON
BANFF	0	265	119	270
BROOKS	265	0	151	182
CALGARY	119	151	0	213
CARDSTON	270	182	213	0
COLD LAKE	515	446	448	623
DRUMHELLER	202	114	91	255
EDMONTON	300	349	274	483
FORT MCMURRAY	682	685	650	848
GRANDE PRAIRIE	495	705	567	772
HIGH LEVEL	821	957	855	1071
JASPER	257	506	355	535
LAC LA BICHE	470	467	429	626
LETHBRIDGE	252	116	170	67
LLOYDMINSTER	452	329	362	512
MEDICINE HAT	363	103	250	212
RED DEER	171	230	129	340
SLAVE LAKE	468	584	486	700
STETTLER	237	204	160	351
WAINWRIGHT	377	264	287	443
WHITECOURT	328	476	360	573

Table 2: Distances calculated using a Pythagorean theorem

Municipality	Banff	Brooks	Calgary	Cardston
Banff	0	324	129	377
Brooks	324	0	197	252
Calgary	129	197	0	248
Cardston	377	252	248	0
Cold Lake	728	542	620	800
Drumheller	231	156	122	295
Edmonton	408	443	299	495
Ft McMurray	908	721	800	977
Grande Prairie	666	996	795	1059
High Level	923	1252	1052	1311
Jasper	363	693	492	754
Lac La Biche	648	473	540	715
Lethbridge	355	161	227	92
Lloydminster	620	434	512	693
Medicine Hat	466	143	338	284
Red Deer	242	321	134	373
Slave Lake	494	767	569	823
Stettler	327	253	219	392
Wainwright	514	327	406	584
Whitecourt	332	660	461	716

Table 3: Distances calculated using the road pattern (rectilinear)

Tables 1 and 2 are very similar, despite long distances in some cells. These calculations support the suitability of either method for calculating distances in Alberta (providing that the 10 degree Transverse Mercator projection coordinates were used for the calculation of the Pythagorean version of the distances). The third table, which adds the sides of the triangle, reports longer distances. It is quite inconsistent with the first two tables, however this does not necessarily imply that its use should always be avoided.

The principal advantage of crow-fly distances lies in computational simplicity. It is very easy to create a circle of a given radius in order to determine the region within a given distance of a starting point. The principal disadvantage, however, lies in the fact that travel must generally follow the road network. These distances may not reflect the true accessibility for a given location with unique characteristics. For example, Ft. McMurray has a single highway that connects it to the rest of the province. Any access calculations that reflect a distance of more than 50 km will not be a proper representation of the access of this community. Similarly, Grande Cache can only be accessed through a single highway. There are many other communities in the province that have these characteristics. (Apparently, even birds follow road pattern.. A summary of research to that effect from Oxford, UK is reprinted in Appendix 2.)

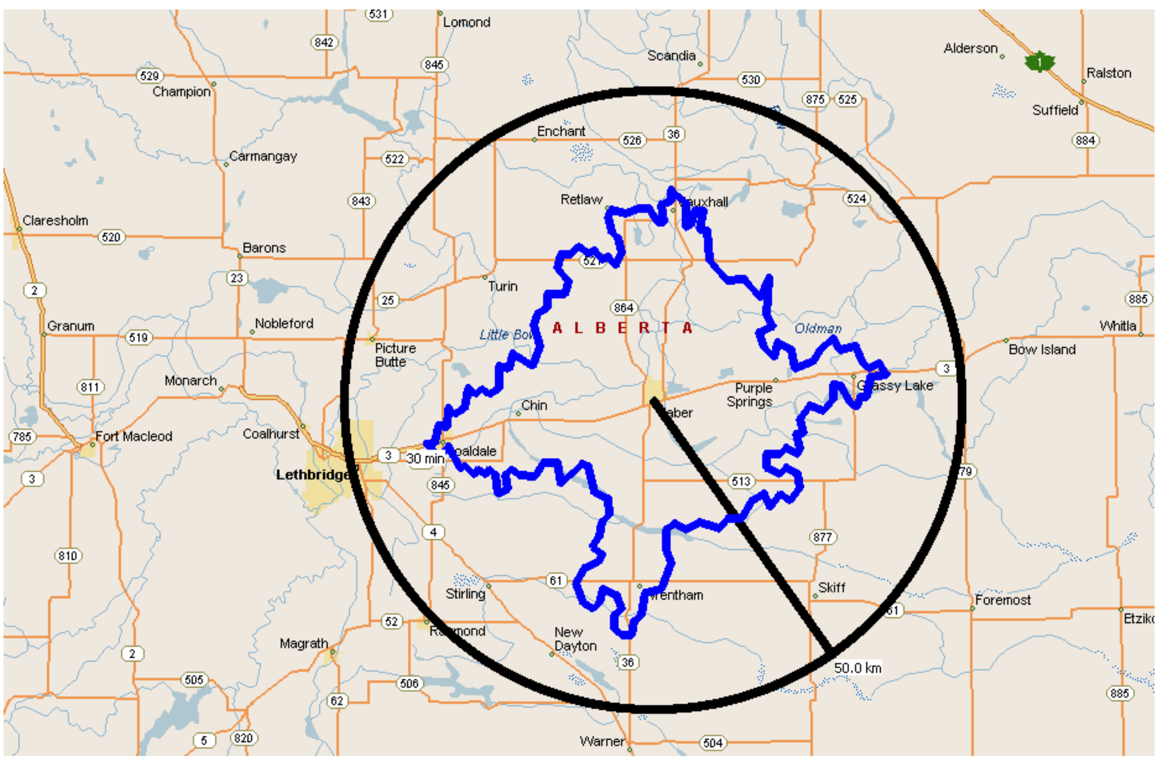

Figure 2: 50 km radius and 30 minute drive-time zone

Figure 2 clearly illustrates the difficulties associated with the use of crow-fly distances in the context of a road network. In this case, the community of Taber serves as an example of a community with few travel restrictions in any direction. However, in reality, roads vary due to the use of construction materials, and road width. These factors determine the maximum speed at which the roads may be safely traveled. In the example above, a round circle is very different than the rough rhomboid shape of the drive-time zone (drive-time zones are explained below). In circumstances where road access is not excellent in all directions, as in the case of the presence of rivers or other barriers, the contrast with a simple circle is even greater.

