# Coverage uncertainty range: A new method for calculating uncertainty around summary statistics in healthcare quality indicators

Kari Krizak Halle[1], Tormod Aarlott Digre[2], Ragna Elise Støre Govatsmark[1] and Torunn Varmdal[1]

*1) Department of Medical Quality Registries, St. Olav's University Hospital, Trondheim, Norway*
*2) InMeta Consulting, Trondheim, Norway*

Correspondence: Kari Krizak Halle, St. Olav's University Hospital, P.O. Box 3250 Torgarden, NO-7006 Trondheim, Norway
E-mail Kari.krizak.halle@stolav.no Telephone +47 916 97 926

## ABSTRACT

Data from clinical health registries, such as medical quality registries, are often used as basis for healthcare quality indicators (QI). To aid the interpretation of quality indicators and support decisions, it is important to quantify the uncertainty around the QI summary statistics. In this paper we suggest a novel method for quantifying such uncertainty: the Coverage uncertainty range. The method is based on the size of the population present in the register relative to the total relevant population and does not make any assumptions about the sampling strategy or the value of the summary statistic. Furthermore, using both simulated data and real-life data from a Norwegian medical quality registry, we illustrate why using confidence intervals when presenting healthcare quality indicators may lead to erroneous conclusions.



## INTRODUCTION

During the past decades, measuring quality in healthcare has become a major topic for policymakers, researchers and health personnel and effort has been made to develop systematic methods of benchmarking quality of care of different healthcare providers (1). A key point of quality indicators is that they apply to the actual, observed results for a given health provider in a given time period, and thus can be classified as descriptive statistics. This utility of health registry data differs from that of epidemiological research where the aim is to generalize findings to a larger, unknown population by use of a sample population and inferential statistics. Confidence intervals are frequently used in inferential statistics to quantify uncertainty around point estimates, but there is a lack of methods for quantifying uncertainty around summary statistics. In this paper, we propose a new method for quantifying uncertainty around summary statistics when the data completeness is known: the Coverage uncertainty range method.

It is well known that data completeness in health registries vary substantially (4,5). Even within the same registry, completeness often varies between different health care providers or between different variables in the registry. Consequently, the data used as basis for quality indicators holds a certain amount of uncertainty. The new method suggested in the present paper is suitable for quantifying the uncertainty around summary statistics of observed results in a known, finite population when the outcome of interest is binary (e.g. yes/no). This new method gives a lower and upper bound for uncertainty around a summary statistic based on the completeness of the data and does not rely on any assumptions about the sampling strategy or the value of the summary statistic, in contrast to standard formulas for confidence intervals.

## MATERIAL AND METHODS

We used simulated data to illustrate the Coverage uncertainty range method. Then we applied the method to real life data from a Norwegian national medical quality registry; the Norwegian Registry of Myocardial Infarction (NORMI). In this study, we only made use of publically available statistics (3), so approval by the Regional Committee for Medical and Health Research Ethics was not required.

### *Norwegian Registry of Myocardial Infarction*

NORMI is one of 58 national medical quality registries in Norway (2023). The registry collects information about all patients hospitalized for acute myocardial infarction (AMI) in Norway, and contains person identifiable information on symptoms, diagnostic procedures, treatment and follow-up after discharge from hospital. All Norwegian hospitals are obliged to report to NORMI, and more than 100 000 cases of acute myocardial infarctions have been included in the registry during the period 2013-2023. Studies have shown that the data quality in NORMI is satisfactory (8, 9). The main purpose of NORMI, as for all national medical quality registries, is to contribute to improved quality of care and reduce non-random variation between health providers (3). NORMI has established a set of 14 quality indicators expressing key elements in the quality of health care for AMI patients. To illustrate the Coverage uncertainty range method, we used the quality indicator *Proportion of patients treated with primary percutaneous coronary intervention (PCI)*

*within 120 minutes after first medical contact*. The quality indicator is calculated as the proportion of cases fulfilling the requirement compared to the total population relevant for the indicator (3).

### *Data completeness*

Data completeness in health registries can be defined on two levels, case completeness and variable completeness (6). Case completeness is often referred to as the coverage, defined as the proportion of the total population present in the register. Variable completeness can be defined as the proportion of cases not containing missing values. Consequently, data completeness can be defined as *case completeness x variable completeness.* For instance, if Hospital A has 80% case completeness and 95% variable completeness for a given variable, data completeness equals 76% (80% x 95%).

