Capabilities and validity of graphical methods in explorative data analysis

Benjamin Mayer[1], Martina Faul[2], Matthias Helm[2]

[1] Institute for Epidemiology and Medical Biometry, Ulm University
[2] Department of Anesthesia and Intensive Care Medicine, Armed Forces Medical Centre Ulm

Correspondence:
Dr. Benjamin Mayer
Institute for Epidemiology and Medical Biometry
Ulm University
Schwabstr. 13
89075 Ulm
Telephone: +497315026896
Fax: +497315026902
benjamin.mayer@uni-ulm.de

Word count: 2138

Abstract word count: 218

Number of figures: 3

Number of tables: 1

Number of references: 14

Abstract

Background: The adequate use of statistical methods is an indispensable aspect of generating novel findings in medical research. In general, there is still a strong focus on presenting the p-values of biometric results, although their validity has to be assessed critically. At the same time, the capabilities of graphical approaches of descriptive statistics are exhausted just rarely to outline the essential findings emphatically.

Methods: By means of a retrospective data sample of emergency medicine, the application of simple, but non-standard graphical methods is demonstrated to highlight the most impressive results. The objective was to identify injury patterns at particular accident localities. Absolute and relative frequencies, as well as weighted kappa coefficients are graphically displayed to present the results.

Results: Appropriate contingency tables enable a rapid identification of the most frequent combinations of site of accident and type of injury. Furthermore, respective spider and star charts illustrate alternatively the most important connections and frequencies. A diagram to display the kappa coefficients with confidence intervals facilitate a fast assessment of the accordance of diagnoses at site of accident and clinic.

Discussion: Graphical methods of descriptive statistics are suitable to work out the essential results of a medical research question in retrospective samples. Effectively, there are no creative limits for their application to illustrate a particular issue in an individual manner.

Key words: retrospective data analysis; graphic; p-value; injury patterns

Introduction

Biometrical evaluation of study data is an indispensable aspect of generating novel findings in medical research. The application of appropriate statistical methods enables to screen and summarize the huge amounts of data in order to uncover the main results and coherencies. The available spectrum of methods is large and has been continually developed on the basis of new research questions [1].

Utilization of descriptive statistical methods should always be the first step of both data analysis and presentation of results, respectively [2]. This gives a first impression of the essential characteristics of the underlying study collective [3]. Specific approaches to answer the primary research hypothesis then usually go beyond the spectrum of descriptive analysis methods. In general, the presentation of these results is primarily focused on displaying p-values, but often still without challenging their validity [4,5].

Interpreting the p-value in a rigorously confirmative manner requests the formulation of the null and alternative hypotheses prior to data collection [1,3,6]. Consequently, the interpretation of a statistical test on the basis of retrospective data can only be explorative and must especially not be considered as a statistical proof. Another example of p-values with debatable validity are correlation analyses. The corresponding statistical test only proves whether the correlation coefficient is different from zero in the default setting. Analyses which rely on a large sample size thus can lead to significant p-values, although the underlying coefficient is close to zero which would indicate no considerable correlation between the investigated variables from a contentual perspective.

An increasing complexity of the investigated medical issue necessarily leads to a more sophisticated biometrical methodology which should be used for the analysis. A concomitant difficulty is often to be able to interpret the results in an intuitive way. Moreover, increased complexity of statistical methodology is associated with the number of assumptions which have be fulfilled by the underlying data to obtain valid results. The application of inappropriate methods therefore restricts interpretation of the results, especially of the p-values [7].

Biometrical methods which attribute to the field of inferential statistics are doubtless of high importance to answer medical research questions. The ability of simple methods of descriptive statistics is often underrated, although especially graphical descriptions provide a virtually unlimited spectrum of options to present the crucial facts of a statistical evaluation in a versatile, individual and vivid manner [8]. This article demonstrates the opportunities of applying graphical methods to show the most important findings by means of a data example from emergency medicine, without the necessity of drawing on complex or hard to interpret statistical methods and non-informative p-values.

