



SOFTWARE TOOL ARTICLE

REVISED **whoishRisk – an R package to calculate WHO/ISH cardiovascular risk scores for all epidemiological subregions of the world [version 2; referees: 3 approved]**

Previously titled: Simple and adaptable R implementation of WHO/ISH cardiovascular risk charts for all epidemiological subregions of the world

Dylan Collins, Joseph Lee, Niklas Bobrovitz , Constantinos Koshiaris, Alison Ward, Carl Heneghan

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

v2 **First published:** 14 Oct 2016, 5:2522 (doi: [10.12688/f1000research.9742.1](https://doi.org/10.12688/f1000research.9742.1))
Latest published: 08 Mar 2017, 5:2522 (doi: [10.12688/f1000research.9742.2](https://doi.org/10.12688/f1000research.9742.2))

Abstract

The World Health Organisation and International Society of Hypertension (WHO/ISH) cardiovascular disease (CVD) risk assessment charts have been implemented in many low- and middle-income countries as part of the WHO Package of Essential Non-Communicable Disease (PEN) Interventions for Primary Health Care in Low-Resource settings. Evaluation of the WHO/ISH cardiovascular risk charts and their use is a key priority and since they only existed in paper or PDF formats, we developed an R implementation of the charts for all epidemiological subregions of the world. The main strengths of this implementation are that it is built in a free, open-source, coding language with simple syntax, can be downloaded from github as a package (“whoishRisk”), and can be used with a standard computer.



This article is included in the **RPackage** channel.

Open Peer Review

Referee Status: 

	Invited Referees		
	1	2	3
REVISED			
version 2	report	report	report
published			
08 Mar 2017			
version 1	?	?	?
published	report	report	report
14 Oct 2016			

- Raivo Kolde**, Philips Research North America USA
- Scott A. Chamberlain**, University of California, Berkeley USA
- Maria Suarez-Diez**, Wageningen University and Research (WUR) Netherlands

Discuss this article

Comments (2)

Corresponding author: Dylan Collins (dylan.collins@phc.ox.ac.uk)

How to cite this article: Collins D, Lee J, Bobrovitz N *et al.* **whoishRisk – an R package to calculate WHO/ISH cardiovascular risk scores for all epidemiological subregions of the world [version 2; referees: 3 approved]** *F1000Research* 2017, 5:2522 (doi: [10.12688/f1000research.9742.2](https://doi.org/10.12688/f1000research.9742.2))

Copyright: © 2017 Collins D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: The WHO Collaborating Centre for Self Care paid for the open access publishing fees. No other funding was provided for this work.

Competing interests: DC has received payment from the WHO for consulting work. AW and CH have received expenses and grant income from the WHO for projects related to CVD and Self Care in NCDs, and direct a WHO Collaborating Centre. CH also receives funding from the National Institute of Health Research (NIHR) School of Primary Care Research. JL, CK, and NB declare no competing interests.

First published: 14 Oct 2016, 5:2522 (doi: [10.12688/f1000research.9742.1](https://doi.org/10.12688/f1000research.9742.1))

REVISED Amendments from Version 1

The main update of this version is the development of an R package using the previously reported R code. The package, called `whoishRisk`, can be downloaded directly from github as described in the paper. The benefit of this is that users now do not need to load the R code or any dependency files in the work space; instead they must download the `whoishRisk` package and load it from the library.

The major update to the code itself is the addition of warning messages when parameters passed to the `WHO_ISH_Risk()` function appear out of range.

The title was changed accordingly to reflect the development of the `whoishRisk` package. Figure 1 was removed as it was no longer relevant, and Table 2 was added to show the new `WHO_ISH_Risk()` warning messages.

See referee reports

Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide, including in many low-and-middle income countries (LMIC)^{1,2}. Preventing CVD is therefore a worldwide priority and the World Health Organisation (WHO) is coordinating a global strategy for LMIC to systematically prevent CVD in primary care³.

In 2007 the WHO and the International Society of Hypertension (ISH) published the WHO/ISH CVD risk charts for all WHO epidemiological subregions of the world⁴. These charts are to be used as part of the WHO's Package of Essential NCD (PEN) Interventions for Primary Health Care in Low-Resource Settings in jurisdictions that do not have their own population-derived risk assessment algorithms. While these charts are a good resource for many health systems, little is known about their validity⁵. Therefore, it is important that jurisdictions that implement these charts conduct operational research and attempt to validate and optimise them for their setting.

