Abstract

Background: Causes of death statistics provide crucial health intelligence in national and international communities. An efficient death registration system provides reliable information for health policy system. In many developing countries, death registration systems face a degree of misclassification and incompleteness. There are many impediments to putting an estimate of cause-specific death rates. Addressing those challenges could prevent misleading results.

Methods: Our data was collected by Ministry of Health and Medical Education, Tehran and Isfahan cemeteries from 1995 to 2010. After converting ICD codes of Iran’s death registration into GBD codes, 170 underlying causes of deaths were recognized in the available data. A wide range of methods were applied for preparing the data. We used several statistical models to estimate mortality rates in age-sex-province groups for all causes of deaths. The considerable number of combinations for age, sex, cause of death, year, and province variables made further complicated model selection and evaluation of the results.

Results: Totally, 58.91% of deaths were related to males. The majority of causes of death were classified as NCDs (77.83%) and injuries (14.80%). We extrapolated 71.76% and 14.71% of causes of death by mixed effect model, spline model with parameter 0.9 and 0.6, respectively.

Conclusion: A comprehensive and unique registration system is able to solve many DRS issues. It is necessary to assess the quality and validity of cause of death data. Scientific methods like analyzing mortality level and cause-of-death data are used to provide an overview for better decisions.

Keywords: Cause of death, international classification of diseases, Iran, misclassification, vital statistics,


Introduction

Civil Registration and Vital Statistics (CRVS) are valuable sources for public health monitoring. Among vital statistics, causes of death statistics provide crucial health intelligence and are used as a snapshot of health status in national and international communities. Moreover, they constitute an important indicator for developing policies. For instance, the Global Monitoring Framework for Non-Communicable Diseases (NCDs) targets 25% relative reduction in risk of premature mortality from cardiovascular disease (CVDs), cancer, diabetes, and chronic respiratory diseases. Knowledge about cause-specific death rates and their distribution in age and sex groups is the main input to policy debates, resource allocation, and planning intervention in health systems. In addition, the trends of cause-specific mortality rates provide an opportunity to study the pattern of risk factors and health outcomes in society over time. On the other hand, cause of death statistics not only indicate major leading causes of death and risk factors in the public health arena, but are also an essential component for estimating the burden of disease in developed societies.

An efficient death registration system provides reliable and up-
to-date information, which is used to make better health and social development decisions.\textsuperscript{8,9} However, Death Registration Systems (DRSs) in many developing countries are imperfect or do not exist at all.\textsuperscript{10} In Iran, occurring deaths are legally registered by death certificates, though the data have been subject to a degree of misclassification and incompleteness since 1995 when the system was established.\textsuperscript{11,12}

The cause of death registration data in Iran is collected by Ministry of Health and Medical Education (MOHME). The Deputy for Research and Technology from 1995 to 2001 and the Deputy for Public Health from 2001 onwards have been responsible for collecting this data in MOHME.\textsuperscript{13} This inconsistency in administration of data during the last 15 years is perhaps one of the reasons for insufficiency of DRS in Iran.

The National and Sub-national Burden of Diseases, Injuries and Risk Factors (NASCOD) is an ongoing project that aims to estimate levels and trends of burden of disease from 1990 to 2015 in Iran.\textsuperscript{13-12} Death statistics, as the main component of the burden of diseases, include levels and trends of mortality and cause-specific death rates. Level and trends of mortality and also incompleteness of death registration system were addressed by another study.\textsuperscript{14} This study is conducted to explain cause of death estimations methods that face misclassification problems. Due to the broad range of subjects that should be discussed in this topic, we only explore the methods that were applied to handle data issues and methodological approaches to estimates cause-specific death rates. Other topics in this field and the results will be published in other papers.

Materials and Methods

Study setting (design)

We aim to estimate cause-specific death rates among Iranian residents from 1990 to 2015 at national and sub-national levels. Target population is divided into 19 age groups consisting of children under one year of age, one to four years of age, and then 5-year age groups up to 85 years of age. All individuals above 85 years of age were classified in 85+ group. All processes were done separately for males and females.