III. Road Network Distances

The availability of data and appropriate software has restricted the use of road network distances. The road network does, however, present the real travel distances on the existing road network. As well, any gaps in the road network will be highlighted by the calculations performed using the network.

Add-on software modules are required of most GIS systems to allow road network distance calculations (e.g. ArcView, ArcGIS, MapInfo and some other platforms). Road network geographic files are also required. (Special care is needed with these files to ensure that all roads are perfectly connected at the intersections. A barrier is assumed to be present if there are any gaps or overlaps are present at any intersection).

Table 4 presents road network distance information for the same communities as tables 1,2, and 3.

Municipality	Banff	Brooks	Calgary	Cardston
Banff	0	313	129	365
Brooks	313	0	187	230
Calgary	129	187	0	240
Cardston	365	230	240	0
Cold Lake	730	545	608	837
Drumheller	260	142	139	318
Edmonton	422	418	299	528
Ft McMurray	872	802	747	976
Grande Prairie	685	928	747	976
High Level	1119	1223	1040	1272
Jasper	292	600	413	648
Lac La Biche	640	512	520	750
Lethbridge	342	154	219	77
Lloydminster	667	445	546	661
Medicine Hat	416	109	290	242
Red Deer	269	326	149	379
Slave Lake	666	726	546	776
Stettler	350	241	230	459
Wainwright	538	338	418	552
Whitecourt	512	648	469	698

Table 4: Distances calculated using the road network

To review, the first two tables were very similar to each other, but the third table reported values greater than for either of the first two tables. This fourth table is far more similar to the third table than to the first two. A number of discrepancies are readily apparent, however. This occurs because the road network analysis takes into account transportation barriers while the rectilinear distance calculations assumes that there is always two roads on a rectilinear grid to link two communities.

IV. Travel Time Analysis

Most of the add-on GIS modules for road network analysis allow for the calculation of the quickest route as well as the shortest route (as long as the speed limit has been entered for every road segment in the network).

For the analysis presented in this report, the following speed limits were used:

Large highways:	110 km/hr
Other primary highways:	100 km/hr
Secondary paved highways:	80 km/hr
Arterial Roads:	60 km/hr
Streets:	50 km/hr

As well, an addition of 20 seconds was made to the time calculations at every level road crossing to account for slowing down and/or waiting at stop signs or traffic signals. These assignments result in a conservative estimate. This is regarded as positive since these estimates must not just reflect travel time under ideal circumstances, but rather should reflect average (or even worst-case) conditions. nt.

The following figures illustrate the differences in road selection when choosing a shortest route vs. a quickest route.

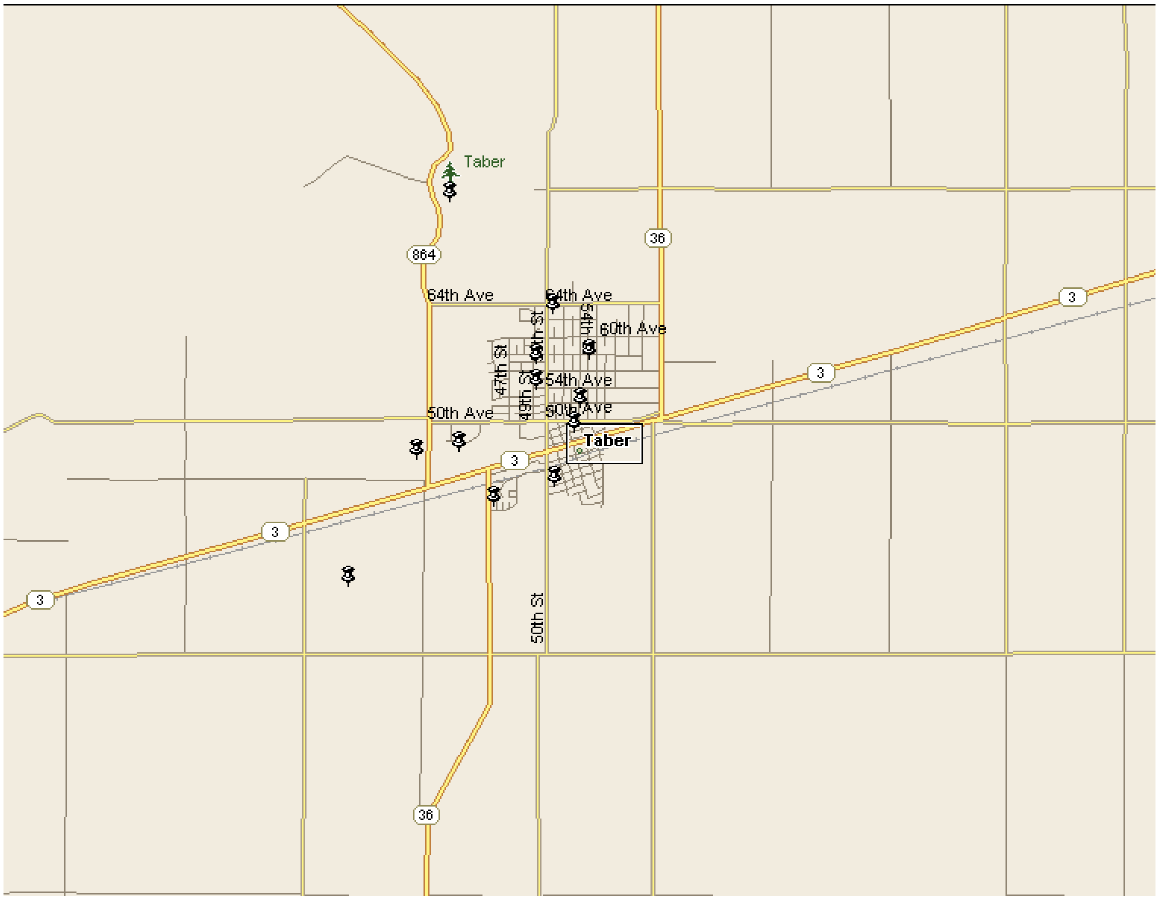

Figure 3: Road network near Taber, AB

In Figure 3, the pushpins are Enumeration Area (EA) locations (the basis for assigning postal codes to points). All streets and roads (major and minor) are used as part of the network. Speed limits have been assigned to ALL roads. The road network examined includes all the streets in all the communities as well as the highways and township-range roads.