Two paper-based versions of WHO/ISH charts are available for each subregion: one that requires measured total cholesterol and one that does not. The latter was made available for use in settings with limited access to laboratory testing or where the cost of cholesterol testing is prohibitive. Both charts require information on age, gender, diabetes status, smoking status, and systolic blood pressure to stratify people into one of five risk categories of 10-year risk of a fatal or non-fatal CVD event. Further instructions for their use have been published, including the definition and classification of the fourteen epidemiological sub-regions³.

Through our experience collaborating with LMIC with the implementation of WHO PEN, we identified a common need for an open-source tool to facilitate the implementation of WHO/ISH risk charts and operational research of WHO PEN at a population level. We therefore developed an R package called `whoishRisk`, which we describe here and make available to researchers in LMIC. R is a statistical computing language and environment which is open source and freely available to anyone⁶.

Methods

Extraction of WHO/ISH cardiovascular risk charts

We extracted all versions of the paper-based WHO/ISH CVD risk charts by hand into a standardized Microsoft Excel template, independently and in duplicate. We compared the duplicate extractions and calculated Cohen's kappa coefficient for inter-rater reliability, using the `irr` package in R⁷. Discrepancies were reviewed by the same two extractors and resolved by referring to the original paper chart.

Development of the WHO/ISH risk function

One author wrote the initial code for the WHO/ISH risk function in R and created the `whoishRisk` package (DC). This was reviewed and adapted by a second author experienced in the R language (CK). Two additional authors (JL, NB), new to the R language, reviewed the code to ensure the syntax was comprehensible.

Validation

A MatLab implementation of WHO/ISH risk charts for epidemiological subregion SEAR D had been previously reported⁸. We used Octave (www.gnu.org/software/octave/) version 8.3.2 to calculate the SEAR D WHO/ISH risk score for every possible combination of risk factors using the previously reported MatLab implementation, and compared the percent agreement to the risk scores generated by our R package, `whoishRisk`.

Results

`whoishRisk` Package

The `whoishRisk` package can be downloaded and installed directly from github using the `install_github()` command in the `devtools` package, with the argument "`DylanRJCcollins/whoishRisk`". The package contains a single function, `WHO_ISH_Risk()`, which calculates the WHO/ISH CVD risk score for any epidemiological subregion of the world based on the parameter values passed to it.

Extraction of WHO/ISH cardiovascular risk charts

All WHO/ISH risk charts were extracted by hand into a single comma delimited file (Dataset 1). The first six columns specify the risk factor values, and the last 14 columns specify the corresponding risk category for a given subregion. `whoishRisk` uses these data internally to calculate the WHO/ISH risk score. Cohen's kappa for initial agreement between the independent extractors was 0.97, indicating excellent agreement. All remaining discrepancies were resolved by consensus.

Dataset 1. WHO_ISH_Scores.csv

<http://dx.doi.org/10.5256/f1000research.9742.d153375>

CSV file of all parameter combinations and corresponding WHO/ISH risk scores used internally by `WHO_ISH_Risk()` to match parameter values passed to it to their respective risk score.

Development of the `WHO_ISH_Risk()` Function

`whoishRisk` contains a single function, called `WHO_ISH_Risk()`, which calculates the WHO/ISH risk score for any epidemiological subregion. The function code is reported herein (Dataset 2).

Dataset 2. R code for the WHO_ISH_Risk() function of the whoishRisk package

<http://dx.doi.org/10.5256/f1000research.9742.d153376>

WHO_ISH_Risk() requires seven parameters: age, gender, smoking status, diabetes status, systolic blood pressure, total cholesterol, and the appropriate WHO epidemiological subregion. The function format in the workspace is: WHO_ISH_Risk(age, gdr, smk, sbp, dm, chl, subregion). These parameters and their abbreviations are summarised in Table 1. No default values are specified for any parameter.

Internally, WHO_ISH_Risk() requires access to the comma delimited file named “WHO_ISH_Scores.csv”, which it calls automatically from within the package but is also included herein (Dataset 1).

The WHO_ISH_Risk() function creates an internal data frame of the risk factor values passed to it. Parameters can be single values or vectors of equal length. It then categorises the continuous parameters age, systolic blood pressure, and total cholesterol. Age and systolic blood pressure were categorised according to WHO guidance¹⁰. Total cholesterol was categorised into one of the five possible categories (<=4,5,6,7 and >=8 mmol/L) according to common clinical practice, rounding up from 0.5 to the nearest integer.

