‘The Seventh Annual International Campbell Collaboration Colloquium’

Enhancing utility of systematic reviews from methodological quality of primary studies. A case study in Training programs.

London, May 14-16, 2007

Salvador Chacón Moscoso (University of Seville)
Susana Sanduvete Chaves (University of Seville)
Milagrosa Sánchez Martín (University of Seville)
Julio Sánchez-Meca (University of Murcia)

Research supported by the *Spanish Ministry of Sciencies and Education* (MEC-SEJ2004-5360/EDUC) and *Andalusian Council of Science, Innovation and Technology* (ref. P06-HUM-01458; org. 1815052701).

INDEX OF CONTENTS

1. INTRODUCTION (I)

- Randomized designs (R.D.) allows unbiased estimates of treatment effects and provides the most precise and unbiased statistical estimate .

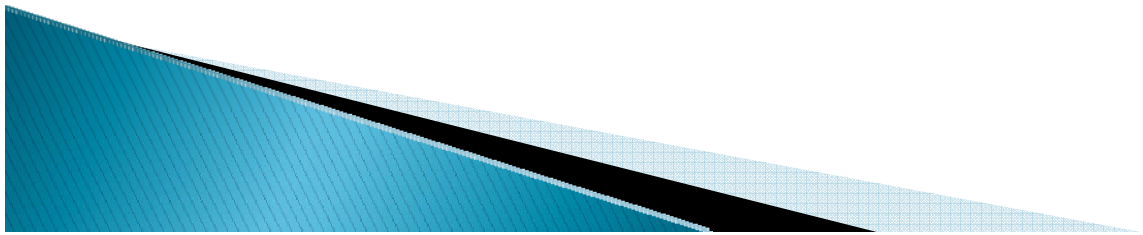
- But, most studies use non-experimental or quasi-experimental methods and would often result in the elimination of such studies from systematic reviews that use strict inclusion criteria (Shadish, Chacón-Moscoso & Sánchez-Meca, 2005).

1. INTRODUCTION (II)

- In most areas, a serious problem of representativeness would appear if considering only Radomized Designs.
- 'quality' standards and the transparency of the systematic review production processes are not developed yet.
- it is necessary to incorporate different types of evidence to address the real needs of intervention program users.
- ***How ?...***
- Based on previous work of our research group about measuring quality of primary studies (Chacón-Moscoso et al., 2004, 2005, 2006). We consider that methodological analysis of primary studies can give an important advance to these problems.

1. INTRODUCTION (III)

- Considering ‘utility’; usefullness as?:
  - It increases the probability of using the results.
  - It facilitates the implementation of the design.
  - it increases the probability of obtaining the ‘valid’ evidences as fostering implementation of previous designed program.



2. OBJECTIVES

- Assess methodological quality in the case of training programs published till September 2006.
- Propose a short scale composed of 19 items in order to provide a comparable index of quality to different studies in the same area (independently of designs/methodologies implemented) increases its probability of use
- Simplify the process identifying designs referents that meet high scientific standards.
- Facilitate communication between policy-academic-real intervention world.
- Easy to share to disseminate and to complete.

3. METHOD

3.1. Sample
3.2. Instruments
3.3. Procedure

3.1. Sample

- We found 2379 articles in database; 445 were accepted because we excluded those whose theme were not training programs in organizations, for example:
  - When parents were trained to improve the education to their children.
  - The training programs for people without a job.
  - Caregivers of elderly people (familiar membership).
  - Courses from the university to their students.

- Finally, we only studied 121 articles because the the full texts of the other papers were not available.

3.2. Instruments

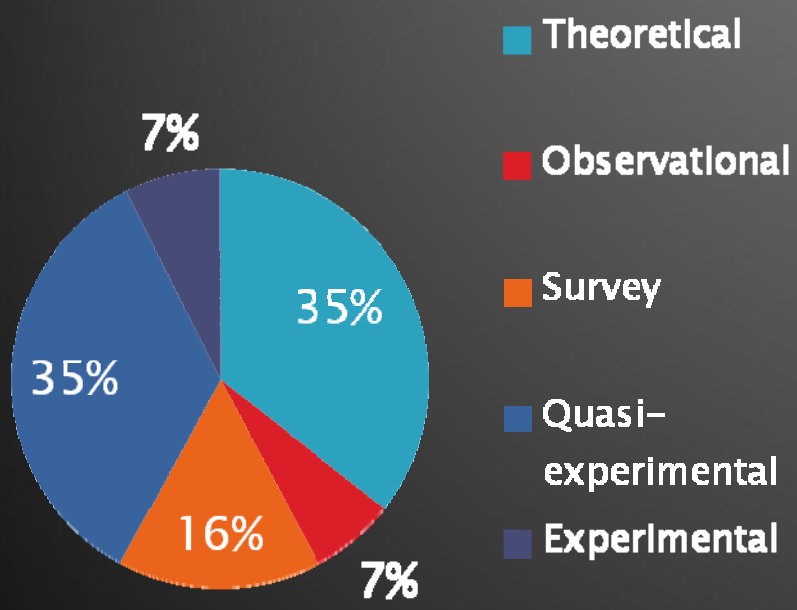
- Available database of interest in the substantive area in the University of Seville to search articles referring to evaluation in training programs: EBSCO Online, Medline, Serfile, CABHealth, CINAHL, PsycINFO, Econlit, ERIC, MathSci, Current Contents and Humanities Index.
- SPSS 14.0 to codify and analyze data.
- Microsoft Excel to calculate the Quality Index.
- A scale to measure quality of programs obtained in a previous study of content validity (Chacón-Moscoso et al., 2006).
  - A more detailed description is going to presented this evening in Parallel Session II, symposium F “Advances in Meta-Analysis”

3.3. Procedure

- We found references of articles according to training programs obtained from available data-bases mainly in University of Seville.
- Data were obtained till September 2006.
- We combined the following keywords: training programs, evaluation and work; they were searched in every document: title, abstract, keywords, complete article.
- Three different coders coded the available studies; an adequate intra-class correlation coefficient of reliability was obtained (0.85).
- We codified the found articles (121) using a scale to measure quality of programs previously elaborated.

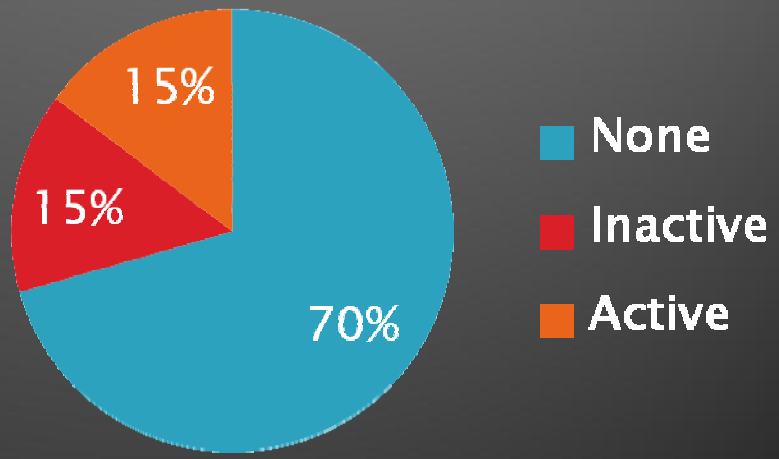
4. RESULTS I

Type of study



70% of the studies are of theoretical and quasi-experimental type.

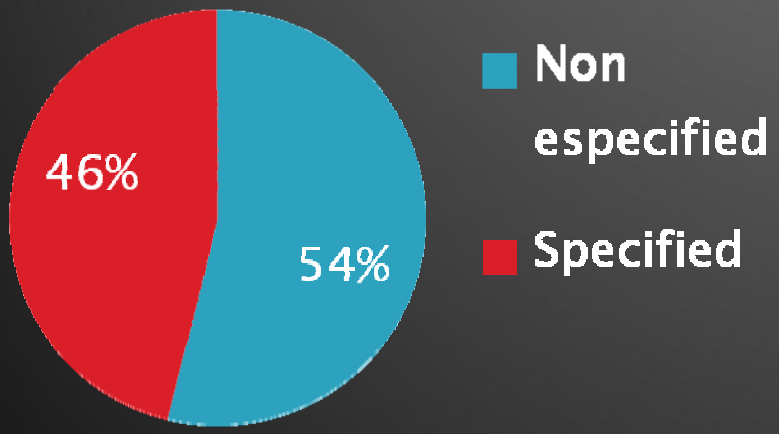
Control group



In 70% of the occasions, there is not control group.

