

Increasing importance of patient ratings

A study on the validity and reliability of ZorgkaartNederland.nl

Rosanne Geesink

Master thesis Policy and Organization of Healthcare

Supervisor Dr. Xander Koolman

Vrije Universiteit Amsterdam

June 2013

Abstract

Objectives: Online patient reviews can be of great importance for enhancing the quality in Dutch hospitals. ZorgkaartNederland.nl (Zorgkaart or ZKN) is a continuous growing platform where patients can rate and review their self-selected healthcare provider. A common criticism is that mainly dissatisfied people use ZKN to evaluate their healthcare provider. The main objective of this study is to gain insight into the validity and selection bias in particular in order to examine the representativeness of the patient ratings from ZKN. Also some attention is paid to the degree of reliability of the ratings.

Study design: The results from ZKN are compared with the validated CQ-index, which uses a standardized sampling frame and contains a question that resembles the ZKN overall rating question. The distributions are compared and independent samples t-tests are used to compare the mean scores of ZKN and the CQI satisfaction question. In addition, a regression analysis is performed to study the effect of chance on the differences between ranking place of the hospitals between the two instruments.

Methods: The CQI satisfaction question used for comparison reads as follows: “How do you rate this hospital?” This one corresponds best with the items from ZKN. Ratings from 75 (out of 94) Dutch hospitals were included. Individual ratings for both instruments of all 75 hospitals together were input for the dependent variable of the t-test. The independent variable indicated the measurement instrument. Therefore the ratings from the CQI ($n=18100$) were coded as 0 and the ratings from ZKN ($n=5703$) were coded as 1. In addition, histograms about the distribution of both variables and correlation coefficients are provided.

Results: The overall mean scores from the ratings on ZorgkaartNederland.nl and those from the Consumer Quality Index are 7.84 and 7.97, respectively. According to the t-test in this study the difference of 0.13 between the two mean scores is significant ($p=0.000$). The regression analysis shows a significant B-value of -0.03. A Pearson correlation coefficient of 0.7 was found.

Discussion: Results show that the difference between ZKN and the satisfaction question from the CQI is very small in relation to the differences between hospitals. The mean scores are very similar, the distributions provide an equally similar picture and the correlation coefficient indicates a strong positive correlation between the hospital ratings of ZKN and the ratings of the CQI satisfaction question. From this study can be concluded that there is no indication that the study population differs importantly between the scientifically gathered sample from the CQI and the self-selection sample of ZKN. The self-selection sample of Zorgkaart leads to representative ratings about healthcare in Dutch hospitals, which are valid and reliable to some extent. Findings from other studies that patients put great value to the availability of reviews in addition to ratings enhance the conclusion that ZKN can play an important role as measurement instrument.

Introduction

Information generated through the measurement of quality of healthcare is very important and serves several stakeholders and user goals. Important stakeholders are the patients, healthcare providers and health insurers. Generally the insight enhances transparency, which is a priority on the political agenda¹. Health consumers (patients) need the information to base their decision for a certain healthcare provider on. This enables the contribution to care tailored to the needs and expectations of patients and the strengthening of the position of patients.

Another goal that can be achieved with this public information about quality is to hand over practical suggestions for improvement to healthcare providers. The information can be of great use to identify where opportunities for improvement can be found^{1,2}.

Also health insurers can contribute to the improvement of quality in healthcare. The quality data can be used to decide which healthcare providers are contracted. Given the structure of the current healthcare system, which is concerned with making care more demand driven, the insurer must also respond to the available information from patients' perspective^{3,4}. People are free to switch from healthcare insurer every year and seek a healthcare policy that suits their needs.

Quality and safety are important aspects for the healthcare sector since it affects people's lives. Nevertheless, with some regularity news items appear about shortcomings and medical errors in Dutch healthcare^{5,6}. Permutations in surgery for example, which can have devastating effects on people's lives, occur too often as was concluded in the recently published report from the Dutch Health Care Inspectorate (IGZ)⁷. Another example is hospital units that needed to be closed or supervised because of substandard quality^{8,9}. So there is some to gain, but therefore insight into the (quality) problems is necessary.

Quality is a quite complex concept and is judged differently by different actors^{1,10}. One general accepted definition is from the Institute of Medicine: 'Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge'. From patients' perspective, quality is defined in terms of their preferences and values about aspects like recovery, mortality or functional status¹⁰. Those can be expressed in a utility, which is a type of preference and consists of different forms¹¹. Health utilities can be used to compare the utilities of health states for any disease or treatment program¹². This makes it possible to compare aspects like disabilities in health and the process of care received between patients. Those outcomes can be used as an indicator for quality of health. Process utility is the quantification of non-health benefits that patients gain. Health consumers can derive benefit like reassurance from healthcare programs¹¹.

Quality from the perspective of care providers is more about the attributes and results of care¹⁰.

To judge and make statements about quality of healthcare, information is needed. Several measurement instruments have been developed and implemented to obtain that information. From the side of providers quality is approached with the use of performance indicators. Those indicators are measurable aspects of healthcare on different levels and consist of three types. Structure indicators measure the structure of care with regard to organizational conditions. Process indicators assess the processes of care. And the outcomes of health (an example is the satisfaction of patients) are measured with outcome indicators¹. In 2007 the IGZ developed the project ZichtbareZorg wherein care providers have to report for these substantive indicators yearly¹³. Scientific research shows that outcomes of care say a lot more about the quality then structure or process does¹⁴. To measure quality of healthcare from patients' perspective, the Consumer Quality Index (CQ-index or CQI) was developed and implemented in 2006¹⁵. It is a scientifically based and validated technique that contains a large questionnaire with 77 questions about different items that underpin care¹⁶. The questions are about hospital stay, care by caregivers and communication around the treatment for example. The questions are largely about patient experiences, but a few others ask a general judgment. In other words the satisfaction of the patient about a care institution or doctor is asked. Those satisfaction questions have to be rated by giving a grade between 0 and 10. The outcomes of the questionnaire are presented in the form of stars. At the end to each hospital a maximum of three stars can be assigned, with two stars corresponding to average healthcare. Only the experience questions are used to create an average view of a care provider. It was a deliberate choice to use experience questions instead of satisfaction questions because the first provide much more information. Respondents are selected through a protocolled way and receive the questionnaire at home. This protocol contains guidelines in order to ensure a representative sample. A CQI-research leader judges the sample for age and gender for example¹⁷.

Another tool used to get insight into the quality of healthcare through patient experiences is ZorgkaartNederland.nl (Zorgkaart or ZKN). It is a website developed by the Patients and Consumer Organizations in the Netherlands (NPCF or in Dutch: Nederlandse Patienten Consumenten Federatie) and Bohn Stafleu van Loghum (BSL) and is in use since 2010. Purposes of the website are to provide patients with information where they can base their decision for a care provider on and hand over practical suggestions for improvement to healthcare providers. Care consumers can rank, by giving a grade from 1 to 10, a care provider on six items: appointments, accommodation, employees, listening, information and treatment². The average from those items form the rating. In addition to those ratings, reviews are present wherein patients share their experience. Patients have to clarify their own rating with a written review. By making this compulsory, a person is not allowed to give unfounded ratings and the comments allow healthcare providers to learn from these evaluations¹⁸.

ZKN also offers some products and services that are available for care providers who bought a “ZorgkaartNederland package”. One of them is a service to collect more ratings and reviews, which is done with the use of so-called RoadTeams. Those teams consist of students who approach health consumers at the entrance of a care institution and questions exact the same items as on the website¹⁹.

Using health- and process utilities is also a manner to compare information about the quality of care delivered¹².

Information being available is not enough; it also needs to be usable²⁰. To use the data appropriate, validity and reliability are considered as being important^{3,21}.

Ideally a measurement instrument really measures what it intends to measure. The extent to which that is the case is called the validity of an instrument¹. Three types of bias can reduce validity: measurement-, confounding-, and selection bias³. Measurement bias is caused by differences in the registration process for example. Mistakes can be made in the measurement and registration of data or in processing the data. Confounding occur when the relation between two factors is influenced or disturbed by another factor, which is called the confounder. The difference in quality between two hospitals may be caused by the fact that one of the two deals with patients who are more sick (older people with comorbidity is an example). In that case the lower quality has nothing to do with the quality of the care delivered in that hospital but is confounded by the fact that more severe people are treated²². Selection bias concern the representativeness of the sample of people used for a measurement instrument. With representativeness is meant that the reviewers (the sample) show the same distribution of relevant characteristics as the whole population of patients²³. When this is not the case it is difficult to identify real quality of care due to selection bias. A sample that consists mainly of the sicker patients mentioned before, will not be seen as representative. There are some other forms of validity that can be looked at in order to say something about the validity of an instrument. In a couple of researches the instrument of interest is compared with another tool that measures the same outcome^{1,15}. With this technique the degree of construct validity can be examined. It is preferable to make such a comparison with an already validated instrument or a golden standard measurement. However, a golden standard for the measurement of quality in healthcare in a perfect way does not exist yet. A strong positive association between the values of both instruments results in high construct validity^{1,15}. Content validity is also a form of validity to guarantee the usefulness of an instrument³. It is the extent to which an indicator truly measures quality of healthcare. One way to ensure reliability is when the outcome of a measurement is not sensitive for chance. The amount of ratings influences the extent to which chance plays a role. Though, indicators where patient characteristics play a role are prone to chance³. From the literature it is known that mistakes which can be contributed to the rater and to environmental factors,

both affect reliability of a measurement instrument¹. Another factor is consistency of the results. When the quality stays the same over time but an instrument measures two different outcomes at two moments, the instrument might be considered as not being reliable.

Recently research showed that the reliability of the registration process for the process indicators from ZichtbareZorg is not valid. Both conscious and unintended errors occur in the registration process¹. Regardless of the type of error, mistakes induce bias that leads to outcomes not reflecting the real quality in a certain healthcare institution. So, the validity of the instrument is reduced. Besides was concluded that the indicators are hardly used by hospitals to monitor and improve quality of healthcare. The Court of Audit (Algemene Rekenkamer) adds to this that besides the disappointing usefulness, the development of those indicators was a very expensive project²⁴.

Taken note of the conclusions about the indicators to provide information about quality of care, it is interesting to look to the side of the care consumer. Between the two instruments that are mentioned before a few differences should be noted. First, ZorgkaartNederland uses only six questions to cover different aspects of healthcare in hospitals instead of the 77 questions of the CQI questionnaire. Second is that the mean score from ZKN is presented through a grade while CQI present the results with stars. Another difference is that people, who use ZKN to value, do this mainly because of intrinsic incentives. Nobody is forced to rate his/her care provider. Therefore the way in how people rate their care provider on ZKN is a spontaneous and anonymous process. This anonymity creates risks concerning the representativeness of the group of people that use the website. The last difference is the presence of reviews in addition to ratings on the website. Previous research shows that the availability of reviews in addition to mean ratings is very important and of great value for patients^{25,26}. Moreover the development and implementation of the CQI questionnaires is much more time consuming and expensive¹⁵. With the rising healthcare costs it is necessary to weigh different instruments and balance the gains and costs²⁷.

ZKN is in use now for 3,5 years and there is a continuous increase in page views, amount of visits and amount of visitors. In the past months there were about 3 million page views and an average of 650.000 individual visitors per month and more than 122.000 ratings²⁸. In order to create a reliable and meaningful platform, Zorgkaart has its own editorial office².

Nevertheless, some people and especially care providers are very critical. Most common is the critic that mainly dissatisfied and frustrated people give their value on the website. This would then be accompanied with extreme negative ratings. Basically, the representativeness of the population is put into question with those statements. With the platform becoming bigger every year it is important to look at the truth about that comment. It is quite important that ZKN reflect the real quality of care and not the individual preferences or dissatisfaction of some patients. In other words, the people that use ZKN have to be representative for the

whole population. Representativeness of the sample has an impact on the instruments reliability and meaningfulness.

The main objective of this study is to gain insight into the validity. In particular selection bias is examined because of the interest in the representativeness of the ZorgkaartNederland sample. It is also tried to view the degree of reliability of patient ratings on Zorgkaart. In order to be able to say something about those subjects, the focus in this research is on how ZKN relates to the validated instrument the Consumer Quality Index (CQI). In addition, there is given some attention to the relationship between the spontaneous ratings and the ratings generated by means of the RoadTeams. This is done to see if the results are consistent with each other, which enhances the reliability of the spontaneous ratings from ZKN.

To shape the research, the following main question is developed: “Is the ZKN study sample representative?” This question is answered through a comparison of the distribution of answers of the ZKN with a golden standard representative sample from the CQI. Both study populations were asked a similar question. If the mean and the distribution of the answers are not meaningfully different, then we conclude that the self-selection sample from the ZKN is comparable to the golden standard sample from the CQI and that the results of the ZKN are not biased as a result of selection bias.

