Use of the National Health Information Database for Estimating Town-Level Mortality in Korea: Comparison with the National Administrative Data, 2014–2017

Ikhan Kim,1 Youngs Chang,1 Hee-Yeon Kang,1 Yeon-Yong Kim,2 Jong Heon Park,2 and Young-Ho Khang1,3

1Department of Health Policy and Management, Seoul National University College of Medicine, Seoul, Korea
2Big Data Steering Department, National Health Insurance Service, Wonju, Korea
3Institute of Health Policy and Management, Seoul National University Medical Research Center, Seoul, Korea

ABSTRACT

Background: This study addressed town-level mortality rates using the National Health Information Database (NHID) of the National Health Insurance Service in Korea in comparison with those derived from the National Administrative Data (NAD) of the Ministry of Interior and Safety.

Methods: We employed the NHID and NAD between 2014 and 2017. We compared the numbers of population and deaths at the national level between these two data sets. We also compared the distribution of the town-level numbers of population and deaths of the two data sets. Correlation analyses were performed to investigate the relation between the NHID and NAD in the town-level numbers of population and deaths, crude mortality rate, and standardized mortality ratio (SMR).

Results: The numbers of population and deaths in the NHID were almost identical to those in the NAD, regardless of gender. The distribution of the town-level numbers of population and deaths was also similar between the two data sets during the entire study period. Throughout the study period, the Pearson correlation coefficients between the two databases for the town-level numbers of population and deaths and the crude mortality rate were 0.996 or over. The correlation coefficients for the SMR ranged from 0.937 to 0.972.

Conclusion: Town-level mortality showed significant correlation and concordance between the NHID and NAD. This result highlights the possibility of producing future analyses of town-level health-related indicators in Korea, including the mortality rate, using the NHID.

Keywords: Correlation of Data; Health Status; Mortality; Population; Republic of Korea

INTRODUCTION

Information on health indicators in small geographical areas may provide better opportunities to address specific local health problems among the most vulnerable populations. More accurate methods for measuring diseases, exposures, behaviors, and
other health measures could facilitate better assessments of population health and the
development of policies and targeted programs for preventing disease.1 In Korea, the Seventh
District Healthcare Plan that every local district government must set up by law requires
information on health indicators stratified by income and geography.2 However, there is a
paucity of information on health indicators at the town (eup/myeon/dong) level in Korea.

In some previous studies, eup/myeon/dong areas, the smallest administrative units in Korea,
were used as the unit of analysis for calculating mortality in Korea by geographical area.3-6
However, those studies only provided town-level mortality within a city3 or for both men and
women combined4,5, mainly due to limitations of the data used in those studies.

The National Health Information Database (NHID) is considered to be a data source suitable
for monitoring mortality at the town level. The NHID utilizes data from the National Health
Insurance Service (NHIS), which covers the entire Korean population.7 The eligibility
database of the NHID includes information on the gender, age, residential area, and death
status of all people enrolled in the national health insurance program.7 Also, various health
and medical information can be obtained through linkage with other databases of the
NHIS.7 A previous study showed that the NHID could be used to estimate mortality and
life expectancy at the national, provincial, and district levels.8 However, it has not yet been
explored whether the NHID can be used for calculating mortality or life expectancy at the
town level.

The purpose of this study was to examine the feasibility of estimating town-level mortality
using the NHID. We compared town-level mortality rates based on the NHID with those
derived from the National Administrative Data (NAD) of the Ministry of Interior and Safety
(MOIS), which constitutes the official national database on vital and internal migration
statistics. In this study, we first compared the overall numbers of population and deaths
between the NHID and NAD during the study period. Second, we compared the distribution
of town-level numbers of population and deaths between the two data sets. Third, correlation
analyses were conducted of the numbers of population and deaths, the crude mortality rate,
and the standardized mortality ratio (SMR) between the NHID and NAD.