NORMI's coverage (case completeness) is calculated annually by comparison to The Norwegian Patient Registry (NPR). NPR includes all patients admitted to Norwegian hospitals, and the population in NPR relevant for NORMI are patients registered with an ICD-10 code I21-I22 (acute myocardial infarction). By the use of personal id-numbers the population registered in NORMI can be compared to the population registered in NPR using the following formula for coverage:

$$\frac{Number\ of\ cases\ registered\ in\ NORMI}{Number\ of\ cases\ only\ registered\ in\ NORMI + Number\ of\ cases\ only\ registered\ in\ NPR + Number\ of\ cases\ registered\ in\ both\ NORMI\ and\ NPR}$$

In 2020, the overall coverage of NORMI was 89%, however, coverage per hospital varied from 77% to 97% (3).

### *Confidence intervals*

A confidence interval is a range of probable estimates for an unknown parameter. The width of a confidence interval depends on both the sample size and the statistical distribution of the unknown parameter (7).

The unknown parameter could for example be the proportion of patients given a specific treatment, represented as a quality indicator. The population present in the registry is used to calculate the summary statistic.

For binomial distributed variables, the formula based on a normal approximation is commonly used when calculating confidence intervals. This formula is based on the central limit theorem, which gives a confidence interval that is invalid when the sample size is small or when the summary statistic is close to 0% or 100%. Perhaps more suited for quality indicators is the Wilson score confidence interval, which is asymmetric and have better performance for small sample sizes or summary statistics close to 0% or 100%. The results in this paper are presented using a 95% Wilson score confidence interval. However, both the confidence interval based on the normal distribution and the Wilson score interval depends on the sample size; when the sample size increases, the confidence interval narrows.

The outcome considered in this paper is binary, e.g. each case (patient) has either met or not met the criteria for the QI, and each patient is either registered or not registered in the registry. The Coverage uncertainty range method is based on identifying a hypothetical lower and upper bound for the true observed value, based on data completeness. Arguably, if the data completeness is low, a large amount of uncertainty may be associated with the summary statistic, while high data completeness means less uncertainty.

For illustration, we use the NORMI quality indicator *Proportion of patients treated with primary percutaneous coronary intervention (PCI) within 120 minutes after first medical contact*. Using standard probability notation, we define the following events:

*B: The patient is treated with PCI within 120 minutes after first medical contact*
*$B^C$: The patient is not treated with PCI within 120 minutes after first medical contact*
*R: The case is present in the register*
*$R^C$: The case is not present in the register*
*V: The variable contains data*
*$V^C$: The variable does not contain data*
*Q: The case is registered and the variable contains data*
*$Q^C$: Either the case is not registered or the case is registered but the variable does not contain data*

Data completeness is defined as $P(Q)=P(R) \bullet P(V)$. We are interested in estimating the true proportion of patients treated with PCI within 120 minutes per hospital. Using the law of total probability,

$$P(B) = P(B \cap Q)+P(B \cap Q^C)$$

and using Bayes' rule

$$P(B) = P(B|Q)\cdot P(Q)+ P(B|Q^C)\cdot P(Q^C)$$

where $P(B|Q)$ is the estimated proportion treated with PCI within 120 minutes based on the data from the register and the data completeness is given by $P(Q) = P(R) \bullet P(V)$, where $P(R)$ is the coverage of the register, and $P(V)$ is the probability that the actual variable contains data. All patients are either included or not included in the register, that is, $P(R^C) = 1 – P(R)$.

*Worst-case and best-case scenarios*

The proportion treated with PCI within 120 minutes among cases where data is missing, $P(B|Q^C)$, is unknown, so instead we consider the two extreme cases. The worst-case scenario is when none of the patients with unknown status were treated with PCI within 120 minutes, that is, $P(B|Q^C)=0$. The best-case scenario is when all patients with unknown status were treated with PCI within 120 minutes, that is $P(B|Q^C)=1$.