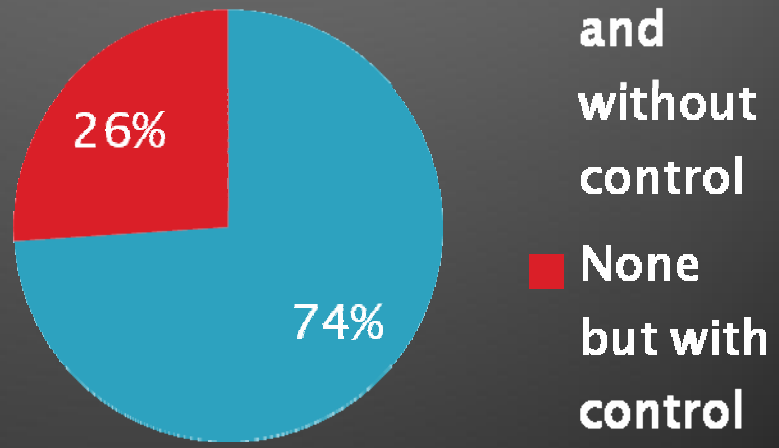
4. RESULTS II

Sample selection criteria



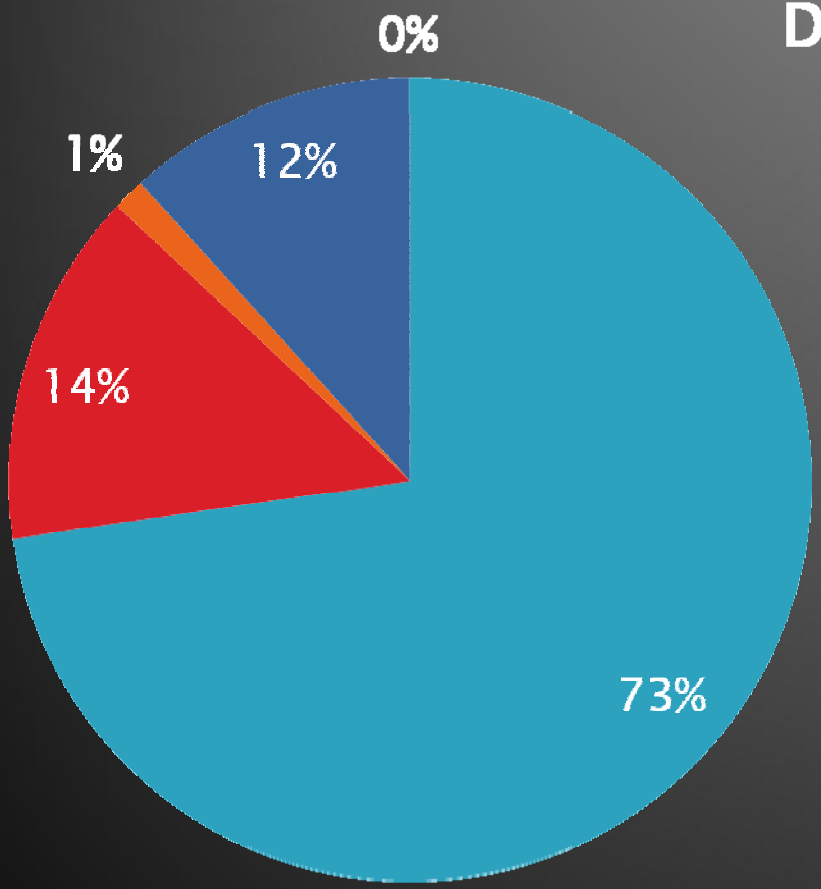
In 54% of the occasions, inclusion and exclusion criteria for units are not specified.

Randomization



In 74% of cases, there is not random assignment and there is not control; in 26%, there is at least control.

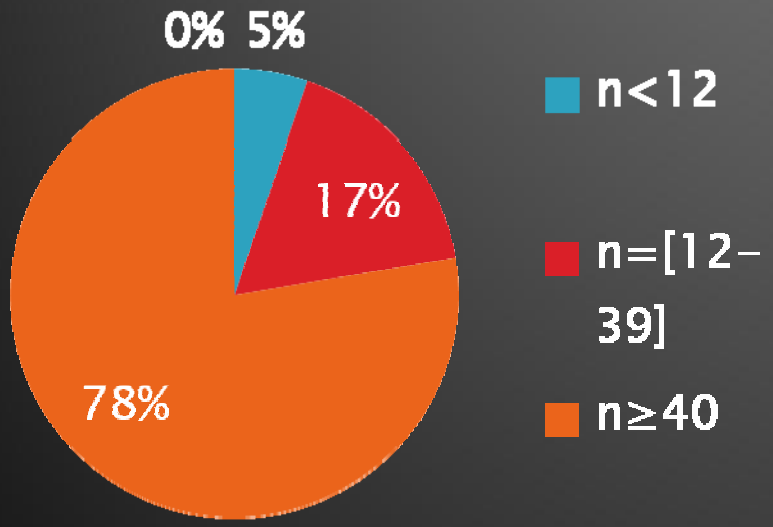
4. RESULTS III



The type of more frequent design, with a 73%, is 1-2 measurements when there is not intervention/ Pre-experimental/ Experimental with a moment of measurement.

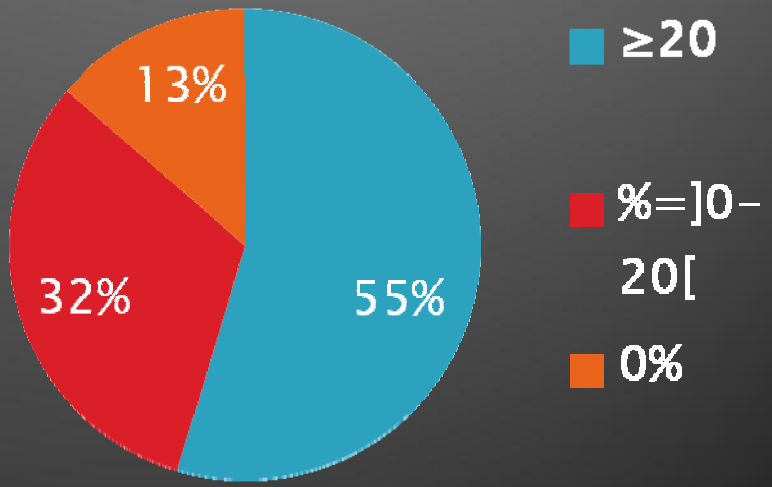
4. RESULTS IV

Sample



In 78% of the cases, sample size is higher than 40.

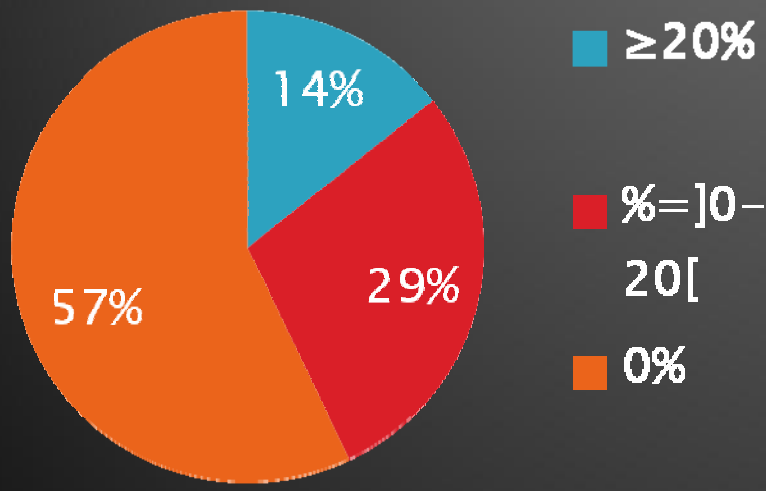
Global attrition



In 55% of occasions, attrition is over 20%; in 32%, between 0 and 20%.

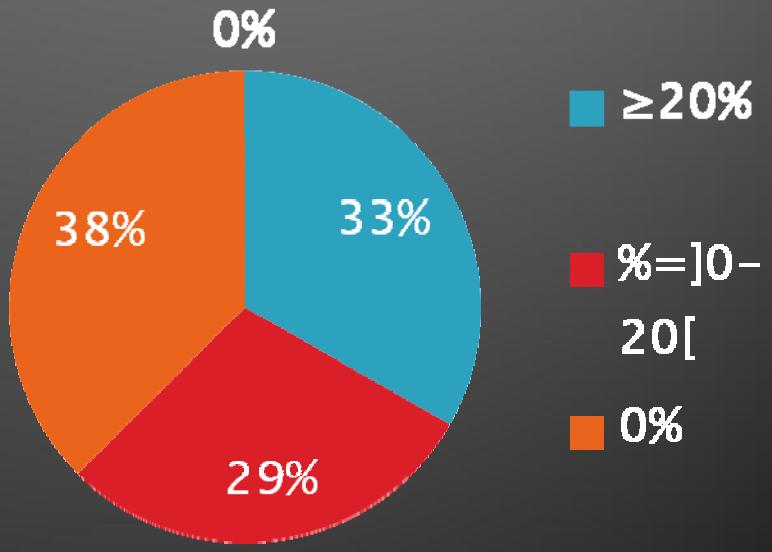
4. RESULTS V

Differential attrition



In the 57% of the cases, differential attrition is between 0 and 20%.