An analysis was performed in order to find the shortest route between Taber and Brooks.

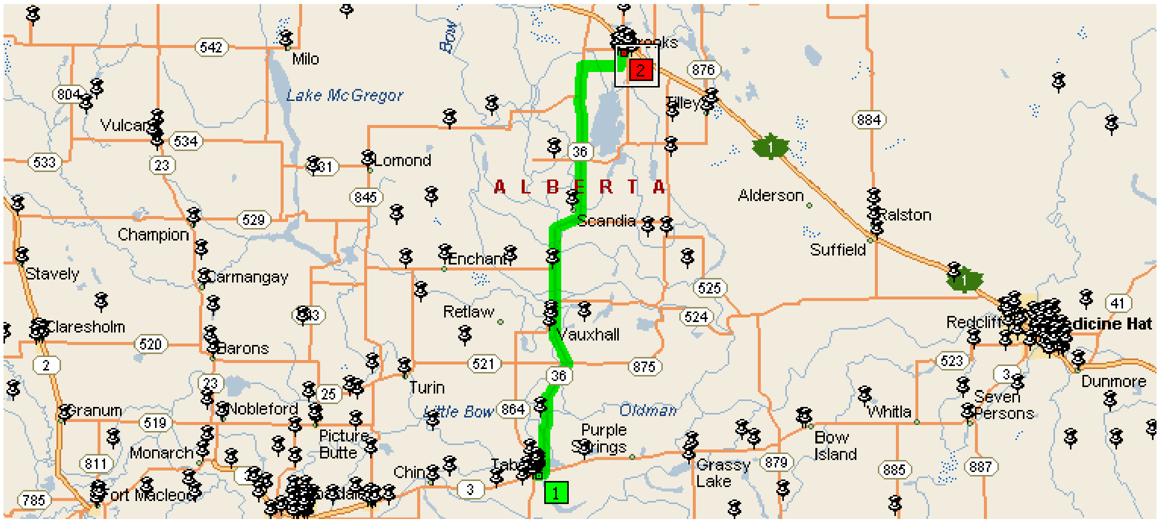

Figure 4: Shortest route between Taber and Brooks

The result of the analysis is presented in Figure 4 where it highlights the roads chosen for the shortest route. The total distance travelled was 102.7 km requiring 1 hr, 25 min of travel time.

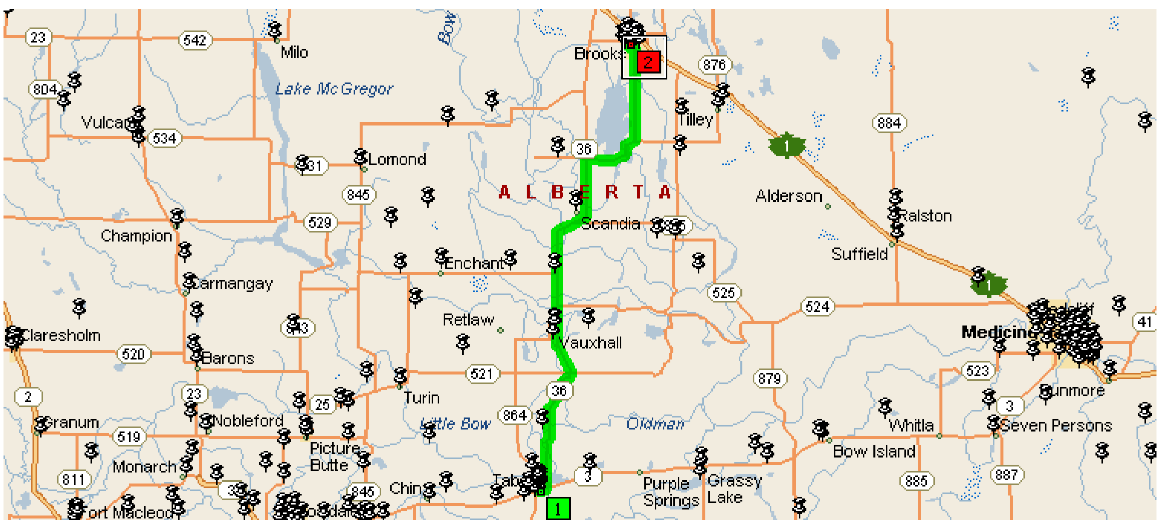

Figure 5: Quickest route between Taber and Brooks

The quickest route is similar to the shortest route, except that a road south of Lake Newell is chosen instead of a road north of the same lake. The total distance travelled was 102.9 km requiring 1 hr, 18 min of travel time. The net difference is travel distance is only 200 metres, but the difference in travel time is 7 minutes. The difference in travel time can be attributed to the differences in speed limits between the roads south and north of Lake Newell. In this case, the model is quite stable and several other road options would result in very similar results.

V. Isochrones (drivetime zones)

Any analysis of access to services will require an analysis of distances from all possible sources to all possible service centres. Within a health context, this results in a matrix of at least 112000 distance pairs (800 origins against 140 destinations). An alternative is to create isochrones (drivetime zones) around each of the service centres. GIS analysis can then be used to determine the status of each origin point against the service isochrones.

In a simple scenario, a speed of 80 km/hr could be assumed to create a circle around Taber. Figure 6 shows the 15 minute isochrone for Taber (with a starting point at the intersection of Hwy 3 and Hwy 36).

