Basics of Statistics for Clinical Research in Hand Surgery

Estadísticas básicas para la investigación clínica en cirugía de mano

Roberto S. Rosales[1] Isam Atroshi[2,3]

[1] Hand Surgery and Microsurgery Unit, GECOT, La Laguna, Tenerife, Spain
[2] Department of Clinical Sciences, Lund University, Lund, Sweden
[3] Department of Orthopedics Hässleholm-Kristianstad, Hässleholm Hospital, Hässleholm, Sweden

Address for correspondence Roberto S. Rosales, MD, PhD, Unidad de Cirurgía de la Mano y Macrocirurgía, GECOT, C/ María del Cristo Osuna, 20, 38204, San Cristobal de La Laguna, Santa Cruz de Tenerife, Spain (e-mail: titorosales@telefonica.net).

Rev Iberam Cir Mano 2018;46:150–171.

Abstract

Statistics, the science of numerical evaluation, helps in determining the real value of a hand surgical intervention. Clinical research in hand surgery cannot improve without considering the application of the most appropriate statistical procedures. The purpose of the present paper is to approach the basics of data analysis using a database of carpal tunnel syndrome (CTS) to understand the data matrix, the generation of variables, the descriptive statistics, the most appropriate statistical tests based on how data were collected, the parameter estimation (inference statistics) with *p*-value or confidence interval, and, finally, the important concept of generalized linear models (GLMs) or regression analysis.

Keywords
- clinical research
- statistic
- carpal tunnel syndrome
- methodology
- hand

Resumen

La "Estadística" es la ciencia de la evaluación numérica que nos permite determinar el valor real de las intervenciones quirúrgicas. La Investigación Clínica en Cirugía de la mano no puede mejorar sin considerar la correcta aplicación de los test estadísticos. El objetivo del presente trabajo de actualización es abordar los principios de los análisis estadísticos usando una base de síndrome de túnel carpiano (STC) para entender la matriz de datos, la generación de variables, la estadística descriptiva, los test más apropiados en base a como han sido recolectada la información, la estimación de parámetros (inferencia estadística) mediante valores "p" o intervalos de confianza, y finalmente, la introducción de los conceptos de "Modelos L
ineales Generalizados".

Palabras clave
- investigación clínica
- estadística
- síndrome del túnel carpiano
- metodología
- mano

Introduction

Hand surgery has experienced important technical advances after the first description of small-artery repair in the digits by Harold Kleinert in 1963.[1] Developments include digital replantation, toe-to-hand transfers, free flaps for hand reconstruction, hand transplantation, and even improvements in nerve microsurgery. However, the quality of our clinical research is still fair.[2] The quality of the reporting in randomized controlled trials (RCTs) is still poor, with a median score of 2 points in the modified Jadad scale, which classifies the quality of RCTs from 0 (worst quality) to 5 (best quality).[3] Despite the fact that the practice of wrist and shoulder arthroscopy started almost at the same time, the efficacy of arthroscopically performed wrist interventions has only been studied in four RCTs; the median modified Jadad score for the wrist RCTs was 0.5 (range 0–1), compared with 3.0 for the 50 RCTs of significantly higher quality assessing

received
October 3, 2018
accepted
October 11, 2018

DOI https://doi.org/10.1055/s-0038-1675587.
ISSN 1698-8396.

Copyright © 2018 Thieme Revinter Publicações Ltda, Rio de Janeiro, Brazil

License terms

interventions performed through shoulder arthroscopy.[4] More than 80% of the papers published in The Journal of Hand Surgery European Volume (JHSE) and 68% of those published in The Journal of Hand Surgery American Volume (JHSA) presented level-IV evidence (that is, mostly case series studies). The percentages of published papers with the highest level of evidence, including high-quality RCTs and systematic reviews (SRs), in the JHSE was 0.9% (level I) and 5.0% (level II), and in the JHSA, they were 8.3% (level I) and 10% (level II), which is very low.[5]

Hand surgery, as a scientific discipline, should require an appropriate and systematic analysis of all variables to demonstrate the significance of the observations that the surgeons face every day in their clinical practice. There are two approaches to this issue: the first one is to use the experience of the hand surgeon to answer questions that arise in daily practice. The second one is to complement the knowledge acquired after years of practice with a scientific evaluation of the observations reached by clinical research based on statistics, the science of numerical evaluation that can thoroughly help determine the real value of a hand surgical intervention. We cannot conduct good clinical researches in hand surgery without considering the application of the most appropriate and accepted statistical procedures.

The purpose of the present paper was to approach the basics of data analysis using a database of carpal tunnel syndrome (CTS) to understand the data matrix, the generation of variables, the descriptive statistics, the most appropriate statistical tests based on how we collect the data, the parameters estimation (inference statistics), the appropriate use of the *p*-value and confidence interval (CI), and, finally, the important concept of generalized linear models (GLMs) or regression analysis.

Before Starting

The main part of clinical research starts with the project before the statistical analysis. Based on our observations in clinical practice followed by a systematic review of the scientific literature, we can establish a research question. After that, we have to select the study population or sample for answering our question, the study design (case series, cross-sectional study, case-control study, cohort study, experimental study), and the appropriate instruments and measurements.[2] Finally, we can start the data analysis.

Example and Data Matrix

1–Study Description

In this part, we present a study using part of a database of CTS originating from an experimental clinical design or RCT designed to analyze the effect of reconstructing the transverse carpal ligament on the outcomes of open carpal tunnel (CT) release. The researcher should preestablish the variables to be analyzed based on the clinical design. The teaching purposes in the present study were to demonstrate the following hypotheses:

Hypothesis 1: Smokers have worse symptoms related to CTS than non-smokers.

Hypothesis 2: Reconstruction of the transverse carpal ligament improves the outcomes of open CT release.

2–Generating the original matrix data

In order to generate the matrix data, it is important to use a well-designed form to collect the data of the variables, such as Access software (Microsoft Corp., Redmond, WA, USA), or to use a second blinded check of the data entry, in case the researcher has used directly the Excel (Microsoft Corp., Redmond, WA, USA) worksheet for collecting the data. Otherwise, problems with the missing values will be encountered, which will also be approached in the present paper.

In ►**Fig. 1**, we can observe the initial matrix data in an Excel worksheet with the following variables: "id", "sex", "smoking" (cigarettes/day [c/d]), "sspre" (severity of symptoms related to CTS before surgery based on the CTS-6 scale), "sspost" (severity of symptoms related to CTS 3 months after surgery), and "reconstruction". The CTS-6 scale is the shorter version of the CTS questionnaire developed by Levine et al in 1993.[6] The CTS-6 scale was developed by Atroshi et al[7,8] and measures the severity of symptoms related to CTS. In the present paper, the Spanish version of the CTS was used,[9] which was developed from the extended Spanish version of the CTS questionnaire,[10] with a good level of reliability and validity for measuring outcomes in patients with CTS.[11]

3–Preparing the data for the analysis: generating and labeling variables

Once we have the main variables and the matrix data, the Excel worksheet can be imported by most statistics computer software. In the present paper, we have used Stata version 14.5 (StataCorp, College Station, TX, USA), and all the charts and graphics have been created using the same software. The statistical techniques for analyzing a matrix data will vary based on the measurement scale (quantitative vs categorical) of the variables. For teaching purposes, in the present paper, we have generated different categorical variables from the variables "smoking," "sspre" and "sspost" to demonstrate the use of different statistical tests. In that way, we have generated the following categorical variables: "HabitSmoke" (Smoking habit), "SmokLevel" (cigarettes/day), and "NivSSpre" (severe level of symptoms before surgery), based on the potential need for surgery, which was defined by Atroshi et al[12] as a CTS symptom severity score of 3.2. We have also generated a new quantitative variable, "sschange", which measures the change in the severity of symptoms between before and after the surgery, based on the variables "sspre" and "sspost"[13]. The software enables us to recode the original string variable "reconstruction" (►**Fig. 1**) to a numerical binary variable: reconst (reconstruction; "0" = No, "1" = Yes). Observe that the statistical software was able to detect the missing values shown as ".", which will be important for the analysis (►**Chart 1**).

An important issue is converting a quantitative variable to a categorical variable. It is generally not recommended to convert a quantitative variable, such as "smoking" and

	A	B	C	D	E	F	G
1	Id	Sex	Age	Smoking	ss pre	ss post	Reconstruction
2	1	0	36	20	3,3	2,6	No
3	2	0	30	25	5,4	3,6	Yes
4	3	0	52	0	5,3	1,6	No
5	4	1	38	0	2,5	1,8	Yes
6	5	1	32	20	4,6	3,1	No
7	6	0	51	0	3,2	1,9	Yes
8	7		34	0	3,9	1,7	Yes
9	8	0	44	0	3,5	1,8	No
10	9	0	27	0	2,3	1,6	Yes
11	10	1	33	10	3,4	2	No
12	11	0	51	0	2,9	1,8	Yes
13	12	1	55	0	2,7	1,1	No
14	13	0	58	5	2,7	1,9	No
15	14	1	48	30	4,9	4,6	Yes
16	15	0	20	0	3,9	3,9	Yes
17	16	1	60	15	4,5	4,7	No
18	17	0	21	35	4,9	5	No
19	18	0	32	20	3,8	1,5	Yes
20	19	1	27	40	4,5	5,4	No
21	20	0	31	10	4,2	3	No
22	21	0	59	15	4	1,1	No
23	22	1	37	25	2,6	1,8	Yes
24	23	0	28	0	4	1,2	Yes
25	24	0	33	30	3,1	1,6	No
26	25	0	34	5	3,7	1,4	Yes
27	26	1	20	25	4,5	1,6	No
28	27	0	66	20	3,3	2,3	Yes
29	28	0	56	0	5,4	4,9	Yes
30	29	0	44	0	3,7	4,6	Yes
31	30				5,1	1,3	No
32	31				2,6	1,4	No

Fig. 1 Original database in Excel worksheet.

"sspre", into a categorical variable because it results in loss of information in the analysis and in the conclusions. This is done here for teaching purposes.

The management of the variable "age" should be appropriate. Usually, we ask the patients about their age and they give us the truncated age (45, 56, 34 years old, etc.). A more correct way is to record the exact time (age, follow-up, etc.). For this purpose, we record the date of birth and the date of entry into the study, and we can calculate "Age" (exact time in years) = (birth date – entry date)/365.25, obtaining the exact "age" variable with decimals. If we use the truncated age, we need to do a correction by adding 0.5 to the truncated age. This correction of 0.5 does not change the dispersion of the data or the standard deviation (SD), but only the mean (►**Chart 2**).