Material and methods

Data example
Although lethality of severely injured persons could be crucially lowered in Germany during the last 20 years, the polytrauma is still the most common reason of death for people aged 44 and younger [9]. Determining basics for a successful care of polytrauma patients are essentially created by an adequate pre and inner clinical

primary care [10]. Due to restrictions regarding the diagnostic capabilities by means of specific devices in the course of preclinical care, the clinical-physiological examination, anamnesis and investigation of cause as well as course of an accident form decisive fundamentals for an optimal emergency care. Ideally, the actively involved emergency doctors have preferably detailed information about the patient's status. Especially in case of traffic accidents, however, this information is often not available. Desirable were therefore indications which injury patterns have to be mainly expected at particular traffic accidents and scenes, respectively, to be able to prepare respective treatment schemes.

Thus, all traffic accidents with emergency doctor operation between May 2005 and October 2009 were retrospectively analyzed at the air rescue center "Christoph 22/Ulm". These included 479 patients with 181 having had a car accident with known accident locality, thereof 112 patients with 223 injuries of clinical AIS score ≥ 3 [11]. This subgroup has been considered for the analysis of injury patterns, since an optimal emergency care has highest priority especially for seriously injured patients. The recorded injuries were divided into 9 categories (intracranial injury, face, neck/cervical spine, thorax, abdomen, thoracic/lumbar spine, pelvis, upper limb, lower limb) and the accident localities into 4 categories (freeway, T-junction, crossing, knot-free route). Of primary interest was if the accident localities came along with particular injury patterns and how the suspected diagnosis of the emergency doctor (severity) was confirmed at the hospital. This second research question based on the entire collective of 479 patients.

Statistical methods
The primary objectives were investigated by means of descriptive analysis approaches. Absolute and relative frequencies as well as weighted kappa coefficients were calculated, and these findings were supplemented by respective graphics and tables. An explorative test with respect to differences in proportions of particular kinds of injuries at distinct accident localities was realized with Fisher's exact test. The strength of accordance of diagnoses at accident scene and hospital was evaluated following [12].

Results

Injury patterns
Preparing a cross table (table 1) which arranges the categories pursuant to the size of their line and column sums enables a clear description of the most often observed combinations of accident scene and type of injury. Overall, there were 223 injuries in 181 patients with AIS ≥ 3. A majority of 161 injuries (72%) could be assigned to knot-fee routes, 31 (14%) to crossings, 20 (95) to T-junctions and 11 (5%) to freeways. Thoracic injuries were primarily observed (79 cases, 35%), as well as intracranial trauma (45 cases, 20%), injuries of the lower limb (34 cases, 15%) and abdominal injuries (24 cases, 11%). However, the frequencies of particular injury patterns at different accident localities were not distinct (p=0.84).

To further analyze and identify injury patterns a spider chart has been created based on the relative frequencies (related to the total number of injuries) of table 1. Figure 1 depicts the frequency of a particular type of injury at an accident locality by means of a connecting line of respective width. Thus, the line width enables a quick

4

identification of frequently observed combinations, e.g. thorax and knot-free route. Moreover, the spectrum of injuries can be swiftly displayed for each of the defined accident localities (number of leaving and arriving connecting lines), or vice versa, the appearance of distinct injury types at particular accident localities. It can be found that injuries of the upper limb with AIS $\geq$ 3 only occurred at knot-free routes.

Figure 2 summarizes the findings from the contingency table and the spider chart once again jointly in a star diagram. The size of the star centers is proportional to the observed rate of all accidents which were allocated to respective localities. Considering each star enables a rapid overview of the represented injury types. Each observed injury is depicted as a single dot according to its category. This also hints at the injuries' distribution to the different categories.

Emergency doctor and hospital diagnoses
The accordance of severity evaluation by means of diagnoses at accident locality made by the emergency doctor (Utstein-Trauma-Style [13]) and the hospital (AIS) afterwards was weak or moderate for distinct body regions (weighted kappa coefficient from 0.02 to 0.67). Only the lower limb was evaluated strong with a 95% probability. To illustrate these findings, the point estimates of the kappa coefficients together with the respective 95% confidence intervals were indicated in a scheme (figure 3) which also included boundaries for different rating categories (weak, light, moderate, good (strong), very good (strong)).

Discussion

The primary objective of this article was to demonstrate how simple methods of descriptive data analysis can be used to identify and display crucial results of a medical research question. The main focus was on the use of inventive graphics which contrast with commonly applied standard approaches of displaying statistical results in practice [8].

