Author's response to reviews

Title: Identification and classification of high risk groups for Coal Workers' Pneumoconiosis using an Artificial Neural Network based on occupational histories: a retrospective cohort study

Authors:

Hongbo Liu (liuhbabc@163.com)
Zhifeng Tang (tzhfcmu@163.com)
Yongli Yang (yl_yang2005@yahoo.com.cn)
Dong Weng (cruise00@gmail.com)
Gao Sun (sungao@mail.cmu.edu.cn)
Zhiwen Duan (dzhiwen@yahoo.com.cn)
Jie Chen (chenjie@mail.cmu.edu.cn)

Version: 2 Date: 14 August 2009

Author's response to reviews: see over
Identification and classification of high risk groups for Coal Workers’ Pneumoconiosis using an Artificial Neural Network based on occupational histories: a retrospective cohort study

Hongbo Liu, Zhifeng Tang, Yongli Yang, Dong Weng, Gao Sun, Zhiwen Duan and Jie Chen

Dear Editor-in-Chief (Dr. Melissa Norton)

Thank you for response to our submission (ref. MS1121233859263884). We are very grateful for the referees’ kindest help with not only science but also English. The authors really appreciated it.

Now, enclosed with a response to the referees, we would like to resubmit the paper to your journal for publication. After carefully studying the comments from the referees, we made a major revision to our manuscript and highlighted all the changes accordingly. From the revised manuscript, you can see that we have rewritten the whole paper on the previous manuscript from the abstract, background through the discussion including reference, table, figure. We also put a lot of efforts into English to improve the readability with concise English. Before the resubmission, we have asked a profession English editor in BioMed Proofreading, LLC to do proofediting in order to keep English mistakes as few as possible. In addition, all the changes made in the manuscript are marked in different color, which may help the reviewer to compare with the previous manuscript. Also, I approve the use of the data in this study. We have revised all detailed through the whole manuscript to conforms to the journal style.

Therefore, we consider this paper to be of significant to the Public Health study and appropriate for the publication in your journal. Should you need further information about this submission, please contact us at your early convenience.

P.S. Enclosed with a response to the reviewers.

Yours sincerely,
Jie Chen, MD & PhD
Response to referee 1:
Dear Dr. Vinicius Antao,

Firstly, we would like to thank you for your kindest comments on our manuscript. Upon getting the comments back, we carefully studied your comments word by word and made major revisions on the previous manuscript.

The following are our responses to your comments accordingly:

1. “First paragraph, line2: CWP is caused by inhalation of coal mine dust (which may contain silica), but not silica alone,… ”: We agree with you. So, we have rewritten this part as “by inhaling respirable coal mine dust.” in Background: First paragraph.

2. “Second paragraph, line5,6: … “efficient to perform” needs further explanation or reference. This technique has been used for many years in different countries as the method of choice for CWP surveillance.”: You are right. Chest X-ray is a gold standard of diagnosing CWP. It is not able to be replaced by any method of predictive models. The neural network designed in this study can predict the risk probability of CWP in coal miners, and further identify and classify different levels of risk for CWP of coal miners. In view of rationale and purpose of this study, we made a major revision to this part in Background: Second and fourth paragraphs.

3. “Last paragraph, line4-7: these conclusions do not belong in the introduction.”: We agree with you. Our purposes are to construct a neural network model to predict the risk probability of CWP in the miners, and recommend different intervals of medical examination for better preventing and controlling CWP. So, we made a major revision of this part in Background: Last paragraph.

4. “Study settings, First paragraph, …it would be helpful to see more details about the operation, …”: Thanks for your suggestions. We have added the information of the operation as “In the Tiefa Colliery, the type of coal mined is
Kennel coal, and the mining technique used is longwall.” in Methods: Study Settings, line 2-3.