4–Handling missing values

The amount of missing values is an index of the quality of the study and of the data collection process in the research. A common mistake in the analysis is to exclude all study participants with missing values (observed not only among young researchers in hand surgery: "If we have enough patients in our study, just go ahead and delete all those with missing values"). Exclusion of the individuals with missing values could be a big problem. If the individuals excluded because of missing values constitute a random subsample of the total study population, the only problem will be decreased power and precision of the analysis. However, when the missing values are related to the values of some of the variables of the study, then we have a more important methodological problem (bias in the parameter estimation

Chart 1 Matrix data from Stata after preparing the data for the analysis

Id	Age	Gender	Smoking	HabitS~g	SmokLe~l	sspre	sspost	sschange	NivSSpre	Reconst
1	36	0	20	1	2	3.3	2.6	.7	1	0
2	30	0	25	1	2	5.4	3.6	1.8	1	1
3	52	0	0	0	0	5.3	1.6	3.7	1	0
4	38	1	0	0	0	2.5	1.8	.7	0	1
5	32	1	20	1	2	4.6	3.1	1.5	1	0
6	51	0	0	0	0	3.2	1.9	1.3	1	1
7	34	.	0	0	0	3.9	1.7	2.2	1	1
8	44	0	0	0	0	3.5	1.8	1.7	1	0
9	27	0	0	0	0	2.3	1.6	.7	0	1
10	33	1	10	1	1	3.4	2	1.4	1	0
11	51	0	0	0	0	2.9	1.8	1.1	0	1
12	55	1	0	0	0	2.7	1.1	1.6	0	0
13	58	0	5	1	1	2.7	1.9	.8	0	0
14	48	1	30	1	2	4.9	4.6	.3	1	1
15	20	0	0	0	0	3.9	3.9	0	1	1
16	60	1	15	1	1	4.5	4.7	-.2	1	0
17	21	0	35	1	2	4.9	5	-.1	1	0
18	32	0	20	1	2	3.8	1.5	2.3	1	1
19	27	1	40	1	2	4.5	5.4	-.9	1	0
20	31	0	10	1	1	4.2	3	1.2	1	0
21	59	0	15	1	1	4	1.1	2.9	1	0
22	37	1	25	1	2	2.6	1.8	.8	0	1
23	28	0	0	0	0	4	1.2	2.8	1	1
24	33	0	30	1	2	3.1	1.6	1.5	1	0
25	34	0	5	1	1	3.7	1.4	2.3	1	1
26	20	1	25	1	2	4.5	1.6	2.9	1	0
27	66	0	20	1	2	3.3	2.3	1	1	1
28	56	0	0	0	0	5.4	4.9	.5	1	1
29	44	0	0	0	0	3.7	4.6	-.9	1	1
30	.	.	.	.	.	5.1	1.3	3.8	1	0
31	.	.	.	.	.	2.6	1.4	1.2	0	0

Observe the new variables generated as the categorical variables: "HabitSmoke" (Smoking habit "0" = NO, "1"= Yes), "SmokLev" (Level of smoking cigarettes/day; "0"= No, "1"= 1–15, "2"= >15), "NivSSpre" (severe level of symptoms before surgery, "0" = No, "1" = Yes), "NivSSpos" (severe level of symptoms after surgery, "0" = No, "1" = Yes) . A new quantitative variable "sschange", which measures the change in the severity of the symptoms between before and after the surgery, was generated from the variables "sspre" and "sspost." Observe that the initial nominal variable "RECONSTRUCTION" in the original matrix data from excel (►**Fig. 1**) has been recoded to a numerical categorical variable: Reconst (Reconstruction; "0" = No, "1" = Yes). Observe that the statistical software was able to detect the missing values represented by a ".", which will be important for the analysis.

process in a multivariate model, for example). Based on the origin of missing values, Stata divides the missing values into: system missing (data without value, values incoherent with the variable format, and impossible values in a data transformation, such as 5/0); and user missing or extended missing values; for example, if a questionnaire item asks about smoking (Yes or No) and the next item asks about the number of cigarettes per day, it is expected to have an empty value or a value of 99 in non-smoking respondents, and it should be assigned as user missing. Based on the degree of possible bias, missing system values can be classified into: missing completely at random (MCAR), such as forgetting to record some data; missing at random (MAR), when the missing value is related to an independent confounder variable (X). For example, when studying whether losing weight could improve the symptoms of CTS, a high number of missing values was observed among younger individuals because younger subjects have less tendency to cooperate in the registration of their habits; and nonrandom missing (NRM), when it is related to the dependent variable (Y). For example, when analyzing the relationship between high blood pressure (HBP) and age group, it is expected that the individuals with HBP and older people present a better cooperation and consequently a lower number of missing values for these variables.

If we look at our data (►**Chart 1**), we have some missing values. The first step before starting our study is to analyze the missing values; probably, they are MCAR. The missing value analysis estimates and describes those values (►**Chart 3**), and it can give information about: total number and percentage of missing values, number of patients with missing values, number of missing values per patient, and the patterns of missing values. At the same time, most of the software generate a new variable called _Nmiss, or similar to it, which can be

Chart 2 Handling the age variable

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	29	39.89655	13.19455	20	66
AgeC	29	40.39655	13.19455	20.5	66.5

Observe the mean and standard deviation (SD). The correction of + 0.5 point in the variable Age only changes the mean but not the SD. The dispersion of the data in the variable AgeC is the same as that of the variable Age.

used to drop the individuals with missing values or to do other analyses to get more information about the missing values.

Most of the statistics software have two different approaches for data analysis with missing values: pairwise (using the maximum number of subjects available for each variable) and listwise (using only subjects with valid values) selection. For example, in multivariate linear regression analysis, Stata always uses listwise selection in the analysis of the model[14] (►**Charts 4A** and **4B**).

There are no rules about the accepted level of missing values. If we have more than 10% of missing values, it is recommended to perform descriptive statistics on the variables of the subjects with complete valid values and those with some missing values. Most of the statistics

Chart 3 Missing values analysis

STATISTICS OF MISSING VALUES

	Missing values			Not	Valid values				
Variable	System	User	Percent	Appl.	Freq.	Percent	Mean	Min	Max
Gender	3	0	9.68	0	28	90.32	.5714286	0	1
AgeC	2	0	6.45	0	29	93.55	40.39655	20.5	66.5
Smoking	2	0	6.45	0	29	93.55	12.06897	0	40
HabitSmoke	2	0	6.45	0	29	93.55	.5862069	0	1
NivTab	2	0	6.45	0	29	93.55	.9655172	0	2
sspre	0	0	0.00	0	31	100.00	3.819355	2.3	5.4
sspost	0	0	0.00	0	31	100.00	2.509677	1.1	5.4
sschange	0	0	0.00	0	31	100.00	1.309677	-.9	3.8
NivSSpre	0	0	0.00	0	31	100.00	.7741935	0	1
NivSSpos	0	0	0.00	0	31	100.00	.2903226	0	1
Reconst	0	0	0.00	0	31	100.00	.483871	0	1
Total	11	0	3.23	0	330	96.77			

Total missing values by subject

	Freq.	Percent
0	28	90.32
1	1	3.23
5	2	6.45
Total	31	100.00

Observe that the total number of missing values was 11. Two missing values in variables: AgeC, Smoking, HabitSmoke and NivTab. Three missing values were observed in Gender. Only 28 patients presented (90.32%) missing values = 0. Two patients (6.45%) presented five missing values, and only one (3.23%) presented one missing value.

Chart 4 Descriptive statistics using pairwise strategy (A) and listwise strategy (B)

A. Descriptive statistics using Pairwise strategy

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	29	39.89655	13.19455	20	66
Smoking	29	12.06897	12.78373	0	40
sspre	31	3.819355	.9221061	2.3	5.4
sspost	31	2.509677	1.356307	1.1	5.4

B. Listwise strategy

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	28	40.32143	13.23311	20	66
Smoking	28	12.5	12.80191	0	40
sspre	28	3.814286	.9103171	2.3	5.4
sspost	28	2.621429	1.380917	1.1	5.4

Observe that the means, standard deviation (SD), and number of observations change based on the strategy of approach of the missing values. In Listwise selection, the software only uses the patients with valid values, which means that Nmisss = 0.

software have some statistical methods to reduce missing values, such as mean imputation and multiple imputations (used in multivariate analysis).[15] The mean imputations process (►**Chart 5**) assigns the mean of the variable to the missing value. Consequently, we will obtain a decreased SD, but no changes in the mean of the variable by the imputation.[14]

Descriptive Statistics

After generating, recoding, labeling and handling missing values in the variables, we have to check the data. We have created a dataset, and now we need to check the work we have done defining the dataset. Checking the accuracy of our data entry is also our first statistical look at the data. Data checking is very important to detect possible errors in data entry.

The description of the data constitutes the starting point of our analysis and sometimes a proper objective of a study. Descriptive statistics are based on the calculation of several indexes and graphics, which give us information about the distribution of the data (central tendency, dispersion, symmetry, etc.) for every variable.

1. Description of quantitative variables.

We have two types of statistical indexes for describing quantitative variables: measures based on momentum (mean, SD, variance, skewness and kurtosis) and indexes based on ordination (median, quartiles, minimum, and maximum).