These two extreme values for $P(B|Q^C)$ give the lower and upper bounds for $P(B)$, the proportion of patients with the outcome of interest, which we define as the Coverage uncertainty range.

$$[P(B|Q)\cdot P(Q), P(B|Q)\cdot P(Q)+(1-P(Q))].$$

We denote $P(B)=p$, $P(R)=d$ and $P(Q)=q$. Then, the coverage uncertainty range is given by

$$[p\cdot d\cdot q, p\cdot d\cdot q+(1-d\cdot q)].$$

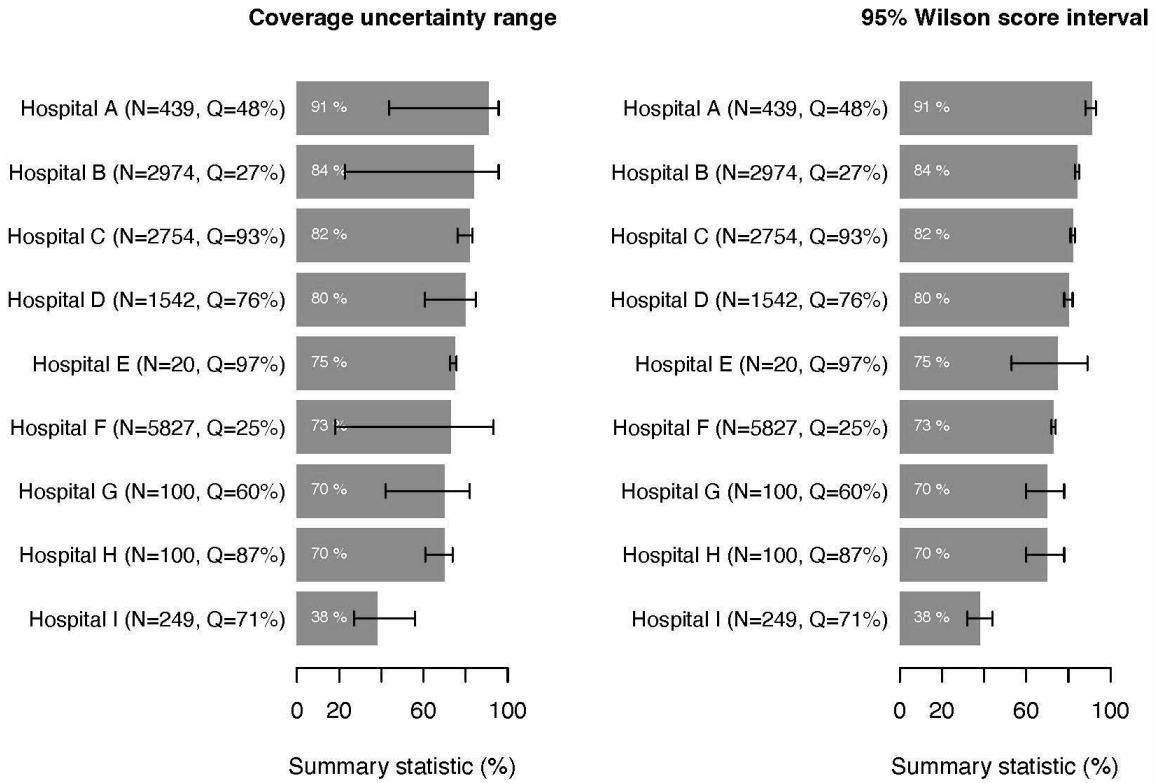


**Figure 1.** Simulated data. Summary statistics with Coverage uncertainty range and 95% Wilson score interval for a hypothetical quality indicator and nine hypothetical hospitals.

**Table 1.** Definitions.

| Term | Definition |
|---|---|
| Quality indicator | A standardized, evidence-based measure of health care quality |
| Coverage | The proportion of the total population present in a registry, also called case completeness |
| Case completeness | Case completeness is often referred to as coverage |
| Variable completeness | The proportion of cases not containing missing values |
| Data completeness | Case completeness x variable completeness |
| Coverage uncertainty range | The range which includes the true value of the observed proportion. |
| Confidence interval | The range, with a given probability, of probable estimates for an unknown parameter. |

## RESULTS

### *Simulated data*

Simulated data for summary statistics and data completeness for nine hypothetical hospitals is provided in Figure 1. The simulated data imitates different real-life cases from medial quality registry data, included hospitals with equal number of included patients but different data completeness and hospitals with a large number of included patients but low data completeness. In our experience, the simulated data is relatively comparable to real life data from a large medical quality registry; however, we have included a couple of extreme examples to better illustrate the difference between the new method and standard confidence intervals.