Several research questions are formulated to answer the main question:

- What is the distribution of the ZKN and the CQI patient satisfaction questions?
- What is the difference in the mean score of both measures and is the difference significant?
- What is the difference between the hospitals in average ZKN score?
- How well do the average hospital scores of both instruments correlate?
- Is the average hospital score correlation sensitive for chance (is the correlation higher if the samples are larger)?
- What is the difference in the mean score between the spontaneous ratings and the RoadTeam ratings from ZKN?
- What is the difference between the hospitals in average score for the spontaneous ratings from ZKN?

This study aims to contribute to create a better understanding about the extent to which ratings on Zorgkaart provide information that is reliable and valid enough to use it for different purposes.

Methods

Explanation of choices made in this study

With ZorgkaartNederland both healthcare institutions and individual care providers can be assessed. However, a lot of individual providers and some institutions have very little or even no review at all. There is a great amount of ratings about hospitals though. Therefore, in order to keep the sample size as large as possible and enhance the relevance of the results, the focus of this research was on ratings about hospitals. The greater numbers of ratings for hospitals also suit the main objective of this study best, which is to get insight into the representativeness of the people that make use of ZKN.

There are a couple of reasons why the CQI was chosen to make the comparison with. Data about the characteristics of the people that rate on ZKN was not available, so assessment of the representativeness could not be based on those characteristics. Therefore the choice was made to compare the population from ZKN to another population from which it is known that it is a representative sample. This is the case for the CQI population. That population is based on a representative sample of the target population and is formed by means of a structured protocol¹⁷. The CQI instrument is validated and scientifically based, which makes it possible to say something about construct validity of ZKN. This is desirable since validity of a measurement instrument has impact on the usefulness of that instrument.

Besides the CQI contains a satisfaction question with comparable content concerning the overall assessment of hospitals and measures this aspect in the same manner. Both generate the same result, namely a grade for a certain hospital. The choice to use this satisfaction question was made because of the good comparability of both content of the question and the similar measurement scale, which is a continuous scale. The experience questions from the CQI are measured on an ordinal and dichotomous scale. Before data analysis it was tried to rescale the answers given on the ordinal- and dichotomous scale but that leads to unsuitable results that are not usable for the comparison of mean scores. It would be interesting though to be able to say something about the relation of ZKN with the total CQI questionnaire.

Therefore a correlation coefficient and regression analysis were performed with both the satisfaction question from the CQI as the total CQI questionnaire with experience questions. For the comparison between ZKN and the satisfaction question from CQI only the spontaneous ratings have been used. The scores are not based on the ratings generated through the use of Road Teams. This choice was made because the spontaneous ratings provide the great majority of input and are the most important for ZKN. Besides, the anonymity of the spontaneous ratings, and not the face-to-face way from the Road Teams, fits the way on how the CQI questionnaire is used best. Only fourteen hospitals do have

RoadTeam ratings so the majority does not. In order to compare apples with apples and not with oranges it was therefore decided to eliminate the RoadTeam ratings for this comparison.

Data collection

Via excel files the required information was available. Stichting Miletus delivered the information about the CQI data. Data from ZorgkaartNederland was supplied by the NPCF and BSL. The data from the Consumer Quality Index used in this research is based on the scores from the satisfaction question number 77 of the questionnaire CQI_Ziekenhuisopname_versie_2.1. It reads as follows: “Welk cijfer geeft u dit ziekenhuis?” or in English: “How do you rate this hospital? Rating is possible on a scale from 0 (very bad) to 10 (excellent). The ratings from the CQ-index date from 2009 because more recent data does not exist, this year is the last year the questionnaire have been implemented. The ratings from ZKN are from the years 2010, 2011 and 2012. The year 2013 was not taken into account because it is only half way and at the start of this research only ratings from January could be included. The amount of hospitals that could be taken into account for the two comparisons differs. In total for 94 hospitals CQI data was available. However, not all hospitals are present on ZorgkaartNederland or have not been assessed. That’s why 75 hospitals with their ratings could be included. A list with the names of the included hospitals for this test can be found in Annex 1. Road Teams are deployed in 14 hospitals; this means that also for only those 14 hospitals spontaneous ratings are included. A list of the names of the included hospitals can be found in Annex 2.

Preparing the data for analysis; creating data sets

In the case of ZorgkaartNederland the data consists of ratings and reviews about all sort of care deliverers placed online. From this file new files were constructed with only the necessary data, which refers to data about hospitals only. In that process a distinction was made for the spontaneous ratings and the ratings generated through the use of Road Teams. The rates from the satisfaction question number 77 were also available in a large excel file from which necessary data had to be filtered.

Before analyses can be done a data set has to be constructed and it was a great task to seek the ratings for each hospital. To begin both instruments handle their own hospital codes. To give an example: Universitair Medisch Centrum Groningen is coded as 113640 for ZKN and coded as 6020101 for the CQI. These codes were needed to search the ratings in the Excel files. Several hospitals have different locations, which are located in different cities. And each of that different location got it’s own hospital code in the case of ZKN. For the CQI data, ratings are not available per location however. So, first the individual locations with their ratings had to be searched and put together in order to get all the ratings from one “main

hospital”. The website from ZorgkaartNederland was used to examine which hospitals are part of the same umbrella hospital.

In the Excel file with data about the CQI, hospitals were expressed by name. There were different names for the same hospital however, an example is the name spelled in capital or just normal (see Annex 1). Here also ratings from the same location had to be searched and put together.

Assumptions

The information on ZorgkaartNederland should give an indication of the real quality of care delivered and should not reflect individual preferences. In order to translate and use the ratings it is therefore important that the people who give their rating are representative for the whole Dutch population. With the use of patient characteristics about age, gender and education level it is possible to see if a group is equally distributed. About the people in this study nothing is known concerning their characteristics. Though, there is information about the grades both populations give to their provider. The CQI population is based on a representative sample of the target population ⁹. Therefore it can be stated that the ratings from the CQI show a representative picture of the quality in Dutch hospitals. Comparing the distribution of the ratings from ZKN with the distribution of the ratings from the CQI provide and compare the means from both is a way to say something about the representativeness of the people that make use of ZKN and the ratings they give. This is also the way to generate insight into the existence or non-existence of extreme (negative) raters.

The main assumption about reliability of ratings is that chance plays an important role. When there are too little ratings the role of chance is often becoming too big³. Hospital scores based on few ratings is sensitive for the accidental composition of the sample. Which makes it more difficult for instruments to show the real quality. Another aspect of reliability is approached to compare the average score of the ratings from the spontaneous ratings on ZKN with the average scores from two other measurement instruments (satisfaction question from CQI and RoadTeam ratings from ZKN). When the scores are almost the same and thus are consistent with each other it strengthens the reliability of the scores provided on ZorgkaartNederland.nl.

Data analysis

The ratings given spontaneous on ZKN were compared to two other measurement methods that are the CQI and the ratings on ZKN generated through the use of RoadTeams. For those comparisons the aggregated scores were used concerning Zorgkaart. This means that the rates for each individual item (appointments, accommodation, employees, listening, information and treatment) were combined to create a mean rating or mean score per person.

SPSS versions 19 and 20 were used for the analyses. To look at the correlation of both variables, an independent samples t-test was chosen. The use of this test is appropriate to compare the mean of two independent groups who do not have a causal relationship with each other. To use the independent samples t-test in the right way, the data should first pass a few assumptions²⁹. The dependent variable is measured on a continuous scale from 0 to 10. The independent variable consists of the two independent instruments ZKN and CQI and there is no relationship between the people who gave their rating as far as known. Only the variable *RoadTeam variabelen* was normally distributed, the other three had a distribution skewed to the left. A nonparametric bootstrap (10.000 replications) allows the comparison of non-normally distributed distributions. Both output from the t-tests with and without added bootstrap show the same result.

From the output of the Levene's test for equality had to be concluded that there is no homogeneity in variances between the variables. SPSS immediately deals with this problem and provides an adapted result that corresponds with the significant F-value. This adaption has to do with a change in degrees of freedom of the test.

For a clear description about the different analyses the methods per comparison are described below.

ZorgkaartNederland versus Consumer Quality Index

In order to examine the representativeness of the spontaneous ratings from ZKN, the mean scores of the ratings were compared with the scores from the ratings of the CQI and also their distributions have been compared.

For this comparison two variables were created. *Score ZKN* contains individual mean scores from the spontaneous ratings from ZorgkaartNederland for the 75 included hospitals and *Vr.77 CQI* contains individual mean scores from the satisfaction question 77 of the CQI questionnaire for the same hospitals. The sample size (N) of the ratings included in this analysis was respectively 5703 and 18100 for Zorgkaart and the satisfaction question from the CQ-index. The distributions were plotted first and are based on the total amount of ratings for both instruments.

Before the independent samples t-test could be used a closer look was taken first to see if the variables meet the assumptions for the test. To use the t-test the variables have to be put together and are merged to create the dependent variable *Scores*. So, all the ratings of both instruments form the basis for the dependent variable. The ratings from the CQI are coded as 0 and the ratings from ZKN are coded as 1 for the independent variable.

To perform a correlation coefficient and create a scatter diagram the aggregated mean score on hospital level is needed. For all 75 hospitals the mean score was calculated for ZKN and the satisfaction question from the CQI. Two variables with both 75 mean scores were

constructed: *GemScoreZKN* and *GemScoreCQIvr77*. They both are normally distributed and therefore a Pearson correlation was done in SPSS (a normal distribution is an assumption for the Pearson correlation). Examining the correlation coefficient also gives insight in the degree of construct validity of the ratings from ZKN.

The aggregated mean scores on hospital level were also used to construct a ranking of the hospitals. This was done for the scores of both instruments.

Regression analyses to test for the effects of chance on the difference in ranking between CQI and ZKN

After the t-test that was done for the comparison between ZKN and the satisfaction question from the CQI, two linear regression analyses were conducted. The aim of those regressions is to examine if the existing differences between ratings from both instruments might be attributable to the role of chance. With this analysis the difference in ranking place per hospital between both instruments were regressed on the smallest sample size of the two samples. This allows the testing of the effect of chance on the differences between ranking place of the hospitals under the assumption that larger sample sizes are less affected by chance and thus correlation should improve if lack of correlation was due to chance.

According to the overall mean score for each hospital a ranking was conducted for both instruments. For the CQI two rankings were made. One with the use of mean scores from the satisfaction question and the other with the use of the total CQI score based on the experience questions. The dependent variable consists of the difference in ranking place. To illustrate: Ziekenhuis Amstelland is number 23 in the CQI satisfaction question ranking and number 37 in the ZKN ranking. The difference of 14 is the number that was put in the dependent variable. The independent variable is the number of the smallest sample. To illustrate: Ziekenhuis Amstelland has been rated 255 times with the satisfaction question from the CQI and 32 times on the website ZKN, then 32 is the rate that is put in the independent variable. The choice to use the smallest sample size was made because that is the determining factor for chance.

The total CQI score based on the experience questions is presented on an ordinal scale from 1-4, which is different from the continuous scale of ZKN (1-10) and therefore not good comparable. Because of this difference it is not possible to view the difference in mean scores. That is the reason the difference in ranking is used as input for the dependent variable. In order to be consistent also for the analysis with the CQI satisfaction question the difference in ranking place was used as input for the dependent variable. In addition a correlation coefficient was performed to view the relation between ZKN and the total CQI.

Spontaneous versus Road Team ratings

In order to examine the consistency of the mean scores from the spontaneous ratings with scores from other measurement instruments the comparison with Road Team ratings was made. Doing this in addition to the comparison of the mean scores with the scores from the satisfaction question of the CQI makes it possible to deepen the conclusion about the degree of reliability.

The variables created for this comparison were: *Spontane waarderingen* containing individual mean scores from the spontaneous ratings for the 14 included hospitals and *RoadTeam waarderingen* containing individual mean scores from the Road Team ratings for the same hospitals. The sample size of the spontaneous- and Road Team ratings was 1314 and 471 respectively. After the assumptions were checked, a t-test with bootstrap was done. Here the independent variable *Gemiddelde_waarderingen* contains the spontaneous ratings and Road Team ratings of the 14 included hospitals. For the dependent variable *RT vs spontaan* the spontaneous ratings are coded as 0 and the ratings generated through Road Teams as 1.

Results

Below, the graphs can be found that show the distribution and dispersion of the ratings from the variables used in this study. Both have a distribution skewed to the left. As can be seen in the figure, the variable from ZKN has a slightly longer tail to the left. However, they provide a similar pattern. For both ZKN and satisfaction question of CQI, most hospitals are rated with a 7 or higher. As the grades become lower, the amount of ratings shows a strong decline for both instruments.