The mean represents the gravity center of the distribution in a specific variable and describes the central tendency if we do not have asymmetry. The SD (S) is the square root of the variance (S^2), which is obtained dividing the sum of squares by the degree of freedom (df) ($S^2 = SS/df$). The SD (S) is interpreted only if the distribution is normal and if it presents symmetry. In this case, the interval mean ± 1 SD contains 68% of the observations in that variable, the interval mean ± 2SD includes 95% of the observations, and the interval mean ± 3SD includes 99.7% of the cases.

$$\mathbf{S}^2 = \mathbf{SS} = \sum_{i=1}^{n}(Xi - \overline{X})^2, \mathbf{df} = \mathrm{n}\text{ -1}, \mathrm{S} = +\sqrt{S}^2$$

Other measures based on moments are skewness and kurtosis. Measures based on ordination are the median and quartiles. Those measures are not affected by the asymmetry of the distribution. The median and the mean are equal when the distribution is symmetric. The symmetry is measured by the skewness index, which will be positive when the mean is higher than the median, and negative when the mean is lower than the median. The degree of flattening of the distribution is measured by the kurtosis index (in Stata, kurtosis > 3 implies a distribution with more sharpness

Chart 5 Descriptive statistics before and after mean imputation for missing values

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	29	39.89655	13.19455	20	66
AgeMean	31	39.89655	12.74714	20	66

Observe that after mean imputation (AgeMean), the mean remained with the same value; however, the standard deviation (SD) decreased.

than a normal distribution). In ►**Fig. 2**, we can observe the P-P normal graph, the box plot and the stem and leaf diagram which give an image of the distribution of the data in the sspre variable. The stem and leaf diagram allows us to calculate in an easy way the median and the Q_1 = percentile 25 and Q_3 = percentile 75 quartiles. The Q_2 = percentile 50 is equal to the median. Besides, we can observe in the box plot that is very useful to see the distribution of the data created by the median and the Q1 and Q3 quartiles, forming the box and the minimum and maximum, which form the whiskers. Box plots can be used to see the symmetry, outliers, and to demonstrate change in a variable such as symptoms related to CTS if we compare the graph plot of the sspre with that of the sspost and see the improvement in symptoms after surgery (►**Fig. 3**). The descriptive statistics alerted us that there were errors during data entry because we could find values > 5 in ►**Fig. 2** (max = 5.4), and, in ►**Fig. 3**, the upper whiskers of the box plots of sspre and sspost were > 5, which is the maximum score that any item of the CTS-6 questionnaire can achieve. Consequently, the researcher should check the original data entry, even the original questionnaires, to

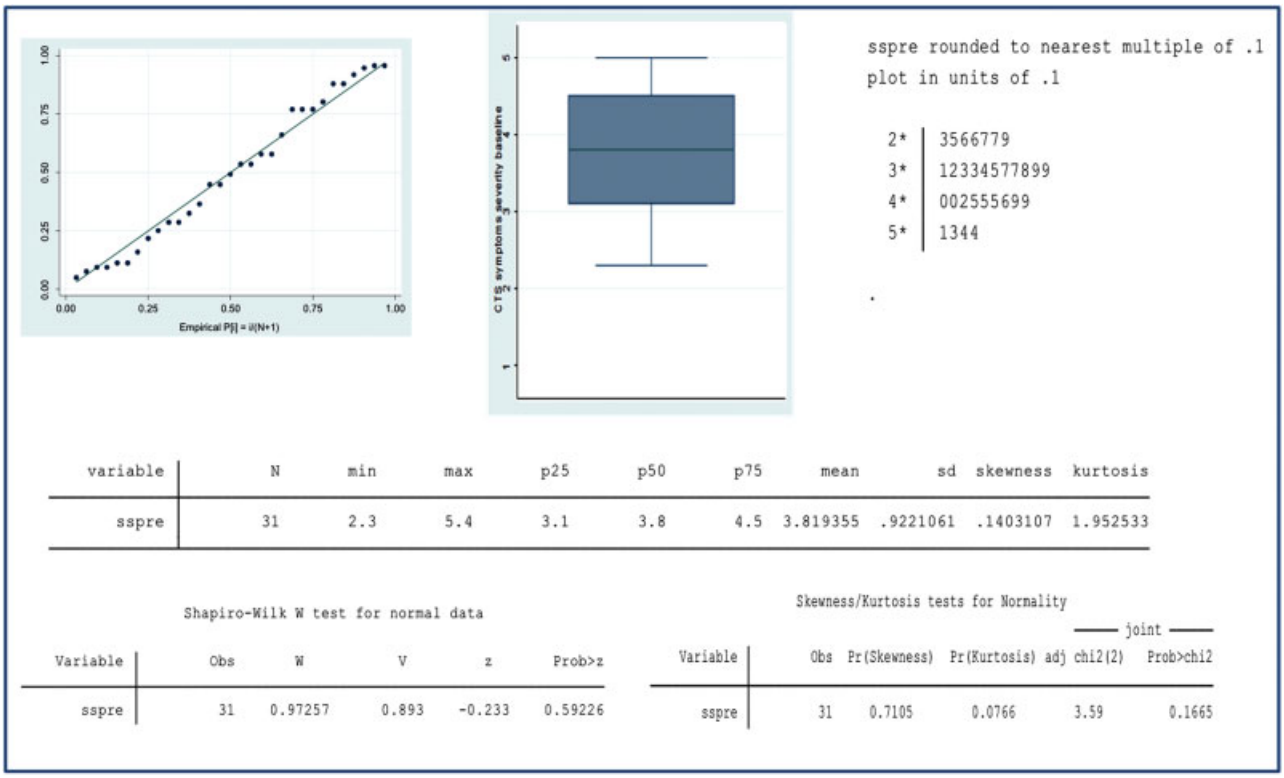


Fig. 2 Descriptive statistics of the quantitative variable "sspre." From left to right: Graphic P-P for normal distribution (observe that the data are almost coincidental with the line); box plot (the box delimitated by the p25 and p75 quartiles, and the median, the whiskers are the minimum and maximum); and the stem & leaf diagram. Descriptive statistics showed data with values higher than the maximum, and the normality tests demonstrated that the data for the sspre variable followed a normal distribution ($p > 0.05$). Abbreviations: sspre, carpal tunnel syndrome symptoms at the baseline before surgery.

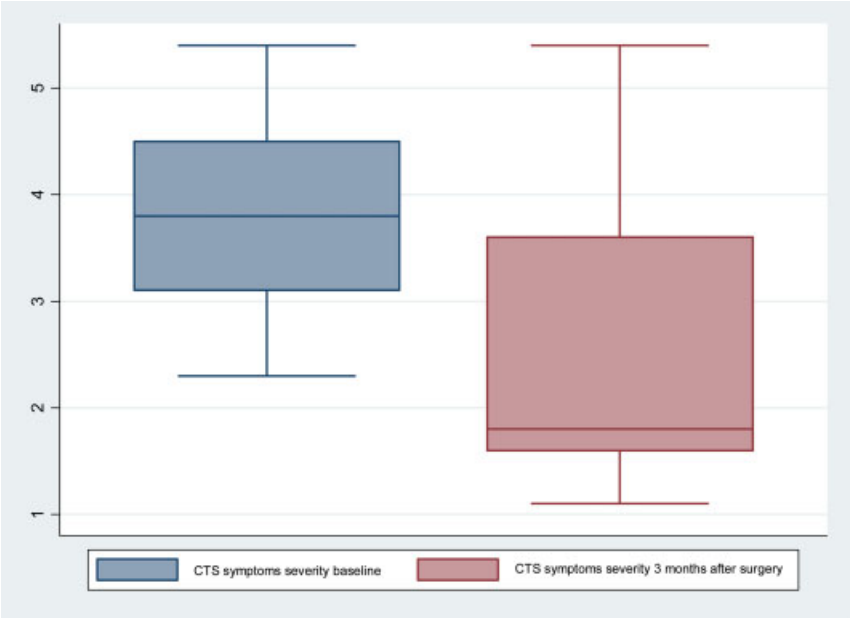


Fig. 3 Box plot of sspre and sspost variables. Observe the change in the severity of the CTS symptoms between before and after the surgery. Notice that the box plot of sspost presents a positive asymmetry. Abbreviations: CTS, carpal tunnel syndrome; sspost, severity of symptoms related to CTS 3 months after surgery; sspre, severity of symptoms related to CTS before surgery based on the CTS-6 scale).

overcome those mistakes. Stata can give us the Id number with sspre and sspost with values > 5 (list Id if sspre >5 & sspre < . | sspost >5 & sspost < .), which were Id: 2, 3, 19, and 30. If the researcher could not access the original matrix or questionnaires or if it is impossible to know the exact values of those variables in those individuals, the researcher should assign a missing value to those entries. The matrix data of the present paper was created for teaching purposes, but the rest of the analyses were done using the correct values in the database

2. Descriptions of categorical variables

Categorical variables (such as "gender", "HabitSmoke", "SmokeLevel", etc.) are described by creating a table of frequencies that classifies the individuals based on the category and calculates percentages (►**Fig. 4**).

The importance of the descriptive statistics is checking our data. The descriptive statistics constitute the main possibility for discovering errors in our data. The main checking list during the descriptive statistics will be:

- **Id**. Check that all variables present an Id number and that there is no duplication of the number.
- **Categorical variables**. Check that the values belong to the set of valid values (e.g., observe values different from 0 or 1 in the "gender" variable)
- **Quantitative variables**. Check that the values are included in the interval that defines the valid values (for

Smoking level	Severe level of symptoms (pre) Negative	Positive	Total
0 c/d	4	8	12
1-15 c/d	1	5	6
>15 c/d	1	10	11
Total	6	23	29

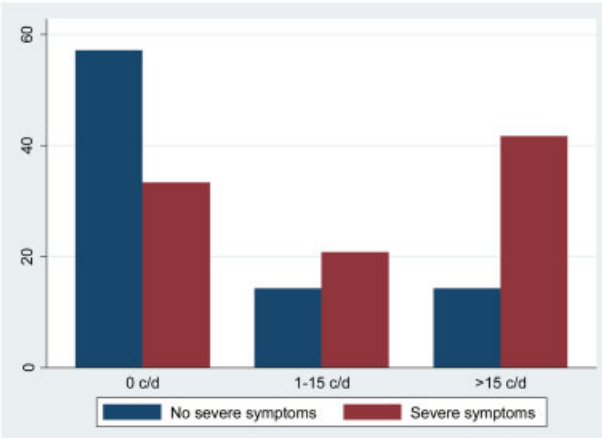


Fig. 4 Descriptive statistics of categorical variables. Cross tabulation of severe level of CTS symptoms (NivSSpre) and smoking level (SmokLevel) and bar graph. Abbreviations: CTS, carpal tunnel syndrome.

example, the sspre variable ranges from 1 to 5, and if the summarized information of the variable detects values over 5, it indicates an error in the collection or in the entry of the data).

- **Dates**. Check that the dates are correct if the researcher has used them for generating variables as "age" or "follow-up" time.
- **Consistency between variables**. Check that the values of a variable are consistent with others (for example, if we observe negative values in "age", it would indicate that the date of birth is posterior to the date of entry in the study).

If we detect errors during the checking process of the descriptive statistics:

- Replace the erroneous data with correct data if the researcher has the original records and information or access to the subjects of the study.
- Replace the erroneous data with missing values if the researcher does not have the correct information.