Internally, a unique identification code is generated corresponding to the combinations of risk factors for each individual. This

Table 1. Description of the WHO_ISH_Risk() function parameters.

Function parameter	Full parameter name	Parameter class	Parameter values
age	Age	numeric	Continuous (years)
gdr	Gender	numeric	Dichotomous (0=female; 1=male)
smk	Smoking	numeric	Dichotomous (0=non-smoker; 1=smoker)
sbp	Systolic blood pressure	numeric	Continuous (mmHg)
dm	Diabetes mellitus status	numeric	Dichotomous (0=not diabetic; 1=diabetic)
chl	Total cholesterol	numeric	Continuous (mmol/L); 0=unknown cholesterol
subregion	WHO epidemiological subregion	character	“AFR_D”, “AFR_E”, “AMR_A”, “AMR_B”, “AMR_D”, “EMR_B”, “EMR_D”, “EUR_A”, “EUR_B”, “EUR_C”, “SEAR_B”, “SEAR_D”, “WPR_A”, “WPR_B”

code is matched to a reference code from the “WHO_ISH_Scores.csv” file. The function stores the risk scores in a data frame that includes the risk factors, and ultimately returns a vector containing the risk scores. The output of the function is one of five different character strings, corresponding to the five different WHO/ISH risk categories: “<10%”, “10 to <20%”, “20 to <30%”, “30 to <40%”, “>=40%”.

Warning messages are included when parameters appear out of range. These messages, their conditions, and their intended interpretation are described in Table 2. Out of range continuous parameters (age, systolic blood pressure, total cholesterol) are non-fatal and a risk score will be generated with a warning message. Out of range dichotomous variables (gender, smoking status, diabetes status) are fatal errors and the output (NA) will be generated with a warning.

Worked example

whoishRisk can be installed in one step using install_github() from the devtools package⁹. We have included a worked example of how to install whoishRisk and use the WHO_ISH_Risk() function to calculate the risk score for five individuals.

```
#Step 1: Install whoishRisk package
> library(devtools)
> install_github("DylanRJCcollins/whoishRisk")

#Step 2: Load whoishRisk package into workspace
> library(whoishRisk)

#Step 3: Load risk factor data
> Age <- c(40, 87, 65, 53, 71) #Age in years
> Gender <- c(0,0,0,1,1) #0=female, 1=male
> Smoking <- c(1,1,0,1,0) #0=non-smoker 1=smoker
> Systolic_Blood_Pressure <- c(129, 157, 134, 189, 141) #SBP in mmHg
> Diabetes <- c(1,1,1,0,1) #0=not diabetic 1=diabetic
> Total_Cholesterol <- c(0, 5.1, 4.5, 0, 8.3) #Total cholesterol (mmol/L, 0=unknown cholesterol)

#Step 4: Pass the risk factor vectors to the WHO_ISH_Risk() function, and set subregion equal to the name of the appropriate epidemiological subregion (e.g. "EMR_B"). This will return a vector of WHO/ISH risk scores.
>WHO_ISH_Risk(Age, Gender, Smoking, Systolic_Blood_Pressure, Diabetes, Total_Cholesterol, "EMR_B")

[1] "<10%" ">=40%" "<10%" ">=40%" ">=40%"
```

Table 2. WHO_ISH_Risk() warning messages, their conditions, and intended interpretation.

Condition of Warning	Warning Message	Interpretation of Warning
age < 19	"At least one age is 18 or younger"	WHO/ISH risk scores are meant for use in adults. At least one age value is equal to 18 or younger. Check for errors in age values.
age > 100	"At least one age is greater than 100"	Age values greater than 100 may indicate a data error. Check to ensure age values over 100 are correct.
gdr > 1	"Gender must be equal to 0 or 1"	A value for gender other than 0 or 1 is included. Check gender values.
smk > 1	"Smoking must be equal to 0 or 1"	A value for smoking other than 0 or 1 is included. Check smoking values.
sbp < 90	"At least one systolic blood pressure is below 90 mmHg"	SBP values below 90 mmHg may indicate a data error. Check to ensure values under 90 mmHg are correct.
sbp > 250	"At least one systolic blood pressure is over 250 mmHg"	SBP values over 250 mmHg may indicate a data error. Check to ensure values over 250 are correct.
dm > 1	"Diabetes status must be equal to 0 or 1"	A value for diabetes status other than 0 or 1 is included. Check diabetes status values.
tc > 10	"At least one total cholesterol is greater than 10 mmol/L. Ensure all values are in units of mmol/L"	Total cholesterol values over 10 mmol/L may indicate a data error. Check to ensure values over 10 mmol/L are correct and in units of mmol/L.