Figure 1. Distribution ScoreZKN

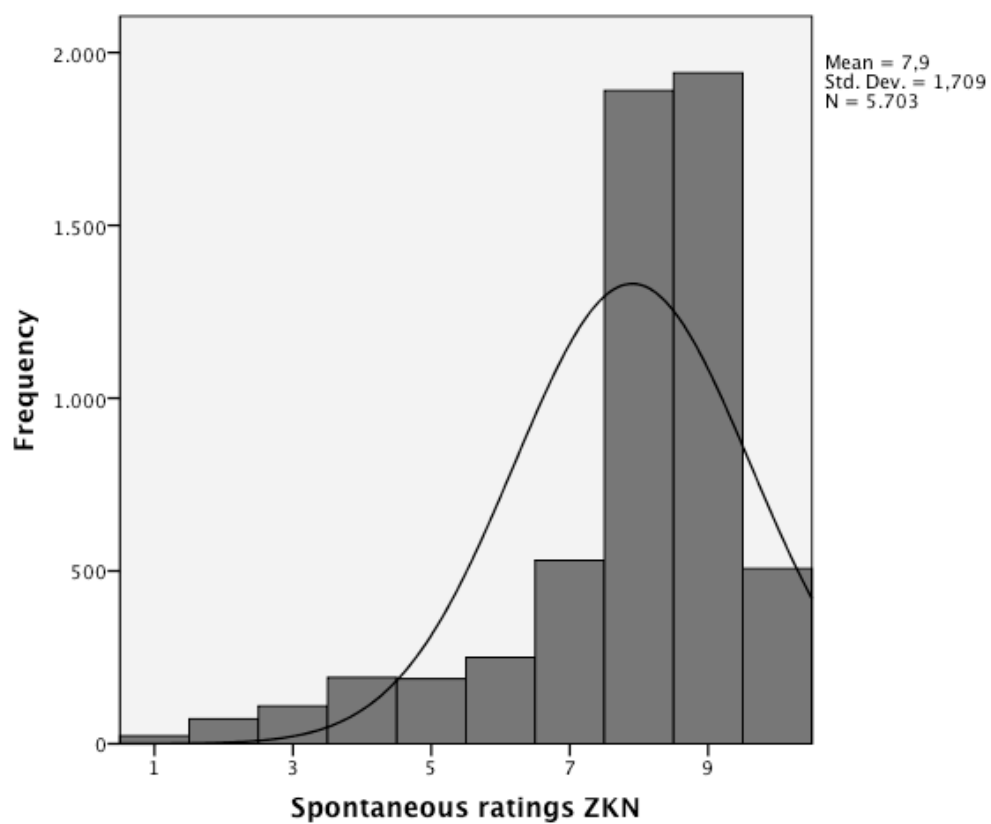


Figure 2. Distribution Vr.77 CQI (satisfaction question CQI)

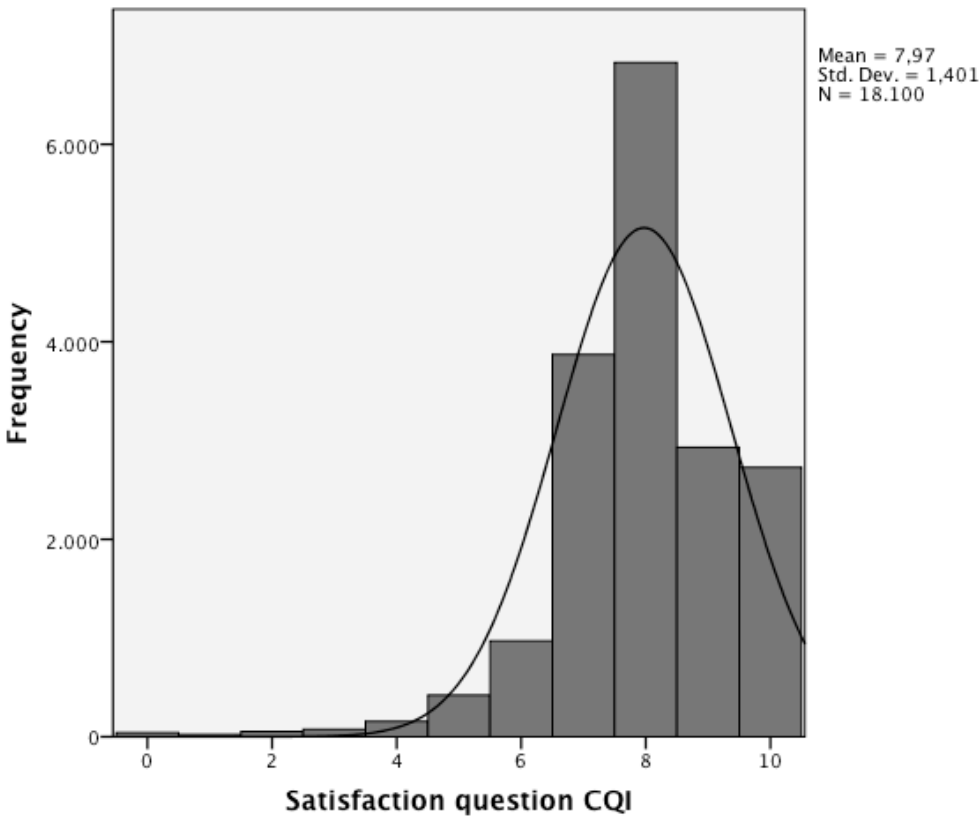
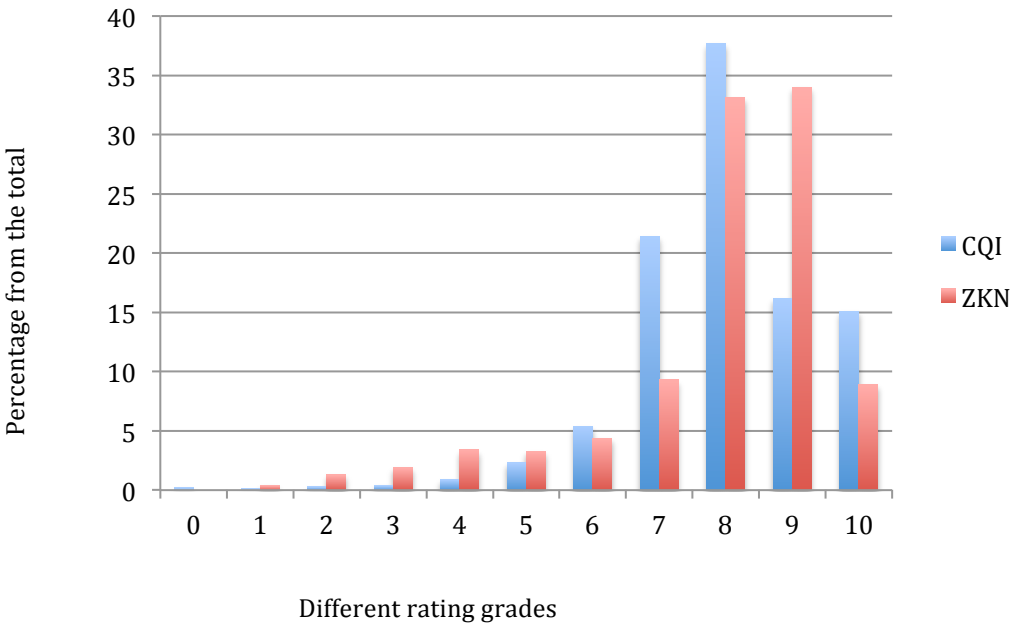


Figure 3 shows the distribution of the ratings from both instruments in one figure. On the y-axis the percentage of ratings with a certain number from the total is displayed.

Figure 3. Distribution ratings from both instruments per number



General information about the variables is presented in table 1. With a sample size of 18100 the Consumer Quality Index has the largest number of ratings and the amount of ratings acquired with Road Teams is the smallest. The ratings of all the 75 hospitals together show a mean score of 7.84 in the case of the spontaneous ratings from ZorgkaartNederland and of 7.97 according to the satisfaction question from the CQ-index. The minimum and maximum score show the variation of mean scores within each instrument.

The ratings given spontaneous on the website from ZKN of the 14 included hospitals together show a mean score of 7.80 and those collected with Road Teams show a mean score of 8.00. The average scores from the spontaneous ratings for the first comparison vary from a minimum of 6.27 to a maximum of 8.77. This means a difference of 2.5 between the hospitals in average ZKN score. The average scores from the spontaneous ratings for the second comparison vary from a minimum of 6.64 to a maximum of 8.49. The difference is therefore 1.9 between the hospitals for the spontaneous ratings.

Table 1. Descriptive information

Variable	N	Mean	Min.	Max.	Std. Deviation
Score ZKN	5703	7.84	6.27	8.77	1.68
Vr.77 CQI	18100	7.97	7.24	8.81	1.40
Spontane waarderingen	1314	7.80	6.64	8.49	1.74
RoadTeam waarderingen	471	8.00	7.74	8.53	0.82

In table 2 the results from both independent samples t-tests are shown.

Table 2. Results from T-tests

T-test	F-test	Mean difference (Std.Error)	Sig.	95% CI Lower Upper	
ZKN vs. CQI	0.000	0.13 (0.03)	0.000	0.09	0.18
Spontaneous vs. RoadTeam	0.000	-0.20 (0.06)	0.001	-0.31	-0.08

The Levene's test for equality of variances shows a significant p-value of .000 for both tests. This means that equality of variances in the variables cannot be assumed. Therefore the

results that correspond with the significant value from the F-test can be found in the lower line in the output table (see Annex 3).

The independent samples t-test in which the spontaneous ratings from ZorgkaartNederland.nl are compared to the ratings from the satisfaction question of the Consumer Quality Index shows a mean difference of 0.13 with a standard error of 0.03. This mean difference has a significant p-value (0.000) and according to the confidence interval the lower and upper ranges are 0.09 and 0.18. With 95% of confidence it can be stated that the difference lies between 0.09 and 0.18. The comparison between the spontaneous ratings and the RoadTeam ratings from Zorgkaart resulted in a significant difference of -0.20. With 95% of confidence it can be said that this difference lies between -0.31 and -0.08.

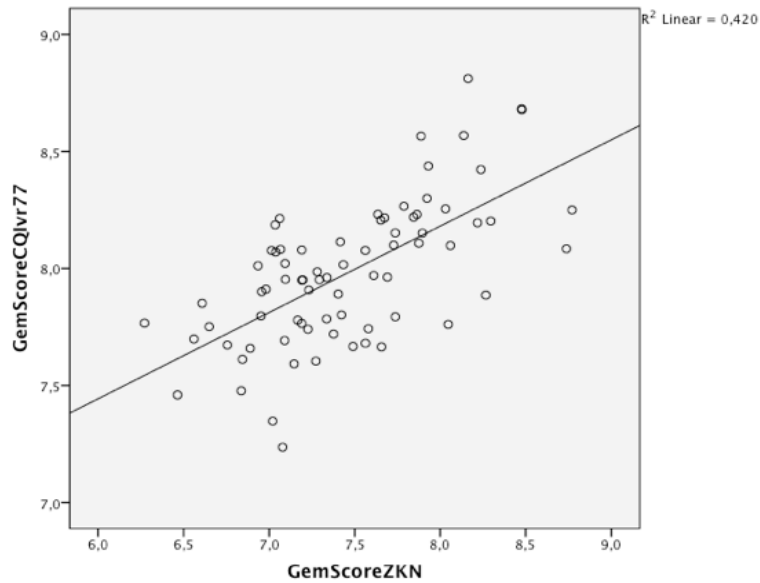
The Pearson correlation coefficient showed in table 3 for the comparison of the mean score per hospital from ZKN with the mean score per hospital of the CQI-satisfaction question is 0.65. A correlation coefficient of 0.55 is found for the ZKN comparison with the total CQI (containing the experience question from the questionnaire).

Table 3. Correlation coefficients

Correlation	ZKN vs CQI 77	ZKN vs CQI total
Pearson	0,65	0,55

For the main comparison a scatter diagram is added to provide a graphic representation of the degree of correlation between the variables. The scatter diagram shows a positive correlation between the ratings from ZKN and the ratings from the CQI.

Figure 3. Scatter diagram for the correlation between spontaneous ratings and satisfaction question



The results from the regression analysis are presented in table 4. The first result is a significant B of -0.03, so when the sample becomes bigger with 1 rating the difference in ranking place between the hospitals reduces with 0.03. When the sample would increase with 100 ratings, hospitals shall become 3 places closer to each other on the ranking, which means that the average score is becoming more similar. For the B of -0.05 it means that when the sample becomes bigger with 1 rating, the difference in ranking place between the hospitals is reduced with 0.05.