Inference Statistics

Decision-making in clinical research implies testing whether our hypotheses are true or false based on empiric results obtained in samples of individuals. The results in a specific research or study are based on the measurements made in one of the infinite samples that can be obtained from the reference population. Consequently, the value of our measurements can vary due to chance. Inference statistics assume that the random variability of sampling follows known laws and allows the researchers to quantify that variability and to facilitate decision-making about the hypotheses established and to draw conclusions.[14]

1. *P*-value and significance test of the null hypothesis (H_0).

Fisher[16,17] established in 1922 the significance test for null hypothesis. For example, our clinical observations allow us to suppose that the reconstruction of the transverse carpal ligament during open CT release could not improve the severity of the symptoms related to CTS. To demonstrate that hypothesis, a sample of 31 individuals was randomized to receive only open CT release (sample A: no reconstruction, $n_A = 16$) versus open CT release plus ligament reconstruction (sample B: reconstruction, $n_B = 15$). The severity of symptoms score ranged from 1 to 5. Group A presented a mean change in the severity of symptoms score between before the surgery and after the surgery (sschange variable) of $\bar{X}_A = 1.48$ (improvement in symptoms); and a SD of $S_A = 1.33$. Group B presented a mean of $\bar{X}_B = 1.13$; and a SD of $S_B = 1.004$. This hypothesis entails that there is a population A: "change in the severity of symptoms after open CT release", and a second population, B: "change in the severity of symptoms in open CT release plus ligament reconstruction." The researcher wants to know if the mean in population A (μ_A) is higher than the mean in population B (μ_B). However, the researcher does not know these values. To know whether or not the reconstruction improved the "sschange," the researcher needs a reference distribution to assess if the observed difference in "sschange" between groups A and B ($d_0 = 0.35$) (►**Fig. 5**) is true or caused by the random fluctuation of the sample. Still, the researcher does not know the magnitude of the difference (δ) between the means of "sschange" in the populations; consequently, this hypothesis cannot be demonstrated. A second option is that the researcher can formulate the null hypothesis that the reconstruction does not improve "sschange" (H_0: $\delta = \mu_A - \mu_B = 0$). Now, it is possible to establish a reference distribution around the H_0, because it is a very specific hypothesis ($\delta = 0$). The reference distribution will be the sample distribution of the observed difference with a mean equal to 0 ($\mu = 0$) and with a SD that will be the standard error (SE_d) of the difference, which is equal to the sum of the SE of the samples A and B $SE_d = \sqrt{(\sigma A/nA)} + \sqrt{(\sigma B/nB)}$.[18] The SDs of the real populations are not known, but if we assumed that difference of the samples presented a normal distribution, we could exchange them by the SD of the samples. From here, we can calculate the t_0 statistic (dividing d_0 by the SEd). Finally, we want to know the probability of finding a value of T similar or higher to t_0, taking into account the degree of freedom (df), which is equal to a total number of individuals ($n = 31$) minus 2 (2 groups) (df = 29). The obtained probability is higher than 0.05, which means that the difference between groups is compatible with the null hypothesis H_0[18]. Consequently, reconstruction of the transverse carpal ligament does not improve the change in the severity of the symptoms after CT release. In this example, given the teaching purpose of this paper, we have not taken into account the sample size, the power of analysis and the type II error, which will be explained below.

2. Null hypothesis ($H_{0)}$ versus alternative hypothesis (H_a). Alpha (α) and Beta (β) Risks

In 1928, Neyman et al,[19–21] based on the Fisher significance test, developed the hypothesis testing under a different point of view, apparently very similar to the one proposed by Fisher but totally different in that Neyman et al. They started with the probability theory to establish a rule for decision-making between two hypotheses that were complementary (H_0 = null hypothesis and H_a = alternative hypothesis). That theory constituted a frontal opposition to the inference proposed by Fisher. Neyman et al proposed the α and β risks and the type I and type II errors.

Type I error happens when H_0 is true and it has been rejected because the risk used was very high. Normally, in clinical research, we use a risk of type I error of 0.05 (α =0.05). The interpretation of α =0.05 is very important. Alpha risk represents the conditional probability of being wrong when the null hypothesis is true. When a researcher rejects the H_0 with a risk α of 0.05, it does not mean that the researcher is wrong 5 out of 100 times, because that would only be true if the null hypothesis were always true.[22]

Type II error occurs when the H_a is true, and the researcher accepted the H_0 because the test was not significant due to a small sample size. The probability of type II error is called Beta risk (β). The complementary probability 1-β represents the probability of accepting the H_a when H_a is true, and it is called power of the test. Usually, in health sciences or in clinical research, we use a power of 80%.

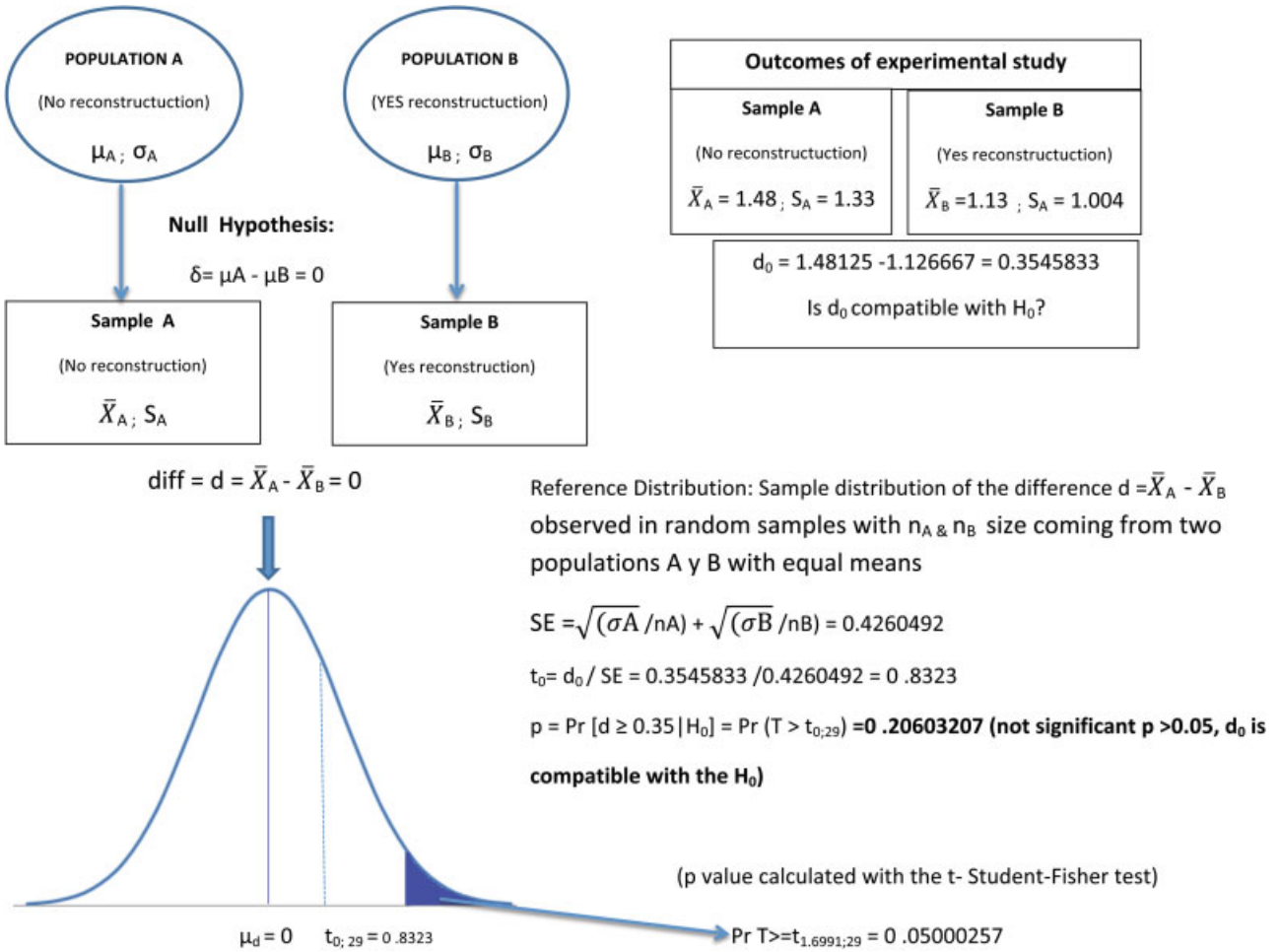


Fig. 5 Comparison of two means. Inference statistic. Note: (μA ; σA) = mean and SD in population A; (μB ; σB) = mean and SD in population B; $\bar{X}_A$; S_A = mean and SD in sample A; $\bar{X}_B$; S_B = mean and SD in sample B; SE = standard error. (*based* on Domenech[18,26]). Abbreviations: SD, standard deviation.

3. Confidence interval versus *p*-value.

In 1934, Neyman,[23] based on the hypotheses test, proposed a new method to answer questions such as the one established in the CTS and ligament reconstruction study explained before (►**Fig. 5**), based on Domenech.[18] A good alternative to the hypothesis testing is the CI around a mean. If the samples are representative of the reference populations, we can calculate an interval that could be considered the set of all hypotheses. In that way, one hypothesis located out of the interval is not credible and then, it can be rejected. It is not necessary to know the *p*-value to conclude whether the difference, observed in ►**Chart 6** and ►**Fig. 5**, is significant. If the H_0 is included in the 95% confidence interval (95% CI), we can conclude that the difference is not significant (mean difference = 0.3545833; 95% CI = -0.5167852 to 1.225952). The interpretation of the 95% CI is: with a confidence of 95%, we can locate the mean difference of "sschange" between -0.52 and 1.23 because the H_0, which is =0, is included in the interval in which the difference is not significant (►**Fig. 6**).