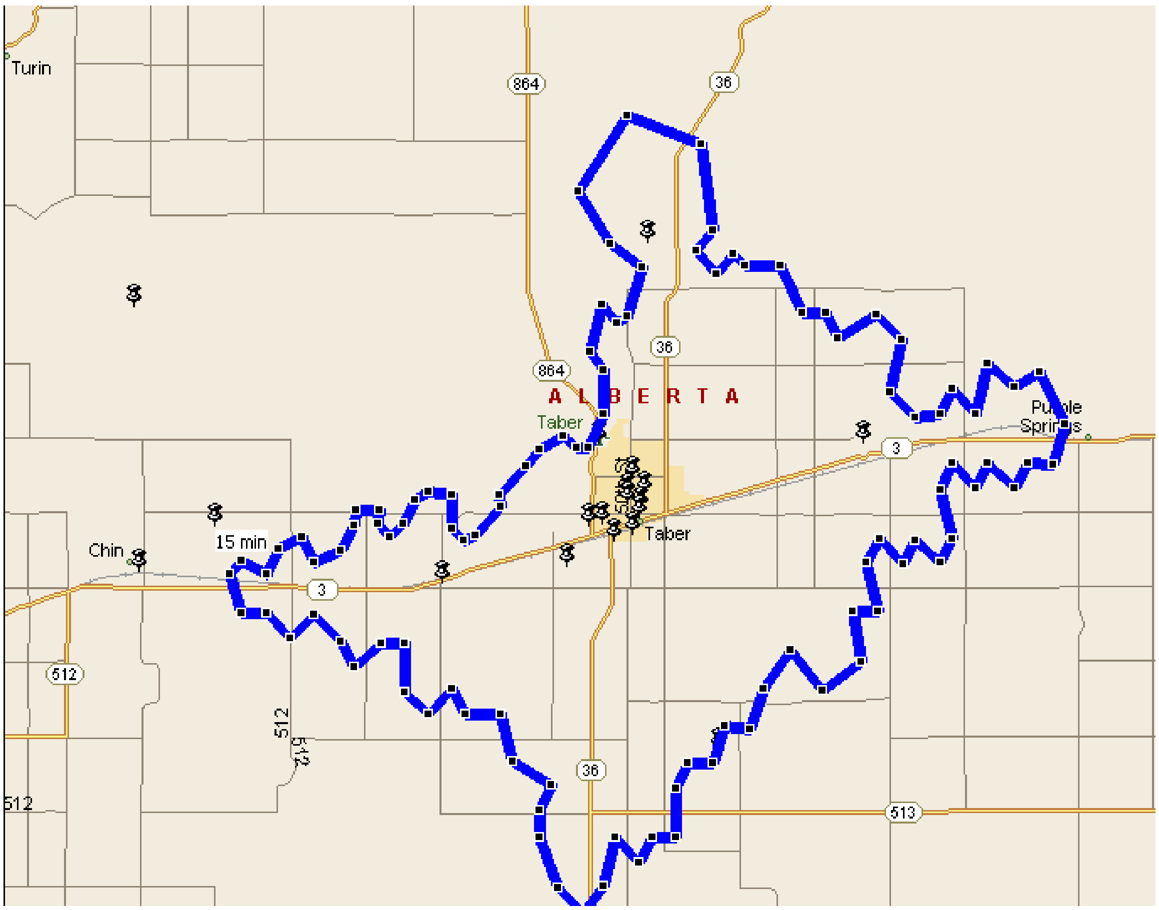

Figure 6: 15 minute isochrone (drivetime zone) from Taber, AB

Note the compressed shape in the northwest portion. This compression exists because the virtual driver must cross the entire town at slow speed with a large number of intersections. The shape stretches on the main highways, such as #3 and #36 due to their higher speed limits.

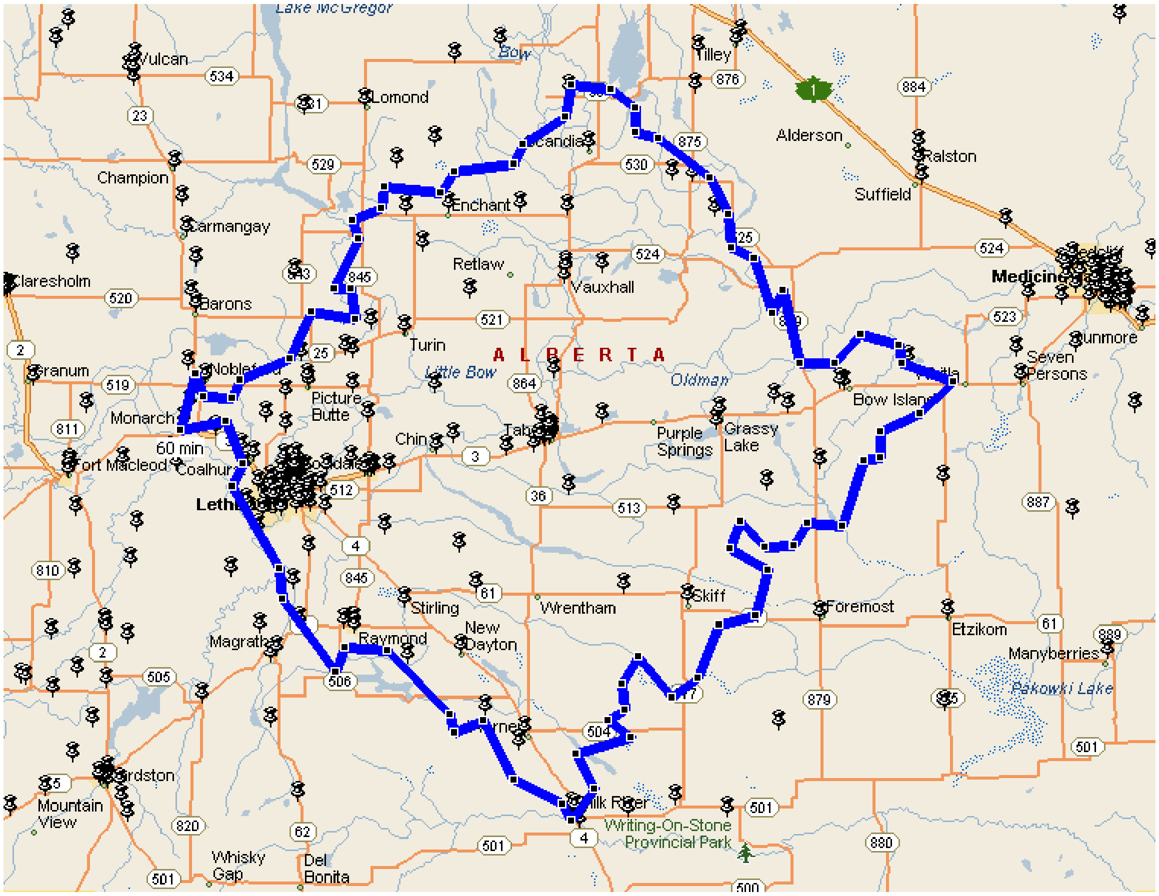

Figure 7: 1 Hour isocrone from Taber, AB

As the drive-time is increased to 60 minutes, the isochrones tend to take a diamond-shape (especially in more densely populated areas). Figure 7 shows the 1 hour isochrone around Taber. It shows a compression on the west side which is a result of crossing through the metropolitan area of Lethbridge. As well, the main highway (#3) veers north and then west again in this region. A quick glance of the shape of the isochrones provides visual evidence of the road connectivity near each centre in every direction.