Table 4. Results from regression analyses

Regression analysis	B (Std.Error)	Sig.
ZKN with	-0.03 (0.030)	0.000
satisfaction question		
ZKN with CQI total	-0.05 (0.034)	0.000

Discussion

Content of the study

The main objective of this study was to gain insight into the validity and selection bias in particular because of the interest in the representativeness of the patient ratings from ZorgkaartNederland.nl (Zorgkaart or ZKN). Therefore Zorgkaart results were compared with the validated CQ-index, which uses a standardized sampling frame and contains a question that resembles the items from ZKN that form a same sort of overall rating question. The mean scores and distribution of the spontaneous ratings from ZKN were compared to the scores and distribution from the satisfaction question of the CQI questionnaire. To do this, independent samples t-tests have been used and distributions have been plotted. In addition to those comparisons a regression analysis was performed to study the effect of chance on the difference between the ranking place of hospitals and the amount of ratings that underlie the mean score (and therefore determine the place on the ranking list). A second comparison was made between the spontaneous ratings on ZKN and the ratings generated through the use of Road Teams from ZKN in order to look at the consistency and therefore degree of reliability of the results.

Strengths and limitations

Strength of this study is that no research has been done before on this particular subject. The results will therefore be of great value to get a better understanding about the instrument ZorgkaartNederland.nl. This is quite important because the role and opinions of the patient in healthcare are becoming increasingly important⁴. Besides, ZKN is becoming bigger and more popular every month what makes it even more necessary that it delivers information that can be used. Knowledge is obtained about the representativeness of the population when compared to a validated instrument. This is important for the usefulness of the ratings. It was possible to make a relevant comparison because of the availability of the data underlying the satisfaction question from the CQI with content comparable to the items from ZKN. The great amount of ratings, which resulted in a huge sample size, for the different variables is contributing to the relevance and generalizability of the results. A large sample size enhances the relevant translation and the reliability of the results to the target population because of the reduced role of chance³.

Some weaker aspects of the study should also be mentioned. For some hospitals there was a major difference between the amount of ratings from ZKN and the CQI. To give an example: the Canisius Wilhelmina hospital was rated 31 times on ZKN and 378 times by people through the CQI questionnaire. Because of the greater role of chance and impact of subjectivity for a small sample size, one can wonder how reliable the mean score for such a hospital is. The proportion in the total mean score is the same for each hospital however; the

mean of a hospital with twenty ratings contribute as much as the mean based on three hundred ratings. The interpretation of the raters about what qualitative good health care is remains an issue and this point may be defined as a form of information bias. However, quality is quite a complex concept and there is no golden measurement instrument to detect it¹. The good thing is that a great amount of individual and subjective ratings create an objective insight and filter out the possible individual preferences. Nothing was known about the background of the participants and therefore it was not possible to assume that there are two fully independent and similar groups. It could be possible that persons who are selected for the CQI questionnaire in 2009 also rate their care provider on ZKN in later years. It is expected though that, when this is the case, it relate to a very small group. To be surer that selection bias does not play a role, the research could be repeated with participants for both groups selected by the researcher. The way in how representativeness is approached with this research makes it only possible to draw conclusions about that representativeness as a result of the relation of both samples to each other. It was not possible to say something about representativeness in the narrow sense of the word.

Findings

Both distributions (see figures 1 and 2) of the ratings are quite consistent and provide a similar picture. This similarity shows that there are no extreme negative ratings at ZKN. At least it indicates that the population from ZKN rate not remarkable more negative than the CQI population does. The fact that the CQI uses a representative sample and that the distributions of ratings for both instruments are very comparable makes the conclusion that the sample for ZKN provides outcomes that are as representative as those from the CQI questionnaire.

The difference between the overall mean score from the spontaneous ratings of ZorgkaartNederland and the overall mean score from the satisfaction question of the Consumer Quality Index is 0.13. All included hospitals score a 7.84 according to ZKN and a 7.97 according to the satisfaction question of the CQI. The significant difference that is found in this study may be due to the great amount of ratings, which produces a large sample size. It is well known that a significant result not necessarily means that it also is an important or relevant one³⁰. The average scores for the hospitals of ZKN vary between 6.3 and 8.8. This means a variation of 2.5 between the hospitals that had the lowest and highest score. When the difference of 0.13 from the t-test is placed in this perspective, it is a very small one and is considered as not being relevant. The general view about how all included hospitals score together is very comparable.

The rounded correlation coefficient of 0.7 and the scatter diagram provides support for the good correlation between the ratings of both instruments on hospital level. This strong

correlation means that when a hospital has a high score for one instrument it is very likely that it also has a high score for the other instrument³¹. According to the rounded correlation coefficient of 0.6, the correlation between ZKN and the total CQI is also quite good.

The strong positive association between the ratings of both instruments also means that the construct validity of ZKN is quite high. On the other hand the coefficient also shows that for a part of the hospitals there are still large differences.

To get more insight into those differences the regression analyses were conducted. They both show that the cause for the differences lies for a great part in the small amount of ratings for some hospitals. The more ratings a hospital on ZKN has, the smaller the difference in ranking between ZKN and CQI becomes. Thus, the ranking is becoming more similar and correlates better when there are more ratings. More ratings mean a reduced influence of chance as regards the results. It demonstrates that existing differences between both instruments are the result of an accidental composition of the small sample that underlies the score. This means that the shortage of ratings for some hospitals and not differences in representativeness contributes to the difference in ranking between ZKN and CQI.

The difference between the spontaneous ratings and the RoadTeam ratings from ZorgkaartNederland is 0.20. All fourteen included hospitals together score a rounded 7.8 in the case of the spontaneous ratings and a rounded 8 in case of the Road Team ratings. The difference in average score for the spontaneous ratings is 1.9. Also here the founded difference of 0.2 can be put in this perspective and it is again a small difference, which is therefore considered as not being relevant. Both methods from ZorgkaartNederland provide a comparable mean score of the fourteen included Dutch hospitals.

Nevertheless, the question rises how the small difference in mean scores and the distributions can be declared. A few factors should be kept in mind and mostly concern the underlying concepts of both instruments. To start, ratings from the CQ-index date from 2009 and are based on one question. From ZorgkaartNederland ratings from three consecutive years (2010, 2011 and 2012) are used and each rating is an average of six items. It is mentioned before that giving a valuation on ZKN is something intrinsic and anonymous. People that have used the CQI- questionnaire did this because they were elected through a protocol. Additionally the response scale is different. On ZKN health consumers can rate on a 10-point scale from 1 to 10, the CQ-index measures on an 11-point scale from 0 to 10. The format of response scales may be influencing the mean or manner in which people give their rate. To the effect on the mean very little attention has been paid but there is one study with some interesting results³². Within this study it was found that a 10-point scale produces a slightly lower mean than a 5- or 7-point scale. Nothing is known about the difference between a 10-point and 11-point scale but research shows that there are differences between different amounts of scale points³².

Years ago it was already found that irrespective of the content, the scale format really

influences response styles and result in different shaped distributions³³. Response styles are a function of individual characteristics but certainly also of the stimuli, like the questionnaire items and format^{34,35}. A rating scale with or without a midpoint also affects response style. Garland found in his study that the denial of a mid-point can lead to more negative ratings than were achieved when a mid-point was available³⁶. Officially the CQI response scale has a mid-point and the scale from ZKN has not, although the mid-point is not a real neutral point because the rates reflect real numbers.

In comparison with the ratings given anonymous and spontaneous via the website, ratings collected through the use of Road Teams are generally higher. Further they show hardly any insufficient scores. That is why the distribution of the ratings for both methods clearly differs. Probably people are more reluctant to review the six items with failures when having face-to-face contact instead of sitting alone behind a computer. This may be an explanation why the Road Team ratings have a higher mean score. It is assumable that socially desirable answers, which do not really reflect the person's opinion, will appear more easily when having face-to-face contact and play a smaller role for electronic surveys³⁷. Like in every interview regardless of the trained interviewers, both interviewer and patient have a certain influence on each other, which can bias the results³⁸.

Conclusions

This study shows that comparing the spontaneous ratings from ZorgkaartNederland to the ratings from the satisfaction question of the validated Consumer Quality Index and to the ratings from ZKN generated through the use of RoadTeams result in mean scores that are very similar. The self-selection sample from the ZKN website has not led to a lower judgment about the care delivered in Dutch hospitals. The existence of small differences is not surprising because of the slightly different underlying procedures from the measurement methods. The consistency of the average mean scores contributes to the reliability of the patient ratings. The same goes for the great sample size and therefore the small role of chance for the overall scores. The distributions of the ratings from both instruments are very similar too. The average score per hospital correlates good in this correlation raises when the outcome is based on a higher amount of ratings. This demonstrates that existing differences are caused by the accidental formation of small sample sizes and that the differences are reduced when the role of chance is reduced. The conclusion from this research is therefore that the self-selection sample of Zorgkaart leads to representative ratings, which are valid and reliable to some extent.

Recommendations

ZorgkaartNederland is a less expensive way to enhance the quality in Dutch hospitals when compared to the two other instruments mentioned in this paper. Edith Schippers, minister of Health, Welfare and Sport, already emphasizes the importance of Internet in healthcare³⁹. The PvdA and the VVD find that there should be one central location where information about the quality of individual healthcare providers and hospitals is available⁴⁰. ZorgkaartNederland.nl fits to both remarks. Very important is the fact that ratings as well as reviews are shown. This is enlightening for the interpretation of the numbers. Besides, patients have indicated that the availability of reviews is of great value for the usability of information about healthcare providers^{25,26}. However, patient reviews always have a portion of subjectivity. It remains important that ratings from both instruments reflect the quality of care that is provided and not variation across health care users⁴. The last quite important remark is that ZKN contains ratings that are recently placed. With the last implementation of the CQI questionnaire about hospitals in 2009 there is no current information. And a critical point for the performance indicators from ZiZo is that it is not clear who is going to finance, collect and process them in the nearby future⁴¹.

It can be stated that ZKN can be of great value in enhancing the quality of care delivered in Dutch hospitals. But therefore it also should be really used. To look at the use of ZorgkaartNederland by patients and providers was considered too for this study, but due to a lack of time this was not feasible. A recommendation for future research would be to look at the real use of the data on ZorgkaartNederland by both health consumers and health providers. That research should include examining the user-friendliness of the website in order to improve the site where needed. Conclusions about quality of care in this study are only based on the numerical ratings and not on the reviews. The incentive of why patients rate with a six or a nine for example has not been taken into account. It is worth using the reviews in future research because it gives more information about the meaning of a number. In addition, it is recommended that hospitals play an active role in stimulating patients to rate their healthcare provider on ZorgkaartNederland.nl.

References

1. Kringos, D.S., Anema, H.A., Asbroek, ten, A.H.A., Fischer, C., Botje, D., Kievit, J., Steyerberg, E.W. & Klazinga, N.S.(2012). Beperkt Zicht, onderzoek naar de betrouwbaarheid, validiteit en bruikbaarheid van prestatie indicatoren over de kwaliteit van de Nederlandse ziekenhuiszorg. Eindrapport in uitvoering van de Nederlandse Federatie van Universitair Medische Centra (NFU).
2. General information about the website (2013). Retrieved March 1th, 2013 from: <http://www.zorgkaartnederland.nl/content/over-zorgkaartnederland>.
3. Koolman, X., Visser, J. & Appelman, M. (2012). Indicatorstandaard, Methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg. Zichtbare Zorg, Ministerie van VWS.
4. Damman, O.C. (2010). Public reporting about healthcare users' experiences: the Consumer Quality Index. Promotores: prof. dr. D.M.J. Delnoij, prof. dr. P.P. Groenewegen, co-promotor: dr. M. Hendriks. Thesis at the NIVEL.
5. Schoemaker, E., Ligthart, M & Stikvoort, M. (2013). IGZ pakt operatieverwisselingen verder aan. Medisch Contact 13 juni 2013 online artikel. Retrieved June 13th from: <http://medischcontact.artsennet.nl/nieuws-26/archief-6/tijdschriftartikel/133363/igz-pakt-operatieverwisselingen-verder-aan.htm>
6. Operatieproces in veel ziekenhuizen nog niet op orde. Zorgvisie 13 juni. Retrieved June 13th from: http://www.zorgvisie.nl/Kwaliteit/Nieuws/2013/6/Operatieproces-in-veel-ziekenhuizen-nog-niet-op-orde-1282151W/?cmpid=NLC|zorgvisie|2013-06-13|Operatieproces_in_veel_ziekenhuizen_nog_niet_op_orde
7. Rapport IGZ (juni 2013). Operatief proces beter gestructureerd, scherpere handhaving op achterblijvende uitvoering blijkt noodzakelijk, 13-25.
8. VUmc onder verscherpt toezicht. Nieuwsbericht 21-08-2012 op de site van IGZ. Retrieved June 13th from: http://www.igz.nl/actueel/nieuws/VUmc_onder_verscherpt_toezicht.aspx
9. Ruwaard sluit ook polikliniek cardiologie. Geplaatst op 20-11-2012 op de site van Skipr. Retrieved June 11th from: <http://www.skipr.nl/actueel/id12862-ruwaard-sluit-ook-polikliniek-cardiologie.html>
10. Pilligrimiene, Z. & Buciuniene, I (2008). Different Perspectives on Health Care Quality: Is the Consensus Possible? ISSN 1392-2785 Engineering Economics. 2008. No 1 (56) Economics of Engineering Decisions.
11. Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J. & Stoddart, G.L (2005). Methods for the Economic Evaluation of Health Care Programmes. Oxford Medical Publications, third edition.
12. Torrance, G.W., Thomas, W.H. & Sackett, D.L (1972). A Utility Maximization Model for Evaluation of Health Care Programs. Health Services Research, Summer 1972; 7(2): 118-133.
13. Information about the project ZichtbareZorg. Retrieved April 19th from: <http://www.zichtbarezorg.nl/page/Ziekenhuizen-en-ZBC-s/Het-project>.
14. Information about the future of the project ZichtbareZorg. Retrieved May 14th from: <http://www.thedecisiongroup.nl/meetbaar-beter-project-description>