Victor et al. [4] stated that publications of medical research are flooded with p-values and the wording "significance", often without a discussion if the conducted statistical test should be interpreted in an explorative or confirmatory manner. The actual meaning of the p-value as decision criterion for rejecting of retaining the null hypothesis takes a backseat in many cases and is rather used to substantiate a statement [4]. There are numerous situations where p-values are not convincing, as for example in figure 3. The corresponding p-values of the displayed kappa coefficients were invariably below the significance level of 0.05, although the coefficients were all rated weak to moderate according to the common interpretation [12,14] with one exception. One has to mind the sample size of the considered data example (473 to 479 cases for the accordance evaluation) which provides the tests with reasonable statistical power. As implemented by default in all statistical software packages, the statistical tests for kappa or correlation coefficients, too, simply prove if the coefficients are different from 0. Of course, the null hypothesis could be adapted to a more reasonable basis (e.g. kappa = 0.5), but this is usually never done. As a result, the p-value based on the default setting cannot be used to support the descriptive findings. In contrast, the application of graphical descriptive methods is recommendable for the considered data example, since figure 3 quickly indicates that

5

there is only a weak or moderate accordance of both diagnoses for the distinct body regions.

The clarity of contingency tables as a standard tool of descriptive statistics to investigate and display relations of two categorical variables decreases as the dimensions of the tables grow. Therefore, the rows and columns in table 1 were additionally arranged according to their row and column sums, respectively. This again enables a rapid identification of frequent (top left) and rare (bottom right) combinations of injury type and accident locality. Furthermore, the spider chart in figure 1 clarifies that only accidents at knot-free routes showed the whole spectrum of injuries, whereas freeway accidents only lead to 3 particular injury types (presence of injury). Additionally, the spider chart depicted the relative number of the combinations (frequency of injury) by means of the line width. The alternative perspective in figure 2 shows again that knot-free routes are the most common accident locality regarding the considered study collective. This is represented by the centers of the stars. Declaring the respective sample size per accident locality is important to be able to qualify the findings once again. The whole spectrum of injuries was indeed observed solely at knot-free routes, but this accident locality had the largest sample size and so the fact that freeway accidents had a smaller spectrum of injuries is may due to the much lower number of cases. Also the finding that injuries of the upper limb could just be observed at knot-free routes is eventually owed to the circumstance that serious injuries of the upper limb are rare and therefore only could be captured if the sample size is large enough.

The sample study showed that the calculation of p-values is not generally appropriate to support a descriptive finding. In case of the reported kappa coefficients the indication of p-values did not improve knowledge and was even misleading, since the corresponding null hypothesis was kept according to the default setting of the software with kappa = 0. Of course, it is possible to adapt the null hypothesis to a particular kappa of interest (e.g. kappa $\geq$ 0.5), but this would have led to non-significant p-values for the sample data. Moreover, the p-value for Fisher's exact test in table 1 does not point to particular relations, though figures 1 and 2 hint at specific findings (e.g. only three injury types were observed at all accident localities).

The presented study also has some limitations. It is important to remember again the fact that the number of cases referred to the number of injuries and not patients. There are probably injuries which come under the same patient, so the application of more complex statistical methods would be required to analyze the data in more detail. Moreover, the underlying sample size for investigating the relation of injury type and accident locality is maybe not sufficient to generate strong conclusions. However, the focus of this article was to present the capabilities of graphical approaches in explorative data analyses, so there was originally no aspiration to conclude in a confirmative manner. This is exactly the problem when results are interpreted (14), i.e. often there is no critical distinction between explorative and confirmative testing.

The capabilities of graphical approaches in descriptive data analyses are unlimited in principle. Of course, if the underlying data were prospectively collected descriptive approaches are not sufficient since the original objective was to generate causal results with stringent conclusiveness. In case of analyzing retrospective data, however, confirmative findings are barred at the outset, so the value of descriptive analysis approaches is increased. The application of graphical methods should

6

therefore be intended in such situations because of their flexible options to display the specific facts of interest.