Validation

Comparison with the published MatLab implementation of the SEAR D risk charts⁸ showed 100% agreement with our R implementation, for all possible combinations of risk factors.

Discussion

To our knowledge, this is the first publically available R implementation of WHO/ISH CVD risk charts for all WHO epidemiological subregions of the world. Our package, *whoishRisk*, may be used for analysis of cardiovascular risk when electronic patient data is available. The code will automatically apply WHO/ISH risk scores to patients based on age, gender, systolic blood pressure, smoking status, diabetes status, total cholesterol, and epidemiological subregion. This code could be used, for example, during a pilot implementation of WHO PEN to audit the accuracy of risk assessment by comparing documented risk scores to actual risk scores calculated using this tool. We have provided a complete worked example.

While WHO PEN guidance specifies the range of systolic blood pressure values for each systolic blood pressure category, it provides no such guidance for categorising total cholesterol. Based on our opinion and clinical experience, and on a previously published implementation in MatLab⁸, we chose to categorise total cholesterol by rounding up at 0.5 to the next integer.

The "WHO_ISH_Scores.csv" file is provided herein for transparency and to promote collaboration and cross validation. While the

risk score values it stores are returned to the workspace as characters (e.g. "10 to <20%"), a user could simply convert these to class numeric or factor. We chose to return them as character strings that are identical to the patient charts in order to produce a literal implementation of the risk charts.

Conclusion

We created an R package called *whoishRisk* to be used for the calculation of WHO/ISH CVD risk charts for all WHO epidemiological subregions of the world. It contains a single function, `WHO_ISH_Risk()`, that requires seven parameters: age, gender, systolic blood pressure, smoking status, diabetes status, total cholesterol, and epidemiological subregion. *whoishRisk* can be used to quickly calculate WHO/ISH risk scores from routinely collected electronic patient data and therefore aid in the implementation and evaluation of these risk charts in low-resource settings.

Data and software availability

F1000Research: Dataset 1. CSV file of all parameter combinations and corresponding WHO/ISH risk scores used internally by `WHO_ISH_Risk()` to match parameter values passed to it to their respective risk score [10.5256/f1000research.9742.d153375](https://doi.org/10.5256/f1000research.9742.d153375)¹¹

F1000Research: Dataset 2. R code for the `WHO_ISH_Risk()` function of the *whoishRisk* package, [10.5256/f1000research.9742.d153376](https://doi.org/10.5256/f1000research.9742.d153376)¹²

Author contributions

DC conceived of the idea, extracted data, wrote the initial code, and manuscript. JL and NB extracted data and contributed to writing the manuscript. CK reviewed and adapted the code and contributed to writing the manuscript. AW and CH reviewed and contributed to writing the manuscript.

Competing interests

DC has received payment from the WHO for consulting work. AW and CH have received expenses and grant income from the

WHO for projects related to CVD and Self Care in NCDs, and direct a WHO Collaborating Centre. CH also receives funding from the National Institute of Health Research (NIHR) School of Primary Care Research. JL, CK, and NB declare no competing interests.

Grant information

The WHO Collaborating Centre for Self-Care paid for the open access publishing fees. No other funding was provided for this work.