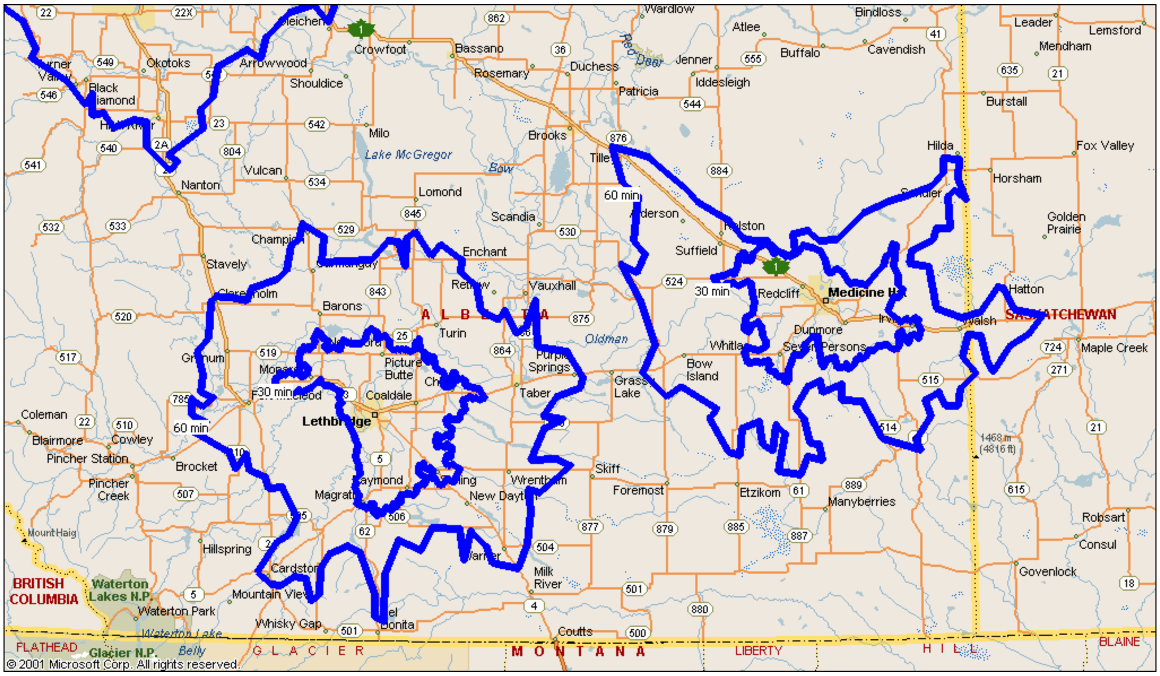

Figure 8: 30 and 60 minute isochrone in Southern Alberta

In figure 8 Isochrones around Lethbridge (west) and Medicine Hat (east) are shown. It is immediately apparent that road connectivity is poor north of Medicine Hat where the Suffield weapons testing range is responsible for a discontinuity in the road network. The most important highway in and out of Medicine Hat is the TransCanada (#1) highway which is responsible for bending the isochrone toward the northwest. Lethbridge also shows some differences away from a true diamond shape as the faster routes toward Calgary are on highway #2 and #23, which are to the northwest of the city. A series of lakes (McGregor Lake and Travers Reservoir) prevent travel in directions purely north from the City of Lethbridge.

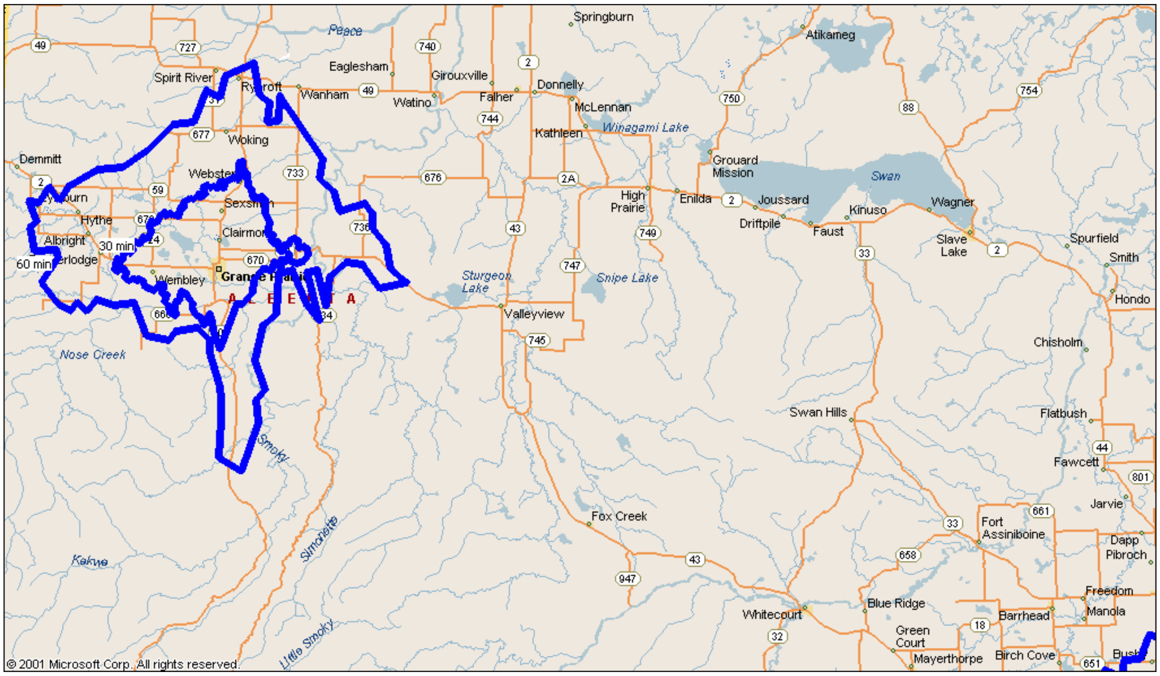

Figure 9: 30 and 60 minute isochrones in Northwest Alberta

Figure 9 presents isochrones around Grande Prairie in Northwestern Alberta. It is clear that road connections are better north of Grande Prairie because agricultural land is present in that direction. Highway #40 provides the main connection south (to Grande Cache and Hinton) which appears as a stretched corridor instead of a diamond.