METHODS

Data
In this study, we used the NHID and the NAD provided by the MOIS in 2014–2017. All subjects
present on January 1 of each year in the NHID eligibility data were followed up for one year.
The numbers of population and deaths by year, gender, 5-year age group (0, 1–4, 5–9, 10–14,
..., 85+), and eup/myeon/dong were obtained. The data set includes both the beneficiaries of
the national health insurance and medical aid programs. We excluded foreigners and cases
where a subject’s gender, age, or address was not known. The NAD is based on the resident
registration statistics provided on the MOIS website, which provides figures since 2008.9
The number of registered residents in the current year was obtained in aggregated form
according to gender, 5-year age group (0, 1–4, 5–9, 10–14, ..., 85+), and eup/myeon/dong at
the end of December of the previous year. The number of deaths was calculated annually
by gender and eup/myeon/dong. The study period was limited to 2014–2017 because the
numbers of deaths in the NAD were not provided separately for men and women between
2008 and 2014.
Unit of analysis

The unit of analysis in this study was eup/myeon/dong, categorized according to the administrative unit classification as of December 31, 2017 of the Korean Administrative Unit Classification issued by the Statistical Policy Bureau of Statistics Korea. However, eight towns (Gunnae-myeon, Jangdan-myeon, Jinseo-myeon in Paju, Gyeonggi and Geundong-myeon, Wondong-myeon, Wonnam-myeon, Imnam-myeon in Cheorwon-gun, Gangwon and Sudong-myeon in Goseong-gun, Gangwon), which are civilian access control areas (for military purposes), were excluded from the analysis. Besides, 26 towns with an average annual population of fewer than 500 persons were integrated with neighboring areas to ensure stable mortality calculations. Towns that were divided or merged during the study period were consolidated into a single unit. Ultimately, a total of 3,431 towns across the nation were analyzed.

Statistical analysis

First, we compared the national-level numbers of population and deaths according to the calendar year and gender between the NHID and NAD. Second, the distribution of the town-level numbers of population and deaths between the two data sets was compared according to year and gender during the whole study period. Third, we analyzed the correlations between the NHID and NAD for the town-level numbers of population and deaths, crude mortality rate, and SMR throughout the entire study period. The crude mortality was defined as the number of deaths per 100,000 population. SMR is a mortality measure derived from an indirect standardization method that can be used when the number of deaths by age group is unknown, or the number of deaths is tiny. This method is considered suitable for measuring differences in health status between small areas. The standard population in the SMR calculation was the total population of the data set.

Ethics statement

This study was approved by the Institutional Review Board (IRB) of Seoul National University Hospital (IRB No. E-1810-008-975) and the National Health Insurance Service of Korea (REQ0000022282). Informed consent was waived for this study because secondary data were used.

RESULTS

Table 1 shows the national-level numbers of population and deaths in the two data sets according to the calendar year and gender for the entire study period. When the data for the entire study period were combined, the resulting total number of population was 203,741,630 in the NHID, which was nearly identical to the corresponding figure of 203,748,512 in the NAD. The NHID recorded 1,101,739 men and women deaths in 2014–2017, 0.7% less than the 1,109,705 deaths recorded in the NAD for the same period. When comparing the two data sets by year, the population numbers were almost the same, and the number of deaths was 0.1%-1.3% less in the NHID than in the NAD. This tendency was similar when men and women were separately examined.

Table 2 shows the distribution of the town-level numbers of population and deaths in the two data sets for the entire study period and each year. During the entire study period, for both men and women, the median population number from the NHID was 44,680 (interquartile range [IQR], 73,004), which was similar to the median population number of 44,661 (IQR, 72,964) from the NAD. The minimum and maximum values were 4,092 and
Table 1. Comparison of numbers of population and deaths at the national level between the NHID of the NHIS and NAD of the Ministry of Interior and Safety, 2014–2017