References

- Mathers CD, Loncar D: **Projections of global mortality and burden of disease from 2002 to 2030.** *PLoS Med.* 2006; 3(11): e442.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- WHO: **Noncommunicable Diseases, Fact Sheet.** 2015.
[Reference Source](#)
- WHO: **Package of essential noncommunicable (PEN) disease interventions for primary health care in low-resource settings.** Geneva, Switzerland; 2010.
[Reference Source](#)
- Mendis S, Lindholm LH, Mancia G, *et al.*: **World Health Organization (WHO) and International Society of Hypertension (ISH) risk prediction charts: assessment of cardiovascular risk for prevention and control of cardiovascular disease in low and middle-income countries.** *J Hypertens.* 2007; 25(8): 1578–82.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cooney MT, Dudina A, D'Agostino R, *et al.*: **Cardiovascular risk-estimation systems in primary prevention: do they differ? Do they make a difference? Can we see the future?** *Circulation.* 2010; 122(3): 300–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
- R Core Team: **R: A language and environment for statistical computing.** 2016.
[Reference Source](#)
- Gamer M, Lemon J, Fellows I, *et al.*: **irr: Various Coefficients of Interrater Reliability and Agreement.** 2012.
[Reference Source](#)
- Raghu A, Praveen D, Peiris D, *et al.*: **Implications of Cardiovascular Disease Risk Assessment Using the WHO/ISH Risk Prediction Charts in Rural India.** *PLoS One.* 2015; 10(8): e0133618.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wickham H, Chang W: **devtools: Tools to Make Developing R Packages Easier.** 2016.
[Reference Source](#)
- WHO/ISH: **World Health organization/International Society of Hypertension (WHO/ISH) risk prediction charts.** 2007.
[Reference Source](#)
- Collins D, Lee J, Bobrovitz N, *et al.*: **Dataset 1 in: whoishRisk – an R package to calculate WHO/ISH cardiovascular risk scores for all epidemiological subregions of the world.** *F1000Research.* 2017.
[Data Source](#)
- Collins D, Lee J, Bobrovitz N, *et al.*: **Dataset 2 in: whoishRisk – an R package to calculate WHO/ISH cardiovascular risk scores for all epidemiological subregions of the world.** *F1000Research.* 2017.
[Data Source](#)

Open Peer Review

Current Referee Status:   

Version 2

Referee Report 14 March 2017

doi:[10.5256/f1000research.11940.r20822](https://doi.org/10.5256/f1000research.11940.r20822)



Scott A. Chamberlain

rOpenSci project, University of California, Berkeley, Berkeley, CA, USA

I'm mostly satisfied with the author's changes, with a few minor changes.

Thank you for making an R package. I downloaded, and checked the package, with a few observations:

- In the package use `utils::read.csv` instead of just `read.csv` - you probably saw the warning if you've run R CMD CHECK or `devtools::check()`
- The package has no examples. I'd strongly suggest including examples.
- The package has no tests. I'd strongly suggest including tests (though if time is limited, at least include examples).
- The authors note that the package now gives warnings when input values are out of range - I'd strongly suggest to document what the correct ranges are for each parameter so users don't have to guess
- In addition, if there are sensible default values I'd use those - if applicable - may not be applicable.
- Urge authors to document what type of value is expected for each parameter, e.g., age (numeric), subregion (character)

Sincerely,
Scott Chamberlain

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 10 March 2017

doi:[10.5256/f1000research.11940.r20821](https://doi.org/10.5256/f1000research.11940.r20821)



Maria Suarez-Diez

Laboratory of Systems and Synthetic Biology, Wageningen University and Research (WUR),
Wageningen, Netherlands

My main concerns have been addressed in this new version of the manuscript.

I think the developed package represents a useful addition to the field. I was able to install and run the

package and indeed the package reports the absence of data points (NA).

Only a minor concern: within R, getting the citation 'citation("whoishRisk")' returns the names of the name of the package, the names of the authors and the R version. I recommend they include a reference to this work.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 09 March 2017

doi:10.5256/f1000research.11940.r20823



Raivo Kolde

Philips Research North America, Cambridge, MA, USA

All my main concerns were satisfied with the update. I was able to install and run the package. I guess from visibility standpoint it would be nice to include it to the CRAN database, but otherwise I approve.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 28 October 2016

doi:10.5256/f1000research.10503.r17004



Maria Suarez-Diez

Laboratory of Systems and Synthetic Biology, Wageningen University and Research (WUR) ,
Wageningen, Netherlands

The authors present an R script and an additional datafile to calculate cardiovascular risk using the Who/ISH risk assessment charts. Previous implementations required the use of MatLab and I believe an R implementation can be an useful addition to the field.

The authors intended to design a code requiring little R expertise. I think the authors have only partly succeeded in accomplishing this goal and the code still needs to be improved.

As mentioned by the other reviewers, RTF is far from an optimal format to distribute the code. The authors should present their code as a package hosted in some repository. Currently the RTF file contains the function definition, data loading and working example in a single view. When creating the package the authors should clearly separate the code from the documentation and provide a working example. If the

authors intent this tool to be used and possibly modified by users with little R expertise then a separate file with detailed instruction should be provided.