15. Kennisvraag – Wat heeft vijf jaar CQ-index opgeleverd? Nivel 2011
16. Koopman, L., Sixma, H., Hendriks, M., Boer, de, D. & Delnoij, D. (2011). Richtlijnen en voorschriften voor de ontwikkeling van een CQI instrument (2011). Handboek CQI Ontwikkeling
17. Werkinstructie Voorbereiding steekproeftrekking uit Handboek Eisen en werkwijzen CQI:
http://www.centrumklantervaringzorg.nl/fileadmin/site/ckz/documenten/Handboek_nieuwe_s_tijl/WIS_02.01_Voorbereiding_steekproeftrekking_dv_20110815.pdf
18. Code of Conduct ZorgkaartNederland (2013). Retrieved March 25th, 2013 from: <http://www.zorgkaartnederland.nl/content/gedragscode>
19. Bierma, L. (2012). Trainingsbijeenkomst ZorgkaartNederland 29 november 2012.
20. Russo, J.E., Krieser, G. & Miyashita, S. (1975). An effective display of unit price information. *Journal of marketing*; 39: 11-19.
21. Wollersheim, H., Hermens, R., Hulscher, M., Braspenning, J., Ouwens, M., Schouten, J., Marres, H., Dijkstra, R. & Grol, R (2007). Clinical Indicators: development and applications. *Neth.J.Med.*; 65:15-22.
22. Pauly, M.V., McGuire, T.G. & Barros, P.P (2012). *Handbook of Health Economics*. Elsevier, first edition 2012.
23. Website from Leiden University. Retrieved March 27th from: <http://www.leidenuniv.nl/fsw/psychologielexicon/index.php3-c=49.htm>
24. Rekenkamer oordeelt hard over Zichtbare Zorg. Nieuwsbericht 28-03-2013 op de site van Skipr. Retrieved March 28th from: <http://medischcontact.artsennet.nl/nieuws-26/nieuwsbericht-1/130270/rekenkamer-ordeelt-hard-over-zichtbare-zorg.htm>
25. Weerdhof, van de, A. (2012). The use of online information to enhance quality in the Dutch health care sector. Master Thesis Tilburg University, 39-42.
26. Damman, O.C., Hendriks, M., & Delnoij, D.M. (2010). Keuze-informatie op basis van patientenervaringen: aanbevelingen en dilemma's. *Tijdschrift voor gezondheidswetenschappen*, 88(7), 396-405.
27. Vermeend, W. (2013). Stijgende zorgkosten vragen om bezuiniging en preventie, Blog Skipr. Retrieved May 11th, 2013 from: <http://www.skipr.nl/blogs/id1376-stijgende-zorgkosten-vragen-om-bezuiniging-en-preventie.html>
28. Statistieken ZorgkaartNederland t/m mei 2013. Supplied by Bohn Stafleu van Loghum. http://www.npcf.nl/?option=com_content&view=article&id=5206&catid=2:nieuws&Itemid=26

29. Retrieved March 2013 from: <https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php>.
30. NRC Stijlboek Statistisch onderzoek. Retrieved April 14th and available via: <http://apps.nrc.nl/stijlboek/statistisch-onderzoek>
31. Kimble, G.A. (1978). How To Use (and misuse) Statistics. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
32. Dawes, J. G. (2008). Do Data Characteristics Change According to the Number of Scale Points Used ? An Experiment Using 5 Point, 7 Point and 10 Point Scales (February 29, 2012). International Journal of Market Research, Vol. 51, No. 1, 2008.
33. Hui, H. C., & Triandis, H. C. (1989). Effects of Culture and Response Format on Extreme Response Style. Journal of Cross-Cultural Psychology, 20 (3), 296-309.
34. Baumgartner, H., & Steenkamp, J.B.E.M. (2001). Response styles in marketing research: A cross-national investigation. Journal of Marketing Research, 38 (May), 143-156.
35. Paulhus, D. L. (1991), Measurement and control of response bias. In: Robinson J.P., Shaver P.R. and Wright L.S (Eds.). Measures of Personality and Social Psychological attitudes (pp. 17-59). San Diego: Academic Press.
36. Garland, R. The Mid-Point on a Rating Scale: Is it Desirable? Marketing Bulletin, 1991, 2, 66-70
37. Kiesler, S. & Sproull, L.S (1986). Response Effects in the Electronic Survey. Public Opinion Quarterly vol.50; 3: 402-413 Oxford Journals.
38. Brink, van den, W.P. & Mellenbergh, G.J (1998). Testleer en testconstructie; 391. Uitgeverij Boom Amsterdam
39. Kabinet zet in op zelfredzaamheid in zorg (2013). Retrieved February 20th from: <http://www.nu.nl/gezondheid/3142563/kabinet-zet-in-zelfredzaamheid-in-zorg.html>
40. Kamer schiet patiënt te hulp bij keuze arts (2013). Retrieved February 20th from: <http://www.nu.nl/politiek/3010638/kamer-schiet-patient-hulp-bij-keuze-arts.html>
41. News item from the website from ZichtbareZorg (2012). Retrieved May 11th from: <http://www.zichtbarezorg.nl/page/Verpleging-verzorging-en-thuiszorg/Nieuws?mod%5BBrickworkNewsModule%5D%5Bn%5D=744>

Annex 1: Hospitals for comparison ZKN and CQI

't Lange Land ziekenhuis, LANGE LAND ZIEKENHUIS 'T, t Lange Land Ziekenhuis
Academisch medisch centrum Amsterdam
Academisch ziekenhuis Maastricht, MAASTRICHT ACADEMISCH ZKH
Albert Schweitzer ziekenhuis
Algemeen ziekenhuis Westfries Gasthuis, WESTFRIES GASTHUIS ALG ZK
Alkmaar Medisch Centrum, Medisch Centrum Alkmaar
Amphia Ziekenhuis
ANNAZIEKENHUIS ST, Sint Anna Ziekenhuis
Antoni van Leeuwenhoek Ziekenhuis
Antonius Ziekenhuis, SINT ANTONIUS ZIEKENHUIS
ATRIUM HEERLEN
BEATRIX STREEKZKHS KON., Streekziekenhuis Koningin Beatrix
Bernhoven Ziekenhuis, Ziekenhuis Bernhoven
Bethesda Ziekenhuis, ZIEKENHUIS BETHESDA
BOVEN IJ ZIEKENHUIS, BovenIJ Ziekenhuis
Canisius-Wilhelmina Ziekenhuis
Catharina Ziekenhuis
DEVENTER ZIEKENHUIZEN ST, Stichting Deventer Ziekenhuizen
Diaconessenhuis Meppel
ELISABETH ZIEKENHUIS ST, Sint Elisabeth Ziekenhuis
Elkerliek Ziekenhuis
Erasmus MC, Erasmus Medisch Centrum
Flevoziekenhuis
FRANCISCUS ZIEKENHUIS ST, Sint Franciscus Ziekenhuis
Gelre Ziekenhuizen
Gemini Ziekenhuis
Groene Hart Ziekenhuis
HAVENZIEKENHUIS
Het Bronovo Ziekenhuis
Het Rode Kruis Ziekenhuis, RODE KRUIS ZIEKENHUIS BEVERWIJK
HOFPOORT ZIEKENHUIS, Zuwe Hofpoort Ziekenhuis
IJSELLAND ZIEKENHUIS
IKAZIA ZIEKENHUIS
Isala Klinieken
Jeroen Bosch Ziekenhuis
Kennemer Gasthuis
Laurentius Ziekenhuis, LAURENTIUS ZIEKENHUIS ST
Leids Universitair Medisch Centrum (Lumc)
LIEVENSBERG ZIEKENHUIS, Ziekenhuis Lievensberg
MAARTENSKLINIEK ST, Sint Maartenskliniek
Maasstad Ziekenhuis
Maasziekenhuis
Martini, Martini Ziekenhuis
Maxima Medisch Centrum
Meander Medisch Centrum
Medisch Centrum Haaglanden, ST. MEDISCH CENTRUM HAAGLANDEN
MEDISCH CENTRUM LEEUWARDEN, Medisch Centrum Leeuwarden/Zorggroep
Noorderbreedte
Medisch Spectrum Twente
NIJ SMELLINGHE, Ziekenhuis Nij Smellinghe
OLVG, Onze Lieve Vrouwe Gasthuis

OMMELANDER ZIEKENHUIS GROEP, LOC. LUCAS
 Ropcke-Zweers Ziekenhuis, Ropcke-Zweers Ziekenhuis
 RADBOUD ACADEMISCH ZIEKENHUIS NIJMEGEN, Universitair Medisch Centrum St.
 Radboud
 REFAJA PROT CHR ZKHS, Refaja Ziekenhuis
 Reinier de Graaf Groep
 RIJNLAND ST ZIEKENHUIZEN, Rijnland Ziekenhuis
 RIJNSTATE ZIEKENHUIS, Ziekenhuis Rijnstate
 RIVIERENLAND ZIEKENHUIS, Ziekenhuis Rivierenland
 Ruwaard van Puttenziekenhuis
 SINT FRANCISCUS GASTHUIS
 Sint Lucas-Andreas Ziekenhuis, ST LUCAS-ANDREAS ZIEKENHUIS
 Slingeland Ziekenhuis
 Slotervaartziekenhuis
 Spaarne Ziekenhuis
 STICHTING ZAANS MEDISCH CENTRUM, Zaans Medisch Centrum
 Tergooiziekenhuizen
 Universitair Medisch Centrum Groningen
 Universitair Medisch Centrum Utrecht
 Viecuri Medisch Centrum Voor Noord-Limburg, VIECURIE MEDISCHC CENTRUM V
 NOORD LIMBURG
 Vu-Ziekenhuis Amsterdam
 Waterlandziekenhuis
 Wilhelmina Ziekenhuis
 Ziekenhuis Amstelland
 Ziekenhuis de Gelderse Vallei
 Ziekenhuis St Jansdal

Annex 2: Hospitals for comparison spontaneous ratings and Road Team ratings

Amphia Langendijk
Amphia Molengracht
Groene Hart Bleuland
Groene Hart Jozef
Jeroen Bosch Ziekenhuis
Laurentius Ziekenhuis
Rijnstate Arnhem
Sint Maartenskliniek
Spaarne Ziekenhuis
St. Elisabeth Ziekenhuis
Streekziekenhuis Koningin Beatrix
UMC Utrecht
Wilhelmina Ziekenhuis Assen
ZorgSaam Ziekenhuis locatie de Honte

Annex 3: Syntax and Output Spss

Comparison ZorgkaartNederland and Consumer Quality Index

FREQUENCIES VARIABLES=AfgerondScoreZKN

/STATISTICS=STDDEV VARIANCE MEAN MEDIAN SKEWNESS SESKEW KURTOSIS

SEKURT

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

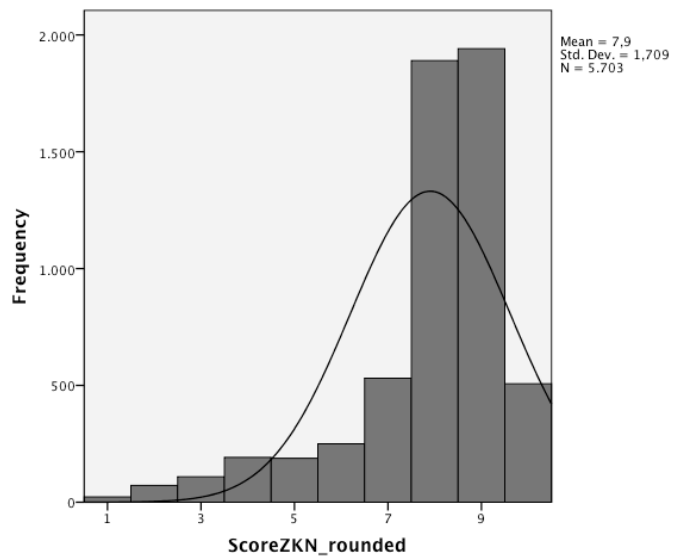
Statistics

AfgerondScoreZKN

N	Valid	5703
	Missing	0
Mean		7,90
Median		8,00
Std. Deviation		1,709
Variance		2,921
Skewness		-1,623
Std. Error of Skewness		,032
Kurtosis		2,698
Std. Error of Kurtosis		,065