References

1. Hartung J. Statistics. *Text and handbook of applied statistics (14th edition).* [*Statistik. Lehr- und Handbuch der angewandten Statistik, 14. Auflage]* Oldenbourg Research Press, Munich, 2005

2. Larson MG. Descriptive statistics and graphical displays. *Circulation 2006, 114:76-81, doi: 10.1161/CIRCULATIONAHA.105.584474*

3. Harms V. *Medical statistics (8th edition).* [*Medizinische Statistik, 8. Auflage].* Harms Press, Lindhöft, 2012

4. Victor A, Elsaesser A, Hommel G, Blettner M. How to evaluate the flood of p-values? Directions to handle multiple testing – part 10 of the series to evaluate scientific publications [Wie bewertet man die p-Wert-Flut? Hinweise zum Umgang mit dem multiplen Testen – Teil 10 der Serie zur Bewertung wissenschaftlicher Publikationen]. *Deutsches Ärzteblatt international 2010, 107(4): 50-56*

5. Gelman A. P values and statistical practice. *Epidemiology 2013, 24(1): 69-72*

6. Goodman SN. Toward evidence-based medical statistics. 1: the p-value fallacy. *Annals of Internal Medicine 1999, 130: 995-1004*

7. Mayer B, Muche R. Chances and obstacles of pilot projects in animal research for statistical sample size calculation. *International Journal of Biological and Medical Research 2012, 3(4): 2319-2323*

8. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical Statistics 2008, 7: 20-35*

9. German Society for Trauma Surgery. Guideline for the care of severely injured persons. [Leitlinie zur Versorgung Schwerverletzter.] *http://www.amf.org*

10. Helm M, Hauke J, Schlafer O, Schlechtriemen T, Lampl L. Extended medical quality management by means of the example of Tracer diagnosis „polytrauma" – a pilot study from air rescue service. [Erweitertes medizinisches Qualitätsmanagement am Beispiel der Tracer-Diagnose „Polytrauma" – eine Pilotstudie aus dem Bereich des Luftrettungsdienstes.] *Der Anästhesist 2012, 61, published online first: DOI: 10.1007/s00101-012-1981-9*

11. Gennarelli TA: *The Abbreviated Injury Scale. 1990 Revision.* Association for the Advancement of Automotive Medicine (AAAM), Des Plains IL, 1990

12. Altman DG: *Practical statistics for medical research.* Chapman & Hall, Boca Raton, 1991

13. Dick WF, Baskett PJ: Recommendations for uniform reporting of data following major trauma – the Utstein style. A report of a working party of the International Trauma Anaesthesia and Critical Care Society (ITACCS). *Resuscitation 1999*, *42:81-100*

14. Viera AJ, Garrett JM: Understanding interobserver agreement: the kappa statistic. *Family Medicine 2005, 37(5): 360-363*

Figure legends:

Figure 1: Spider chart of combinations of accident locality and injury type (line scale: 0.1pt = 1% of the total amount of observed combinations; ICI = intracranial injury, FI = face injury, NCSI = neck/cervical spine injury, TI = thoracic injury, AI = abdominal injury, TLSI = thoracic/lumbar spine injury, PI = pelvis injury, ULI = upper limb injury, LLI = lower limb injury, FW = freeway, TJ = T-junction, C = crossing, KFR = knot-free route)

Figure 2: Star chart of combinations of accident locality and injury type (absolute frequencies per combination; size of star centers proportional to the number of observed injuries per accident locality; ; ICI = intracranial injury, FI = face injury, NCSI = neck/cervical spine injury, TI = thoracic injury, AI = abdominal injury, TLSI = thoracic/lumbar spine injury, PI = pelvis injury, ULI = upper limb injury, LLI = lower limb injury, FW = freeway, TJ = T-junction, C = crossing, KFR = knot-free route)

Figure 3: Weighted kappa coefficients and 95% confidence intervals of accordance of diagnosis of severity at accident locality and clinic (kappa: 0-0.2 = weak, 0.2-0.4 = light, 0.4-0.6 = moderate, 0.6-0.8 = good, 0.8-1 = excellent)