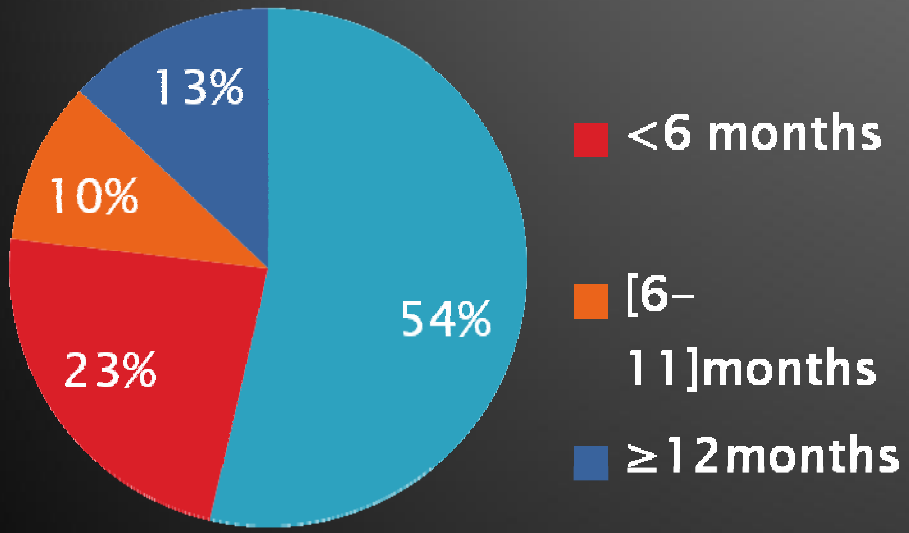
Exclusions after sample assignment



In 33% of cases, exclusions after randomization are more than 20%.

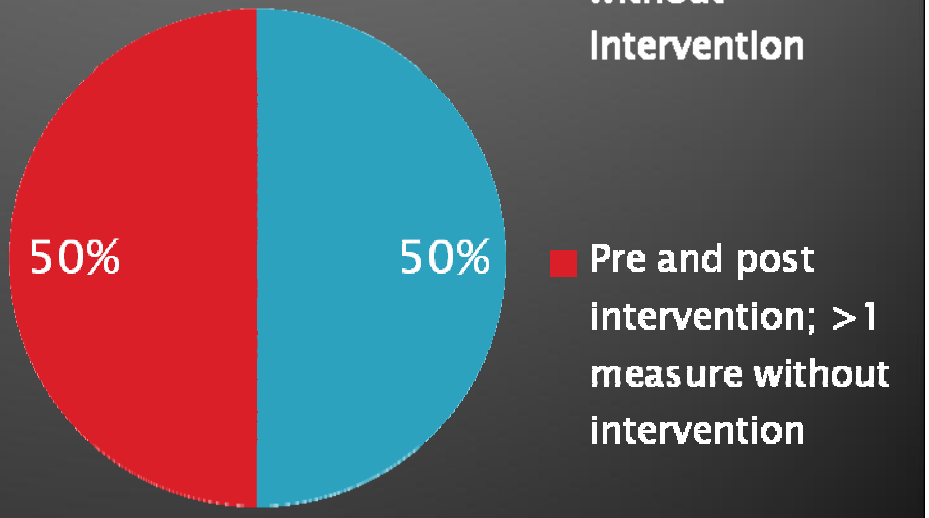
4. RESULTS VI

Follow-up



In54% of occasions, there is not follow-up period.

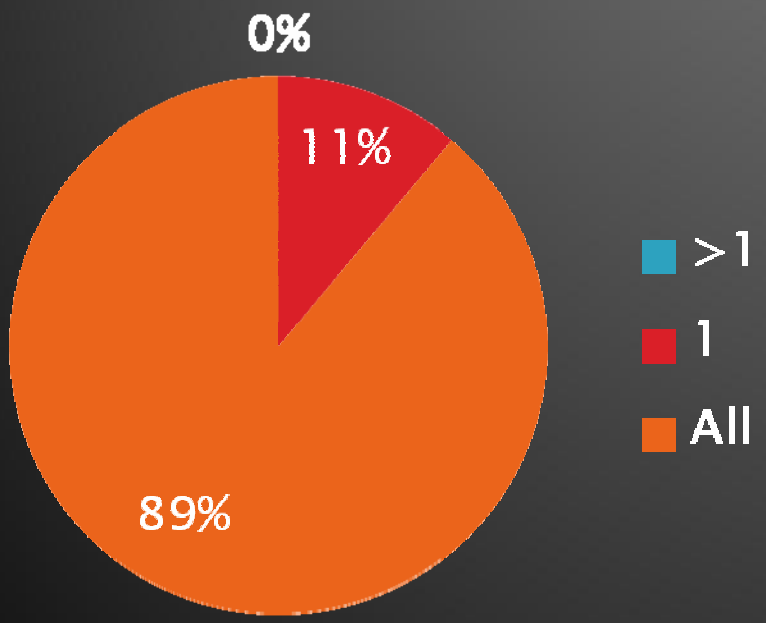
Moments of measurement



The different types from measures occur in the same percentage of cases.

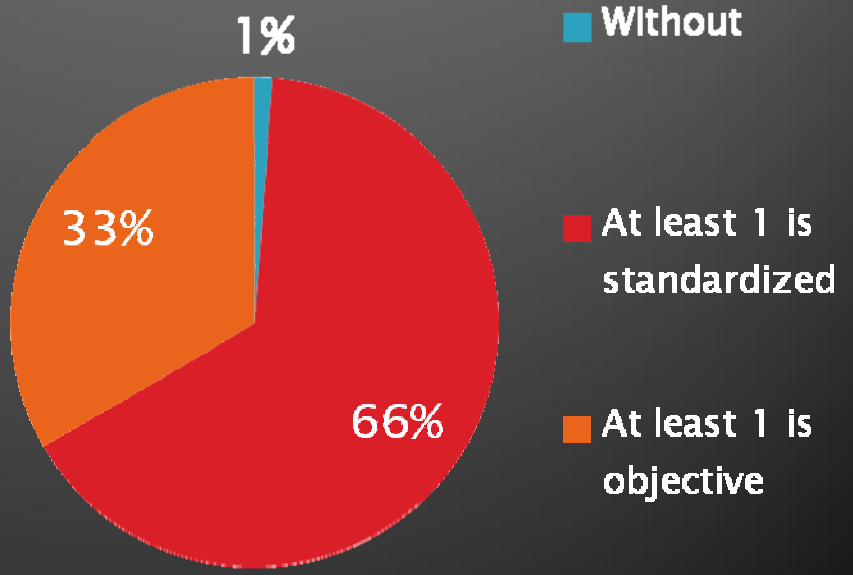
4. RESULTS VII

Measurements in every moments



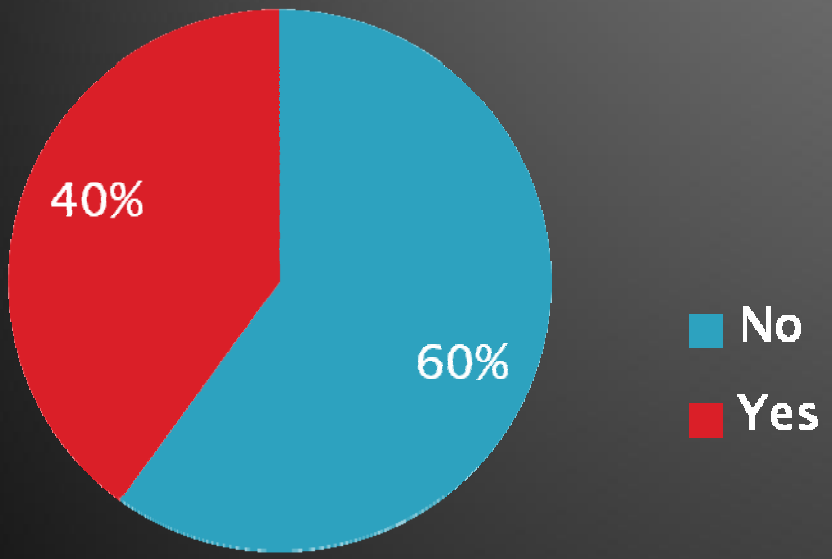
In 89% of cases, all the measurements appear in every measure moments.

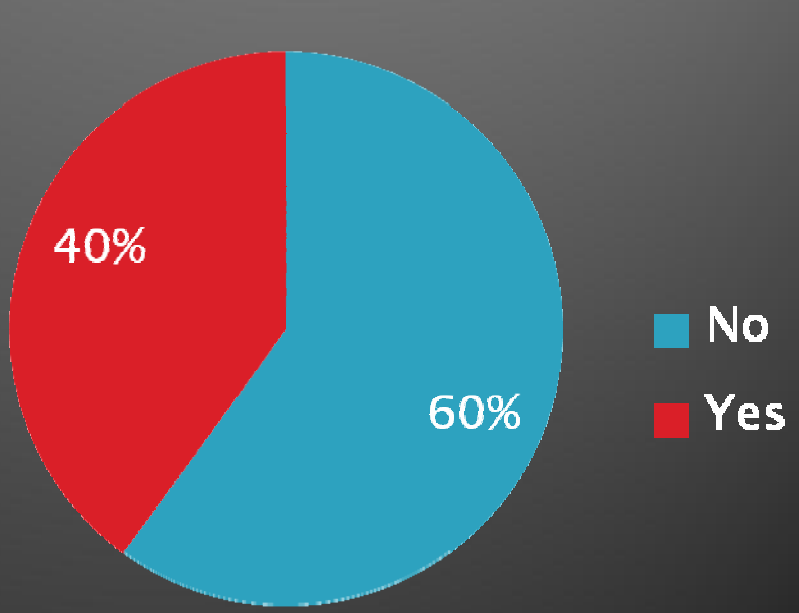
Normalized dependent variables



In 66% of cases, at least one dependent variable is standardized .