The authors state that users can modify the WHO_ISH_Scores.csv file to represent other risk categories. That would require explanation of the content of the file. Although the used abbreviations gdr for gender might seem obvious, they should be explained. Also when downloading the file from F1000, the file name is changed and appears as "5ae9107XXX..._WHO_ISH_Scores.csv" so that the reading fails. This could be fixed by including the file as a dataset in the corresponding package.

When data from multiple patients are provided (as in the example) the output is a factor. I think it would help the users match input and output values is some tabular format or some patient identifier is provided in the output.

If any of the input parameters are missing (NA values) the code outputs "NA" but no information regarding the missing values is provided. It might be helpful if the code were to report which value was actually missing.

Fig. 1 shows the code and contains mainly information on how some continuous variables (age, cholesterol and systolic blood pressure are discretized). I think a more efficient way of conveying this information is by including it in Table 1.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 06 Mar 2017

Dylan Collins, University of Oxford, UK

Dear Maria Suarez-Diez,

Thank you for your valuable comments. In response we have made the following changes:

- We created an R package including all documentation files which can be downloaded from github
- We no longer suggest users to change the "WHO_ISH_Scores.csv" files and we have updated the file name hosted by F1000 to be "WHO_ISH_Scores.csv". This file, as suggested, is internal to the package and while we report it here is not necessary to download.
- We have added a series of warning messages to help explain outputs of "NA" and to help catch errors in the input values (e.g. out of range parameters)
- We have described how continuous variables are discretized in the main text

Every Best,

Dylan Collins

Competing Interests: No competing interests were disclosed.

Referee Report 25 October 2016

doi:10.5256/f1000research.10503.r17006



Scott A. Chamberlain

rOpenSci project, University of California, Berkeley, Berkeley, CA, USA

The authors describe some R code for helping to calculate cardiovascular risk scores. I think the code needs significant work.

- The code is in an .rtf file. This is very bad software practice. The authors should at the very very least put the code in a file with .R extension.
- Ideally the code should be put in an R package. It is relatively straight-forward to make an R package these days. See <http://r-pkgs.had.co.nz/> for help. This makes it easy to add documentation (of which there is currently essentially none for the current function included in the manuscript), tests, etc. Looks like the authors have a Github repository (https://github.com/DylanRJCollins/WHO_ISH_R_Implementation) - this could be made into an R package.
- Another benefit of making a package, is that R packages have a way of including datasets in them. The csv file is small enough that you can include it in the package if that's appropriate for the use case here. If the `WHO_ISH_Scores.csv` dataset is not likely to change, or not likely to change very often, they can include it in the package.
- The dataset `WHO_ISH_Scores.csv` has two columns that are probably row names that should be removed. If they aren't data, remove them.
- The dataset `WHO_ISH_Scores.csv`: the data is not in a tidy format that is easy to work with. There shouldn't be numeric data combined with symbols (e.g. $\geq 40\%$). If possible, I urge authors to find a way to make these into numeric columns while retaining the same information. However, it may be that it's too difficult to separate numeric values from the greater than / less than symbols etc.
- R packages should be cited by reference. e.g. "using the irr package (version 0.84)". It is good they cite the version, but put a reference in in your references. Run `citation("irr")` in an R session to get a reference for it.
- "We used RStudio (version 0.99.489) to compare the duplicate extractions ...": Rstudio is just an IDE. Say that you used R, not RStudio. It's fine to cite RStudio, but also cite R. Run `citation()` in an R session to get the citation for R.
- Octave is an open source language, and thus the authors should share the Octave code.
- "[...] we intentionally sought to use simple syntax in the base package" - the authors script is in fact not a package, but if they do make a package this wording is good.
- In the Conclusion: "We created a simple R implementation of WHO/ISH CVD risk charts for all WHO epidemiological subregions of the world, requiring only the base R package." - refer to this as

"base R", not "the base R package".

- If the dataset "WHO_ISH_Scores.csv" can be modified, thus different versions of it may be used by the user, it makes more sense to pass in the dataset as a parameter, instead of hard coding reading in the data in the function.
- I urge authors to include a license with their R package (to be created) - and to submit the package to CRAN so it's easy for all to install.

