Author's response to reviews

Title: Spatial variation in the management and outcomes of acute coronary syndrome

Authors:

Alain Vanasse (alain.vanasse@usherbrooke.ca)
Theophile Niyonsenga (theophile.niyonsenga@chus.qc.ca)
Josiane Courteau (josiane.courteau@chus.qc.ca)
Jean-Pierre Gregoire (jeau-pierre.gregoire@pha.ulaval.ca)
Abbas Hemiari (abbas.hemiari@chus.qc.ca)
Julie Loslier (julieloslier@hotmail.com)
Goze B Benie (Goze.bertin.benie@usherbrooke.ca)

Version: 3 Date: 26 May 2005

Author's response to reviews: see over
Sherbrooke, May 26th 2005

BMC Cardiovascular Disorders
Submission of the revised manuscript: « Regional variation in the management and outcomes of acute coronary syndrome ».

To the editor,

On behalf of my colleagues, I am pleased to submit this revised manuscript. We want to thank the reviewers for their valuable and pertinent comments. We seriously considered each comment. We are confident that this revised manuscript will be at their satisfaction and will meet the high standard quality of your journal. You will find below a point-by-point description of the changes made. Should you have any question, please feel free to e-mail the first author. We are looking forward to hearing from you.

Alain Vanasse, MD, PhD
Department of Family Medicine
Faculty of Medicine
Université de Sherbrooke

RESPONSES TO THE REVIEWER BRAHMAJE NALLAMOTHU:

Minor Essential Revisions
1. Another explanation for the lack of association between distance and in-hospital death might be survival bias. That is, individuals in rural areas with ACS may be less likely to survive to hospitalization. Sicker patients in urban areas would be able to make it to the hospital. The inability to adjust for co-morbidities and acute presentations makes this a real possibility in this dataset. The authors should consider mentioning this in their discussion also.

   We totally agree with this comment and this was added to the discussion.

2. Can the authors expand their legend for Figure 4 to include descriptions of the x, y and z axes?

   The x and y axes represent longitude and latitude. The z axe represents the local $\beta$ parameters corresponding to the local relationships between ICP and the log-transformation of length of stay. Since dependent and independent variables are
often denoted by x and y in statistical models, longitude and latitude were denoted by u and v in Figure 4. This is now better indicated in the figure.

RESPONSES TO THE REVIEWER JOANNA MOSCHANDREAS:

As it stands, the statistical analysis appears inadequate in certain respects, and revision is required. It is essential that the three points provided below are addressed:

1. The regression models (logistic and linear regression). In the Discussion (pg 15) and the Conclusion (pg 18-19) it is stated that the main findings of the study are essentially those presented in Table 2. In the corresponding statistical models, however, no 2nd order interactions were considered. The authors themselves note that the paradoxical finding between LoS, ICP and proximity may be explained by interaction between the covariates. These models need to be refitted, including interaction terms. If interactions are present, then the main effects cannot be interpreted so the conclusions resulting from main effects analyses (presented in Table 2) may be misleading. In addition, the residuals will obviously change if the models change. These analyses need to be rerun initially including 2nd order interaction terms in all models.

   As suggested by the reviewer, we allowed the models to integrate the interaction terms gender x ICP, Gender x distance and ICP x distance, if they were statistically significant at a 0.05 level. As expected, the interaction term ICP x distance appears significant in the LoS model, and thus Table 3 has been removed as the information is provided in Table 2. Table 2 now presents the \( \beta \) parameters together with their 95\% confidence intervals instead of the odd ratios. All analyses have been rerun including statistically significant interaction terms, and residuals have been recalculated.

2. The hierarchical cluster analysis (HCA) undertaken. In the application of HCA to group the standardized ratios in the 16 Quebec regions it is not clear which proximity measure was used (for the distance matrices), why the choice of four clusters was made for each of the ICP, HD & EHR ratios or if, in fact, there was any statistical basis to the approach. The last sentence of paragraph 1 on page 8 (‘This method was used instead of …’) needs to be clarified. It would appear preferable not to perform the HCA, and simply to present the ratios in each region in Figure 2.

   The cartographic representation of continuous data often imposes a succession of choices which affect the results obtained. The conversion of continuous to discrete data is an important step in cartography. This ‘discretization’ reduces a part of the
information contained in the data, but results in a more readable and interpretable map. In most of the methods for converting continuous to discrete data (equally spaced, quintiles, natural breaks, mean and standard deviation method, etc), the number of groups must be *a priori* fixed, and the choice of the method often depends on the distribution of the data (*Rican S. Mapping epidemiological results: The principles behind methods to make continuous data discrete and their importance in cartographic representation. Cahier Santé 1998; 8:461-70*). We chose to use a statistical cluster analysis, the HCA, instead of usual grouping methods, because HCA can be used regardless of the data distribution and because we can choose the appropriate number of clusters after a graphical analysis of the dendrogram. We also think that the HCA method is appropriate in our particular study and results in additional information since regions within a same cluster resemble and regions in different clusters differ from one another.

Although we can understand the concern of the reviewer, we are convinced that breaking categories with HCA brings valuable information based on sound statistical methods.

3. **GWR and binary response.** For the logistic regression models, it was not possible to statistically assess spatial variation using the package GWR. The ad-hoc approach used here cannot lead to any conclusions about spatial variability or the lack of it (pg 12). The title does not adequately reflect the content of the manuscript (as spatial heterogeneity is only statistically assessed for LoS).