4. RESULTS VIII



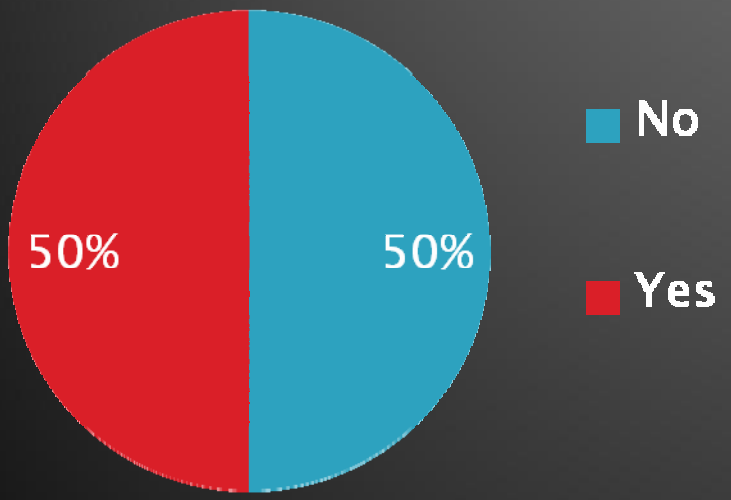


About the unique five cases in which this data is specified, in three there (60%) is mask in the evaluator.

Of the five specified data, in three of them there is mask in the participants.

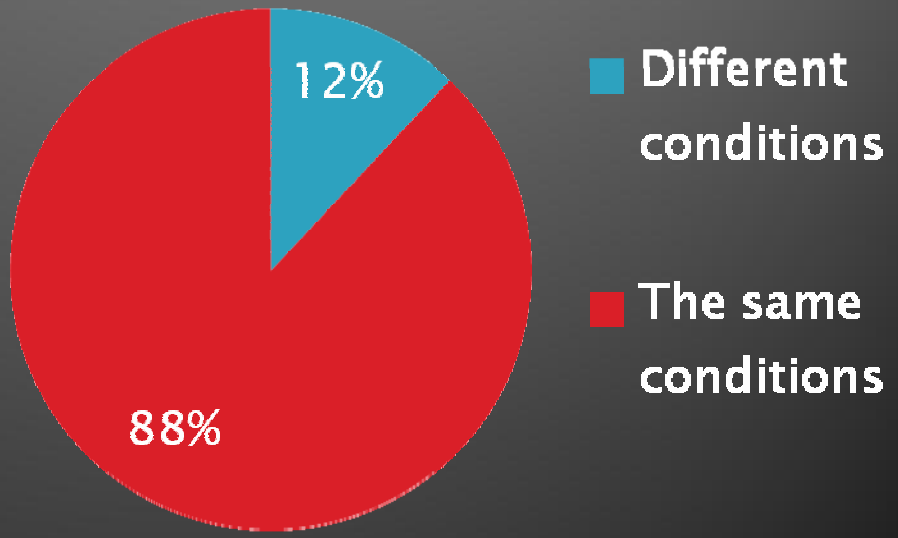
4. RESULTS IX

Mask in professional of intervention/ internal evaluator



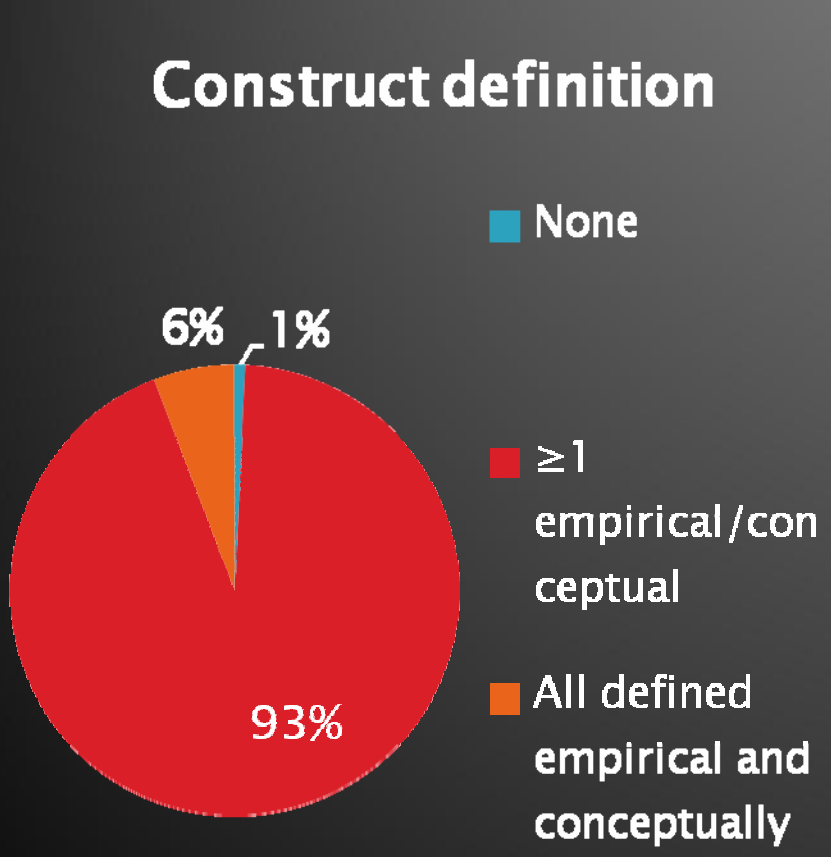
In two of the four detected data is mask in professional of intervention.

Homogeneity in process: intensity, duration and professionals

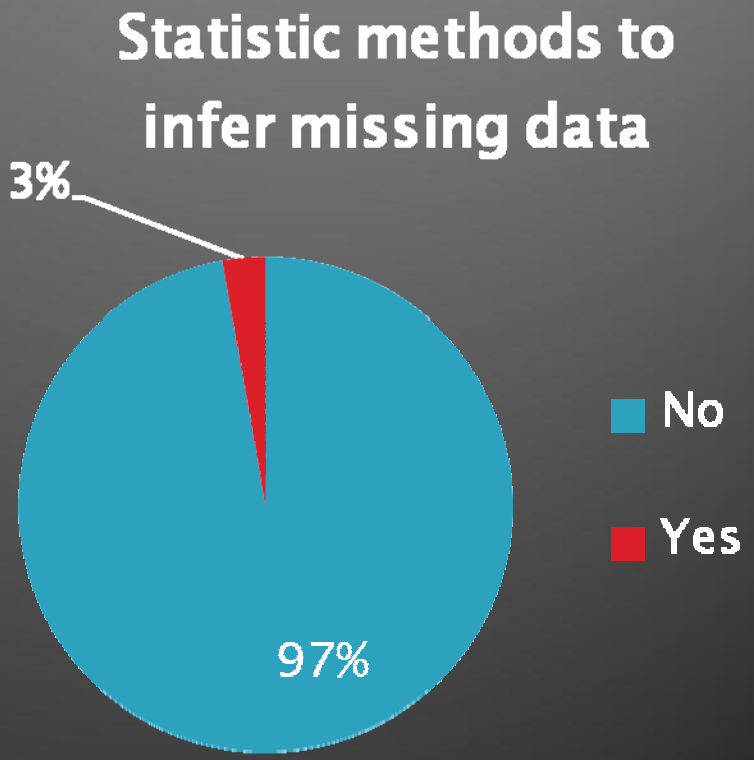


In 88% of cases, conditions are the same for every participants.

4. RESULTS X



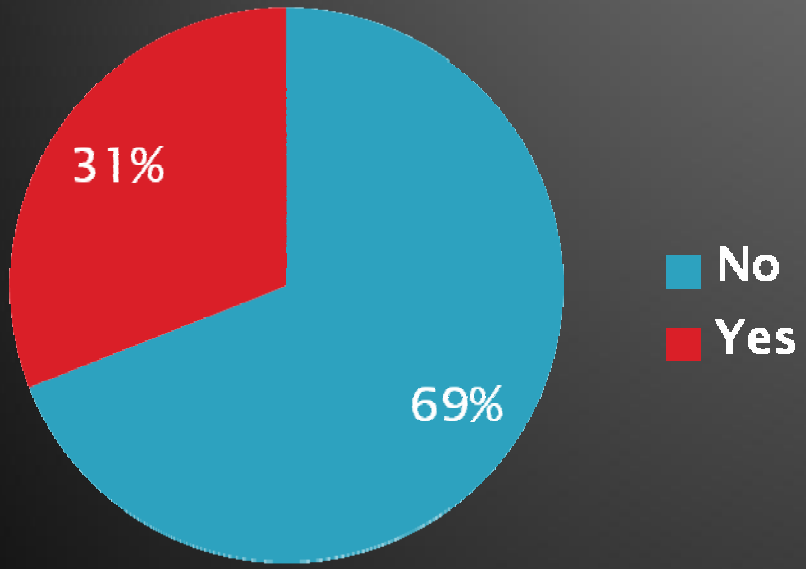
In 93% of the cases, constructs are defined at least empirical or conceptually.