AfgerondScoreZKN

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	23	,4	,4	,4
2	72	1,3	1,3	1,7
3	109	1,9	1,9	3,6
4	192	3,4	3,4	6,9
5	188	3,3	3,3	10,2
6	250	4,4	4,4	14,6
7	531	9,3	9,3	23,9
8	1890	33,1	33,1	57,1
9	1941	34,0	34,0	91,1
10	507	8,9	8,9	100,0
Total	5703	100,0	100,0	

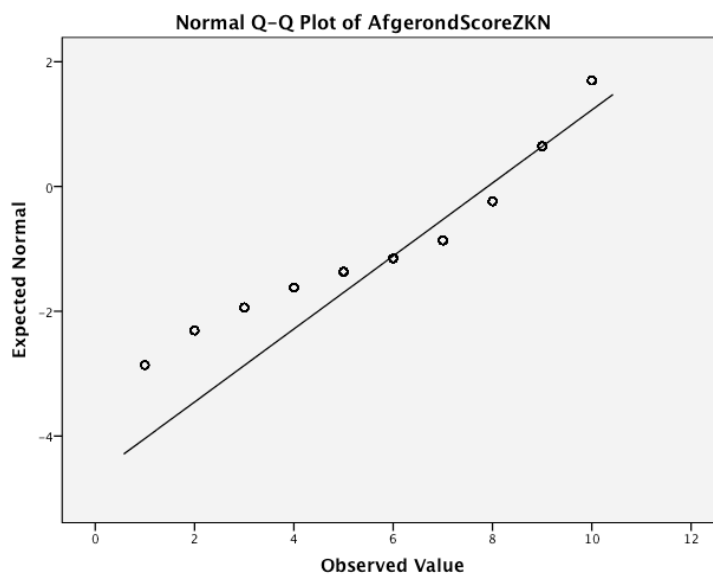


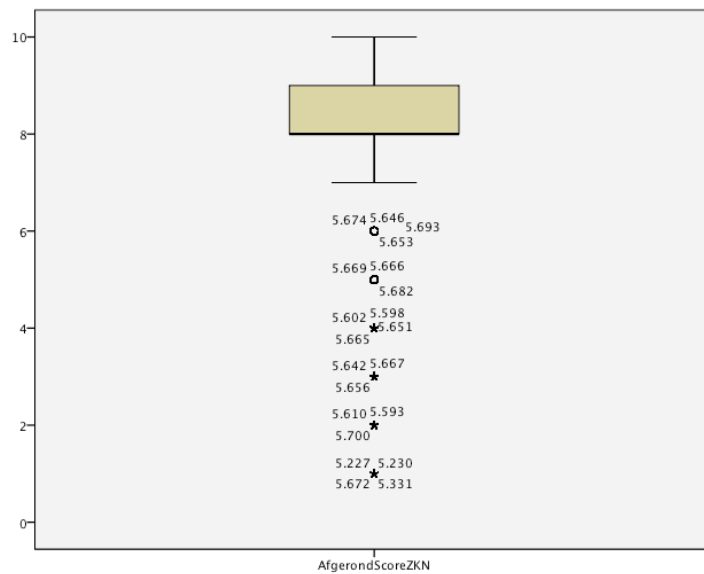
```
EXAMINE VARIABLES=AfgerondScoreZKN
/PLOT BOXPLOT STEMLEAF NPLOT
/COMPARE GROUPS
/STATISTICS NONE
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
AfgerondScoreZKN	,283	5703	,000

a. Lilliefors Significance Correction





FREQUENCIES VARIABLES=vr77CQI
 /STATISTICS=STDDEV VARIANCE MEAN MEDIAN SKEWNESS SESKEW KURTOSIS
 SEKURT
 /HISTOGRAM NORMAL
 /ORDER=ANALYSIS.

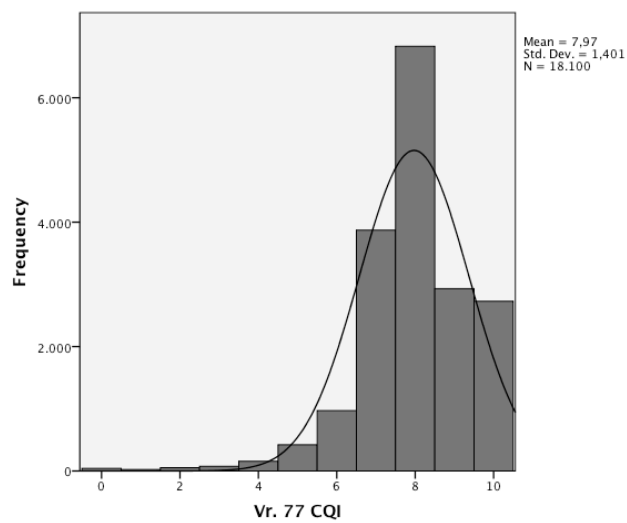
Statistics

vr 77 CQI

N	Valid	18100
	Missing	0
Mean		7,97
Median		8,00
Std. Deviation		1,401
Variance		1,963
Skewness		-1,121
Std. Error of Skewness		,018
Kurtosis		3,861
Std. Error of Kurtosis		,036

vr 77 CQI

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	41	,2	,2	,2
1	25	,1	,1	,4
2	54	,3	,3	,7
3	75	,4	,4	1,1
4	155	,9	,9	1,9
5	421	2,3	2,3	4,3
6	970	5,4	5,4	9,6
7	3871	21,4	21,4	31,0
8	6828	37,7	37,7	68,7
9	2931	16,2	16,2	84,9
10	2729	15,1	15,1	100,0
Total	18100	100,0	100,0	

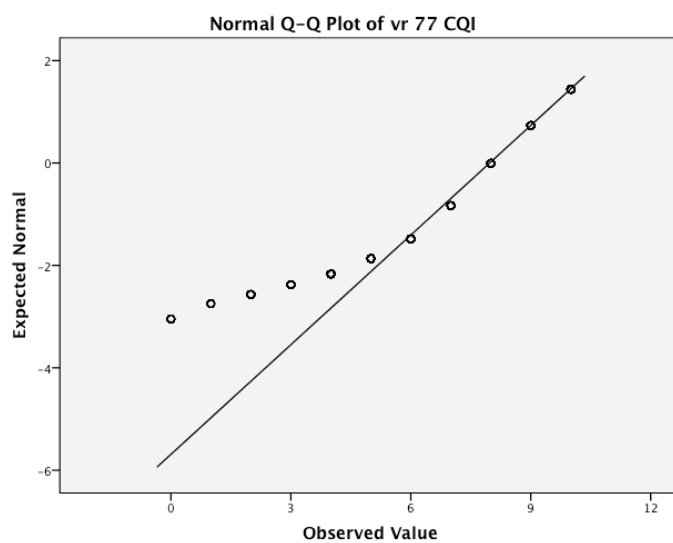


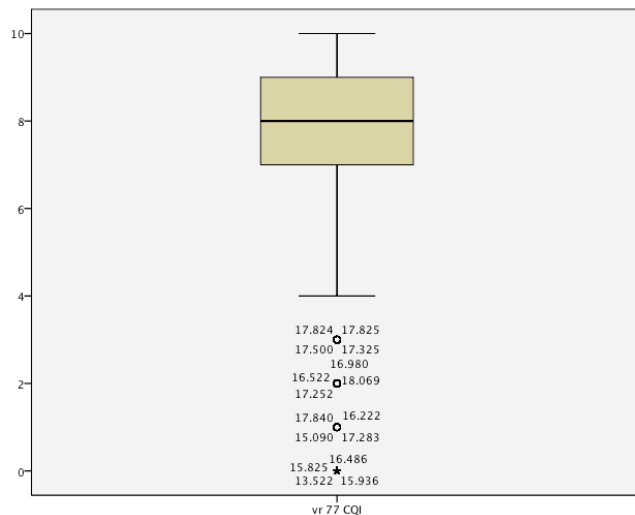
```
EXAMINE VARIABLES=vr77CQI
/PLOT BOXPLOT STEMLEAF NPLOT
/COMPARE GROUPS
/STATISTICS NONE
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
vr 77 CQI	,198	18100	,000

a. Lilliefors Significance Correction





```
RECODE ScoreZKN (1 thru 10=1) INTO TypeWaardering.
VARIABLE LABELS TypeWaardering 'ZKNvsCQI'.
EXECUTE.
```

```
RECODE vr77CQI (0 thru 10=0) INTO TypeWaardering.
VARIABLE LABELS TypeWaardering 'ZKNvsCQI'.
EXECUTE.
```

```
T-TEST GROUPS=TypeWaardering(0 1)
/MISSING=ANALYSIS
/VARIABLES=Scores
/CRITERIA=CI(.95).
```

Group Statistics

ZKNvsCQI		N	Mean	Std. Deviation	Std. Error Mean
Scores	,00	18100	7,97	1,401	,010
	1,00	5703	7,84	1,684	,022

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Scores	Equal variances assumed	200,582	,000	5,891	23801	,000	,132	,022	,088	,17
	Equal variances not assumed			5,357	8334,397	,000	,132	,025	,084	,18

```
BOOTSTRAP
/SAMPLING METHOD=SIMPLE
/VARIABLES TARGET=Scores INPUT=TypeWaardering
/CRITERIA CILEVEL=95 CITYPE=BCA NSAMPLES=10000
/MISSING USERMISSING=EXCLUDE.
```

Group Statistics

			Statistic	Bootstrap ^a			
				Bias	Std. Error	BCa 95% Confidence Interval	
						Lower	Upper
Scores	,00	N	18100				
		Mean	7,97	,00	,01	7,95	7,99
		Std. Deviation	1,401	,000	,013	1,377	1,426
		Std. Error Mean	,010				
	1,00	N	5703				
		Mean	7,84	,00	,02	7,80	7,88
		Std. Deviation	1,684	-,001	,024	1,636	1,731
		Std. Error Mean	,022				

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

Bootstrap for Independent Samples Test

		Mean Difference	Bootstrap ^a				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
Scores	Equal variances assumed	,132	,000	,025	,000	,085	,179
	Equal variances not assumed	,132	,000	,025	,000	,085	,179

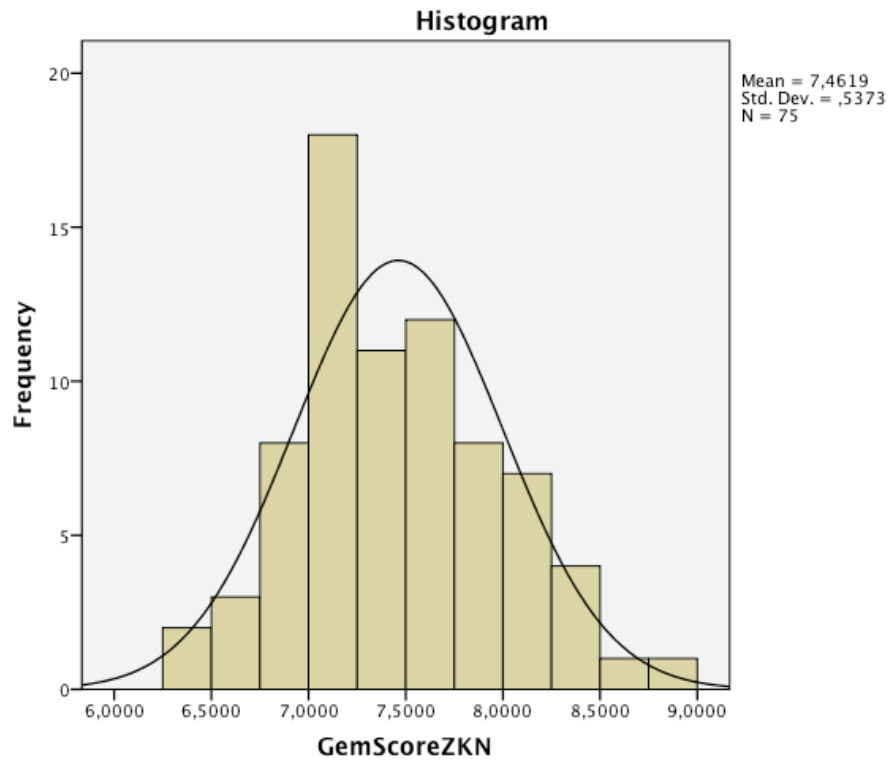
a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

FREQUENCIES VARIABLES=GemScoreZKN
 /STATISTICS=STDDEV VARIANCE MEAN MEDIAN SKEWNESS SESKEW KURTOSIS
 SEKURT
 /HISTOGRAM NORMAL
 /ORDER=ANALYSIS.