By using the GWR approach, the focus was not to assess spatial variability in dependent variables LoS, ICP, HD and EHR, but instead assess the spatial variability in the RELATIONSHIPS (whose magnitude is quantified by the regression coefficients, the β parameters) between dependent variables of interest and covariates or potential predictors. Thus, the spatial variability in the dependent variables is not compromised by the fact that the GWR approach did not allowed us to use a statistical test in the local parameters of the logistic GWR models. In fact, the spatial or regional variability in the dependent variables is assessed by the Pearson χ² test for comparisons between regional proportions and the Fisher F-statistics for comparisons of regional means (ANOVA after a logarithmic transformation). We could also have used regression models with the variable “region” as the independent variable to capture the regional heterogeneity or geographic coordinates as independent variables to capture some form of spatial heterogeneity or trends such as “gradients”. We clarified and added the
following information regarding the GWR approach in the ‘Statistical Analyses’ section:

The GWR approach extends the traditional global regression framework

\[ y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \]

where regression parameters are constant over the whole study region, by allowing local rather than global parameters to be estimated so that the model is rewritten as:

\[ y_i = \beta_0(u_i,v_i) + \sum_k \beta_k(u_i,v_i) x_{ik} + \varepsilon_i \]

where \((u_i,v_i)\) denotes the coordinates in the \(i\)th point in space and the parameters now vary over the study region with geographic coordinates (spatial variability in parameters). In the case of the logistic regression, the left side of the equations above \((y_i)\) is replaced by logit \([\operatorname{Prob}(Y_i = 1| x_i],\ where \logit (p) = \log[p/(1-p)]\). The difference between traditional regression and GWR is that, as opposed to traditional regression, GWR assumes implicitly that observed data near to location \(i\) have more of an influence in the estimation of the local parameters than do data located farther from \(i\). In essence, the GWR model measures the relationships inherent in the model around each location \(i\).

Specific comments & minor corrections

1. pg 7. It is not clear why the variable representing distance to cardiology center requires categorization. In addition, it is not certain that the categories chosen (<32km, 32-64 km, 64-105 km, >105 km) are indeed the most appropriate. It is stated that the categories chosen were based on aerial transportation times of 60, 90 & 120 minutes. Reference 11 provides the same cut-offs but the transportation times were based on ground ambulance data and round-trip helicopter distances provided by the University of Michigan (unpublished data).

   In the first version of the manuscript, the distance variable was defined as a binary variable. As suggested by the previous reviewers, we added further categories. In order to choose categories that were based on previous studies, we considered the same categories as those used in references 11 and 18. We could have considered the variable ‘distance’ as continuous, but then this would result on the assumption that each 1 km increase results in the same risk change.
2. pg 7. The distance was ‘defined by the centroid of the postal code area’. How is ‘centroid’ defined (presumably not in the mathematical sense). For example, is it the largest concentration of dwellings?

The centroids of the postal codes are provided by DMTI Spatial data (ref 15). These centroids are the geometrical centers of the polygons of the postal codes. This is specified in the revised manuscript.

3. pg 8. 1st sentence of paragraph 2. Replace ‘multiple log-normal analysis was …’ by ‘multiple linear (log-normal) regression analysis was…’

We replaced the sentence in the text.

4. pg 8. State which package was used for the cluster analysis.

We used SPSS release 11.0 for the hierarchical cluster analyses. This was clarified in the revised manuscript.

5. pg 9. Global estimates and local estimates. Please explain this terminology. Are the global estimates those obtained by the ordinary least square approach?

This terminology was introduced to differentiate between estimates that depend on the location of individuals, like GWR models, and estimates that does not, like traditional regression models. We provided more in-depth explanations in the ‘Statistical Analyses’ section.

6. pg 9. State the version of ArcGIS used and the producers of the package. State the versions of GWR, SAS & SPSS used (and producers). State which package was used for the regression analyses, and therefore the computation of residuals.

We detailed which package was used for which analysis and added their versions. The producers of the packages figure at their reference. In the revised version, we added the following paragraph:

The global regression analyses and residuals analyses were done using SAS Release 8.02 [26], the hierarchical cluster analyses were done using SPSS Release 11.0.1 [27] and the local estimates of the relationships between outcomes and predictors were estimated using GWR Release 3 [25]. Cartographic representations were done using ArcGIS Release 8.3 [23].

7. pg 11. 2nd paragraph. There is an unjustified etiologic emphasis on the interpretation of the results. e.g. ‘The presence of an ICP reduces HD rates’ – could it not be subjects who are too ill are not subjected to ICP?

We agree with the reviewer that the cohort design do not allow for causal inference. We changed the wording to reflect the correlation between ICP, LoS, HD and EHR.
8. pg 11. 2nd paragraph. It is noted that women had decreased odds of receiving ICP. Possible interactions between gender and ICP on the odds of HD and EHR should be considered. See general comment 1.

Here again and according to the answer to the general comment 1, when statistically significant, the interaction term gender x ICP was integrated in the models.

9. pg 11. Were any statistical tests performed to confirm the postulated relationships between LoS, ICP and distance (Table 3)? If not, the last section of paragraph 2 and Table 3 should be removed.

Since we added the interaction terms in the regression analyses, Table 3 provided redundant information so we removed it from the manuscript. Nevertheless, the interpretation remains the same, as the interaction term in the LoS model reflects that, for patients with ICP, the LoS increases as we move away from a specialized cardiology center.

10. pg 18. The word ‘robust’ should be avoided as there is no indication that the models are indeed ‘robust’ in the statistical sense.

We removed the word ‘robust’, as suggested.