In 97% of the cases, statistic methods to input missing data are not used.

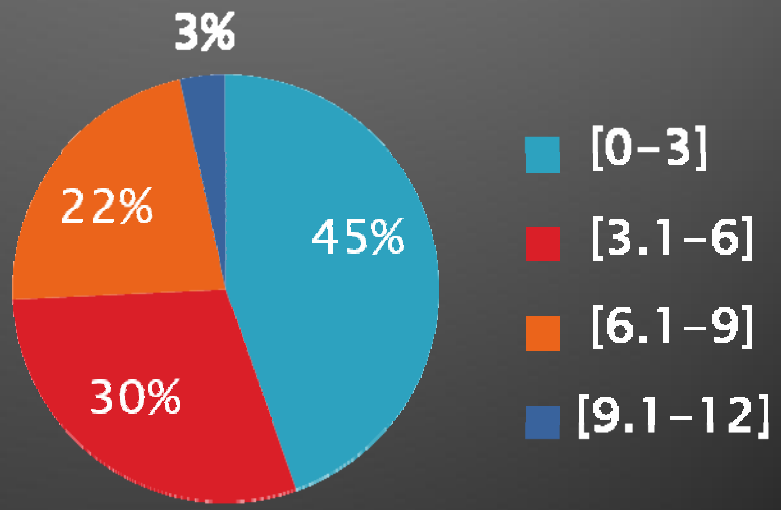
4. RESULTS XI

Effect size and value



In 69% of cases, effect size value is not specified.

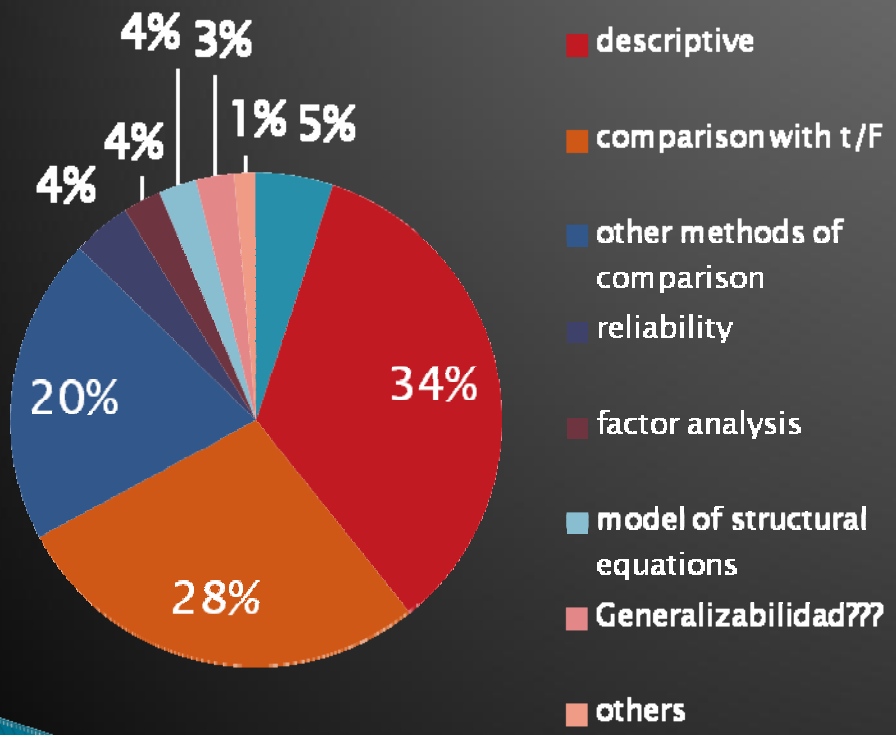
Index of quality



75% of studies have an index of quality lower than 6.

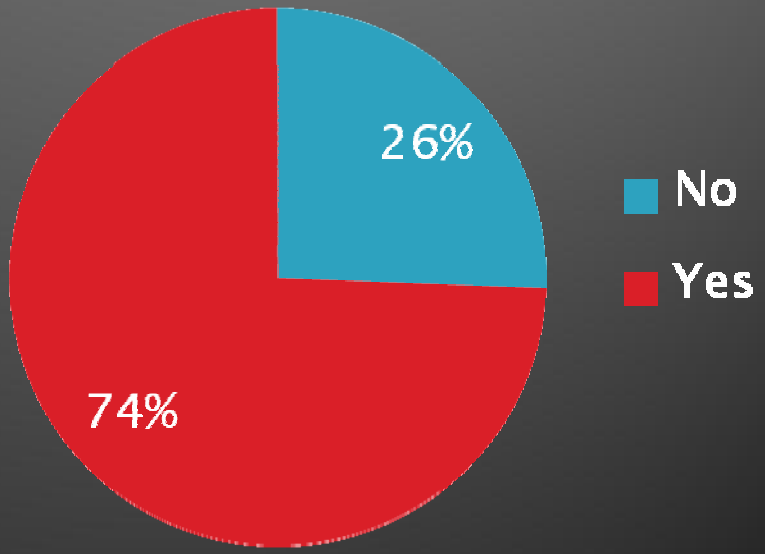
4. RESULTS XII

Statistic index calculated



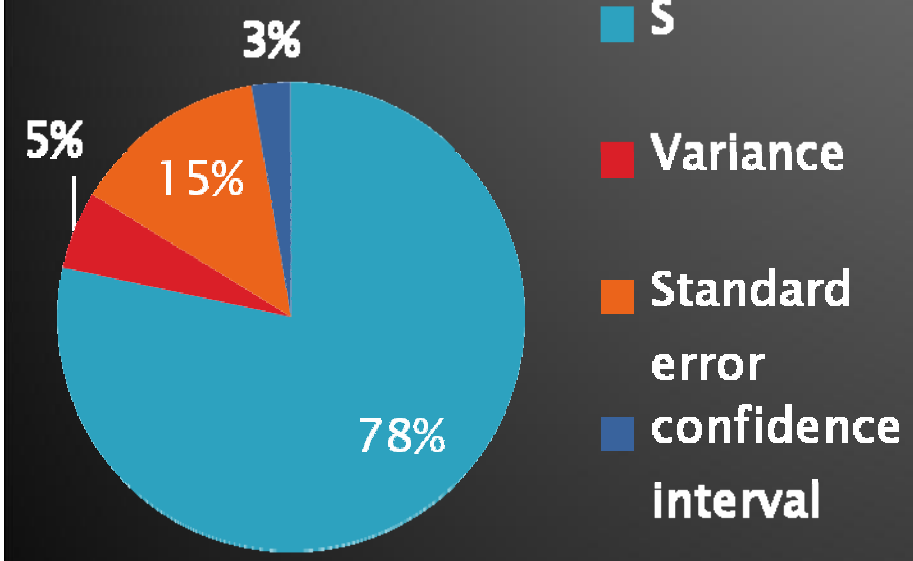
In 34% of cases, specified values are descriptive; in 28%, t or F .

Significant differences between measures

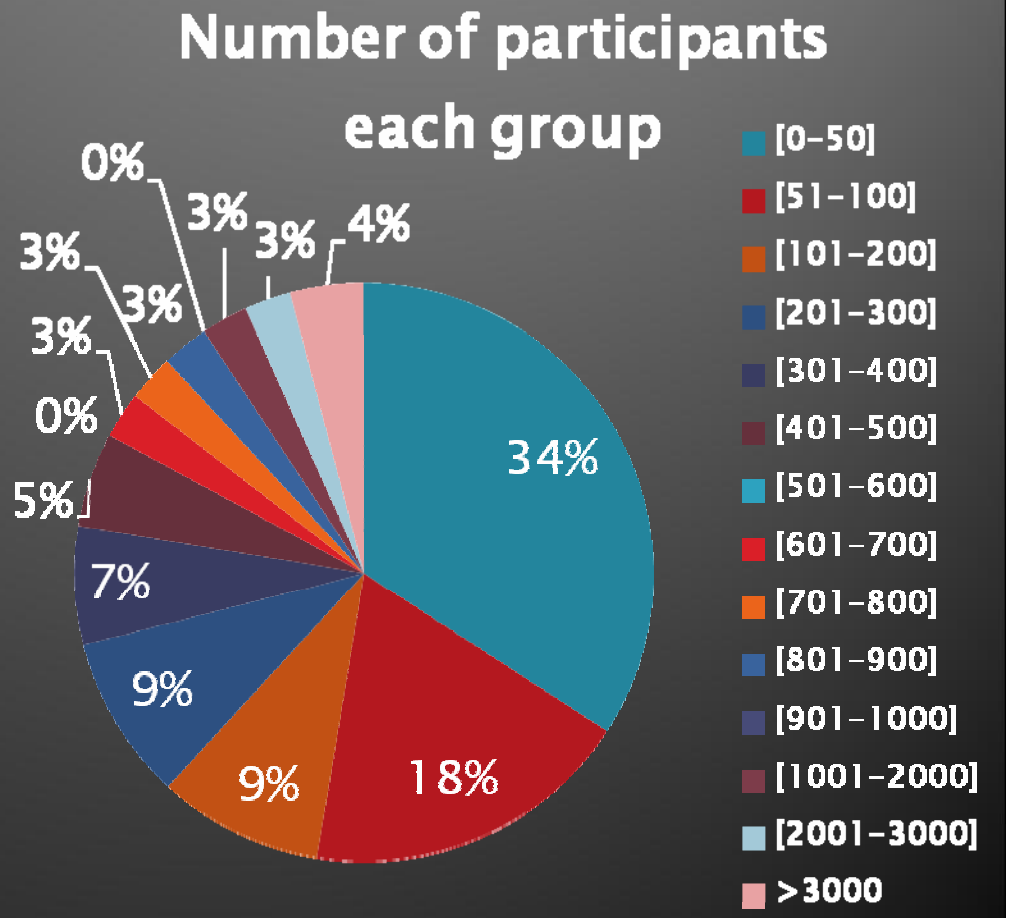


In 74%, results showed statistically significant differences.