Statistics

GemScoreZKN

N	Valid	75
	Missing	0
Mean		7,461897
Median		7,403600
Std. Deviation		,5372740
Variance		,289
Skewness		,303
Std. Error of Skewness		,277
Kurtosis		-,269
Std. Error of Kurtosis		,548



```

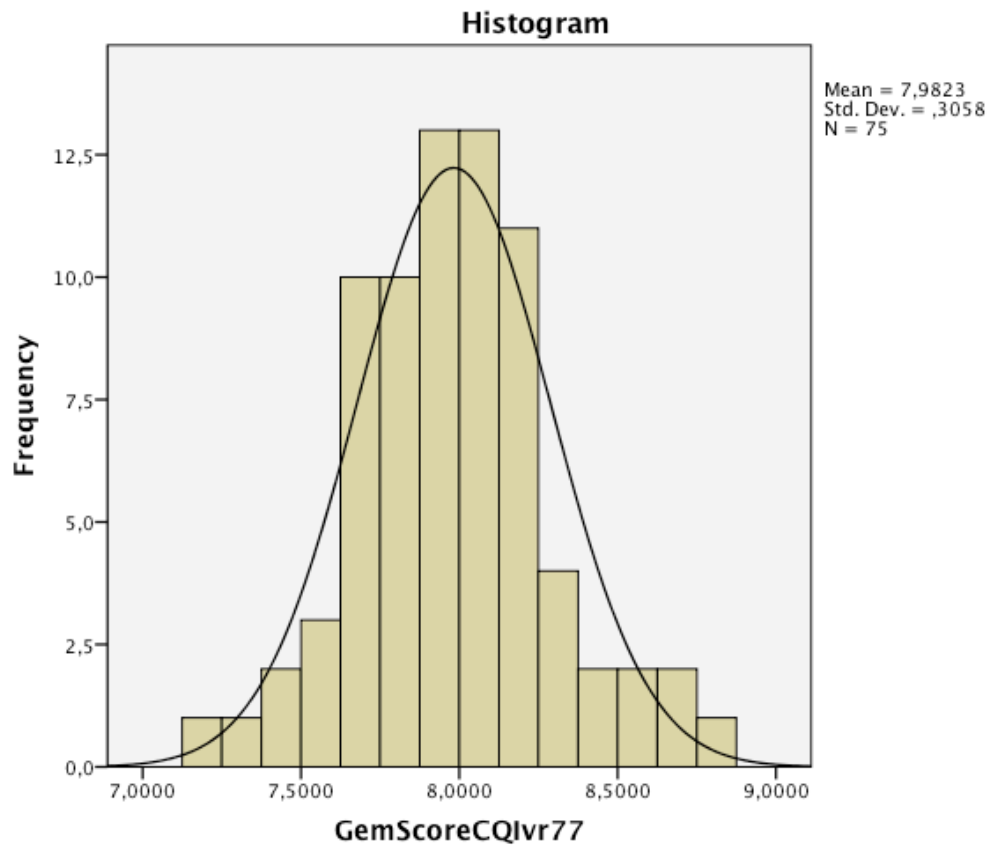
FREQUENCIES VARIABLES=GemScoreCQIvr77
  /STATISTICS=STDDEV VARIANCE MEAN MEDIAN SKEWNESS SESKEW KURTOSIS
  SEKURT
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

```

Statistics

GemScoreCQIvr77

N	Valid	75
	Missing	0
Mean		7,982264
Median		7,962700
Std. Deviation		,3057644
Variance		,093
Skewness		,299
Std. Error of Skewness		,277
Kurtosis		,361
Std. Error of Kurtosis		,548



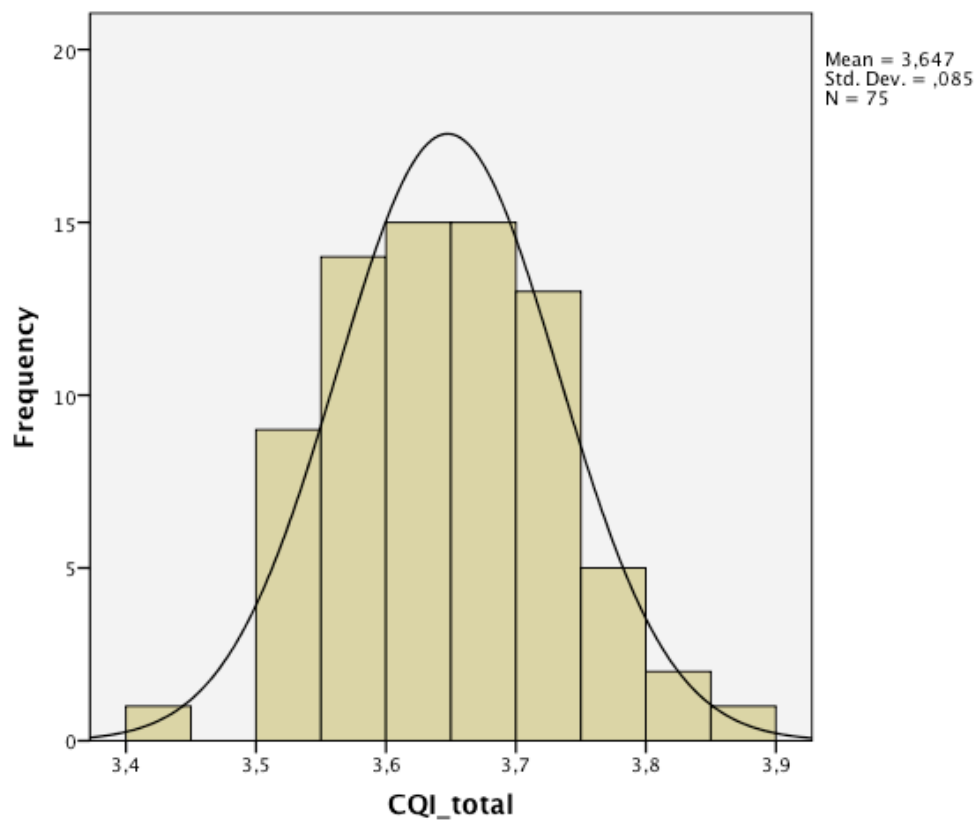
```

FREQUENCIES VARIABLES=CQI_overallscore CQI_omgeschaaldeoverall
/STATISTICS=STDDEV VARIANCE SEMEAN MEAN MEDIAN
/HISTOGRAM NORMAL
/ORDER=ANALYSIS.

```

Statistics

		CQI_overallscore	CQI_omgeschaaldeoverall
N	Valid	75	75
	Missing	0	0
Mean		3,64748	8,13886
Std. Error of Mean		,009835	,026081
Median		3,64380	8,11000
Std. Deviation		,085172	,225871
Variance		,007	,051



CORRELATIONS

/VARIABLES=GemScoreCQIvr77 GemScoreZKN
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.

Correlations

		GemScoreCQI vr77	GemScoreZK N
GemScoreCQIvr77	Pearson Correlation	1	,648**
	Sig. (2-tailed)		,000
	N	75	75
GemScoreZKN	Pearson Correlation	,648**	1
	Sig. (2-tailed)	,000	
	N	75	75

** . Correlation is significant at the 0.01 level (2-tailed).

CORRELATIONS

```
/VARIABLES=CQI_overallscore ZKN_overallscores
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

Correlations

		CQI_overallscore	ZKN_overallscores
CQI_overallscore	Pearson Correlation	1	,547**
	Sig. (2-tailed)		,000
	N	75	75
ZKN_overallscores	Pearson Correlation	,547**	1
	Sig. (2-tailed)	,000	
	N	75	75

**. Correlation is significant at the 0.01 level (2-tailed).

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Gem.verschil
/METHOD=ENTER KleinsteSample.
```

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	,729	,086		8,526	,000	,559	,900
Kleinste Sample	-,003	,001	-,289	-2,579	,012	-,006	-,001

a. Dependent Variable: Gem.verschil

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Rankverschil_totalCQIandZKN

/METHOD=ENTER KleinsteSample.

Coefficients^a

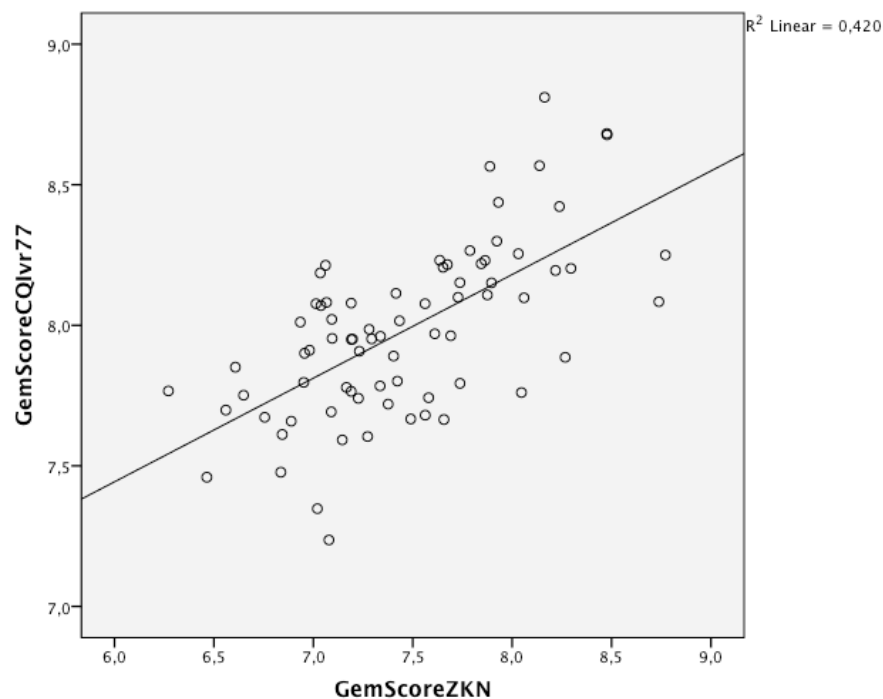
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18,253	2,366		7,713	,000
	Kleinste Sample	-,052	,034	-,174	-1,512	,135

a. Dependent Variable: Rankverschil_totalCQIandZKN

GRAPH

/SCATTERPLOT(BIVAR)=GemScoreZKN WITH GemScoreCQIvr77

/MISSING=LISTWISE.



Comparison Spontaneous ratings and RoadTeam ratings

DESCRIPTIVES VARIABLES=Gemiddelde

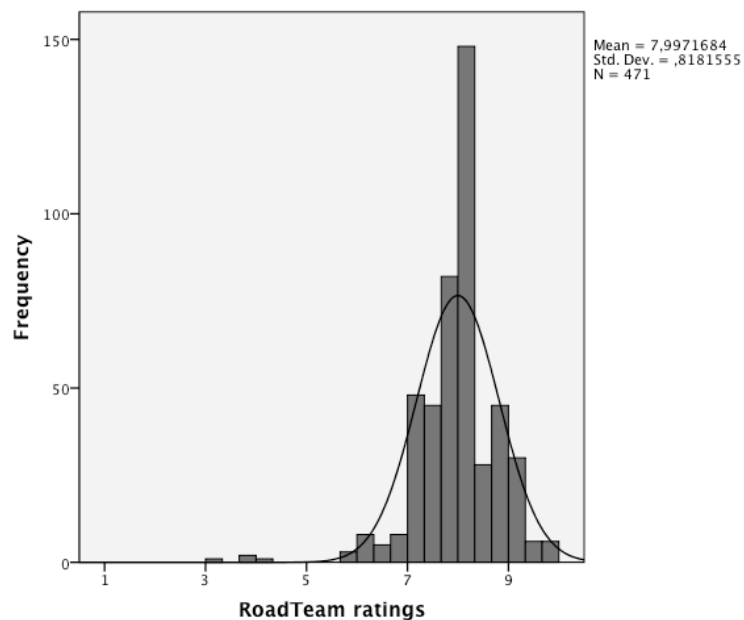
/STATISTICS=MEAN STDDEV MIN MAX SEMEAN.

Descriptive Statistics						
	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Gemiddelde	471	3,0000000	10,0000000	7,99716843	,037698613	,818155455
Valid N (listwise)	471					

FREQUENCIES VARIABLES=Gemiddelde

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.