Table 1: Frequency of injury types at distinct accident localities

	Knot-free route	Crossing	T-junction	freeway	Σ rows
Thorax	52 (65.8%)[a] (32.3%)[b] (23.3%)[c]	13 (16.5%) (41.9%) (5.8%)	7 (8.9%) (35.0%) (3.1%)	7 (8.9%) (63.6%) (3.1%)	79 (100%)
Intracranial	31 (68.9%) (19.3%) (13.9%)	7 (15.6%) (22.6%) (3.1%)	5 (11.1%) (25.0%) (2.2%)	2 (4.4%) (18.2%) (0.9%)	45 (100%)
Lower limb	26 (76.5%) (16.1%) (11.7%)	2 (5.9%) (6.5%) (0.9%)	4 (11.8%) (20.0%) (1.8%)	2 (5.9%) (18.2%) (0.9%)	34 (100%)
Abdomen	17 (70.8%) (10.6%) (7.6%)	5 (20.8%) (16.1%) (2.2%)	2 (8.3%) (10.0%) (0.9%)	0 (0%) (0%) (0%)	24 (100%)
Pelvis	11 (84.6%) (6.8%) (4.9%)	2 (15.4%) (6.5%) (0.9%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	13 (100%)
Neck/cervical spine	8 (72.7%) (5.0%) (3.6%)	1 (9.1%) (3.2%) (0.4%)	2 (18.2%) (10.0%) (0.9%)	0 (0%) (0%) (0%)	11 (100%)
Upper limb	9 (100%) (5.6%) (4.0%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	9 (100%)
Face	4 (80.0%) (2.5%) (1.8%)	1 (20.0%) (3.2%) (0.4%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	5 (100%)
Thoracic/lumbar spine	3 (100%) (1.9%) (1.3%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	0 (0%) (0%) (0%)	3 (100%)
Σ columns	161 (100%)	31 (100%)	20 (100%)	11 (100%)	223