4. RESULTS XIII

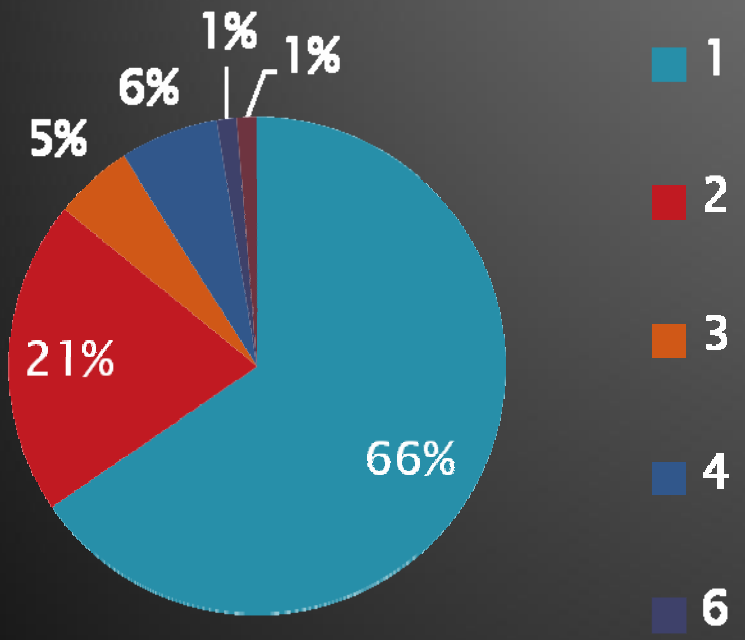


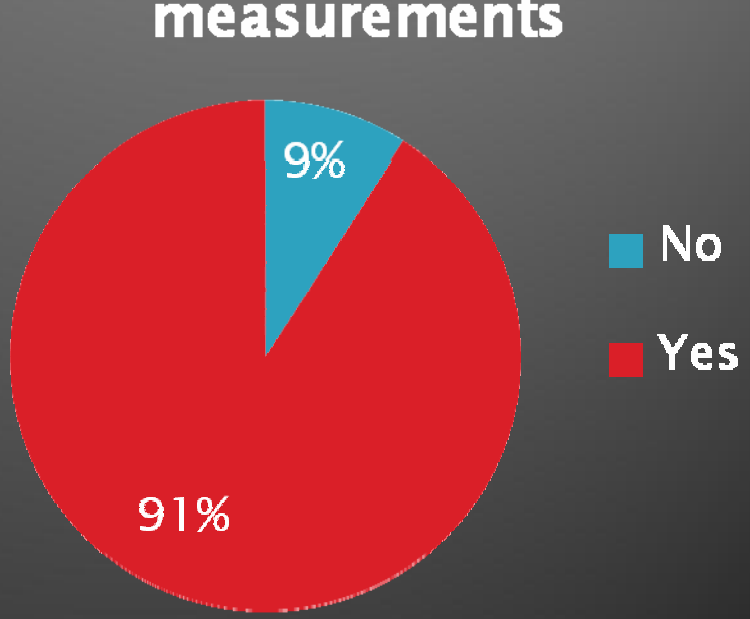
In 78% of occasions, deviation is the variability index used.



In 34% of the studies the number of subjects by group is between 0 and 50.

4. RESULTS XIV

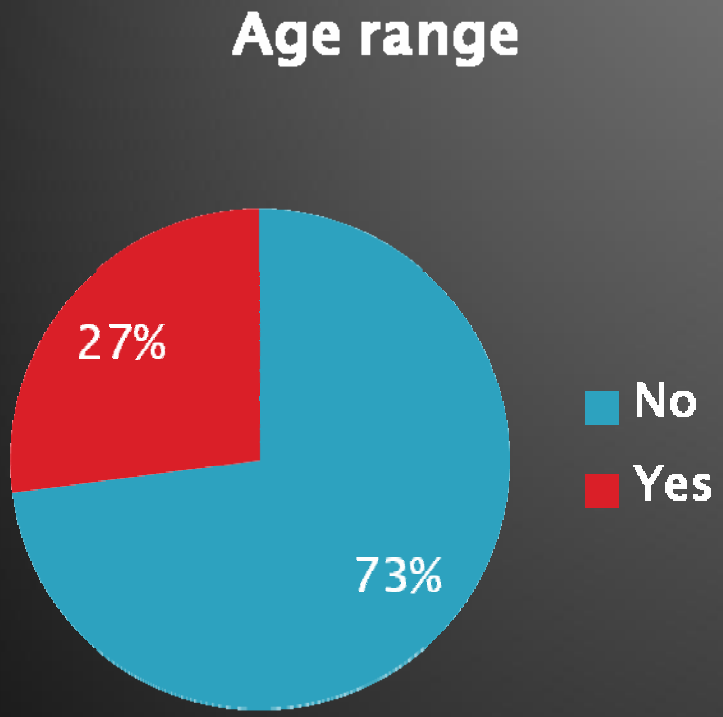




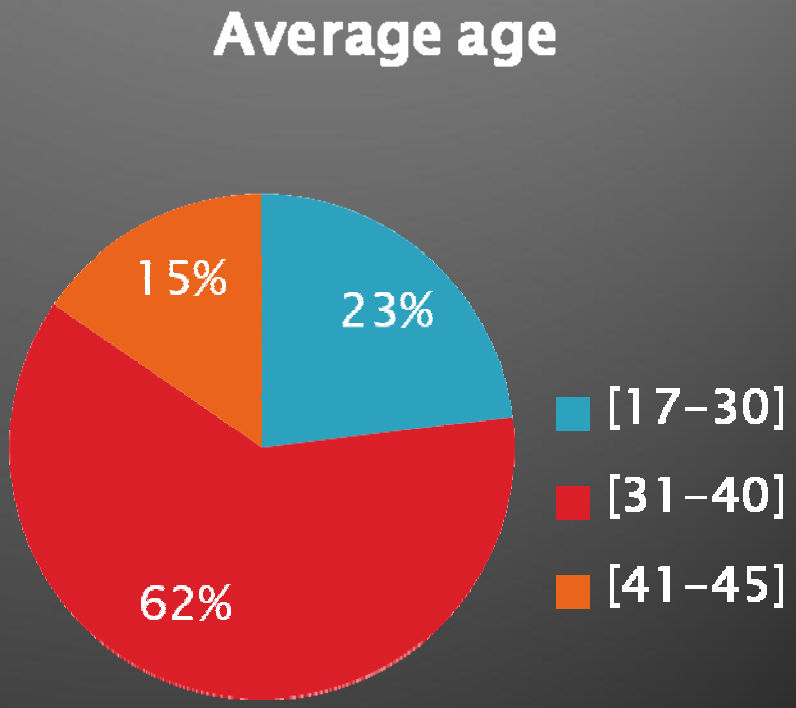
66% of the studies have only one group.

in 91% of the cases there are exclusions after measurement.

4. RESULTS XV

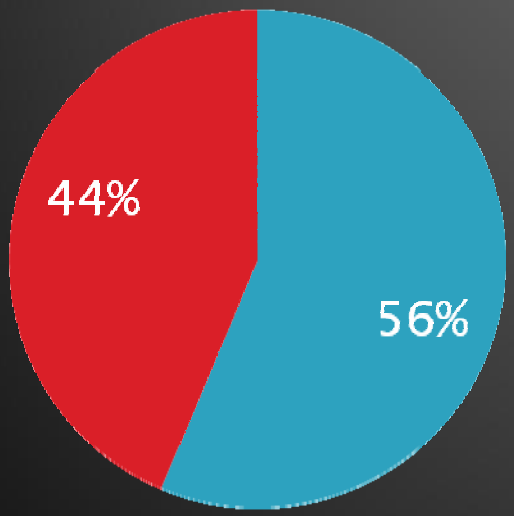


In 73% of cases, the age of the participants is not reported.