This concept, introduced by Newman, was forgotten for a long time. Today, however, it is very common to use 95% CI instead of presenting many *p*-values in result charts. The 95% CI has the advantage of giving information about the magnitude of the effect or of an important clinical change. If we already know the minimal important clinical difference (MICD) expected in our outcome variables (such as sschange) in a specific population, we cannot only determine whether the difference observed in the 95% CI is significant, but also assess if that difference (effect) is clinically important. If we look at ►**Fig. 7**, we can see the relationship between the significance and the effect assessed by different 95% CIs. The criterion for MICD on change in the severity of symptoms after CT release, which is ~ 0.9 to 1, has been measured using the CTS-6 PRO instruments,.[8,24] The first CI (A) is located above the MICD, which means that the 95% CI is significant (it does not include the H_0) and presents an important effect. Confidence interval "B" includes the MICD in the interval; however, it does not include the H_0; consequently, this 95% CI is inconclusive about the effect; however, it is statistically significant. Confidence interval "C" includes the criterion and the H_0, and it constitutes an inconclusive and not significant 95% CI. Confidence interval "D" is located below the criterion, but it does not include the H_0; consequently, it is a significant interval, but with a non-important clinical effect. Finally, CI

Chart 6 T-Student-Fisher test for comparing two means from independent samples. Change of symptoms severity (ChangeSS) in reconstruction vs no reconstruction of the transverse carpal ligament in open carpal tunnel release

A

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No	16	1.48125	.3330689	1.332276	.7713304	2.19117
Yes	15	1.126667	.2593888	1.004608	.5703331	1.683
combined	31	1.309677	.2118203	1.179366	.8770826	1.742272
diff		.3545833	.4260492		-.5167852	1.225952

```
    diff = mean(No) - mean(Yes)                                   t =   0.8323
Ho: diff = 0                                     degrees of freedom =       29

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.7940         Pr(|T| > |t|) = 0.4121          Pr(T > t) = 0.2060
```

B. t-Student test is a particular form of the regression analysis.

Source	SS	df	MS
Model	.973388441	1	.973388441
Residual	40.7537083	29	1.40530029
Total	41.7270968	30	1.39090323

Number of obs	=	31
F(1, 29)	=	0.69
Prob > F	=	0.4121
R-squared	=	0.0233
Adj R-squared	=	-0.0104
Root MSE	=	1.1855

sschange	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Reconst						
No	.3545833	.4260492	0.83	0.412	-.5167852	1.225952
_cons	1.126667	.3060828	3.68	0.001	.500657	1.752676

(A) Observe that the results of the one-tailed comparison of the two means is similar to the one observed in ►**Fig. 5**. Conclusion: the results of the one-tailed and two-tailed tests were not significant, assuming equal variances and normal distribution of the differences in sschange. (B) Observe the same results (diff of means = 0.3545833 = β coefficient [No reconstruction]) of comparing two means using regression analysis.

"E" is below the criterion and it does include the H_0; therefore, this 95% CI is not significant and the effect is not important.[18] In ►**Fig. 6**, the readers can understand, based on the 95% CI of the "sschange" difference between open CT release and open CT release plus ligament reconstruction, that the difference was not significant because it includes the H_0, and the CI is inconclusive considering to the magnitude of the effect because it includes the criterion.

Another aspect of the 95% CI is the precision of the interval. A narrower 95% CI implies a better precision because the SE is lower. If we have to select between two 95% CIs that are significant (that is, which do not include the H_0) and with an

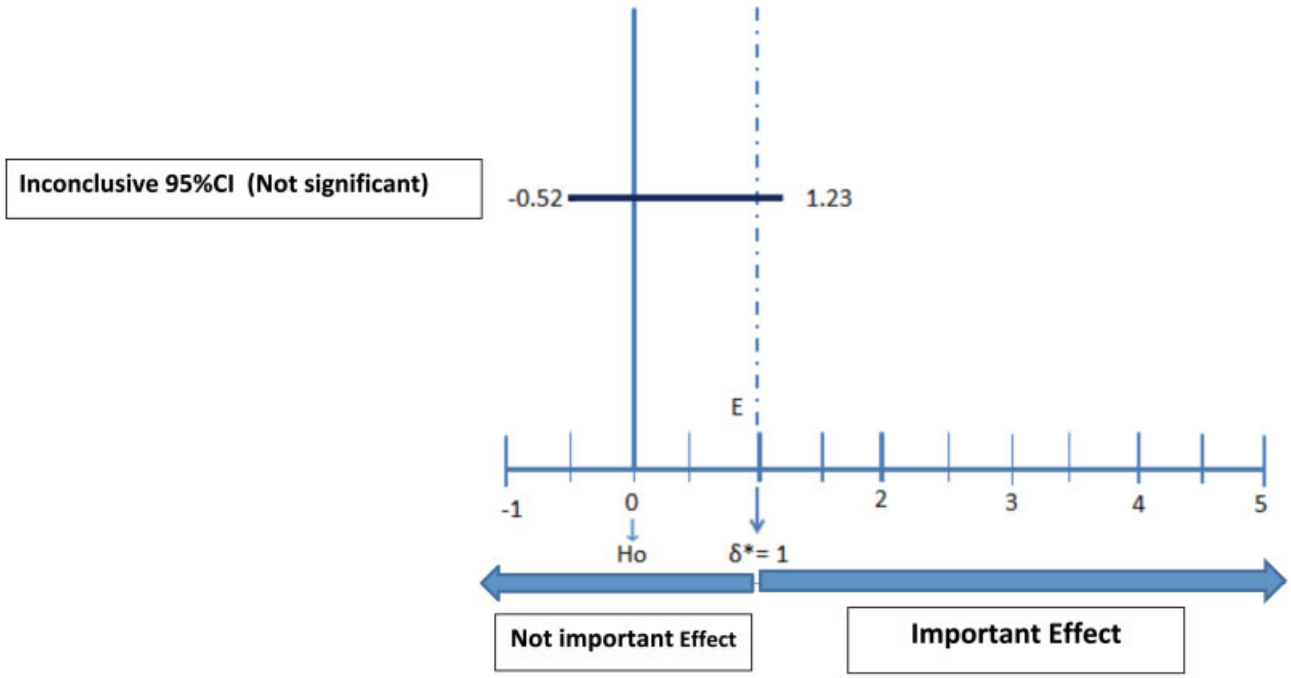


Fig. 6 Change in CTS symptoms in ligament reconstruction vs open CT release. Interpretation of the results using 95% CI (diff Mean = 0.3545833; 95% CI = - 0.5167852 to 1.225952). The H_0 is included in the 95% CI (results not significant). The criterion is included in the CI (inconclusive CI). Criterion of 1 point in the CTS-6 questionnaire, which measured the severity of the symptoms related to CTS. (based on Atroshi et al[8] and Ozyürekoğlu et al[24]). Abbreviations: CI, confidence interval; CT, carpal tunnel; CTS, carpal tunnel syndrome.

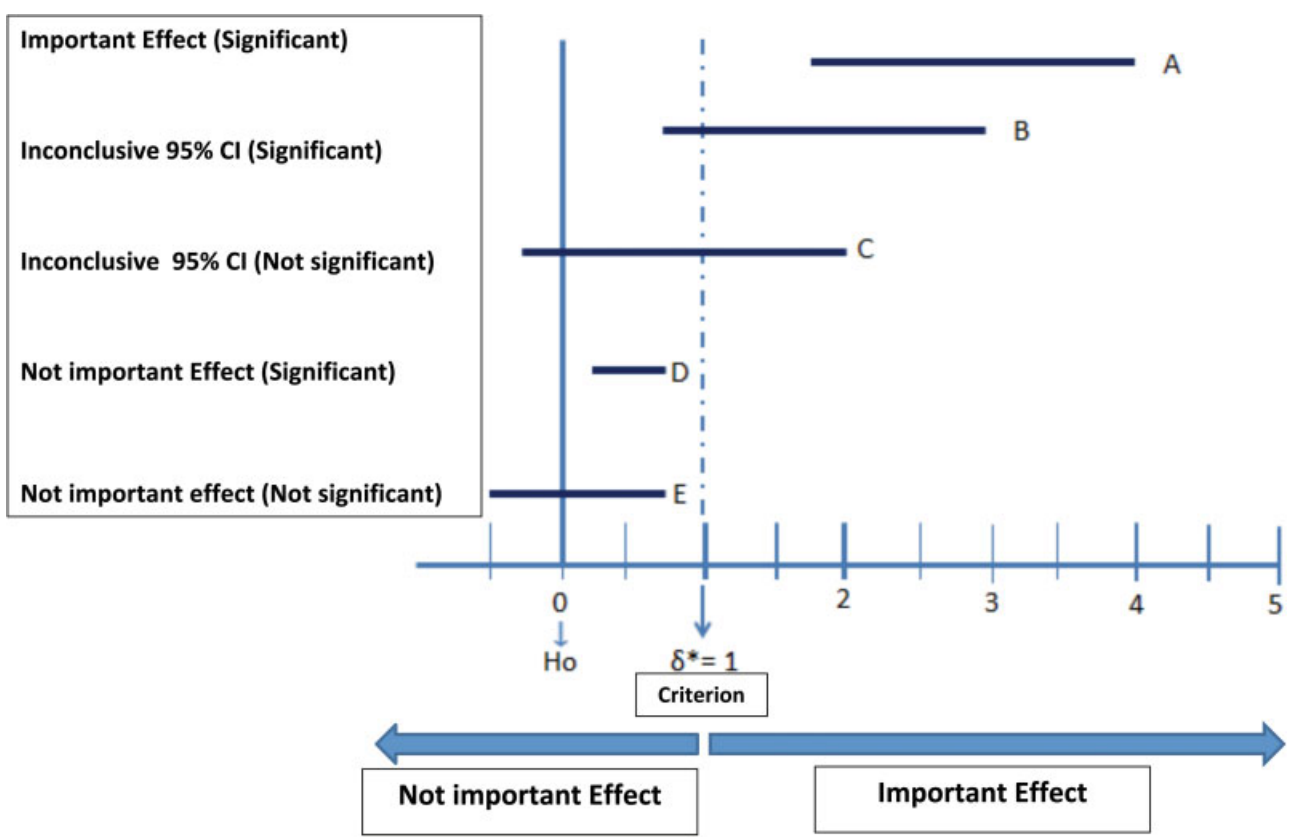


Fig. 7 Comparison of the results of different confidence intervals. Interpretation of the results based on the 95% CI and a criterion (δ*) or minimal important clinical effect. (based on Domenech[18]). Abbreviations: CI, confidence interval.

important effect (both CIs above the criterion or of the MICD), we will choose the narrower 95% CI.

Statistics Test for Data Analysis

Based on the hypotheses established in the Study Description section of the present paper, we will have different statistical tests for answering those research questions.

Before starting the proper data analysis, we need to explore the sample by the fit-goodness test. An important issue, especially in a sample size < 30, is to assess if the distribution of the data follows a normal distribution in the population. We have different statistical tests and graphics for that purpose. For testing the sample for normality, the most commonly used test is the Shapiro-Wilk test, with which we can establish the H_0 = no difference between the sample and a normal distribution. Observe in ►**Fig. 2** that all tests of normality (Shapiro-Wilk, skewness and kurtosis tests for normality) were not significant ($p > 0.05$); therefore, we accept the H_0 and can conclude that the variable "sspre" follows a normal distribution.