Causes of death were registered based on the tenth International Classification of Disease (ICD10) codes, but we converted the ICD 10 into Global Burden of Disease (GBD) study codes. Exploration of GBD codes shows that 165 of Iran’s ICD 10 codes are eligible to be considered GBD causes of death for the Iranian population.

The study included all deaths related to Iranian residents living in Iran. Moreover, records related to cases of mutilation (buried limbs), cases of abortion (those results in neonatal death) or stillbirth, foreigners (who were not residents of Iran), and duplicates were all eliminated.

Data source

Death Registration System (DRS) data from 1995 to 2010 was used in this study. Furthermore, we added data from Tehran and Isfahan cemeteries (Behesht-e-Zahra and Bagh-e-Rezvan, respectively) to the DRS data set. The information of other cemeteries all around the country were also registered in national DRS. Therefore, we used the most comprehensive data set. In fact, this data set included the five sub-data sets below:

1) DRS data from 1995 to 2001, which were collected by the Deputy for Research and Technology at provincial level (Data set 1995 – 2001)
2) DRS data from 1995 to 2001, which were collected by the Deputy for Public Health at provincial level (Data set 2001 – 2004)
3) DRS data from 2006 to 2010, which were collected by the Deputy for Public Health at provincial and district levels (Data set 2006 – 2010)
4) Behesht-e-Zahra cemetery data from 1995 to 2010 (Tehran data)
5) Bagh-e-Rezvan cemetery data from 2007 to 2010 (Isfahan data)

Finally, 3,645,608 individual records of death formed the primary data set.

Calculating the cause-specific death rates needs population at risk by sex and age groups in each province. We have extracted such data from the national censuses of 1996, 2006 and 2011. For other years, the following growth model was used to estimate population data:\textsuperscript{16}

\[ P = P_0 e^{rt} \]

Growth rate (\( r \)) was calculated by considering the first census as initial population (\( P_0 \)) and the second as final population (\( P \)). Then, population between these censuses were obtained by applying annual growth rate to the above formula. This process was conducted in each sex and age group and all provinces.

Censuses, Household Expenditure, and Income Survey (1985 – 2013) provide useful information about distribution of socioeconomic and cultural variables across the country. Among them, we created and applied covariates such as wealth index, years of schooling and urbanization in the statistical modeling. These are variables which affect cause of death and were available to be included in our study.\textsuperscript{17}

Data issues and Data preparation

Although DRS has become an ongoing process in the recent years, it is yet not appropriate to use it for scientific purposes. The following issues are those that should be solved before conducting any analysis.

Inconsistency in DRS administration

There have been several institutions which have administered DRS in the last 20 years. The Deputy of Research and Technology in MOHME established DRS and administered data for seven years since 1995. In 2001, the Deputy of Public Health in MOHME took responsibility for administering DRS. Tehran and Isfahan cemeteries administer their deaths data independently, as well. Coding systems, diseases categories, age groups, and coverage of DRS vary from one institution to another. In order to aggregate data sets, all variables which are essential in cause of death study were selected. Then, other information in each data set was used to improve the validity of these variables. Finally, the variables were recorded based on the same codebook and the main data set was formed by combining identical data sets.

Duplicates

There are several kinds of duplicates in Iran DRS Data sets. The first type occurs when an individual death is registered several times. It usually happens when a person who is resident of city A travels to city B in order to receive medical treatment and dies.
there. In this situation, one record will be usually recorded as
two separate death events in both cities. On the other hand, it is
usual to bury limbs with its owner characteristics; mutilation is,
therefore, another reason for registering a case more than once.
The other type is duplication among data sets. For instance, a
high percentage of Tehran and Isfahan cemeteries records are
found in MOHME data sets, as well. It is sometimes possible to
recognize duplicates and remove them from data set, especially
for individual records. For the latter type of duplicates, we had to
select more reliable data set.

Misalignment
Administrative divisions of Iran’s provinces and districts have
been reformed several times in the last 25 years; that is, some
adjacent districts have become independent from one province
and formed a new province, or merged with an existing province.
Therefore, we chose administrative division in 2011 (31 provinces)
as the reference, and restructured data sets of other years to have
all 31 provinces for all years.