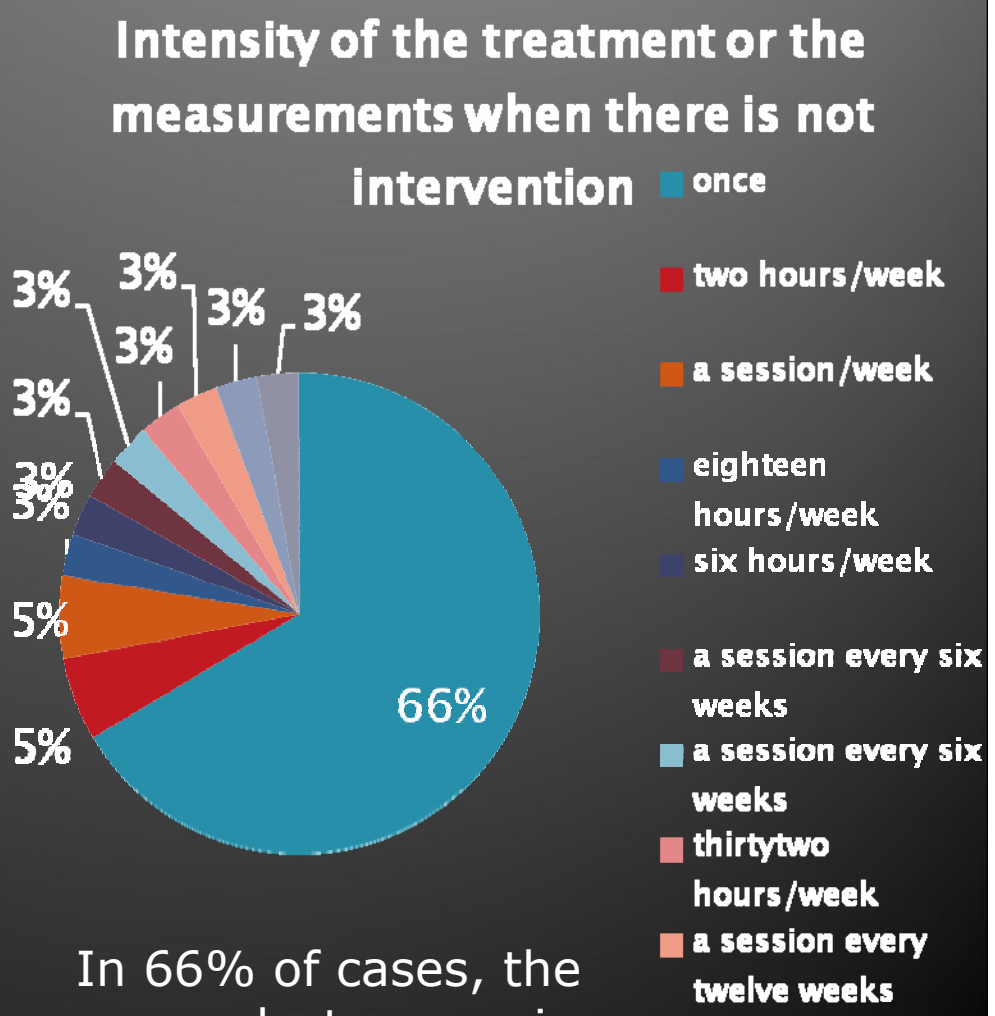


In 62% of occasions, the mean of the age of the participants is between 31 and 40; 23% between 17 and 30.

4. RESULTS XVI

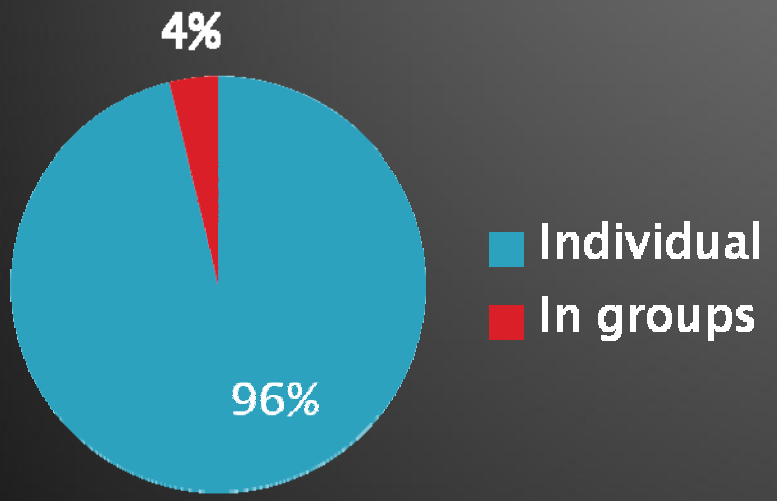


In 56% of occasions, period of treatment are 6 months or less.

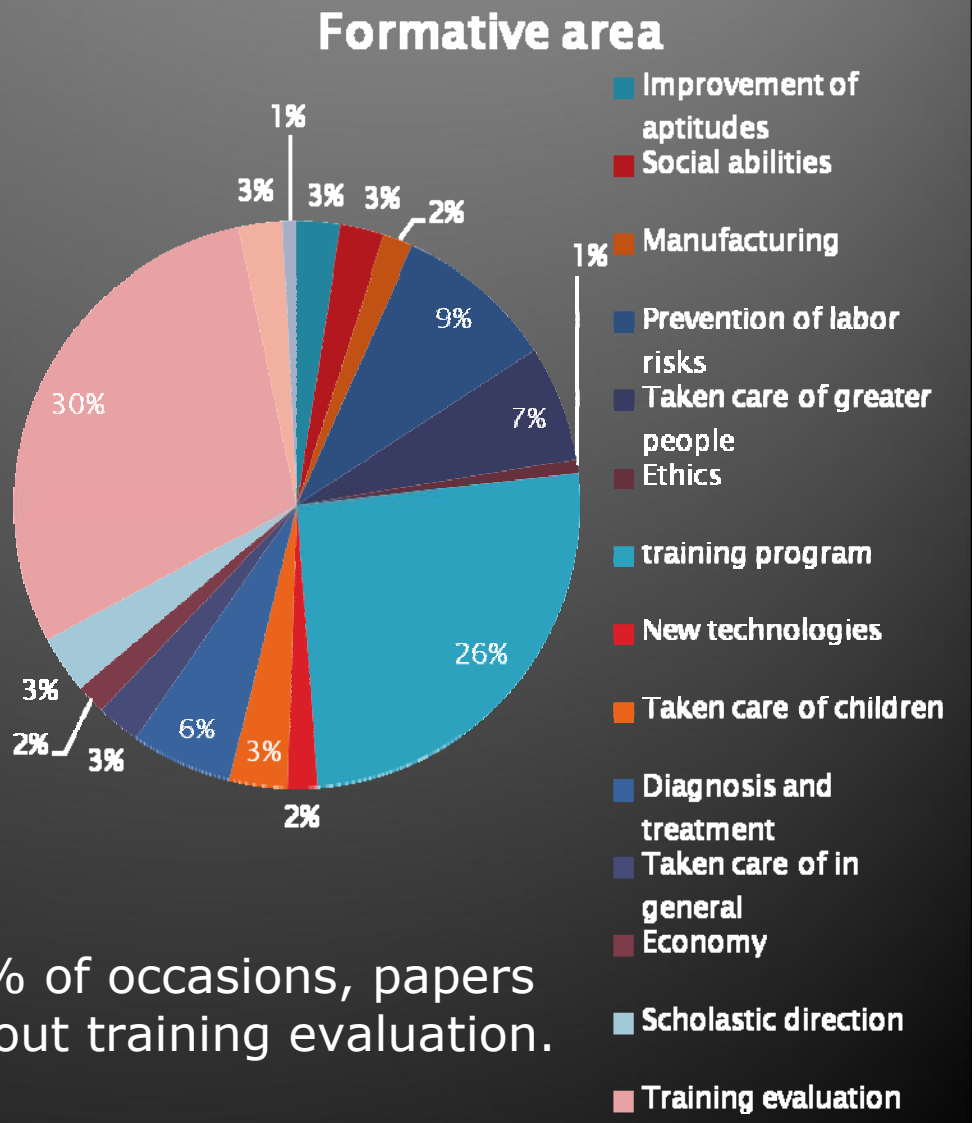


In 66% of cases, the program lasts a session.