The statistical tests to be used will depend on how the researcher recorded the variables in the study. Following with the example and the database used in the present paper, the researcher can face different scenarios based on the exposure and outcomes variables:

1.–Exposure and Outcomes (dichotomous or binary variable). When the exposure and the outcomes are binary variables, we have to create a crosstab table in which the exposure or independent variable (IV) is located in columns, and the outcomes or dependent variable (DV) in rows. The proportions in each column are compared by the chi squared (χ^2) or by the Fisher exact statistics tests, and the clinical relevance of association is assessed by the proportion rate (PR) and odds ratio (OR).[25] For example, association between severe level of CTS symptoms (NivSS) and smoking habit: two proportion comparison, PR or OR > 1 means positive association, PR or OR $= 1$ constitutes the H_0 (no association), and PR or OR < 1 means negative association (►**Fig. 8**) (►**Chart 7**). The proper use of OR or of PR will depend on the clinical design of the study. In a cross-sectional study (►**Chart 8**), the results come in terms of prevalence difference (0.215686), OR of prevalence (3.75), and prevalence ratio (1.3235), with the 95% CI. For cohort studies or experimental studies (►**Chart 9**), the results come in terms of risk difference (0.215686), relative risk or risk ratio (RR) (1.3235), and OR (3.75), with the 95% CI. However, for case-control studies, the results come only in terms of OR (►**Chart 9**). Observe that the results obtained by Stata in ►**Charts 8**, **9** and **10** are similar to those shown in ►**Fig. 8** done by hand calculator.

2. Exposure (categorical variable) and Outcomes (dichotomous or binary variable). When the exposure variable presents several categories ($c > 2$) and the outcomes variable is binary, we have to check if the IV is an ordinal categorical variable, because in that case we have to assess the tendency by a trend test or by a Mantel Haenszel test and by the deviation from linearity test. If we look at ►**Fig. 8** (association between NivSSpre and smoking level: comparison of several proportions), the proportions in columns of NivSSpre (severe level of symptoms related to CTS) is increasing with the number of cigarettes/day (c/d) (smoking level). But the linear trend analysis was not significant (►**Chart 10**). Otherwise, if we do not have an ordinal aspect of the categories, the statistics test will be again a χ^2 test for multiple comparisons or by regression model[18,25] (►**Chart 10**), which showed that there was not an association between severe level of symptoms and smoking level.

3. Exposure (binary variable) and Outcomes (quantitative variable). When the IV is a binary variable (smoking habit) and the outcomes is quantitative (SSpre = severity of symptoms related to CTS), the statistics test compares two means by the Student-Fisher *t*-test and the magnitude of the effect by the 95% CI (►**Fig. 8**). Apart from the test for normality distribution of the sample when the sample size is not large ($n < 30$), we need to test for homogeneous variances or perform the Levine test, which establishes the H_0 that the 2 samples compared presented an equal variance. If the Levine test is significant, we have to use the Student *t*-test for unequal variance, in which one only changes the dfs.[26] Besides, we have to differentiate between the Student *t*-test for independent samples, as in this case (comparison of sspre in smoking and non-smoking samples) (►**Fig. 8**), or in the example in ►**Fig. 3** and ►**Chart 6** (comparison of sschange in a ligament reconstruction sample with only a CT release sample); and a paired Student *t*-test for dependent samples (for example, the comparison of sspre with sspost to know if there was a significant improvement in the symptoms between before the surgery and 3 months after the surgery).

When the sample is small or not normally distributed, the non-parametric tests used for comparing two quantitative variables are the Wilcoxon signed-rank test, in cases of dependent samples, and the U-Mann Whitney test (T Wilcoxon test), in cases of independent samples.

4. Exposure (categorical variable) and outcomes (quantitative variable). When the IV is a categorical variable with more than 2 categories (smoking level) and the DV is a quantitative variable, the analysis of variance (ANOVA) is used, defining the reference category for the analysis (that is, association between SSpre and smoking level [cigarettes/day]) (►**Fig. 8**).[26] The ANOVA is a generalization of the Student-Fisher *t*-test in which it is assumed that the data of "c" categories are random samples of "c" populations with equal variance and normal distribution if some of the samples present a size < 30. Consequently, it is mandatory for ANOVA to test the samples for normality (Shapiro-Wilk Test) and for homogeneity of the variances (Levine Test). A non-parametric version of ANOVA is the Kruskal-Wallis test. The ANOVA can be analyzed using a linear regression model (►**Chart 11**).[26]

Id	Smoke (SmokLev)		SSpre (NivSS)	
1	20	>15 c/d	3.3	Positive
2	25	>15 c/d	4.4	Positive
3	0	0 c/d	4.3	Positive
4	0	0 c/d	2.5	Negative
5	20	>15 c/d	4.6	Positive
6	0	0 c/d	3.2	Positive
7	0	0 c/d	3.9	Positive
8	0	0 c/d	3.5	Positive
9	0	0 c/d	2.3	Negative
10	10	1-15 c/d	3.4	Positive
11	0	0 c/d	2.9	Negative
12	0	0 c/d	2.7	Negative
13	5	1-15 c/d	2.7	Negative
14	30	>15 c/d	4.9	Positive
15	0	0 c/d	3.9	Positive
16	15	1-15 c/d	4.5	Positive
17	35	>15 c/d	4.9	Positive
18	20	>15 c/d	3.8	Positive
19	40	>15 c/d	4.5	Positive
20	10	1-15 c/d	4.2	Positive
21	15	1-15 c/d	4	Positive
22	25	>15 c/d	2.6	Negative
23	0	0 c/d	4	Positive
24	30	>15 c/d	3.1	Positive
25	5	1-15 c/d	3.7	Positive
26	25	>15 c/d	4.5	Positive
27	20	>15 c/d	3.3	Positive
28	0	0 c/d	5	Positive
29	0	0 c/d	3.7	Positive
30	.	.	5	Positive
31	.	.	2.6	Negative

Two proportions comparison

(χ^2 Test) (Z Test)

NivSS	Smoking habit: No fuma	Fuma	Total
Negativo	4	2	6
Positivo	8	15	23
Total	12	17	29

$p_0 = 8/12 = 0.67$ $p_1 = 15/17 = 0.88$; $d = 0.88 - 0.67 = 0.21$; $PR = 0.88/0.21 = 1.32$

Several proportions comparison (χ^2 Test)

NivSS	Smoking level: 0 c/d	1-15 c/d	>15 c/d	Total
Negativo	4	1	1	6
Positivo	8	5	10	23
Total	12	6	11	29

$p_0 = 8/12 = 0.67$ $p_1 = 5/6 = 0.83$ $p_2 = 10/11 = 0.91$; $PR_1 = 0.83/0.67 = 1.24$ $PR_2 = 0.91/0.67 = 1.36$

Two means comparison

(t –Student-Fisher Test)

	Smoking Habit: No	Yes
SSpre	3.7	3.1
	3.9	3.3
	2.5	4.5
	3.2	3.8
	.	.
	.	
n (total)	12	17

Means: $y_0 = 3.491667$ (n = 12) $y_1 = 3.905882$ (n=17); d = 0.4142157

"c" means comparison

(ANOVA)

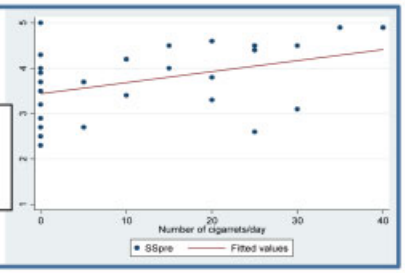
	Smoking level (c/d): 0	1-15	>15
SSpre	3.1	4	3.3
	3.3	4.2	4.9
	4.5	3.7	4.9
	3.8	4.5	4.5
	.	2.7	4.6
		3.4	.
		.	.
		.	
Total n (29)	12	6	11

$y_0 = 3.491667$ $y_1 = 3.75$ $y_2 = 3.990909$; $d_1 = y_1 - y_0 = 0.2583333$ $d_2 = y_2 - y_0 = 0.4992424$

Simple Linear Regression

$y_{(SSpre)} = A + Bx_{(c/d)}$

$y = 3.44058 + 0.0243519(c/d)$



Logistic Regression

	Smoking level (c/d = cigarettes /day): 0	5	10	15	20	25	30	35	40	Total
NivSS+	4	1	0	0	0	1	0	0	0	6
NivSS-	8	1	2	2	4	2	2	1	1	23
Total	12	12	2	2	2	4	3	2	1	29

Fig. 8 Statistic tests based on the exposure and outcome variables. Comparison of two proportions when the exposition, and the outcomes are binary variables. Comparison of several proportions when the exposition is categorical, and the outcome is a binary variable. Comparison of two means when the exposition is binary, and the outcome is a quantitative variable. Comparison of several means when the exposition is categorical, and the outcome is a quantitative variable. Simple linear regression when both (exposition & outcome) variables are quantitative. Logistic regression model when the outcome is binary, and the exposition is quantitative. (based on Domenech[26])

Chart 7 The association between severe level of carpal tunnel syndrome symptoms and smoking habit

A

```
Two-sample test of proportions                   Yes: Number of obs =       17
                                          No smoking: Number of obs =       12

------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Yes |   .8823529   .0781425                      .7291965    1.035509
  No smoking |   .6666667   .1360828                      .3999494     .933384
-------------+----------------------------------------------------------------
        diff |   .2156863   .1569228                     -.0918768    .5232493
             |  under Ho:   .1527304     1.41   0.158
------------------------------------------------------------------------------
        diff = prop(Yes) - prop(No smoking)                       z =   1.4122
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.9211         Pr(|Z| > |z|) = 0.1579          Pr(Z > z) = 0.0789
```

B

```
                 | Somking habit          |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
           Cases |        15           8  |         23
        Noncases |         2           4  |          6
-----------------+------------------------+-----------
           Total |        17          12  |         29
                 |                        |
            Risk |  .8823529    .6666667  |   .7931034
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+-----------------------
 Risk difference |         .2156863       |   -.0918768    .5232493
      Risk ratio |         1.323529       |    .8557255     2.04707
 Attr. frac. ex. |         .2444444       |   -.1685991     .511497
 Attr. frac. pop |         .1594203       |
      Odds ratio |             3.75       |    .6322621    21.53979 (Cornfield)
                 |------------------------+-----------------------

                          Pearson chi2(1) =   1.9943   Pr = 0.158
                  likelihood-ratio chi2(1) =   1.9778   Pr = 0.160
                           Fisher's exact =                 0.198
                    1-sided Fisher's exact =                 0.172
```

C

```
Logistic regression                               Number of obs     =         29
                                                  LR chi2(1)        =       1.98
                                                  Prob > chi2       =     0.1596
Log likelihood = -13.795749                       Pseudo R2         =     0.0669

------------------------------------------------------------------------------
    NivSSpre | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  HabitSmoke |       3.75   3.638982     1.36   0.173     .5597956    25.12078
       _cons |          2   1.224745     1.13   0.258     .6022511    6.641747
------------------------------------------------------------------------------
```

(A) Comparison of two proportions using the Z statistic test, which in acceptable sample sizes follows a normal distribution (Wald test). (B) The same comparison using the chi squared (χ^2) test. Observe that $Z^2 = 1.412^2$ is equal to the Pearson statistic $\chi^2 = 1.9943$. The proportions of severe symptoms in smoking (0.88253) and in non-smoking (0.666667) are similar to those observed in the χ^2 cross-tab describes as risk. The risk differences in cross-tab 0.215683 are similar to the one shown in the Z test (diff). (C) The odds ratio (OR) in the logistic regression analysis is similar to the OR observed in B, with a similar likelihood ratio ($\chi^2 = 1.98$).