Misclassification
DRS in Iran encounters high rates of misclassification of
different types, the most important of which is misclassification
of cause of death. Age-sex restriction (mismatch between registered
age, sex, and cause), garbage, and ill-defined codes are other kinds
of misclassification. Geographical misclassification (provinces)
is fairly related to misalignment. Using prior information is a
common and recommended approach to solve misclassification
problem.11

Missing values
Dealing with missing values is critical to the future analysis.
We should explore the nature of missing values, including their
pattern, type, and cause.

Incompleteness
Incompleteness is the main issue, not only in DRS but also in
all registration systems. Even the best DRS cannot capture all
deaths; thus, incompleteness is an important criterion in validity
assessment of DRSs. Mohamadi, et al. are conducting a parallel
study on level and trend of mortality in Iran from 1990 to 2015;
assessment of DRSs. We should explore the nature of missing values, including their
pattern, type, and cause.

Incompleteness is the main issue, not only in DRS but also in
all registration systems. Even the best DRS cannot capture all
deaths; thus, incompleteness is an important criterion in validity
assessment of DRSs. Mohamadi, et al. are conducting a parallel
study on level and trend of mortality in Iran from 1990 to 2015;
assessment of DRSs. We should explore the nature of missing values, including their
pattern, type, and cause.

Lost space-time data points
In addition to the foregoing issues, there are space and time
points without any valid registered data. For instance, national
DRS data set is not available for 2005. Moreover, there are several
provinces whose DRS has not registered death records in some
years. Furthermore, the present registration system did not collect
data for all periods of study (from 1990 to 1994), thus forcing us
to estimate results for these lost space-time data points.
Due to issues in DRS data, studying cause of death is a critical
and time-consuming process. Firstly, a wide range of methods
was applied for preparing data for future analysis. Secondly, we
used several statistical models to estimate mortality rates by age-
sex-province groups for all causes of death. Knowledge about
all steps of these processes is useful in similar experiments, and
is presented in detail. A graphical illustration of these steps is
provided in Figure 1.

Step 1: Key variables of this study, including age, sex, cause of
death, residency area, and nationality were extracted from all data
sets. Sex and nationality variables were recoded to unique values.
Unusual values were registered for age in all data sets because
of insensitive data entry process. That is why we recoded ages
above 115 into missing values. The threshold 115 years of age
was chosen as longevity above that is unlikely among the Iranian
population.18
We reformed new provinces based on the administrative division
of 2011 for each data set:

1) In data set 1995 – 2001, which was based on coverage areas
of each Medical University, it was assigned to existing provinces
in 2011.
2) In data set 2001 – 2004 provinces were reformed to the same
as data set 1995 – 2001 and there was no case of mismatch.
3) In data set 2006 – 2010, we used district variable to reshape
provinces based on administrative division in 2011. Province of
residency or Medical University that registered death were used
respectively to fill missing district.
4) In Tehran data set, we assumed that all records were related
to Tehran or adjacent cities so we considered them attributable to
residents of Tehran province.
5) In Isfahan data set, we assumed that all records were related
to Isfahan or adjacent cities so we considered them attributable to
residents of Isfahan province.

Finally, there was no missing value in province variable.
An expert physician in ICD10 explored causes of death.
Seventeen chapters of ICD10 were eligible to be considered in
studying cause of death including chapters 1 – 19 except chapter
7 (diseases of eye and adnexa) and chapter 18 (Symptoms, signs
and abnormal clinical and laboratory findings, not elsewhere
classified).19 Wrong codes like “it will be specified later” or
“YYY” and garbage codes in chapter 18 of ICD10 were recoded
into missing value. In Tehran, causes of death were registered
according to physician’s diagnosis, not ICD10. A team of
physicians assigned ICD10 codes to diagnosis and several
possible codes in unclear cases.