| Gender | Year | NHID | NAD | NHID/NAD | No. of population | NHID | NAD | NHID/NAD | No. of deaths |
|--------|------|------|------|----------|------------------|------|------|----------|--------------|
| Total  | 2014–2017 | 203,741,630 | 203,748,512 | 1.000 | 1,000 | 203,741,630 | 203,748,512 | 1.000 | 1,000 |
|        | 2014   | 50,655,308  | 50,662,752  | 1.000 | 265,928 | 50,655,308  | 50,662,752  | 1.000 | 267,683 |
|        | 2015   | 50,853,193  | 50,861,629  | 1.000 | 273,965 | 50,853,193  | 50,861,629  | 1.000 | 277,472 |
|        | 2016   | 51,039,935  | 51,039,939  | 1.000 | 278,773 | 51,039,935  | 51,039,939  | 1.000 | 279,017 |
|        | 2017   | 51,193,194  | 51,184,192  | 1.000 | 283,073 | 51,193,194  | 51,184,192  | 1.000 | 285,533 |
| Men    | 2014–2017 | 101,799,286 | 101,802,715 | 1.000 | 1,000 | 101,799,286 | 101,802,715 | 1.000 | 1,000 |
|        | 2014   | 25,326,070  | 25,329,566  | 1.000 | 145,812 | 25,326,070  | 25,329,566  | 1.000 | 146,857 |
|        | 2015   | 25,414,403  | 25,418,435  | 1.000 | 148,812 | 25,414,403  | 25,418,435  | 1.000 | 150,651 |
|        | 2016   | 25,496,699  | 25,496,737  | 1.000 | 150,901 | 25,496,699  | 25,496,737  | 1.000 | 151,091 |
|        | 2017   | 25,562,114  | 25,557,977  | 1.000 | 152,578 | 25,562,114  | 25,557,977  | 1.000 | 153,870 |
| Women  | 2014–2017 | 101,942,344 | 101,945,797 | 1.000 | 1,000 | 101,942,344 | 101,945,797 | 1.000 | 1,000 |
|        | 2014   | 25,329,238  | 25,333,186  | 1.000 | 120,116 | 25,329,238  | 25,333,186  | 1.000 | 120,826 |
|        | 2015   | 25,438,790  | 25,443,194  | 1.000 | 125,153 | 25,438,790  | 25,443,194  | 1.000 | 126,821 |
|        | 2016   | 25,543,236  | 25,543,202  | 1.000 | 127,872 | 25,543,236  | 25,543,202  | 1.000 | 127,926 |
|        | 2017   | 25,631,080  | 25,626,215  | 1.000 | 130,495 | 25,631,080  | 25,626,215  | 1.000 | 131,663 |

NHID = National Health Information Database, NHIS = National Health Insurance Service, NAD = National Administrative Data.

Table 2. Distribution of numbers of population and deaths among 3,431 towns in Korea: findings from the NHID of the NHIS and the NAD of the Ministry of Interior and Safety, 2014–2017

| Gender | Year | Population | Death |
|--------|------|------------|-------|
|        | 2014–2017 | Median (IQR) | Minimum | Maximum | Median (IQR) | Minimum | Maximum |
| Total  | 2014   | 44,680 (73,004) | 4,092 | 476,523 | 280 (229) | 18 | 1,965 |
|        | 2015   | 11,177 (18,242) | 487 | 96,566 | 68 (57) | 0 | 444 |
|        | 2016   | 11,053 (18,270) | 1,019 | 119,778 | 69 (57) | 3 | 493 |
|        | 2017   | 9,707 (18,318) | 960 | 148,556 | 72 (62) | 5 | 538 |
| Men    | 2014   | 22,571 (36,193) | 2,040 | 233,607 | 152 (127) | 9 | 1,102 |
|        | 2015   | 5,638 (9,082) | 229 | 48,586 | 37 (31) | 0 | 253 |
|        | 2016   | 5,610 (9,065) | 507 | 58,729 | 38 (33) | 1 | 277 |
|        | 2017   | 5,579 (9,005) | 508 | 62,476 | 38 (33) | 1 | 278 |
| Women  | 2014   | 22,038 (36,767) | 2,005 | 242,916 | 128 (105) | 9 | 863 |
|        | 2015   | 5,554 (9,195) | 258 | 49,140 | 31 (27) | 0 | 191 |
|        | 2016   | 5,468 (9,227) | 501 | 61,049 | 32 (26) | 0 | 216 |
|        | 2017   | 5,485 (9,182) | 499 | 65,004 | 33 (27) | 1 | 212 |

NHID = National Health Information Database, NHIS = National Health Insurance Service, NAD = National Administrative Data, IQR = Interquartile range.