4. RESULTS XVII

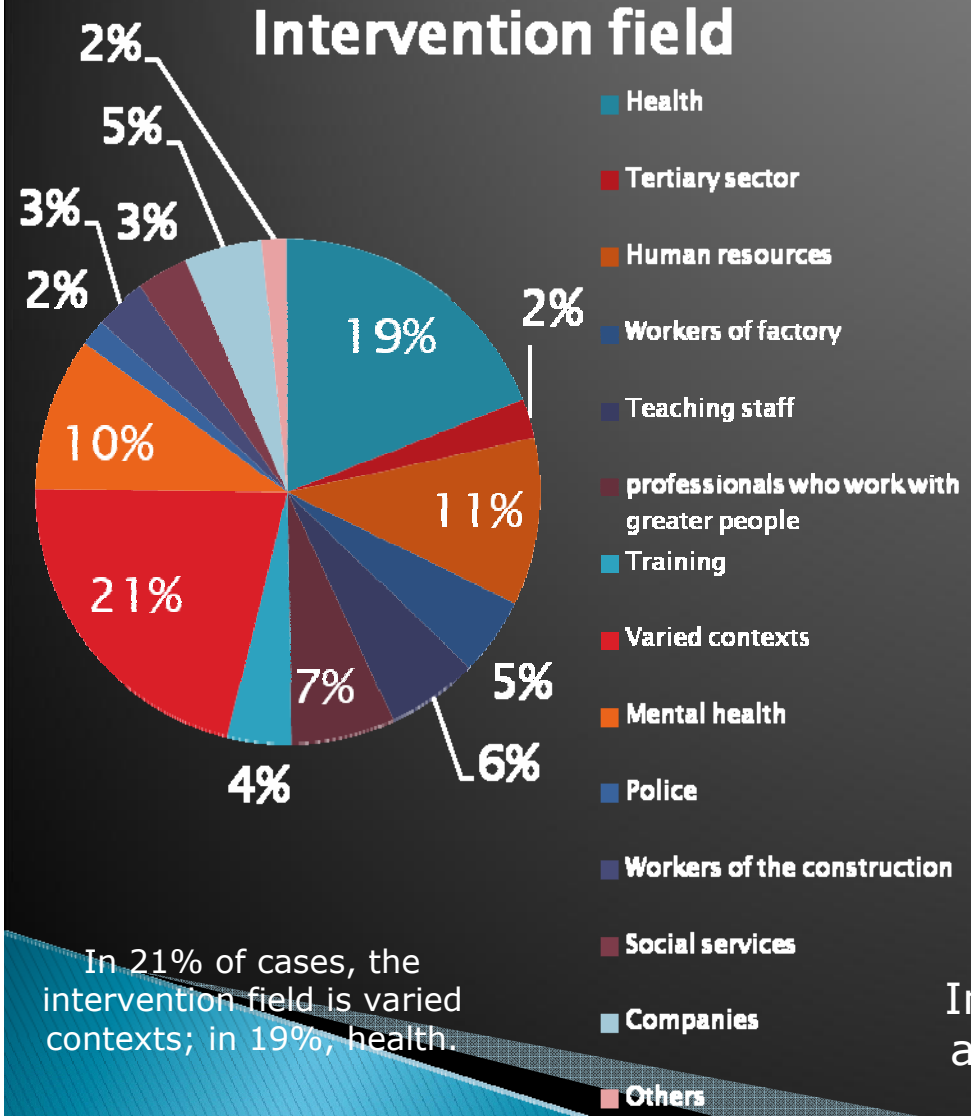


In 96% of occasions, beneficiaries participate individually in the program.

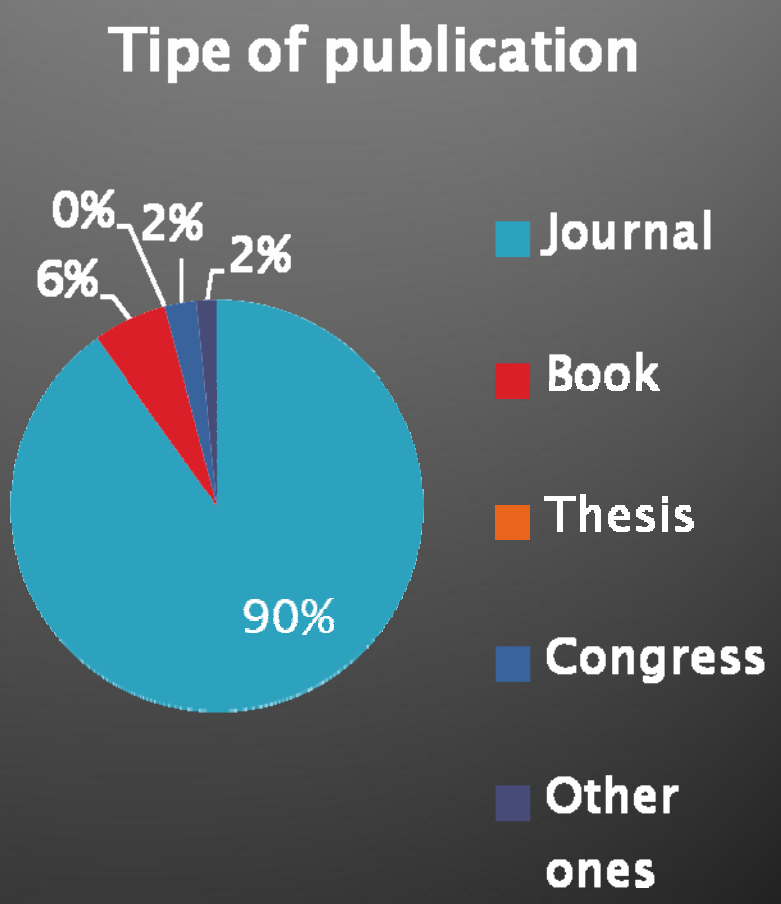


In 30% of occasions, papers talk about training evaluation.

4. RESULTS XVIII



In 21% of cases, the intervention field is varied contexts; in 19%, health.



In 90% of occasions, found studies are papers from different journals.

5. RESULTS AND CONCLUSIONS (I)
How to improve quality of designs in the evaluation of training programs

- In the 35% of collected papers, studied data were not specified because they were theoretical (not empirical).
- When data is specified, it usually shows a medium level of standardization and specification.
- Efficacy of training programs is assessed in a descriptive way.
- Quality index is very low in most of cases.

5. RESULTS AND CONCLUSIONS (II)
How to improve quality of designs in the evaluation of training programs

- **Assignment procedure of units (subjects) to conditions (causal effects):**
  - Should be clearly specified (randomly if possible –unbiased estimation of the effect size–).
  - Alternative: use similar comparison groups (using matching of units before assignment or cohort groups).
- **Pretest observations (observations previous to program implementation):**
  - Enhance using multiple pretest observations and trying to use high quality measures (for example physiological and standardized ones).
  - We must use at least one pretest observation (to test effects of interventions).
  - Alternative to pretest observations: pretest of independent samples, retrospective measures or proxy pretest of outcomes.

5. RESULTS AND CONCLUSIONS (III)
How to improve quality of designs in the evaluation of training programs

- Post-test observations:
  - We will always have a posttest observation, but we should add multiple posttest observations, equal or similar to pretest ones.
  - Enhance normalized post-test observations.
  - Alternative: we can combine post-test observations with non-equivalent dependent variables.
- Comparison groups:
  - More extensive information about sampling features (selection, error, bias, attrition, etc.) should be detailed.
  - Randomly conformed groups should be enhanced.
  - Multiple comparison groups should be used.
  - In extreme cases we can obtain comparison groups from regression extrapolation, or by using secondary data to make comparisons.

5. RESULTS AND CONCLUSIONS (IV)
How to improve quality of designs in the evaluation of training programs

- **Implementation of the program:**
  - Efficient follow-up procedures.
  - Alternatives (in some contexts): switching replications design; reversal design.
- **Control techniques.**
- **Combine different data analysis -not only descriptive-.**

5. RESULTS AND CONCLUSIONS (V)
How to improve quality of designs in the evaluation of training programs

- High percentage of missing data in many items
  - many studies were of theoretical type
  - low specification of the required data

- Is relationship between type of design and the quality of the study?, there are statistically significant differences in favor of the experimental studies.

- Low quality index (specially in non-experimental designs), it indicates the necessity to improve the design and the beginning of the works (emphasizing the specification of the excellent questions, use of comparison groups, explicitación of the allocation method,etc.)

- Via of immediate future work: meta-analytical study to verify the eficacy of trainning programs.

6. REFERENCES

- Chacón-Moscoso, S., Sánchez-Meca, J., Alarcón-Rubio, D., Marín-Martínez, F., Sanduvete-Chaves, S., and Huedo-Medina, T. (2004, February). *Quality of intervention programs. A comparative study between USA and Europe(European Union Countries)*. Paper presented at the 4th Campbell Collaboration Colloquium. Washington D.C.
- Chacón, S., Sánchez, J., Sanduvete, S. & Alarcón, D. (2006, February). A scale to measure the quality of primary studies for meta-analysis. A content validity study. Paper presented at the *VI Annual Campbell Collaboration Colloquium.* Los Angeles.
- Chacón-Moscoso, S., Sanduvete-Chaves, S. and Alarcón-Rubio, D. (2005, February). *Towards the Validation of a Scale to Measure the Quality of Primary Studies for Meta-analysis.* Paper presented at the 5th Campbell Collaboration Colloquium. Lisbon.
- Shadish, W., Chacón-Moscoso, S. & Sánchez-Meca, J. (2005). Campbell Collaboration and Evidence-Based Decision-Making: Enhancing systematic reviews of program evaluation results in Europe. *Evaluation. The International Journal of Theory, Research and Practice, 11,* 95-109.
- Shadish, W. R., Cook, T. D. & Campbell D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston & New York: Houghton Mifflin Company.

Thanks for your attention

schacon@us.es
sussancha@us.es
evaluacionforpas@us.es

Http://innoevalua.us.es