In many cases, garbage codes, or aggregation of several codes
were registered as underlying cause. Possible ICD10 codes were
considered in these situations. We provided a frame of age and sex
restriction for all causes. It was assumed that registration of wrong
cause of death is more likely than wrong age or sex. Based on this
frame, we recoded impossible causes of death into missing values.
We made an exception in data set 2006 – 2010 to deal with sex
restriction by checking first name if necessary.
Lastly, we reached the three following kinds of causes of death:
1) Specified underlying causes with ICD10 codes, 2) Missing
values (when there was no information about underlying cause)
and 3) Several possible ICD10 codes according to incomplete
information.

Step 2: In step 1, all data sets were combined to form the main
data set. In the present step, we aimed to impute all missing
values. We used a multiple imputation approach using Amelia
package20 in R statistical software for age and sex variables
which did not exceed 5% of missing values in these variables. In
addition, a multinomial imputation was performed using STATA
11 software21 for causes of death.
Amelia performs multiple imputations, a general-purpose approach to deal with missing values. Multiple imputations can reduce bias and increase efficiency compared with other methods. Amelia also uses a bootstrap and EM algorithm to impute missing values from a data set. This approach produces multiple output data sets for analysis. We introduced the main data set to Amelia as a time series data. Year was considered as time variable and province as cross section variable. Imputation was performed in the presence of age, sex, and dummy variables of ICD10 chapters as covariates. The value of 115 was considered as maximum level of age variable and number of imputation was set to five.

After each imputation, we checked age and sex restrictions. Imputed values of age or sex were recorded into missing value if there were any restriction. Amelia imputed missing values of rechecked data set again. This process was repeated until no missing value and age-sex restriction existed in the data set. At the end of this stage, we combined outputs to achieve age and sex imputed versions of the primary data set.

We applied a multinomial imputation, a two-stage approach to impute cause of death variable. At first stage, we assigned ICD-chapters into cause of death variable; then, missing ICD-chapters
were imputed using the following multinomial logistic model:

\[
\Pr(x_i = k | z_i) = \begin{cases} 
\frac{1}{1 + \sum_{l=2}^{K} \exp (z_i^l \beta_l)}, & \text{if } k = 1 \\
\frac{\exp (z_i^k \beta_k)}{1 + \sum_{l=2}^{K} \exp (z_i^l \beta_l)}, & \text{if } k > 1 
\end{cases}
\]

Where \( Z_i \) is the vector of predictors for observation \( i \) and \( \beta_i \) is regression coefficient for outcome \( l=2,...,K \). Here we used age, sex, and year as covariates. At the second stage, the related causes were imputed in each ICD chapter. To avoid imputing unacceptable causes, a process of checking age-sex restriction was done after second stage. Causes of death detected in this process were recoded into missing value. We repeated this approach to the extent that there was no missing value in cause of death.

**Step 3:** As mentioned above, there were three situations for cause of death variable after step 1: first, cause of death matched to specific underlying cause and ICD10 code; second, missing values in cause of death (that were imputed in step 2) and finally, uncertainty in cause of death according to garbage codes. In this step, we aimed to redistribute deaths of the latter situation to other specified cause of death. First, we calculated number of deaths by age, sex, province, cause, and year. Then, we redistributed deaths of garbage codes on deaths of probable cause using proportionality approach in related age, sex, and cause combination in each year. There were situations where proportion of death could not be calculated in some age and sex combinations; we, therefore, eliminated age effect on death distribution.

**Step 4:** Obtaining comparable results requires standard classification codes for causes of death. Although ICD10 is the most popular classification disease approach, GBD codes were preferable as they enable us to compare results with the GBD study. An expert physician familiar with definition of diseases and causes of death explored the definitions of all codes in GBD and ICD10. As a result, a map was created to identify relations between codes in the two classification systems. Finally, comparing GBD and ICD10 showed three kinds of relations between codes in these systems:

1) ICD10 in Iranian DRS classified cause of death in more detail than did GBD. In this case, we had several codes in ICD10, which corresponded to one code in GBD. We called them many to one (MTO) codes. Dealing with MTO codes was easily achieved by aggregating deaths related to many codes in ICD10 into one code in GBD.