476,523, respectively, in the NHID, which were also similar to the corresponding values of 4,093 and 476,375 in the NAD. The median number of deaths in the NHID was 280 (IQR, 229), which was also nearly identical to the value obtained using the NAD, and the minimum and maximum values were not significantly different. This was also true when analyzed by year, gender, or both year and gender. Supplementary Figs. 1 and 2 depict graphical representations of the very similar distribution of the town-level numbers of population and deaths for the whole study period between the two data sets.

Table 3 shows the results of the correlation analysis between the NHID and NAD for the numbers of population and deaths, crude mortality, and SMR throughout the study period. In both gender, the Pearson correlation coefficient was 1.000 for the number of population, and 0.998 for the number of deaths and crude mortality, respectively. The same findings were obtained in a sub-analysis by gender. The Pearson correlation coefficient of the SMR was 0.960 for men and women combined, 0.972 for men, and 0.937 for women.

https://jkms.org
Fig. 1 presents the scatter plots and correlation coefficients of the town-level crude mortality of the two data sets according to gender throughout the study period. Regardless of gender, the town-level crude mortality of the two data sets showed an almost perfect correlation (a 45° diagonal line) for almost all towns, implying that the crude mortality rates from the NHID and NAD were nearly identical.

Fig. 2 shows the scatter plots and correlation coefficients of the town-level SMR from the NHID and NAD according to gender throughout the study period. Overall, the town-level SMRs of the two data sets were relatively more different than the crude mortality rates, but most showed a close relationship. Supplementary Figs. 3 and 4 show the scatter plots and correlation coefficients of the town-level numbers of population and deaths according to gender throughout the study period. As for crude mortality and SMR, the town-level numbers of population and deaths in both data sets were almost on the 45° diagonal line, regardless of gender.

Table 3. Results of the correlation analysis between the number of population and deaths, crude mortality rate, and SMR according to gender: findings from the NHID of the NHIS and the NAD of the Ministry of Interior and Safety, 2014–2017

| Period   | Gender | No. of towns | No. of population | No. of deaths | No. of deaths per 100,000 | SMR |
|----------|--------|--------------|-------------------|---------------|-------------------------|-----|
| 2014–2017 | Total  | 3,431        | 1.000             | 0.998         | 0.998                   | 0.960 |
|          | Men    |              | 1.000             | 0.998         | 0.997                   | 0.972 |
|          | Women  |              | 1.000             | 0.996         | 0.996                   | 0.937 |

SMR = standardized mortality ratio, NHID = National Health Information Database, NHIS = National Health Insurance Service, NAD = National Administrative Data.
DISCUSSION

In this study, the NHID from the NHIS and the NAD from the MOIS showed similar numbers of population and deaths at the national level. In particular, the number of population in the two data sets was approximately the same. The distribution of the number of population and the number of deaths at the town level was also similar between the two data sets. The Pearson correlation coefficients of the town-level numbers of population, numbers of deaths, and crude mortality from the NHID and NAD pooled over the entire study period were 0.996 or over. The Pearson correlation coefficient of the SMR ranged from 0.937 to 0.972.

Bahk et al.\(^8\) reported that crude mortality and life expectancy at the national, provincial, and municipal levels estimated using the NHID were highly correlated and concordant with data from the Korean Statistical Information Service (KOSIS). Regardless of gender, the national-level crude mortality of the NHID was 0.99 fold the KOSIS data. The concordance correlation coefficient of crude mortality at the municipal level of the two data sets ranged from 0.997 to 0.999, and that of life expectancy was 0.914 to 0.990 depending on the period and gender. This study showed that the NHID presented very similar results to the NAD of the MOIS, in addition to its high level of similarity with the KOSIS data. It has also been found that the NHID is useful for estimating regional mortality rates not only at the national, provincial, and municipal levels but also at the level of towns.\(^8\)

The NHID has several strengths over the NAD. First of all, the NAD only provides total numbers of death while the NHID provides age-specific numbers of deaths. Using the NAD data, mortality rates in each town can be derived only by the indirect standardization method while both direct and indirect standardization methods can be applied to the NHID data. Town-level life expectancy can be also calculated using age-specific mortality rates from the NHID. Second, the NHID provides information on the numbers of death according to individual linkage to mortality registration data from Statistics Korea, using unique personal identification numbers. However, the numbers of death in the NAD are aggregate data which are not based on individual linkage. Numerator-denominator bias in the calculation of town-level mortality rate would be an unresolved problem in the NAD. Finally, the NHID contains various health indicators. If the NHID is accepted as a reliable source for town-level health indicators, many health indicators could be produced and used in the future.