[a] row percentages, add up to 100%
[b] column percentages, add up to 100%
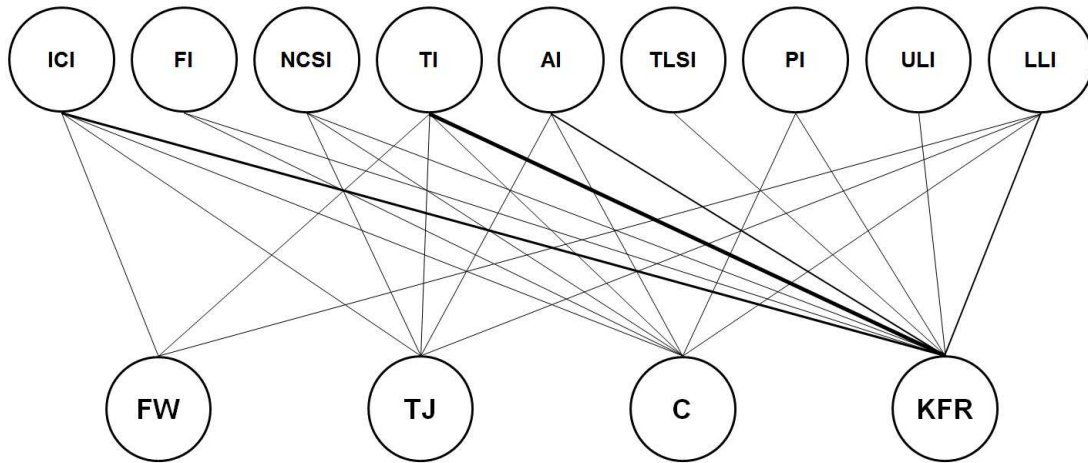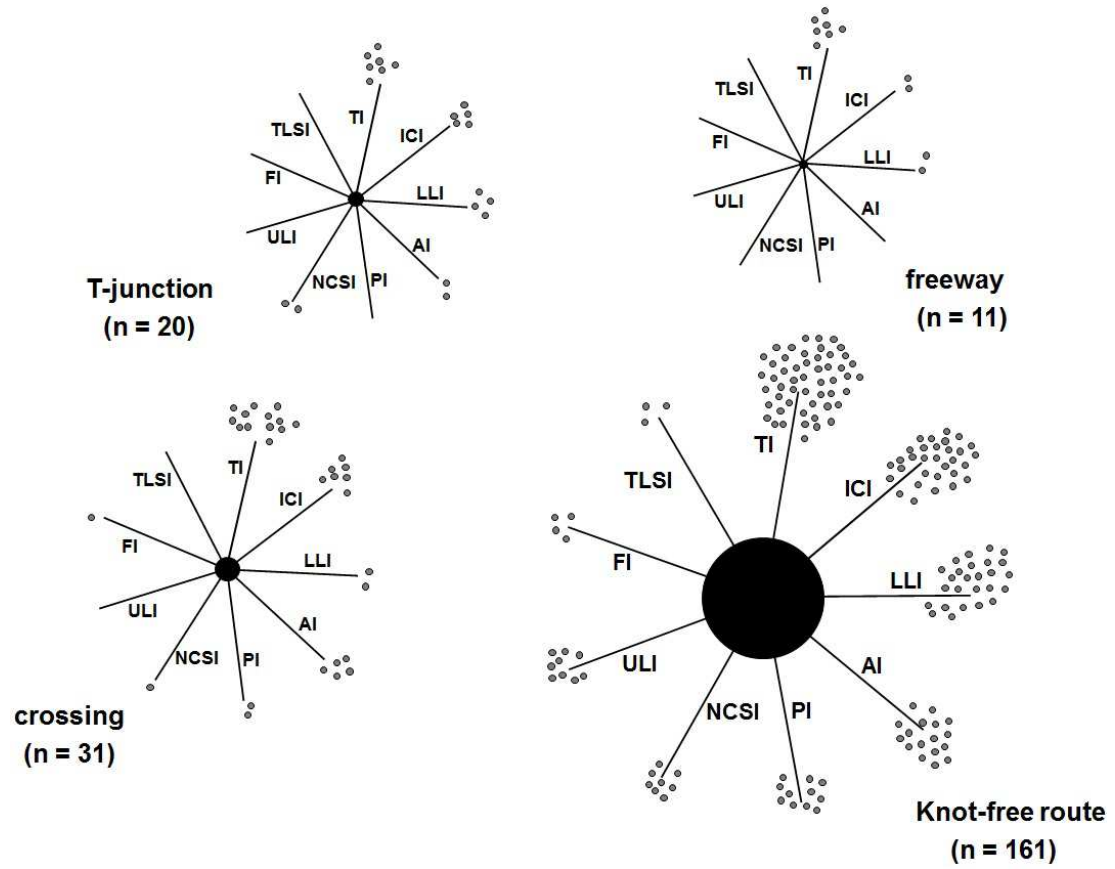[c] overall percentages, related to the total number of injuries (n=223)
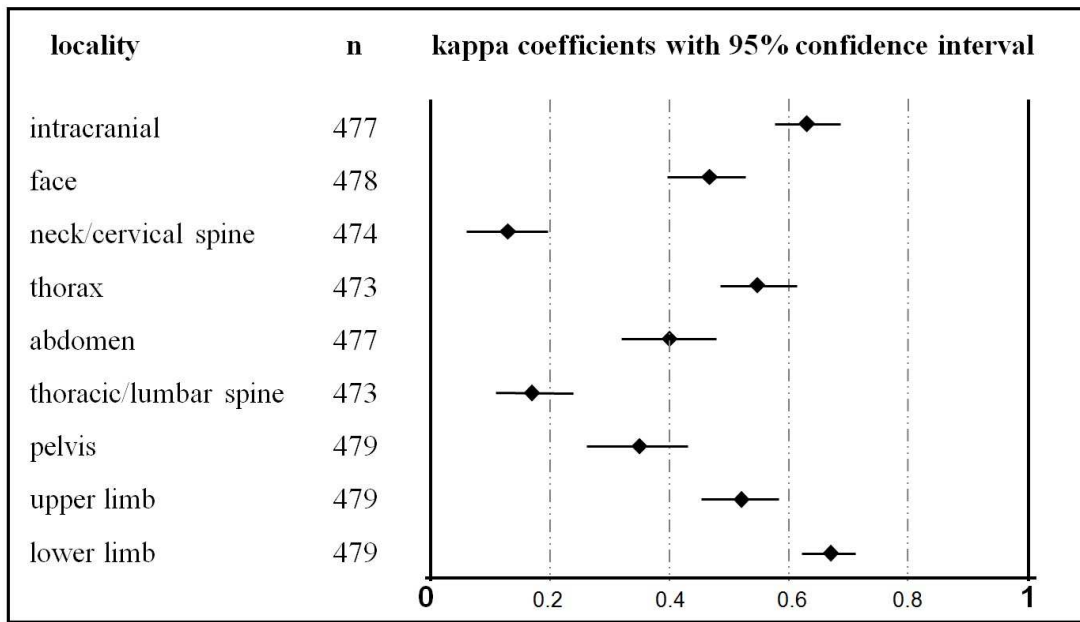
Figure 1

Figure 2

locality	n	kappa coefficients with 95% confidence interval
intracranial	477	
face	478	
neck/cervical spine	474	
thorax	473	
abdomen	477	
thoracic/lumbar spine	473	
pelvis	479	
upper limb	479	
lower limb	479	

Figure 3