EXAMINE VARIABLES=Gemiddelde

/PLOT BOXPLOT STEMLEAF NPLOT

/COMPARE GROUPS

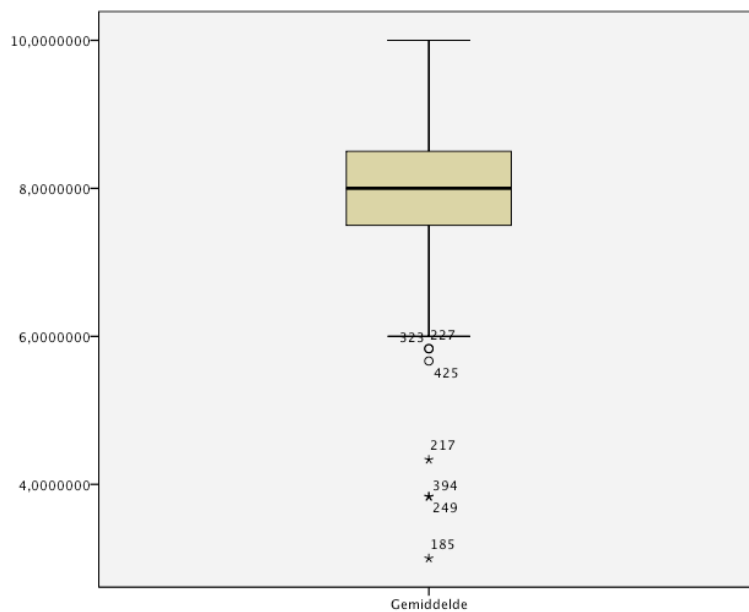
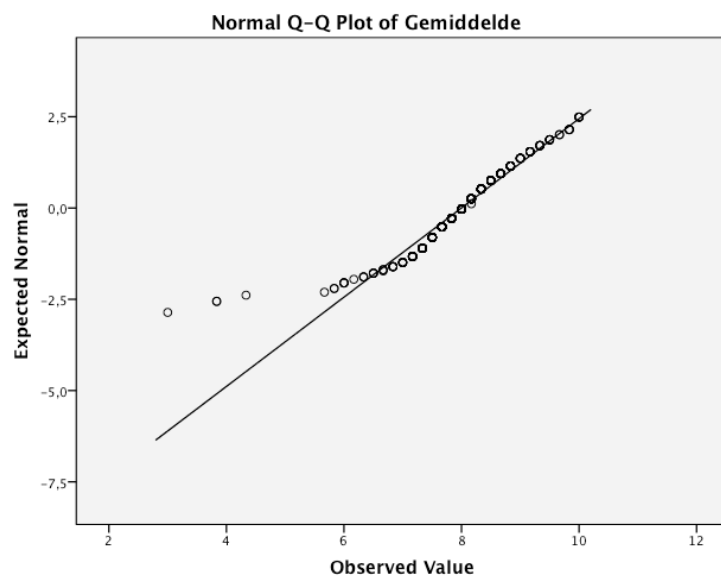
/STATISTICS DESCRIPTIVES EXTREME

/INTERVAL 95

/MISSING LISTWISE

/NOTOTAL.

Descriptives			Statistic	Std. Error
Gemiddelde	Mean		7,99716843	,037698613
	95% Confidence Interval for Mean	Lower Bound	7,92308975	
		Upper Bound	8,07124712	
	5% Trimmed Mean		8,02516340	
	Median		8,00000000	
	Variance		,669	
	Std. Deviation		,818155455	
	Minimum		3,00000000	
	Maximum		10,00000000	
	Range		7,00000000	
	Interquartile Range		1,00000000	
	Skewness		-1,165	,113
	Kurtosis		6,034	,225



DESCRIPTIVES VARIABLES=Gemiddelde
 /STATISTICS=MEAN STDDEV MIN MAX SEMEAN.

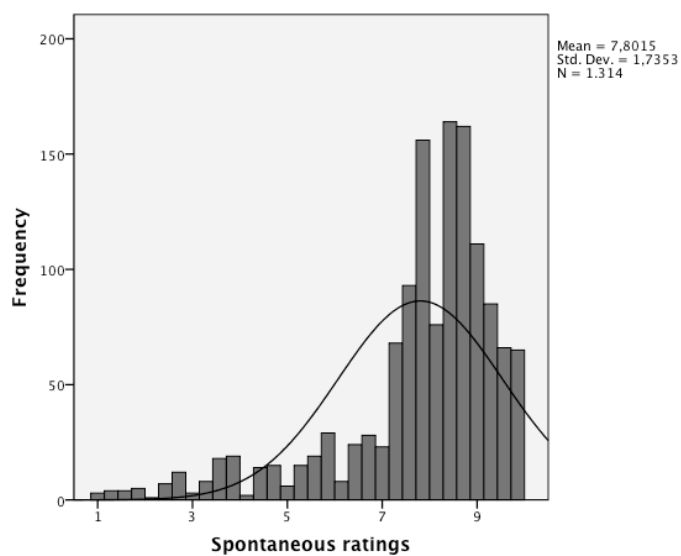
Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Gemiddelde	1314	,0000	10,0000	7,801548	,0478721	1,7353227
Valid N (listwise)	1314					

FREQUENCIES VARIABLES=Gemiddelde

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.



EXAMINE VARIABLES=Gemiddelde

/PLOT BOXPLOT STEMLEAF NPLOT

/COMPARE GROUPS

/STATISTICS DESCRIPTIVES EXTREME

/CINTERVAL 95

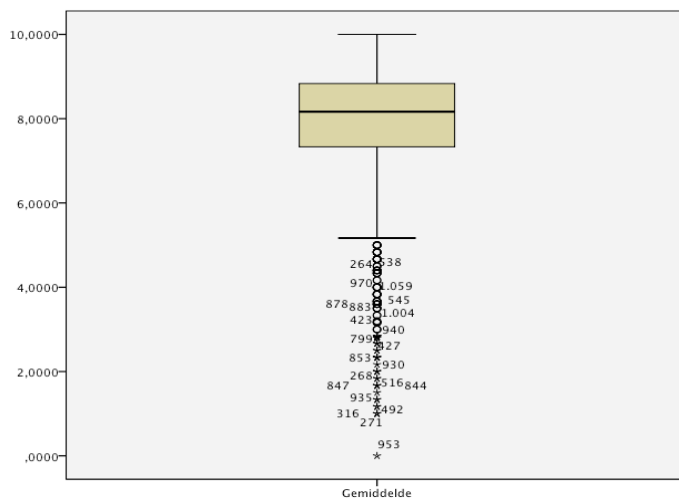
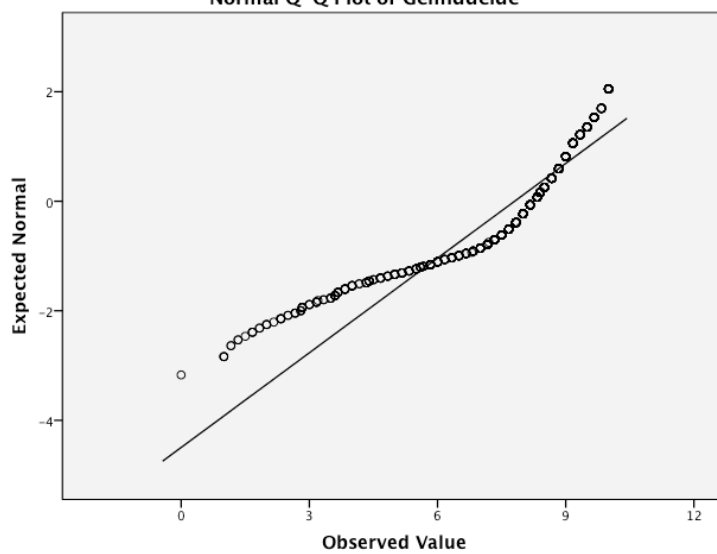
/MISSING LISTWISE

/NOTOTAL.

Descriptives

			Statistic	Std. Error
Gemiddelde	Mean		7,801548	,0478721
	95% Confidence Interval for Mean	Lower Bound	7,707634	
		Upper Bound	7,895463	
	5% Trimmed Mean		7,966995	
	Median		8,166700	
	Variance		3,011	
	Std. Deviation		1,7353227	
	Minimum		,0000	
	Maximum		10,0000	
	Range		10,0000	
	Interquartile Range		1,5000	
	Skewness		-1,638	,067
	Kurtosis		2,687	,135

Normal Q-Q Plot of Gemiddelde



BOOTSTRAP

/SAMPLING METHOD=STRATIFIED(STRATA=Ziekenhuis)
 /VARIABLES TARGET=Gemiddelde_waarderingen INPUT=RoadTeam
 /CRITERIA CILEVEL=95 CITYPE=BCA NSAMPLES=10000
 /MISSING USERMISSING=EXCLUDE.

T-TEST GROUPS=RoadTeam(0 1)

/MISSING=ANALYSIS
 /VARIABLES=Gemiddelde_waarderingen
 /CRITERIA=CI(.95).

Group Statistics

			Statistic	Bootstrap ^a			
				Bias	Std. Error	BCa 95% Confidence Interval	
RT vs spontaan						Lower	Upper
Gemiddelde_waarderingen	Spontaan	N	1314				
		Mean	7,801548	,000128	,046119	7,709534	7,892663
		Std. Deviation	1,7353227	-,0010788	,0506528	1,6350276	1,8326364
		Std. Error Mean	,0478721				
	RoadTeam	N	471				
		Mean	7,997168	,000077	,037840	7,921133	8,071759
		Std. Deviation	,8181555	-,0026271	,0529302	,7260510	,9150893
		Std. Error Mean	,0376986				

a. Unless otherwise noted, bootstrap results are based on 10000 stratified bootstrap samples

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Gemiddelde_waardering en	Equal variances assumed	130,677	,000	-2,354	1783	,019	-,1956200	,0830947	-,3585932	-,0326469
	Equal variances not assumed			-3,210	1661,455	,001	-,1956200	,0609338	-,3151351	-,0761050

Bootstrap for Independent Samples Test

		Mean Difference	Bootstrap ^a				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
Gemiddelde_waarderingen	Equal variances assumed	-,1956200	,0000518	,0596203	,001	-,3114715	-,0784992
	Equal variances not assumed	-,1956200	,0000518	,0596203	,001	-,3114715	-,0784992

a. Unless otherwise noted, bootstrap results are based on 10000 stratified bootstrap samples

FREQUENCIES VARIABLES=Gemiddelde_spontaan Score_RT

Statistieken Spontaan vs RT vergelijking op ziekenhuisniveau

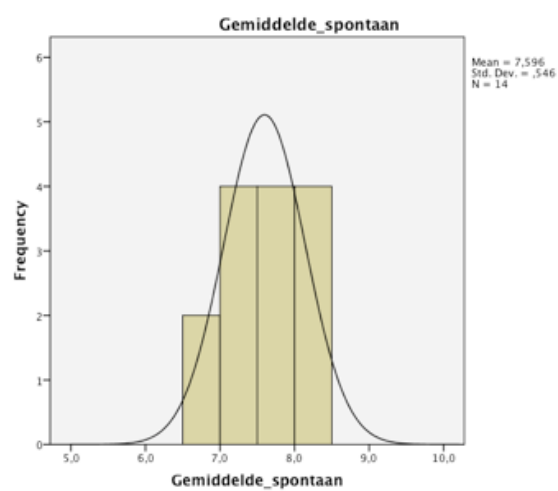
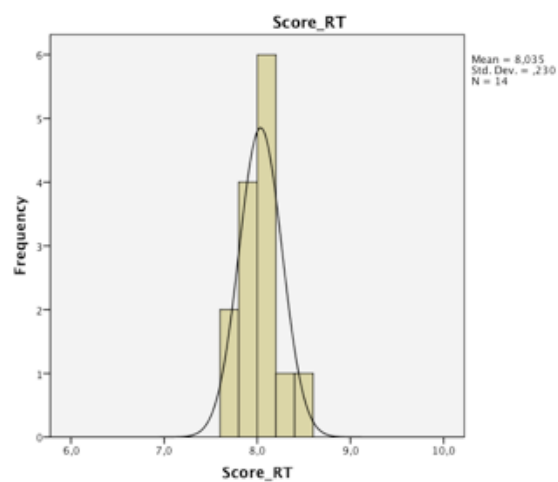
/STATISTICS=STDDEV VARIANCE MEAN MEDIAN

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

Statistics

		Gemiddelde_spontaan	Score_RT
N	Valid	14	14
	Missing	0	0
Mean		7,59636	8,03521
Median		7,55600	8,05100
Std. Deviation		,546449	,229957
Variance		,299	,053



CORRELATIONS

```
/VARIABLES=Gemiddelde_spontaan Gemiddelde_RT
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

Correlations

		Gemiddelde_spontaan	Gemiddelde_RT
Gemiddelde_spontaan	Pearson Correlation	1	,036
	Sig. (2-tailed)		,903
	N	14	14
Gemiddelde_RT	Pearson Correlation	,036	1
	Sig. (2-tailed)	,903	
	N	14	14

GRAPH

```
/SCATTERPLOT(BIVAR)=Score_RT WITH Gemiddelde_spontaan
/MISSING=LISTWISE
/TITLE='Spontaan vs RoadTeam'.
```