The correlation coefficients of the town-level crude mortality rates and SMRs between the NHID and NAD were very high, but the magnitudes of the correlation coefficients for the SMRs were smaller than those of the crude mortality rates. The tendency was particularly evident in women. The SMR is calculated by dividing the actual number of deaths by the number of expected deaths in each region. The number of expected deaths is estimated by multiplying the nationwide average age-specific mortality rate by the age-specific population in each region. Therefore, even if a small absolute difference in the number of deaths exists between the two data sets, the difference in the SMR can be relatively large compared to the difference in the crude mortality rate.

This study has limitations. First, the MOIS only provides the combined number of deaths for both gender in the data from 2008 to 2013, while the number of deaths for each gender is only available starting in the 2014 data. Therefore, the period of this study was limited
to 2014 and later. The results of correlation analysis for both gender combined for the four years from 2010 to 2013 are shown in Supplementary Table 1. The results for the years 2010–2013 were very similar to the results for the years 2014–2017. Second, there is uncertainty about the validity of using the SMR to measure mortality, as was explored in this study. Some studies have shown that a significant difference in SMR could occur if the structures of two study populations are significantly different. However, the town-level population by 5-year age groups in the two data sets was entirely consistent. When we compared the number of population, and the population proportion by 5-year age group between the two data sets used in this study, the concordance correlation coefficient in all age groups was above 0.999.

In this study, the town-level mortality calculated using the NHID of the NHIS was found to be nearly identical to and highly correlated with the town-level mortality obtained using the NAD of the MOIS. This result shows that the NHID can be used to estimate town-level mortality in Korea in the future. It is necessary to monitor the health status of each eup/myeon/dong in Korea through the SMR and life expectancy calculated using the NHID. In addition, if the NHID is considered as a reliable source for town-level health indicators other than mortality rates, many health indicators can be produced and used in the future.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Results of the correlation analysis between the numbers of population and deaths, crude mortality rate, and SMR for men and women combined: Findings from the National Health Information Database of the National Health Insurance Service and the National Administrative Data of the Ministry of Interior and Safety, 2010–2013

Click here to view

Supplementary Fig. 1
Histograms and probability density functions of the town-level numbers of population from the NHID of the National Health Insurance Service and the NAD of the Ministry of the Interior and Safety among 3,431 towns in Korea, 2014–2017. (A) Histogram and probability density function of the numbers of the population in both men and women combined. (B) Histogram and probability density function of the numbers of the population in men. (C) Histogram and probability density function of the numbers of the population in women.

Click here to view

Supplementary Fig. 2
Histograms and probability density functions of the town-level numbers of deaths from the NHID of the National Health Insurance Service and the NAD of the Ministry of the Interior and Safety among 3,431 towns in Korea, 2014–2017. (A) Histogram and probability density function of the numbers of deaths in both men and women combined. (B) Histogram and probability density function of the numbers of deaths in men. (C) Histogram and probability density function of the numbers of deaths in women.

Click here to view
Supplementary Fig. 3
Scatter plots and Pearson correlation coefficients \( r \) of the town-level numbers of population from the NHID of the National Health Insurance Service and the NAD of the Ministry of the Interior and Safety among 3,431 towns in Korea, 2014–2017. (A) Correlation of the numbers of the population in both men and women combined. (B) Correlation of the numbers of the population in men. (C) Correlation of the numbers of the population in women.

Click here to view

Supplementary Fig. 4
Scatter plots and Pearson correlation coefficients \( r \) of the town-level number of deaths from the NHID of the National Health Insurance Service and the NAD of the Ministry of the Interior and Safety among 3,431 towns in Korea, 2014–2017. (A) Correlation of the numbers of deaths in both men and women combined. (B) Correlation of the numbers of deaths in men. (C) Correlation of the numbers of deaths in women.