For the simulated data, we calculated data completeness using hypothetical coverage and variable completeness for all hospitals, and data completeness was calculated as variable completeness x coverage ($P(Q)=P(R) \cdot P(V)$). Observe that hospitals G and H both have N = 100 patients and summary statistic p = 70%, but different data completeness. Also note that hospital B have summary statistic p = 84% and N = 2974 patients.

Moving on, we add the Coverage uncertainty range (Figure 1, left panel) and 95% Wilson score intervals (right panel). Hospitals H and G both have N=100 cases included in the registry and the Coverage uncertainty range for hospital H (60% data completeness) is wider than the range for hospital G (87% data completeness), reflecting a greater amount of uncertainty around the summary statistic for the hospital with lowest data completeness. The Wilson score intervals for hospitals G and H, on the other hand, depends on the summary statistic and the sample size and have therefore the exact same width.

Furthermore, hospital B has N=2974 cases included in the registry but a data completeness of merely 27%, indicating that there is a large amount of uncertainty in the summary statistics for this hospital. The difference between the Coverage uncertainty range and the 95% Wilson score interval is illustrated in Figure 1. The left panel shows that the Coverage uncertainty range for

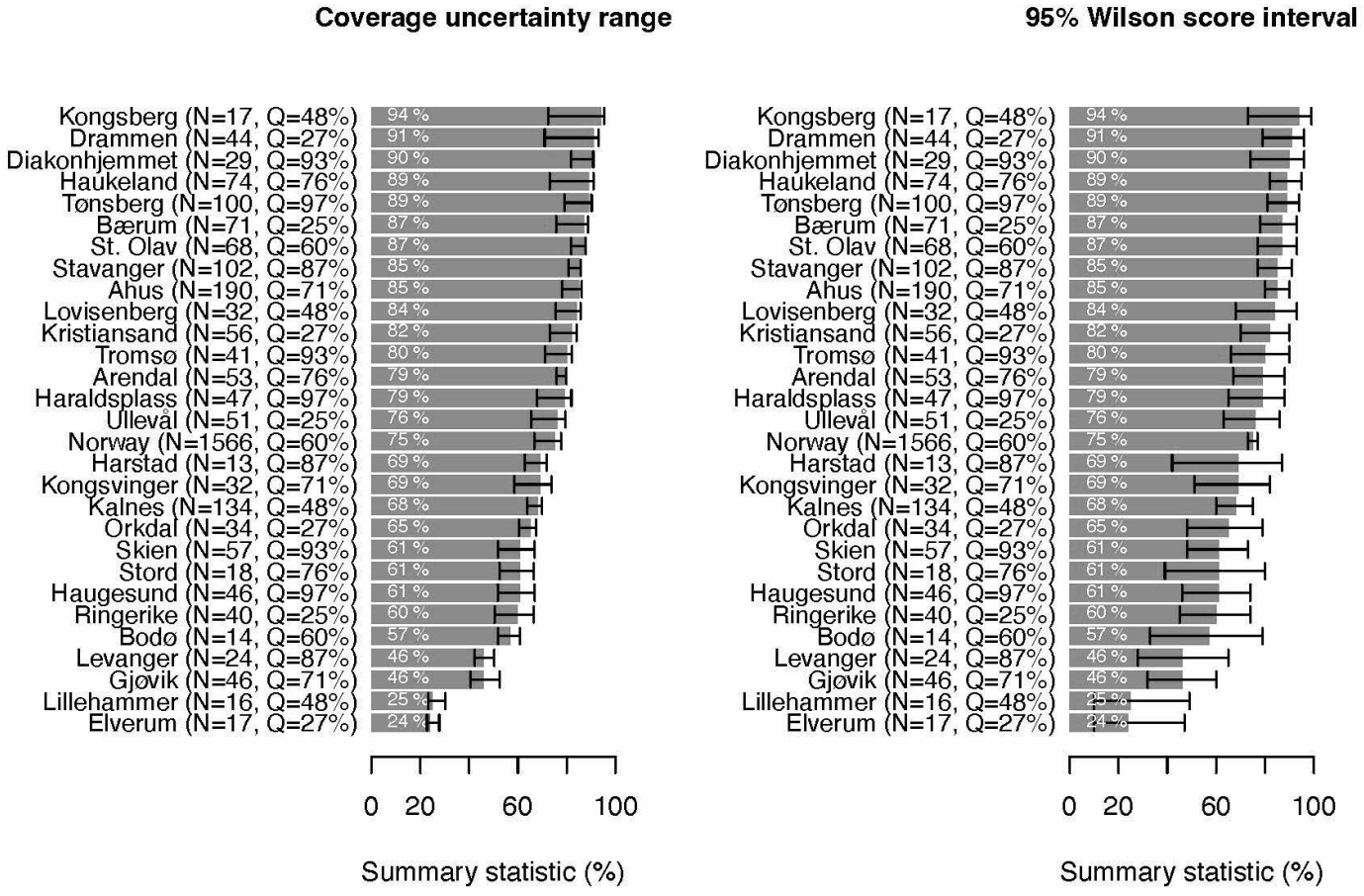


**Figure 2.** Real life data. Summary statistics with Coverage uncertainty range and 95% Wilson score interval. Data from the QI "Primary PCI within 120 min", the Norwegian Myocardial Infarction Registry, 2020.

hospital B is wide, while the right panel shows a narrow confidence interval around the summary statistic for this hospital. This illustrates that using confidence intervals as a measure of uncertainty when the data completeness is low and the sample size is large may cause the summary statistic to appear more robust than it really is, which may lead to erroneous conclusions.

### *Real life example: The Norwegian Myocardial Infarction Registry*

We used publically available, anonymous, aggregated data from the Norwegian Myocardial Infarction Registry to illustrate the new method. Figure 2 presents the actual results for the NORMI quality indicator *Proportion of patients treated with primary percutaneous coronary intervention (PCI) within 120 minutes after first medical contact* for the year 2020. Summary statistics with Coverage