2) ICD10 in Iranian DRS and GBD had the same definition for a cause of death and allocated one code for that. We called these codes one to one (OTO) codes. We simply recoded these codes.

3) GBD classified cause of death in more detail than did ICD10 in Iranian DRS. The main issue in recoding causes of death into GBD codes was related to this situation. Thus, we had an ICD10 code (OTM codes) that had to be assigned to several GBD codes proportionally.

A proportional redistribution approach was considered to deal with OTM codes. This approach needed the prior distribution of GBD cause of death in each age, sex, and year. This information was available for some years in the GBD study. For other years, we predicted number of deaths using the following Poisson regression model with logarithm link function. In this model, years of schooling (YOS), wealth index (WI) and urbanization ratio were considered as covariates. In addition, logarithm of population was added to model as offset to adjust results:

\[
\log(\text{number of death for cause } i) = a + \log(\text{population}) + \beta_1 \text{sex} + \beta_2 \text{age} + \beta_3 \text{year} + \beta_4 \text{YOS} + \beta_5 \text{WI} + \beta_6 \text{urbanization ratio}
\]

**Step 5:** Mohamadi, et al. have conducted a parallel study on the same data sets in order to predict levels and trends of child and adult mortality rates in the Islamic Republic of Iran from 1990 to 2015. They applied a Spatio-temporal model and Gaussian process regression to predict rates of death for all provinces of Iran.\(^4\) We used their results to deal with incompleteness of DRS. According to their results, we just had all-causes mortality rates in 31 provinces for males and females. Therefore, we had to assume that incompleteness of all causes was equal and the rates of death by the same scale parameter for all causes of death were changed. All cause specific mortality rates were rescaled by the above-mentioned approach except for rates of death in Alborz province, which has separated only recently from Tehran in 2010. Therefore, many cases of death related to this province were registered in Tehran. Exploring trend of incompleteness in Alborz confirms this issue. Unfortunately, the high rate of misclassification in most periods of study caused sparse or empty cells in the Alborz data set. In this situation, scale up was unable to solve this problem. We used cause distribution of Tehran and Alborz in each age and sex groups and applied predicted rate of death for Alborz to identify death distribution in this province. A necessary assumption in this approach was equality of death distribution according to age, sex, and cause in both provinces. We hope that this assumption is justified due to the adjacency of the two provinces and similarity of death covariates.

**Step 6:** After the above five steps, we had a unique data set. Death rates by age, sex, province and cause were determined based on locations and years which DRS captured data points. For estimating lost location-time data points, two-stage modeling approach was applied consisting of a random intercept mixed effect model and a spatial temporal model. This approach was applied and advised by Foreman, et al. in order to model cause of death.\(^5\)

Random effect
For all combinations of age, sex, province and year, we modeled cause fraction of death using the following mixed effect model:

\[
\logit(\text{cause fraction}) = a + b_1 \text{year} + b_2 \text{YOS}_{ij} + b_3 \text{WI}_{ij} + b_4 \text{urbanization ratio}_{ij} + b_5 \log(\text{Population}_{ij}) + \epsilon_{ij} \\
\sim N(0, \sigma^2_{\epsilon}), \text{ } \epsilon_{ij} \sim N(0, \sigma^2_{\epsilon})
\]

Where \( \epsilon_{ij} \) is the residual for province \( i \) and year \( j \) and \( b_i \) is the random effect related to province \( i \). By using the random effects
In LOESS local regression, we weighted all observations in time: the lower and upper bounds of uncertainty interval, respectively. Therefore, predicted cause fractions applied to total mortality rates are derived from Mohamadi, et al. study. In this model, we were able to explore the effect of each province separately. Then, the predicted values and residuals were estimated with standard deviations.

Spatio-temporal model
The effects of covariates on dependent variable were explored and considered by mixed effect model. Spatial temporal model was utilized to take into account how the dependent variable varies further across time, space, and age. The model assumes that residuals contain valuable information that is not directly observed, but varies systematically across space, time, and age group nonetheless.