Code comments:

- I'm not sure the authors meant to do this, but the use of single ampersand in the `ifelse` statements (e.g., `ifelse(df\$age > 17 & df\$age < 50, 40, df\$age)`) means that they get a vector of logicals (e.g., `df\$age > 17 & df\$age < 50` evaluates to `TRUE FALSE FALSE FALSE FALSE`) using their example, when I think what they want is a single logical return value. If so, use double ampersand instead: `&&` instead of `&`.
- I'm pretty sure, but have not tested thoroughly, that the long series of if statements in the section "Match the look up value with the reference value" can be replaced with just this:
`ref[[subregion]][match(df\$luv, ref\$refv)]`.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 06 Mar 2017

Dylan Collins, University of Oxford, UK

Dear Scott A. Chamberlain,

Thank you for your thoughtful and considered feedback. We have responded in the updated version as detailed below.

- We have created an R package which can be downloaded from github
- We provide the code from the R package as a .R file
- We have included the "WHO_ISH_Scores.csv" data file in the R package
- We tidied the "WHO_ISH_Scores.csv" files but have kept the risk scores as character strings to retain a literal implementation of the risk charts, as described in the updated manuscript
- We have cited R packages and R as suggested
- The MatLab code used in Octave has previously been published as cited in the manuscript and is freely available
- We no longer suggest users to change the "WHO_ISH_Scores.csv" files
- We have included a GPL-3 license for the R Package

Every Best,

Dylan Collins

Competing Interests: No competing interests were disclosed.

Referee Report 20 October 2016

doi:10.5256/f1000research.10503.r17007



Raivo Kolde

Philips Research North America, Cambridge, MA, USA

The paper describes a software tool for calculation of WHO/ISH cardiovascular risk scores for different epidemiological subregions of the world. This could be a useful piece of software for researchers working with cardiovascular epidemiological data and make the practices for calculating such scores more reproducible over multiple studies. However, in the present form I found the paper to be severely lacking in both implementation and substance.

First, the code should not be distributed as an RTF file. R packages are well accepted standard for reproducibly distributing R code that also forces to provide some rudimentary documentation and examples. As an additional benefit, hosting the package in a repository like CRAN, Bioconductor or even Github, makes it easy to install and update the program. There is no reason to not follow this convention for this particular project, thus, the code should be converted into package format and hosted in one of the above-mentioned services to be any use to the community.

The argument of the code in the present form being simply adjustable is not valid. One has to spend quite a bit of time understanding what is happening in the code, especially when not an R expert. For the code to fulfill its purpose it has to be rewritten in a way that adjustable parts are presented to the user as function parameters (with reasonable defaults). This way users have to know even less R to use the code effectively.

Also, having a figure depicting the code is uncommon to say the least and not useful in any way. It would make much more sense to include a worked out example to the text, where you would describe a common use case of your R package.

To enhance usefulness of the package and make it more universally applicable it would be good to see some more risk prediction models like Framingham and SCORE added to the software. This would make it easy to compare the validity of different scores over a study population.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 06 Mar 2017

Dylan Collins, University of Oxford, UK

Dear Raivo Kolde,

Thank you for your insightful review and comments. In response we have made the following changes:

- As suggested, we created an R package which can be downloaded directly from github
- We no longer advocate for the code to be adjustable and through the development of a package have made it simpler to use
- We removed the figure of the code, and have added a worked example in the text
- The main difference between scores like Framingham, SCORE, and QRISK2 is that they have underlying cox model equations which can be implemented in R, whereas WHO/ISH risk charts do not – hence the rationale for the development of this package. While the additional of further risk scores was outside the scope of this work, we welcome collaborators who might want to contribute in this respect.

Every Best,

Dylan Collins

Competing Interests: No competing interests were disclosed.

Discuss this Article

Version 1

Author Response 20 Oct 2016

Dylan Collins, University of Oxford, UK

Dear Rajarshi Guha,

Thank you for reading our work. I've created a plain text format version of Dataset 2 which is the dataset that contains the R code and worked example.

It can be found on my github repository [here](#). Should you have any difficulty downloading from github or any further questions please let me know.

Every Best,

Dylan

Competing Interests: No competing interests were disclosed.

Reader Comment 18 Oct 2016

Rajarshi Guha, National Center for Advancing Translational Sciences, National Institutes of Health, USA

Please make the R code available in plain text format.

Competing Interests: No competing interests were disclosed.