uncertainty ranges and 95% Wilson score intervals are shown in the left and right panel, respectively. From the publically available data, only information about the summary statistics and coverage were available, while the variable completeness was unknown. Consequently, for simplicity, we used the coverage as an estimate of the data completeness. The data completeness for the different hospitals varied from 77% at Kongsberg hospital to 96% at Arendal Hospital. At Kongsberg Hospital, 94% of the patients got primary PCI within 120 min. However, the data completeness for Kongsberg Hospital was only 77%, hence; data for 23% of the patients were missing.

Using the formula described in the methods section of this paper, we calculated that if all the patients missing from Kongsberg Hospital did *not* receive primary PCI within 120 minutes, the summary statistic for Kongsberg would be 72% instead of 94%. Similarly, if all missing patients did get primary PCI within 120 minutes, the summary statistic for Kongsberg Hospital would be 95%. Consequently, the Coverage uncertainty range for Kongsberg Hospital is [72%; 95%]. Figure 2 also shows that the confidence intervals overlap for Lillehammer and Gjøvik hospitals, while the Coverage uncertainty range identifies that there truly is a difference in the summary statistics between the two hospitals. For Elverum hospital, data completeness of 95% gives a narrow Coverage uncertainty range, while the 95% Wilson score interval indicates considerable uncertainty around the summary statistic, due to a low number of patients.

## DISCUSSION

In this paper, we have presented the Coverage uncertainty range method for quantifying uncertainty in healthcare quality indicator summary statistics. The

method can be used when the outcome of interest is binomial (e.g. yes/no) and data completeness is known or can be satisfactorily estimated.

Healthcare quality indicators are measured and published regularly, and their main purpose is informing decision makers, healthcare providers and the public about the quality of care of different healthcare providers. In the world of descriptive statistics, we argue that there is no uncertainty around the observed result for a given quality indicator if data completeness is 100%. In such cases, the results are valid independent of the sample size even if the sample size is small, as all observed cases are registered for the relevant variables and consequently we have information about the entire population.