5. “Study population, second paragraph:… “coal miners would be defined as combining…” is hard to understand”: Thank you for your comments. We divided occupational category into four groups according to the way a miner exposed to dust in his work and the composition of the dust. Coal miners in combining are those that their duration of tunneling were more than 2 years, but not more than one-half of the time exposed to dust. We have rewritten this part in Methods: Study population, second paragraph: line 11-15. and We made a major revision about the Study population and occupational categories to make them clear.

6. “Discussion, first paragraph, line 6,7: reference 25 is about silicosis, …”: Thank you for your suggestion. We have deleted the reference 25, and rewritten this part in Discussion, first paragraph.

7. “Discussion, second paragraph, line 5-9:...why cumulative dose exposure was not calculated. …”: Thank you for your comments. It is the best for this study to add the detailed dust concentrations at the work place, and calculate cumulative dose exposure. In Tiefa Colliery, dust concentrations at the work place are measured by the department of environmental protection of this colliery, and for the colliery, the dust concentrations are secret. In our manuscript, most of data in the study were derived from personnel records in the Manpower Resource Section and the department of industry hygiene and occupational disease of Tiefa Colliery. So, we was not able to obtained the results of the detailed dust concentrations from the department of environmental protection of this colliery. This is a limitation of our study. We discussed this issue and to reflect dust exposure level and to show the change of dust level, we think that occupational histories of coal miners are closely related to dust concentration and cumulative dose exposure, especially duration of dust exposure and occupational category. They could be used to reflect dust exposure level.

To make this clear, we deleted some descriptions in Discussion: second paragraph, line 1-18, and added these explaination in Limitation: second paragraph.
8. Some minor errors We have made them revised. All the changes made in the manuscript are marked in different color.
Response to referee2:

Dear Dr. Yohana Mashalla,

Firstly, thank you for your kindest comments on our manuscript. We really appreciate your patient help with not only scientific critiques but also our some minor corrections including spelling and statistical descriptions. We carefully studied your comments word by word as soon as getting your comments back, and made major revisions on the previous manuscript accordingly.

The following are our responses to your comments accordingly:

1. “On paragraph one of the background, …. More literature review on the prevalence of CWP in developing countries must be provided to justify the study for the benefit developing countries.”: Thank you for your comments. We searched for many references and discussed this issue, then we made a major revision according to your suggestions. We described as: in developing countries, owing to the growing world economy, the increase in coal production and utilization results in numerous miners exposed to the health hazards of coal mine dust. In China, coal is the major energy resource (about 70% of electricity is generated in coal-burning power plants). The estimated number of underground miners at present is > 6 millions. It has been reported that the number of new CWP patients is > 4000 cases per year and CWP accounts for about 48% of the total number of cases of pneumoconiosis in China. We also added more references in this part. Please see the detailed in Background, first paragraph: line 7-14.

2. “Rationale for the study is not clearly described. …”: Thank you very much for your comments. We discussed deeply about this issue. It is important to clarify the rationale of the study. According to your suggestion, we not only rewrite this question in Background part, but also made many major revisions about this issue through the whole manuscript.

We considered as : At present, a chest X-ray is the gold standard of monitoring and diagnosing CWP. In the US all miners working in an underground coal mine must be offered a chest X-ray every 5 years by the mines. Current regulations in China
require that coal miners must be offered a chest X-ray every 2-3 years (GBZ188-2007). Coal miners with different occupational categories and durations of dust exposure may be at different levels of risk for CWP [14, 15]. It is necessary to identify and classify different levels of risk for CWP of coal miners with a different work history. In this way, we can recommend different intervals of medical examinations according to different levels of risk for CWP.

Our aim is to: We constructed a neural network model based on occupational histories to predict the risk for CWP in miners. We classified different levels of risk for CWP of coal miners and recommended different intervals of medical examinations according to different levels of risk for CWP. It could provide the basis for further emending the measures of CWP prevention and control and it is important for strengthening the surveillance of occupational hazards in coal mines.