In order to implement the spatio-temporal model, we assumed that each observation affected all other observations. These relations and influences were calculated by a weighting system. As mentioned above, three levels were considered and a weight matrix was defined for each dimension.

The first dimension calculated weights of age groups using a simple decay function:

\[ W_{a_{ij}} = \frac{1}{e^{\omega(\text{age group}_{i} - \text{age group}_{j})}} \]

Where \( \omega \) controls the smoothing level over age. It could be decreased to have more degree of smoothness.

Using a weighting scheme similar to the Tricubic weights used in LOESS local regression, we weighted all observation in time:

\[ W_{t_{ij}} = (1 - \frac{|\text{year}_{i} - \text{year}_{j}|}{\text{argmax}(|\text{year}_{i} - \text{year}_{j}|) + 1})^3 \]

Where \( \lambda \) was defined as a parameter of smoothing across time. We used \( \lambda \) of 2 in order to avoid issues of compositional bias and sparse data.

Step 7: After modeling cause fractions in all combination, it is necessary to calculate mortality rates in these subgroups. Therefore, predicted cause fractions applied to total mortality rates are derived from Mohamadi, et al. study.

All estimations should be reported with a value of uncertainty. Variance of mixed effect predictions were calculated using common likelihood methods. To expand these variations for results, simulated data was generated and explored. For each combination, we generated 1000 random normal values from distribution of predicted parameters. Spatio-temporal model was applied on these values and the 2.5th and 97.5th percentiles of results formed the lower and upper bounds of uncertainty interval, respectively.

Results
In this study, we explored model assessments and data issues. Other topics including reports of cause-specific mortality rates will be presented in other articles. Percentage of missing values, crude death rates (CDR), distribution of garbage codes, and proportion of three major cause of death types including communicable disease (CD), non-communicable disease (NCD) and injuries were reported to show data attributes.

In the DRS data set, 58.91% and 41.09% of deaths pertained to males and females, respectively. Most cases of death were classified as NCDs (77.83%) and injuries (14.80%). Exploring number of deaths showed that 56.53% of CDs, 56.05% of NCDs and 75.12% of injuries pertained to males according to DRS. In addition, the proportion of under-five-year deaths was 7.92% of all deaths.

Percentages of missing values in age and sex are depicted in Figure 2. Although missing patterns within data sets are the same, there are several differences between data sets. It appears that all systems had achievements to control missing errors according to the decreasing trend observed in these plots. For instance, in the Isfahan cemetery, 0.51% of missing values were decreased to 0.1% from 2007 to 2010, and in the same period, missing values in age were reduced from 2.39% to 0.84%. Moreover, national data registry had a lower percentage of missing values compared with cemetery registration. Tehran cemetery registration had the highest missing values in sex in comparison with national values in all years. On the other hand, the range of missing proportion was different among deputies of MOHME. For instance, the percentage of missing values in sex varied from 0.06% to 0.51% in the deputy for public health compared with 0.43% to 3.78% in the deputy for research and technology.

Figure 3 shows CDRs for males, females, and both sexes in 1995 (first death registration year, administered by the deputy for research and technology), 2001 (first death registration year, administered by the deputy for public health), 2004 (reform in DRS), and 2010 (final year with available data). The national registered CDR per 1000 population increased from 2.87 in 1995 to 4.31 in 2010. Under-five values for CDR are considered as incompleteness of DRS. Therefore, the majority of provinces were subject to incompleteness in DRS. Despite this issue, we could observe a progressive trend over the years of the study. The minimum values of CDR pertained to Golestan (1.04) and Kohgiluyeh and Boyer-Ahmad (2.89) in 1995 and 2010, respectively. The maximum values in these years are 4.99 in 1995 (Sistan and Baluchestan) and 6.08 in 2010 (Isfahan).