Chart 8 Measuring of the association between NivSSpre and smoking habit for a cross-sectional clinical design

ASSOCIATION MEASURES: CROSS-SECTIONAL STUDY

	Unexposed	Exposed	TOTAL	Exposed proportion 95% CI (Wilson)
Cases	8	15	23	0.652174 (0.448903 to 0.811887)
NonCases	4	2	6	0.333333 (0.096771 to 0.700007)
TOTAL	12	17	29	0.586207 (0.407378 to 0.744868)
Prevalence	0.666667	0.882353	0.793103	
95% CI (Wilson) Lower	0.390622	0.656636	0.616099	
95% CI (Wilson) Upper	0.861880	0.967120	0.901539	

	Estimate	95% Confidence Interval		
Prev. Diff. (PD)	0.215686	-0.082736	0.504453	Newcombe[a]
		-0.091877	0.523249	Wald
Prev. Ratio (PR)	1.3235	0.8557	2.0471	
Odds Ratio (OR)	3.7500	0.6323	21.5398	Cornfield[a]
		0.5598	25.1208	Woolf
Association				
Pearson Chi2	1.9943**	p= 0.157890 (2-sided)		
Corrected	0.8965	p= 0.343732 (2-sided)		
Fisher Exact Test		p= 0.198063 (2-sided)		

[a]Recommended CI

Abbreviation: NivSSpre, severe level of symptoms before surgery. *recommended confidence interval (CI).
Observe that the results in a cross-sectional study come in terms of prevalence, prevalence difference, prevalence ratio (PR) and odds ratio (OR) of prevalence. For the PR and OR, the $H_0 = 1$, and the intervals include the H_0, consequently, the association was not significant.

5.–Exposure and Outcomes (quantitative variables). When both variables are quantitative, the linear regression analysis is the selected test. A linear equation ($y = A + BX$) is calculated, in which the slope, or β coefficient, gives us information about the contribution of the smoking level (cigarettes/day) to the severity of the symptoms related to CTS (sspre) (►**Fig. 8**).[26]

6. Exposition (quantitative variable) and Outcomes (categorical variable). In the final part of ►**Fig. 8**, the variable "outcomes" is categorical. In the present case, it is binary (NivSS = severe level of symptoms), and the exposition is quantitative (smoking: number of cigarettes/day); the statistical test will be a logistic regression.[26]

Generalized Linear Models (GLMs)

We have studied the association between symptoms related to CTS and smoking. This association has been analyzed in different ways based on the type of exposure and outcome variables. In ►**Chart 10**, we observed a non-significant association between the severe level of CTS symptoms (NivSSpre) and smoking level (SmokLevel; c/d), with an OR = 2.5 (1–15 c/d versus no smoking); and an OR = 5 (>

Chart 9 Measuring of the association between NivSSpre and smoking habit for a cohort and experimental clinical design

ASSOCIATION MEASURES: EXPERIMENTAL STUDY

		Treat.0	Treat.1	TOTAL
	NonEvents	8	15	23
	Events	4	2	6
	TOTAL	12	17	29
	Risk	0.666667	0.882353	
95% CI	Lower	0.390622	0.656636	
(Wilson)	Upper	0.861880	0.967120	

	Estimate	95% Confidence Interval		
Risk Diff. (RD)	0.215686	-0.082736	0.504453	Newcombe[a]
		-0.091877	0.523249	Wald
Risk Ratio (RR)	1.3235	0.8557	2.0471	
Odds Ratio (OR)	3.7500	0.6323	21.5398	Cornfield[a]
		0.5598	25.1208	Woolf
Association				
Pearson Chi2	1.9943**	p= 0.157890 (2-sided)		
Corrected	0.8965	p= 0.343732 (2-sided)		
Fisher Exact Test		p= 0.198063 (2-sided)		

[a]Recommended CI

ASSOCIATION MEASURES: CASE-CONTROL STUDY

	Unexposed	Exposed	TOTAL	Exposed proportion 95% CI (Wilson)
Cases	8	15	23	0.652174 (0.448903 to 0.811887)
Controls	4	2	6	0.333333 (0.096771 to 0.700007)
TOTAL	12	17	29	

	Estimate	95% Confidence Interval		
Odds Ratio (OR)	3.7500	0.4082	47.7554	Exact[a]
		0.5598	25.1208	Woolf
Association				
Pearson Chi2	1.9943**	p= 0.157890 (2-sided)		
Corrected	0.8965	p= 0.343732 (2-sided)		
Fisher Exact Test		p= 0.198063 (2-sided)		

[a]Recommended CI

Abbreviation: NivSSpre, severe level of symptoms before surgery. *recommended confidence interval (CI).
For experimental and cohort clinical designs, the results are shown in the same way, based on risk ratio (RR) and odds ratio (OR). However, in case-control studies, the results of the same analysis come only in terms of OR.

Chart 10 Association between smoking level (cigarettes/day) and severe level of carpal tunnel syndrome symptoms

A. TREND TEST (Prevalence data)

Valid cases: 29 (93.5%)
Total cases: 31

Expo Metric	Cases	Total	Preva-lences	Wilson [95% Conf. Int.]		Ref. category: First PR	95% Conf. Inter.	
0	8	12	0.66667	0.39062	0.86188	1.0000	.	.
1	5	6	0.83333	0.43650	0.96995	1.2500	0.7308	2.1381
2	10	11	0.90909	0.62264	0.98377	1.3636	0.8769	2.1207
TOTAL	23	29	0.79310	0.61610	0.90154			

MH Test for linear Trend: Chi2(1) = 1.9988 (p= 0.1574)
Deviation from linearity: Chi2(1) = 0.1313 (p= 0.7171)

B. Chi2 Tests

Severe level of symptoms (pre)	Smoking level 0 c/d	1-15 c/d	>15 c/d	Total
Negative	4	1	1	6
	33.33	16.67	9.09	20.69
Positive	8	5	10	23
	66.67	83.33	90.91	79.31
Total	12	6	11	29
	100.00	100.00	100.00	100.00

$$OR_1 = \frac{5/1}{8/4} = 2.5$$

$$OR_2 = \frac{10/1}{8/4} = 5.0$$

Pearson chi2(2) = 2.1301 Pr = 0.345
likelihood-ratio chi2(2) = 2.1842 Pr = 0.336
Fisher's exact = 0.442

C. Logistic Regression

Logistic regression

Number of obs = 29
LR chi2(2) = 2.18
Prob > chi2 = 0.3355
Pseudo R2 = 0.0739

Log likelihood = -13.692534

NivSSpre	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
NivTab						
1-15 c/d	2.5	3.137475	0.73	0.465	.2136439	29.25428
>15 c/d	5	6.072479	1.33	0.185	.4625826	54.0444
_cons	2	1.224745	1.13	0.258	.6022511	6.641747

(A) Observe that the prevalence of severe level of symptoms before surgery (NivSSpre) is increasing with the level of smoking. The deviation from linearity test presented a non-significant result ($p = 0.7171$), which implies that the proportions are located in a straight line. The Mantel Haenszel (MH) test showed that the prevalence increase of symptoms with the smoking level was not significant ($p = 0.1574$), which means that there is no linear trend because the line is horizontal and not ascending. (B) The same analysis of association when there is not an ordinal categorization. (C) The same analysis with similar odds ratios using a logistic regression model.

Chart 11 Association between the symptoms related to carpal tunnel syndrome (sspre) and smoking level (SmokLevel)

A. ANOVA

```
                    Number of obs =         29    R-squared     =  0.0590
                    Root MSE      =    .900011    Adj R-squared = -0.0134
```

Source	Partial SS	df	MS	F	Prob>F
Model	1.320849	2	.6604245	0.82	0.4535
NivTab	1.320849	2	.6604245	0.82	0.4535
Residual	21.06053	26	.8100204		
Total	22.381379	28	.79933498		

	Contrast	Std. Err.	t	P>|t|	[95% Conf.	Interval]
NivTab						
(1-15 c/d vs 0 c/d)	.1416667	.4500057	0.31	0.755	-.7833332	1.066667
(>15 c/d vs 0 c/d)	.4734848	.3756859	1.26	0.219	-.2987486	1.245718

B. Linear Regression Model

```
. regress sspre i.NivTab
```

Source	SS	df	MS
Model	1.32084901	2	.660424504
Residual	21.0605303	26	.810020396
Total	22.3813793	28	.799334975

```
Number of obs   =        29
F(2, 26)        =      0.82
Prob > F        =    0.4535
R-squared       =    0.0590
Adj R-squared   =   -0.0134
Root MSE        =    .90001
```

sspre	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
NivTab						
1-15 c/d	.1416667	.4500057	0.31	0.755	-.7833332	1.066667
>15 c/d	.4734848	.3756859	1.26	0.219	-.2987486	1.245718
_cons	3.608333	.2598109	13.89	0.000	3.074284	4.142382

(A) The analysis of variance (ANOVA) for assessing the association between a quantitative outcome variable (sspre) and a categorical exposition variable (NivTab). (c/d = cigarettes/day.) The results demonstrated that there was no association. The contrast analysis assessed the association sspre in each group of NivTab, using as reference the "no smoking" group. (B) The ANOVA analysis using regression model. Observe that the F statistics is the same (0.82), that the *p*-value is the same ($p = 0.4535$), and that the t-values (0.31; 1.26) and the difference mean values (0.1416667; 0.4734848) in the contrast analysis are coincidental with those in the regression model for the β coefficients.