The detailed of revising in **Background: second paragraph, line 4-9 and fourth paragraph;** and in **Discussion: first paragraph, line 7-12.**

3. “Is the question posed by the author well defined?”: We agree with you. For this issue, we not only rewrite this question in Background part, but also made many major revisions about this issue through the whole manuscript. Even we revised the title as: **Identification and classification of high risk groups for Coal Workers’ Pneumoconiosis using an Artificial Neural Network based on occupational histories: a retrospective cohort study.**

Also, to better monitor coal miners, especially in developing countries, It is necessary to identify and classify different levels of risk for CWP of coal miners with a different work history. In this way, we can recommend different intervals of medical examinations according to different levels of risk for CWP. Also, We classified different levels of risk for CWP of coal miners and recommended different intervals of medical examinations according to different levels of risk for CWP. It could provide the basis for further emending the measures of CWP prevention and control and it is important for strengthening the surveillance of occupational hazards in coal mines.

Like above question, the detailed of revising in **Background: second paragraph, line 4-9 and fourth paragraph;** and in **Discussion: first paragraph, line 7-12, and other parts.**
4. “Methodology,...the study fails to define whether the study attempted to validate superiority of the neural Network method against the “Gold Standard”. …”: Thank you for your comments. According to your suggestion, we made major revision like above 2 questions. We reconsidered as: Chest X-ray is a gold standard of diagnosing CWP. It is not able to be replaced by any other method. Any predictive model is not able to compare against the “Gold Standard”. In this revision of this manuscript, we predicted the risk probability for CWP of coal miners using the neural network model. We just identified high risk groups for CWP according to the risk probability. And recommended different intervals of medical examinations (Chest X-ray, Gold Standard) according to different levels of risk for CWP. To make it clear, we have added and revised some information in Background: second paragraph and fourth paragraph.

5. “Influence of cigarette smoking: The many studies of CWP have demonstrated significant influence of cigarette smoking on earth crust crushing employees e.g. coal miners. In this study, there is no mention on the smoking habits of the studied population, duration of smoking and number of cigarette smoked. It would have added value to the study if this important confounder was taken into account in the analysis of the data.”: We agree with you. The habit, duration and number of smoking influence on coal miners’s health. Smoking mainly influence on pulmonary function and respiratory symptoms. CWP is a lung disease caused by dust, which has significant changes of pathology, mainly including lung fibrosis, which smoking influence is weakly. In our study, we want to make use of occupational histories to predict the risk for CWP, so we didn’t take into account the influence of cigarette smoking in this study.

6. “Are limitations of the work clearly stated?”: Thank you very much for your suggestions. We didn’t state the limitations of the work clearly. In this revision manuscript, we added “Limitations of the study” after Discussion. Please see the detailed in Discussion: last paragraph- Limitations of the study.

7. “Are the discussions and conclusions well balanced and adequately supported
by the data?”: Thank you for your comments. For this issue, we not only rewrite this question in Discussion and Conclusion parts, but also made many major revisions about this issue through the whole manuscript, including Background. Even we revised the title as: Identification and classification of high risk groups for Coal Workers’ Pneumoconiosis using an Artificial Neural Network based on occupational histories: a retrospective cohort study. Please see the detailed in Discussions: first and sixth paragraph, and Conclusions part.

8. Some minor errors We have made them revised. All the changes made in the manuscript are marked in different color.
Response to referee 3:

Dear Dr. Mattias Ohlsson,

Firstly, thank you for your kindest comments on our manuscript. We really appreciate your patient help with not only scientific critiques but also our some minor corrections. Certainly, main question is statistical analysis. We carefully studied your comments word by word as soon as getting your comments back, and made major revisions on the previous manuscript accordingly.