Click here to view

REFERENCES

1. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *Am J Prev Med* 2016;50(3):398-401. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/27216861) | [CROSSREF](https://doi.org/10.1016/j.amepre.2016.03.031)

2. Korea Health Promotion Institute. *Guideline for Establishment of the 7th Local Healthcare Plan*. http://www.khealth.or.kr/kps/publish/view?menuId=MENU00890&page_no=B2017003&pageNum=1&siteId=&srch_text=&srch_cate=28&srch_type=ALL&str_clft_cd_list=&str_clft_cd_type_list=&board_idx=10127. Updated 2018. Accessed December 8, 2018.

3. Choi MH, Cheong KS, Cho BM, Hwang IK, Kim CH, Kim MH, et al. Deprivation and mortality at the town level in Busan, Korea: an ecological study. *J Prev Med Public Health* 2011;44(6):242-8. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/21795363) | [CROSSREF](https://doi.org/10.3961/jpmph.11.0508)

4. Kim JH, Yoon TH. Comparisons of health inequalities in small areas with using the standardized mortality ratios in Korea. *J Prev Med Public Health* 2008;41(5):300-6. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/19028293) | [CROSSREF](https://doi.org/10.3961/jpmph.07-0799)

5. Shin H, Lee S, Chu JM. Development of composite deprivation index for Korea: the correlation with standardized mortality ratio. *J Prev Med Public Health* 2009;42(6):392-402. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/19585801) | [CROSSREF](https://doi.org/10.3961/jpmph.09-0352)

6. Yoon TH, Noh M, Han J, Jung-Choi K, Kang YH. Deprivation and suicide mortality across 424 neighborhoods in Seoul, South Korea: a Bayesian spatial analysis. *Int J Public Health* 2015;60(8):969-76. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/26653769) | [CROSSREF](https://doi.org/10.1007/s10795-015-0754-4)

7. Cheol Seong S, Kim YY, Khang YH, Heon Park J, Kang HJ, Lee H, et al. Data Resource Profile: The National Health Information Database of the National Health Insurance Service in South Korea. *Int J Epidemiol* 2017;46(3):799-800. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/28209633) | [CROSSREF](https://doi.org/10.1093/ije/dyw159)

8. Bahk J, Kim YY, Kang HY, Lee J, Kim I, Lee J, et al. Using the National Health Information Database of the National Health Insurance Service in Korea for monitoring mortality and life expectancy at national and local levels. *J Korean Med Sci* 2017;32(11):1764-70. [PUBMED](https://www.ncbi.nlm.nih.gov/pubmed/28745150) | [CROSSREF](https://doi.org/10.3346/jkms.2017.32.11.1764)

9. Ministry of the Interior and Safety. *Resident Registration Statistics*. http://27.101.213.4/index.jsp. Updated 2018. Accessed December 3, 2018.

10. Statistics Korea. *Korean Administrative Unit Classification as of December 31, 2017*. http://kssc.kostat.go.kr/ksscNew_web/kssc/common/CommonBoardList.do?gubun=1&strCategoryNameCode=019&strBbsId=kascrr&categoryMenu=014. Updated 2018. Accessed December 4, 2018.
11. Jonker MF, van Lenthe FJ, Congdon PD, Donkers B, Burdorf A, Mackenbach JP. Comparison of Bayesian random-effects and traditional life expectancy estimations in small-area applications. *Am J Epidemiol* 2012;176(10):929-37.

12. Gordis N. *Epidemiology*. 5th ed. Philadelphia, PA: Elsevier Saunders; 2014.

13. Goldman DA, Brender JD. Are standardized mortality ratios valid for public health data analysis? *Stat Med* 2000;19(8):1081-8.

14. Julious SA, Nicholl J, George S. Why do we continue to use standardized mortality ratios for small area comparisons? *J Public Health Med* 2001;23(1):40-6.

15. Silcocks PB, Jenner DA, Reza R. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. *J Epidemiol Community Health* 2001;55(1):38-43.