The use of isochrones provides an efficient method to determine the travel time estimates from service centres to large numbers of potential client communities.

VI. Software Choice

A set of 300 community pairs (representing a variety of distance ranges) was used to test the results obtained from some of the available software packages. The street network file for Alberta was obtained from Health Canada's Spatial Data Warehouse, and the file was created by DMTI, based on Statistics Canada's street network file. The distance calculations obtained from ArcView 3.2 (with the network analysis add-on module) were compared against the distances obtained from performing the same calculations using PCI's SPANS 7.0. The same speed limits and waiting times were used and the same road network file was used. The results were almost identical.

In order to determine if a different road network file would provide a different set of answers, the distances between the 300 community pairs were calculated using a different road network file. The analysis was performed using Microsoft's MapPoint 2001 (which includes its own road network file). The results were almost identical once the same speed limits and waiting periods were entered into MapPoint. The analysis of road network distance calculations indicates that software choice has little effect on the results obtained from the analysis.

However, any analysis of access to services will require an analysis of distances from all possible sources to all possible service centres. Within a health context, this results in a matrix of at least 112000 distance pairs (800 origins against 140 destinations). An alternative is to create isochrones (drivetime zones) around each of the service centres. GIS analysis can then be used to determine the status of each origin point against the service isochrones. Some of the GIS software add-ins also provide this functionality. The results obtained using MapInfo and the ArcView 3.2 network analysis were compared.

Isochrones (drivetime zones) were calculated using MapPoint 2001 and the results for a set of 20 communities by 20 communities were compared with the results obtained from the ArcView network analysis module. As expected, the results were almost identical. For remote communities, MapPoint was far more conservative and required more intervention than ArcView. MapPoint was selected to do all the isochrone (drivetime) zone calculations for this very reason. Human intervention was required for the more remote communities, thus allowing for an examination of the type of road (gravel, ice road, etc) in order to calculate the drivetime zones. Wabasca is an example of one of these remote communities.

VII. Conclusion

In Alberta, travel times should be calculated using the road network and the assigned speed limits. This procedure closely resembles the real travel patterns of humans (and birds –see appendix 2). The use of simple circles to calculate "crow flies" distances should be avoided unless no other means of obtaining distances is available. The choice of software has little influence on the results obtained.

References

Ellehoj, E.A., Schopflocher, D, et al (2004). *Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas Boundaries-Geographic Methodology Series No. 1*.

Ellehoj, E.A., Schopflocher, D, et al (2004). *Using the AH&W Lookup File to Aggregate Data to Different Geographic Boundaries-Geographic Methodology Series No. 2*.

Ellehoj, E.A., Schopflocher, D, et al (2004). *Health Rate Mapping Template-Geographic Methodology Series No. 3*.

Ellehoj, E.A., Wilson, S, et al (2005). *Definition of Rural and Urban Areas for Health Services Provision-Geographic Methodology Series (in preparation)..*

Ellehoj, E.A., Schopflocher, D, et al (2005). *Geographic Activities in AH&W - Geographic Methodology Series (in preparation)..*

Health Trends in Alberta, 2001. Alberta Health Surveillance, Alberta Health and Wellness. ISSN 1480-6657

Appendix 1:

This appendix consists of an edited summary of methods that can be used to calculate crow-fly distances. It was originally prepared by Robert G. Chamberlain of Caltech (JPL), <rgc@jpl.nasa.gov>, and reviewed on the comp.infosystems.gis newsgroup in Oct 1996. It was revised in November 1997 and February 1998.

The inverse tangent formula and the pointer to Ed Williams' formulas were added in November 1997. Mikael Rittri, <Mikael.Rittri@carmenta.se>, provided further enhancements to the formulae. Steven Michael Robbins' discussion of distance computation on an ellipsoid was added in May 1999.

Calculating distances between two points (summary)

The Earth is round, but big, so we can consider it flat for short distances. But, even though the circumference of the Earth is about 40,000 kilometers, flat-Earth formulas for calculating the distance between two points start showing noticeable errors when the distance is more than about 20 kilometers. Naturally, allowable error depends on its intended application.

If many distances are going to be computed, the amount of calculation has to be considered. Cartesian coordinates express distances in two different directions, such as north-south for one direction and east-west for the other. The straight line distance between two points can then be thought of as the long side of a right triangle with one of the short sides being the north-south distance between the points and the other being the east-west distance. (A right triangle is one that has a square corner.) The usual formula for computing the length of the long side of a right triangle is the Pythagorean Theorem. Using this formula from geometry requires knowing about square roots.

Near the North Pole and near the South Pole, the longitude lines, which appear in a north-south direction and are called the meridians, approach each other noticeably - in fact, they meet at the pole. The latitude lines, which appear in an east-west direction, are circles around the pole. Treating differences in locations along these directions as if they were the sides of a right triangle leads to errors in the computation of distance. Very near the pole, the answer could be incorrect - but a different flat-Earth approximation, obtained from plane trigonometry, can be used for short distances: the Polar Coordinate Flat-Earth Formula.

Latitude and longitude are spherical coordinates, based on recognition that the Earth is round. Their definition does not require that the Earth be exactly spherical, but approximating the Earth as a sphere is satisfactory for most needs. Map projections are used to convert from spherical coordinates to flat (Cartesian) coordinates.

The Law of Cosines for Spherical Trigonometry appears as a very suitable candidate for calculating distances on a sphere, however it is not suitable for very short distances on these surfaces. The problem is as follows: Suppose you have a right triangle with a very small angle. The ratio between the short side and the long side is very close to 1.0 (they are almost the same length). The formula computes that ratio first, then requires the computer to find the angle that has that ratio. In principle, the computer can do so - after all, the formula is mathematically correct - but ordinary computers approximate all numbers to a certain number of significant digits. With 7 or 8 significant digits, the computer cannot distinguish between the ratios for angles smaller than about a minute of arc (a minute is 1/60 of a degree). Since the angle being computed has its apex at the center of the Earth, a minute of arc corresponds to almost 2 km on the surface.