The following are our responses to your comments accordingly:

1. “It is not clear why such a complex (15 hidden nodes) neural network model was needed….. On what dataset is this performance measured? Clarifications are need.”: Thank you for your comments. In the manuscript, we indeed ignored the clarification of determining the optimal number of neurons. We randomly split all data into 5 subsets of equal size. At any number of neurons, it was trained on all but one subset, and tested on the remaining one. From 6 neurons, we gradually increased the number of neurons. When there were 15 neurons, the performance of the trained neural network on output sample in the testing set began to deteriorate.

   To make it clear, we have added the information in Development of the neural network: second paragraph, line 3-8. All the changes made in the manuscript are marked in different color.

2. “It is stated that the neural network model had 6 inputs, corresponding to the variables given i Table 1….Why is there only 6 inputs? How are the categorical variables encoded? If only 6 variables are used then a justification is needed…..”: Thank you for your comments. In our study, the neural network had 6 input variables. Of those, “Job” variable is the occupational category which included tunneling, combining, mining, and helping. We thought that the dust exposure level should declined with four occupational categories (tunneling, combining, mining, and helping). So, we didn’t consider it as a categorical variable. In the manuscript, we ignored the question and the clarification of the variables’ normalization. In fact, the
variable values in the global dataset were normalized to $[0, 1]$ according to the equation in Matlab: $x(i,:)=(x(i,:)-\text{min}(x(i,:)))/(\text{max}(x(i,:))-\text{min}(x(i,:)))$.

In the revised manuscript, we have added the information in Methods: Study Population, second paragraph, last line. All the changes made in the manuscript are marked in different color.

3. “Regarding the sensitivity analysis:…What does a reasonable interval means? How much does the result depends on the size of this interval. …”: Thank you for your comments. The sensitivity analysis used in our study is a common method which often were used by some researchers. The basic idea of this method is that the inputs to the network is varied within a reasonable interval (or between its mean plus (minus) a user-defined number of standard deviation), and the network output is recorded as a percentage deviation. The size of this interval certainly influence on the result of the output. But the order of the change in the output can not be changed. So, in our study, we selected mean minus a standard deviation and mean plus a standard deviation as a reasonable interval.

In the manuscript, we didn’t clearly clarify this question, and now have revised it in Methods: Sensitivity analysis, line 4-6. All the changes made in the manuscript are marked in different color.

4. “Although the data set is large (14655) there are very few positive cases (236) which makes this a very unbalanced data set….. Why they did not use e.g. K-fold cross validation to estimate the test performance.”: We agree with you. K-fold (generally 10-fold) cross-validation is a good method to test the performance of the neural network. In k-fold cross-validation, overall dataset is randomly split into k subset, and the model is trained and tested k times. Every time, it is trained on all but one subset, and tested on the remaining subset. The average of the k individual accuracy is calculated to minimize the bias associated with the random sampling of the taining and testing data samples.

We didn’t take an account of this method, but used the ROC to assess the test performance. The main reason is that we need a reasonable predictive probability value of coal miners to identify different levels of risk for CWP, and further recommended different intervals of medical examinations according to different levels of risk for CWP. The ROC plot could derived the classification value from a
trusted gold standard. It also has a very important connection to neural network applied to classification applications.

5. **Some minor essential revisions** We have made them revised. All the changes made in the manuscript are marked in different color.

6. “A very common approach when using neural networks is to use an ensemble approach to increase the performance of the network. The authors should perhaps consider to test ensembles to see if an additional performance increase can be obtained.”: We agree with you. Ensemble learning methods is a popular approach in the machine learning and data processing of some researches. It can improve the generalization of single learning machine, and obtain better generalization performance with nonlinear models. In this study, we didn’t take an account of this method, but used a Bayesian learning algorithm by introducing probabilistic treatment of the Bayesian inference technique. As a simple and feasible method, it can overcome some difficult problems, and also can improve the generalization better than other learning algorithm. You give us a kindly suggestion. We were going to use the ensemble approach to increase the performance of the network in a next study.