However, 100% data completeness in health registries is extremely rare, and there is a need for a valid measure of summary statistics uncertainty in order to correctly interpret and use quality indicator results, for example for benchmarking purposes. Data can be missing at random, or there can be a systematic error in a health registry leading to selection bias (2). A great advantage of the Coverage uncertainty range is that the method is not affected by potential selection bias.

The results presented in this paper show that using a method depending on the sample size and not the data completeness to quantify the uncertainty in quality indicator summary statistics may lead to erroneous conclusions. A hospital with a large number of patients but low data completeness will have a narrow confidence interval, even if this situation indicates a large amount of uncertainty in the results. For two hospitals with the same number of patients, the width of the confidence intervals will be equal, even if data completeness for the hospitals differs.

Figure 2 shows the results using either Coverage uncertainty range or a 95% Wilson score interval to describe the uncertainty around a summary statistic. Even if the bounds of the Coverage uncertainty range may be extreme (based on "worst-case" and "best-case" scenarios), the method successfully identifies significant differences in hospital performance, also when the Wilson score interval fails to do so.

A limitation with the Coverage uncertainty range method is that it requires that data completeness is known or can be satisfactorily estimated. Furthermore, the method only considers the two extreme cases, best-case versus worst-case scenario. The patients with unknown status may or may not be missing at random, but the degree of selection bias is often unknown. However, an advantage of the Coverage uncertainty range is that the method is not affected by potential selection bias. The results presented in this paper illustrate that the Coverage uncertainty range method adds valuable information when comparing hospital performance.

For cases with incomplete data on a certain quality indicator, supplementary information (such as age and gender) may still be available within the registry or by linkage with other data sources. Consequently, an interesting topic for future work will be to make use of such supplementary information to further develop the Coverage uncertainty range to present a more precise summary statistic. Furthermore, investigating potential use of the method on nominal data, listing each missing case with the lowest/highest value, could be of interest. For example, the 99$^{th}$ percentiles for blood tests or other physiological measurements (eg. height or weight) where a normative consensus for extreme values exists.

## CONCLUSION

The Coverage uncertainty range is a measure of uncertainty around a summary statistic that only depends on the data completeness and not the sample size, the value of the summary statistic or the sampling strategy. The results presented in this paper illustrate that the Coverage uncertainty range adds valuable information when comparing hospital performance by healthcare quality indicators. Furthermore, we have shown that using Wilson score confidence intervals for quality indicators may lead to erroneous conclusions. Using this new method we show that registries can safely present their results even if some of the hospitals have low data completeness or few patients.

## ETHICS

The Regional Committee for Medical and Health Research Ethics, REC Mid Norway, has confirmed that The Health Research Act only applies to studies where non-anonymous data is used. The results presented in this paper only use aggregated, anonymous publically available data.

## REFERENCES

1. Quentin W, Partanen VM, Brownwood I, et al. Measuring healthcare quality. In: Busse R, Klazinga N, Panteli D, et al., editors. Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies [Internet]. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2019. (Health Policy Series, No. 53.) 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK549260/)
2. Ferreira-Gonzalez I, Marsal JR, Mitjavila F, et al. Patient registries of acute coronary syndrome: assessing or biasing the clinical real world data? *Circ Cardiovasc Qual Outcomes* 2009;**2**:540-7.
3. Annual Report, Norwegian Myocardial Infarction Registry, 2020, Årsrapport 2020.pdf (stolav.no)
4. Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of acute cardiovascular outcome diagnoses recorded in european electronic health records: A systematic review. *Clin Epidemiol* 2020;**12**:1095-111.

5. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Asso*c 2002;**9**(6):600-11.
6. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;**31**(4):337-50.
7. Rothman K, Greenland S, Lash T. Modern Epidemiology, 3rd edition, Lippincott Williams & Wilkins, 2008.
8. Govatsmark RE, Janszky I, Slordahl SA, et al. Completeness and correctness of acute myocardial infarction diagnoses in a medical quality register and an administrative health register. *Scand J Public Health* 2020;**48**(1):5-13.
9. Govatsmark RE, Sneeggen S, Karlsaune H, Slordahl SA, Bonaa KH. Interrater reliability of a national acute myocardial infarction register. *Clin Epidemiol* 2016;**8**:305-12.