Since the formula is mathematically correct, it can be manipulated into other forms. The Haversine Formula is one result of such manipulations. It has a similar problem, but it is "poorly conditioned" when the two points are all the way around the Earth from each other, rather than when they are close to each other. The discussion below gives a second version of the haversine formula that is easier to program on some computers.

Calculating distances between two points (detailed discussion)

The distances considered here are along the surface of the Earth, deliberately ignore the effect of differences in elevation. Distances on the surface of the terrain, whether geodesic, on roads, or cross-country, depend on relief (including elevation differences), the status of engineering projects, and perhaps even route selection. Hence, computation is idiosyncratic and not well suited to simple approximations. If the distance is less than about 20 km and the locations of the two points in Cartesian coordinates are X_1, Y_1 and X_2, Y_2 then the Pythagorean Theorem

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

will require the least amount of computation and will be in error by

- less than 30 meters for latitudes less than 70 degrees
- less than 20 meters for latitudes less than 50 degrees
- less than 9 meters for latitudes less than 30 degrees

(These error statements reflect both the convergence of the meridians and the curvature of the parallels. The error is non-linear with distance; shorter distances will have better percentage errors.)

The flat-Earth distance d will be expressed in the same units as the coordinates.

If the locations are not already in Cartesian coordinates, the computational cost of converting from spherical coordinates and then using the flat-Earth model may exceed that of using the more accurate
spherical model.

Otherwise, presuming a spherical Earth with radius R (see below), and the locations of the two points in spherical coordinates (longitude and latitude) are lon_1,lat_1 and lon_2,lat_2 then the

Haversine Formula (from R.W. Sinnott, "Virtues of the Haversine", Sky and Telescope, vol. 68, no. 2, 1984, p. 159):

dlon = lon_2 - lon_1
dlat = lat_2 - lat_1

$$a = (\sin(\frac{dlat}{2}))^2 + \cos(lat_1) * \cos(lat_2) * (\sin(\frac{dlon}{2}))^2$$

$$c = 2 * \arcsin(\min(1, \sqrt{a}))$$

d = R * c

will give mathematically and computationally exact results. The intermediate result c is the great circle distance in radians. The great circle distance d will be in the same units as R.

When the two points are antipodal (on opposite sides of the Earth), the Haversine Formula is ill-conditioned (see the discussion below the Law of Cosines for Spherical Trigonometry), but the error, perhaps as large as 2 km, is in the context of a distance near 20,000 km. Further, there is a possibility that round-off errors might cause the value of sqrt(a) to exceed 1.0, which would cause the inverse sine to crash without the bulletproofing provided by the min() function.

Most computers require the arguments of trigonometric functions to be expressed in radians. To convert lon1,lat1 and lon2,lat2 from degrees, minutes, and seconds to radians, first convert them to decimal degrees. To convert decimal degrees to radians, multiply the number of degrees by pi/180 = 0.017453293 radians/degree.

Inverse trigonometric functions return results expressed in radians. To express c in decimal degrees, multiply the number of radians by 180/pi = 57.295780 degrees/radian. (But be sure to multiply the number of RADIANS by R to get d.)

The Haversine Formula can be expressed in terms of a two-argument inverse tangent function, atan2(y,x), instead of an inverse sine as follows (no bulletproofing is needed for an inverse tangent):

dlon = lon2 - lon1
dlat = lat2 - lat1

$$a = (\sin(\frac{dlat}{2}))^2 + \cos(lat_1) * \cos(lat_2) * (\sin(\frac{dlon}{2}))^2$$

$$c = 2 * \arctan(\sqrt{a}, \sqrt{(1-a)})$$

d = R * c

The problem of determining the great circle distance on a sphere has been around for hundreds of years, as have both the Law of Cosines solution (given below but not recommended) and the Haversine Formula.

The Pythagorean flat-Earth approximation assumes that meridians are parallel, that the parallels of latitude are negligibly different from great circles, and that great circles are negligibly different from straight lines. Close to the poles, the parallels of latitude are not only shorter than great circles, but indispensably curved. Taking this into account leads to the use of polar coordinates and the planar law of cosines for computing short distances near the poles:

The Polar Coordinate Flat-Earth Formula

$$a = \frac{\pi}{2} - lat_1$$

$$b = \frac{\pi}{2} - lat_2$$

$$c = \sqrt{(a^2 + b^2 - 2 * a * b * \cos(lon_2 - lon_1))}$$

d = R * c

is computationally only a little more difficult than the Pythagorean Theorem and will give smaller maximum errors for higher latitudes and greater distances. The maximum errors, which depend upon azimuth in addition to separation distance, are equal at 80 degrees latitude when the separation is 33 km, 82 degrees at 18 km, 84 degrees at 9 km. But even at 88 degrees the polar error can be as large as 20 meters when the distance between the points is 20 km.

The latitudes lat_1 and lat_2 must be expressed in radians (see above); pi/2 = 1.5707963. Again, the intermediate result c is the distance in radians and the distance d is in the same units as R.

An UNRELIABLE way to calculate distance on a spherical Earth is the Law of Cosines for Spherical Trigonometry ** NOT RECOMMENDED **

$a = \sin(lat_1) * \sin(lat_2)$
$b = \cos(lat_1) * \cos(lat_2) * \cos(lon_2 - lon_1)$
$c = \arccos(a + b)$
$d = R * c$

Although this formula is mathematically exact, it is unreliable for small distances because the inverse cosine is ill conditioned. The following calculations illustrate the point:
cos (5 degrees) = 0.996194698
cos (1 degree) = 0.999847695
cos (1 minute) = 0.9999999577
cos (1 second) = 0.9999999999882
cos (0.05 sec) = 0.999999999999971
A computer carrying seven significant figures cannot distinguish the cosines of any distances smaller than about one minute of arc.

What is the radius of the earth, R?

The historical definition of a "nautical mile" is "one minute of arc of a great circle of the earth". Since the earth is not a perfect sphere, that definition is ambiguous. However, the internationally accepted (SI) value for the length of a nautical mile is (exactly, by definition) 1.852 km. Thus, the implied "official" circumference is 360 degrees times 60 minutes/degree times 1.852 km/minute = 40003.2 km.
The implied radius is the circumference divided by 2 pi:

R = 6367 km

When must the flatness of at the poles be considered?