Quality of registering cause of death is crucial in this study. Therefore, we present the national and sub-national distribution of cause of death in Figures 4 and 5. According to Figure 4, communicable, maternal, neonatal, and nutritional diseases are responsible for 7.98% of deaths with valid causes of death in 2010. The majority of deaths pertained to the category of non-communicable diseases with 78.42%, and 13.60% of deaths were attributed to injuries in this year. Although the proportion of all three main categories have increased as a result of the decreasing proportion of garbage and missing codes, NCDs spread on a larger scale due to global changes in cause of death. The percentage of missing values in cause of death variable decreased from 63.83% in 1995 to 13.28% in 2010. Tehran was notably different from other provinces with the highest proportion of garbage and missing codes with 38.29% in 2010. Other provinces with the highest proportion of garbage and missing codes with 38.29% in 2010.

We explored 567 ICD codes including underlying cause of death and garbage codes in the primary data set. Mapping ICD to GBD resulted in 165 possible causes of death according to GBD classification of diseases.

Discussion
Developing global assessments for causes of death commenced in 1980 with some studies on several causes. The comprehensive,
Figure 2. Distribution of missing values in age and sex in DRS from 1995 to 2010.

Figure 3. Provincial distribution of crude death rates for males, females and both sexes in 1995, 2001, 2004, and 2010.
**Figure 4.** Trend of proportional death by major class of causes of death.

**Figure 5.** Provincial proportional death by major class of causes of death in 1995, 2001, 2004, and 2010.
independent and evidence-based approach to public health policy was utilized in the GBD study in 1990. GBD is a global effort to quantify the magnitude of loss of health and the contributing key causes. In addition, during the last decade, some studies were conducted on specific diseases. However, there is no comprehensive and sub-national study on burden of diseases except in a few developed countries.

The burden of disease study in Iran was first conducted in 2003. It is important to mention that this study was at national level and six provinces without any estimates for attributed burden to risk factors and any trend analysis over time. The necessity of assessing the burden of diseases and their distribution motivated the Ministry of Health to conduct a new modified study on this topic called NASBOD in order to enhance the burden of disease studies. This study, as the main component of NASBOD, is a pioneering effort to report causes of death at sub-national level among West Asian countries.

We attempted to use all available facilities to increase the validity of results. Tehran and Isfahan cemetery data sets were added to yield adequate and comprehensive information coverage. There is no similar study that has used this amount of data. All missing values and garbage codes were exactly defined and appropriate methods were applied to deal with them. Since age and sex restrictions were checked in all steps, no information was lost or assigned to invalid groups. Cause specific death rates seem to be more reliable because of rescaling them according to incompleteness. Moreover, uncertainty intervals were measured.

In addition to the above-mentioned features of this study, there are still certain limitations, as follows: Because of computational constraints, our modeling was conducted by causes and provinces. An extended approach which considers correlation between these variables may improve the results. Furthermore, several sources of uncertainty, including redistributing and cleaning, were ignored, due to computational constraints. We only applied available covariates. It is suggested that a combination of covariates may be gathered and then a subset of them extracted using a covariate selection algorithm.

The usefulness of cause of death statistics depends on the accuracy of DRS. Several factors influence this accuracy, ignoring which may result in incorrect inferences. Among them, altering the coding technique, physician’s abilities to state the underlying causes of death and inconstancy in death registration system are critical. Unfortunately, there are inconsistencies, discontinuities, and under-reporting in Iran’s DRS, which should be modified. However, the WHO has advised countries to register the underlying cause of death in death certificates based on ICD codes; DRS in Iran did not follow this rule but fortunately many progressive reform are being conducted now. A high percentage of garbage codes is observed in data sets and Tehran death registry does not register ICD codes at all.

Comparing Tehran with national death registry revealed that registering death out of accepted and defined framework decreases the validity of results. Thus, a comprehensive and unique registration system is able to solve many presented issues. In addition, it is necessary to assess the quality and validity of cause of death data in Iran. Scientific methods like analyzing mortality level and cause-of-death data could be used to provide an overview for better decisions.

Competing interest

The authors declare that they have no competing interests.

Author’s contributions


Acknowledgments

We would like to thank the Institute for Health Metric and Evaluation (IHME) team for providing the results of the GBD study. We also thank the Ministry of Health and Medical Education of Islamic Republic of Iran, and Setad-e-Ejraie Farmane Imam for their kind help and supports.

References