15 c/d versus no smoking) in the cross tab analysis. The questions are: are those results the real effect of smoking on the severity level of the symptoms? Is there another variable which could affect this association? The same applies for the example exposed in ►**Chart 6**, the effect of ligament reconstruction on the change in CTS symptoms in open CT release. Using regression models, the observed effect can be adjusted to modifiers (confounder and/or interaction) variables. The decision of adjusting for a confounder variable in a regression model should not be taken based on significant statistical test, but on the basis of important changes in the effect.

The analysis of the association between severe level of symptoms and smoking level (cigarettes/day) (►**Chart 10**)

Chart 12 Advantage of the generalized linear models (GLMs)

A

NivSSpre	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
NivTab						
1-15 c/d	2.5	3.137475	0.73	0.465	.2136439	29.25428
>15 c/d	5	6.072479	1.33	0.185	.4625826	54.0444
_cons	2	1.224745	1.13	0.258	.6022511	6.641747

```
Logistic regression                               Number of obs     =         28
                                                  LR chi2(3)        =       5.33
                                                  Prob > chi2       =     0.1493
Log likelihood = -11.883961                       Pseudo R2         =     0.1831
```

NivSSpre	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
NivTab						
1-15 c/d	4.966647	7.276107	1.09	0.274	.2812278	87.71387
>15 c/d	13.60615	20.43082	1.74	0.082	.717103	258.1598
Gender						
Male	.1447207	.1828532	-1.53	0.126	.0121631	1.72193
_cons	2.545505	1.802837	1.32	0.187	.6352082	10.20074

% of change (1-15 c/d) = abs((2.5 -4.966647) /4.966647)*100 = 49.66423

% of change (> 15 c/d) = abs((5 - 13.60615) /13.60615)*100 = 63.251912

B

sschange	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Reconst						
No	.3545833	.4260492	0.83	0.412	-.5167852	1.225952
_cons	1.126667	.3060828	3.68	0.001	.500657	1.752676

sschange	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Reconst						
No	.3992063	.4403245	0.91	0.373	-.507659	1.306072
Gender						
Male	-.5296296	.4714144	-1.12	0.272	-1.500526	.4412665
_cons	1.163492	.319445	3.64	0.001	.5055828	1.821401

% of change = abs((0.3545833 - 0.3992063) / 0.3992063)*100 = 11.17793

(A) Analysis of the association of the severe level of symptoms (LevSSpre) with smoking level (cigarettes per day [c/d]) shown in ►**Chart 11**. The logistic regression allows the inclusion of modifier variables in the model and to know the association of ssppre and SmoKLev adjusted to gender. (B) The effect of ligament reconstruction on the change of carpal tunnel syndrome (CTS) symptoms using the regression linear model (shown in ►**Chart 6**). The same effect analysis adjusted to the variable "Gender."
Observe that the % of change in the effect (49.66423% in 1–15 c/d) (63.251912% in > 15 c/d) (11.17793% in no reconstruction) is higher than 10%. Consequently, the variable "gender" is a modifier variable of the association between severe level of CTS symptoms and smoking level; as well as of the effect of the ligament reconstruction on the change in symptoms related to CTS.

demonstrated that there was no significant association based on the Mantel-Haenszel trend analysis and on the χ^2 test. The $OR_1 = 2.5$ (95% CI: 0.2136439 to 29.25428) and $OR_2 = 5$ (95% CI: 0.4625826 to 54.0444) were not significant because the 95% CI included the $H_0 = 1$. A similar analysis using, in this case, a logistic regression model, allows us to include, in the model, more variables that can affect the effect of smoking level on the level of severity of the symptoms. If we observe ►**Chart 12**, we can find that the same model adjusted by gender increased the effect (OR) > 10%, which is clinically important, despite that the adjusted ORs ($OR_1 = 4.966647$; $OR_2 = 13.60615$) are still non-significant. Consequently, the association between severe level of CT symptoms and smoking level (cigarettes/day) should be adjusted for gender; and gender constitutes a variable that modified the association between smoking level and severe level of CTS symptoms.[18,26,27]

In a similar manner, we can adjust for modifier variables in a different regression model as in the analysis of the effect of ligament reconstruction on the change in symptoms after open CT release (►**Chart 6**). The analysis can be done by a Student *t*-test but using a multiple linear regression model the effect (difference on the mean change of symptoms) increased from 0.3545833 to 0.3992063 in the model adjusted by gender, which constituted a change in the effect by 11.17793%, also considered as clinically important.

The possibility of adjusting the models for different modifier variables is the most important advantage of using regression models to do the same analysis. In 1989, based on the work developed by McCullagh et al,[28] the GLMs were introduced: a set of models constituted by a linear combination of predictor variables ($X_1, X_2, X_3, \ldots X_i$) that can be a mix of quantitative (continuous and discrete) and categorical variables; and a dependent variable (Y) that can be quantitative (linear regression model), binary (logistic regression model), ordinal (ordinal logistic regression model), nominal (multinomial logistic regression model), count (Poisson, negative binomial, and zero-inflated Poisson regression models), etc. Consequently, the GLMs are a broad class of models that include linear regression, ANOVA, Poisson regression, log-linear models, etc.

Conclusions

The aim of the present update article is to serve as a gentle introduction to data analysis for clinical research in hand surgery. Early steps in statistics are important to improve the quality of our scientific papers. Clinical practice in hand surgery can be improved by good clinical research, and statistics is a fundamental support tool.

Acknowledgments

The authors wish to thank Dr. D. Jose M. Domenech, professor of statistics at the Universitat Autònoma de Barcelona. Without his teachings, the present paper could not have been written. The authors also wish to thank Dr. Fernando Corrella, chief editor of the Ibero-American Journal of Hand Surgery. Without his enthusiasm, this paper would not have been written.

References

1 Kleinert HE, Kasdan ML, Romero JL. Small blood-vessel anastomosis for salvage of severely injured upper extremity. J Bone Joint Surg Am 1963;45-A:788–796
2 Rosales RS. Clinical research in hand surgery. J Hand Surg Eur Vol 2015;40(05):546–548 Doi: 10.1177/1753193415583624
3 Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg Am 2004;29(04):727–734, discussion 735–737
4 Tadjerbashi K, Rosales RS, Atroshi I. Intervention randomized controlled trials involving wrist and shoulder arthroscopy: a systematic review. BMC Musculoskelet Disord 2014;15:252
5 Rosales RS, Reboso-Morales L, Martin-Hidalgo Y, Diez de la Lastra-Bosch I. Level of evidence in hand surgery. BMC Res Notes 2012; 5:665
6 Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg Am 1993;75(11):1585–1592
7 Atroshi I, Lyrén PE, Gummesson C. The 6-item CTS symptoms scale: a brief outcomes measure for carpal tunnel syndrome. Qual Life Res 2009;18(03):347–358
8 Atroshi I, Lyrén PE, Ornstein E, Gummesson C. The six-item CTS symptoms scale and palmar pain scale in carpal tunnel syndrome. J Hand Surg Am 2011;36(05):788–794
9 Rosales RS, Atroshi I. Spanish versions of the 6-item carpal tunnel syndrome symptoms scale (CTS-6) and palmar pain scale. J Hand Surg Eur Vol 2013;38(05):550–551
10 Rosales RS, Delgado EB, Díez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. J Hand Surg Am 2002;27(02): 334–343
11 Rosales RS, Martin-Hidalgo Y, Reboso-Morales L, Atroshi I. Reliability and construct validity of the Spanish version of the 6-item CTS symptoms scale for outcomes assessment in carpal tunnel syndrome. BMC Musculoskelet Disord 2016;17:115 Doi: 10.1186/s12891-016-0963-5
12 Atroshi I, Gummesson C, Johnsson R, McCabe SJ, Ornstein E. Severe carpal tunnel syndrome potentially needing surgical treatment in a general population. J Hand Surg Am 2003;28 (04):639–644
13 Rosales RS, Diez de la Lastra I, McCabe S, Ortega Martinez JI, Hidalgo YM. The relative responsiveness and construct validity of the Spanish version of the DASH instrument for outcomes assessment in open carpal tunnel release. J Hand Surg Eur Vol 2009;34 (01):72–75
14 Doménech JM, Navarro JB. Regresión Lineal Múltiple con predictores categóricos y cuantitativos. 11ª ed. Barcelona: Signo; 2018
15 Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10(04): 585–598
16 Fisher RA. On the mathematical foundations of theorical statistics. Philos Trans R Soc Lond A. 1992;222A:309–368
17 Fisher RA. Statistical Methods for Research Workers. 5th ed. Edinburgh: Tweeddale Court London. 33 Paternoster Row; 1934
18 Doménech JM. Fundamentos de de Diseño y Estadistica. UD7. Comprobación de hipótesis: Pruebas de significación, pruebas de hipótesis y tamaño de los grupos. 18ª ed. Barcelona: Signo; 2017
19 Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purpose statistical inference, Part I. Biometrika 1928; 20A:175–240
20 Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purpose statistical inference, Part II. Biometrika 1928; 20A:263–294
21 Neyman J, Pearson ES. On the problem of the most efficient test of statistical hypotheses. Philos Trans R Soc Lond A. 1933;231:289–337

22 Schwartz D. Métodos estadísticos para médicos y biólogos. Barcelona: Herde; 1985
23 Neyman J. On two different aspects of the representative method; the method of stratified sampling and the method of purposive selection. J R Stat Soc 1934;97(04): 558–625
24 Ozyürekoğlu T, McCabe SJ, Goldsmith LJ, LaJoie AS. The minimal clinically important difference of the Carpal Tunnel Syndrome Symptom Severity Scale. J Hand Surg Am 2006;31(05):733–738, discussion 739–740
25 Doménech JM. Fundamentos de de Diseño y Estadística. UD9. Comparación de dos proporciones. Medidas de asociación y de efecto. 18ª ed. Barcelona: Signo; 2017
26 Doménech JM. Análisis estadístico de un estudio sobre Tabaco y Carboxihemoglobina. 4 ed. Barcelona: Signo; 2017
27 Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Applied Regression Analysis and other Multivariate Methods. 4th ed. Pacific Grove (CA): Duxburry Press; 2008
28 McCullagh P, Nelder JA. Generalized linear model. 2nd ed. London: Chapman and Hall; 1989