A quick test is to compare the results produced by using the two extreme values of the radius of curvature for the Earth:

minimum radius of curvature: 6336 km
maximum radius of curvature: 6399 km

If the results are different enough to cause differences large enough to affect the results obtained from the calculations (unlikely in a health setting, more likely in a surveying setting) then assuming the Earth is spherical is not appropriate.

The shape of the Earth is well approximated by an oblate spheroid. The radius of curvature varies with direction and latitude. According to formulas given on pages 24 and 25 of the book by Snyder, "Map Projections - A Working Manual", by John P. Snyder, U.S. Geological Survey Professional Paper 1395, United States Government Printing Office, Washington DC, 1987,

the radius of curvature of an ellipsoidal Earth in the plane of the meridian is given by

$$R' = a * \frac{1 - e^2}{1 - e^2 * (\sin(lat))^2)^{\frac{3}{2}}}$$

where a is the equatorial radius,
b is the polar radius, and
e is the eccentricity of the ellipsoid

$$e = \sqrt{\frac{1 - b^2}{a^2}}$$

and the radius of curvature in a plane perpendicular to the meridian and perpendicular to a plane tangent to the surface is given by

$$N = \frac{a}{\sqrt{1 - e^2 * (\sin(lat)^2))}}$$

A Swedish book by Ilmar Ussisoo, Kartprojektioner [map projections] (published by the National Land Survey, Sweden, Professional papers 1977/6) suggests use of the geometric mean of these two radii of curvature for all azimuths. The use of these produces errors of order of magnitude 0.1% for distances within 500 km at 60 degrees latitude.

The formula for that average is no more complicated than either of its components. That is,

$$r = \sqrt{R' * N}$$

becomes

$$r = \frac{a * \sqrt{1 - e^2}}{1 - e^2 * (\sin(lat))^2}$$

Using these formulas with

a = 6378 km Equatorial radius (surface to center distance)
b = 6357 km Polar radius (surface to center distance)
e = 0.081082 Eccentricity

gives the following table of values for the

Radii of Curvature:

Latitude	r	R'	N
00 degrees	6357 km	6336 km	6378 km
15 degrees	6360 km	6340 km	6379 km
30 degrees	6367 km	6352 km	6383 km
45 degrees	6378 km	6367 km	6389 km
60 degrees	6388 km	6383 km	6394 km
75 degrees	6396 km	6395 km	6398 km
90 degrees	6399 km	6399 km	6399 km

Note that the radius of curvature for an ellipsoid is not the same as the distance from the surface of the ellipsoid to the center. In fact, the radius of curvature increases as the radius decreases. Also, be aware that a variety of ellipsoids with slightly different parameters have been fit to the Earth; the preferred ellipsoid may depend on the region in which you are most interested.

Also note that spherical earth computations will provide underestimates of real world distances measured in the direction of the equator (and especially for trans-equatorial links) and overestimates for those measured in the direction of the poles (and especially for trans-polar ones).

For most purposes, it is quite satisfactory to treat the Earth as a sphere. If not, an ellipsoid can provide a better approximation. Some standard textbooks that may be helpful appear below:

Bomford, Guy 1980 Geodesy Clarendon Press, Oxford ISBN 0-19-851946-X

Vanicek, Petr, and Krakiwsky, Edward 1986 Geodesy, the Concepts North-Holland, Amsterdam
ISBN 0-444-87775-4

Torge, Wolfgang 1980 Geodesy de Gruyter, Berlin (translated to English by C. Jekeli)
ISBN 3-11-007232-7

Software for solving distance and azimuth problems on the ellipsoid can be obtained by anonymous ftp from several sources, two of which are listed below:

The URL of the National Geodetic Survey (of the National Oceanic and Atmospheric Administration in the US Department of Commerce) is: ftp://www.ngs.noaa.gov/pub/pcsoft/for_inv.3d/

See the read.me file for explanations. The NGS software directory may contain other listing of interest. Its URL is: http://www.ngs.noaa.gov/PC_PROD/pc_prod.shtml/ (case sensitive)

The NGS provides FTP access at: ftp://ftp.ngs.noaa.gov/pub/pcsoft

Another anonymous ftp source for ellipsoid software is the US Geological Survey (of the US Department of the Interior), at: http://kai.er.usgs.gov/pub/Proj.4/

Again, see the readme.txt file for explanations. The URL for the USGS home page is: http://kai.er.usgs.gov/homepage.html

Appendix 2:

Pigeons take the high road
From The Times, February 06, 2004

LONDON: The secret of the carrier pigeon's uncanny ability to find its home coop has been revealed by British scientists: they do it by following roads.
When the birds are released far from home they navigate back in remarkably similar fashion to their human owners - choosing the trunk routes recommended in road atlases, a major satellite tracking study has shown.

Homing pigeons often cruise down a motorway before turning onto city ring roads and leaving at major junctions, even when such a route adds kilometres to their journey.

Pigeons hardly ever travel as the crow flies, preferring to take the easy option of following the roads, even when it involves much greater physical exertion.

Just like drivers, they select straight main roads rather than twisting country lanes, choosing economy of thought over fuel efficiency.

The findings, from a team at Oxford University, indicate homing pigeons do not always navigate by taking bearings on the sun, as had been previously believed, but instead seek out routes that make the journeys less taxing to work out.

"It really has knocked our research team sideways to find pigeons appear to ignore their inbuilt directional instinct and follow the road system," said Tim Guilford, Professor of Zoology, who led the study.

In the study, Professor Guilford and his colleague, Dora Biro, attached miniature global positioning satellite tracking devices, each weighing just 18 grams, to homing pigeons. These were then released up to 32km away from their home coops in Oxfordshire.

While the birds initially used the sun to get their bearings, they rapidly learned the layout of the road network and used it as a guide to getting home.

Different pigeons developed different favourite routes, but all of them tended to follow linear features on the landscape wherever possible - roads, railway lines, hedgerows and rivers.

"It is striking to see the pigeons fly straight down the A34 Oxford bypass, and then sharply curve off at the traffic lights before curving off again at the roundabout," Professor Guilford said.