

Methodological quality of clinical trials of keloid scars

Calidad metodológica de los ensayos clínicos de cicatrices queloides

Martha A. Morales-Sánchez*, Andrea De Villafranca-Dugelby, Fermín Jurado-Santa Cruz, and María L. Peralta-Pedrero

Research Unit, Centro Dermatológico Dr. Ladislao de la Pascua, Servicios de Salud Pública de la Ciudad de México, Mexico City, Mexico

Abstract

Background: There is heterogeneity in the design of clinical trials (CTs) for the treatment of keloid scars that compromise the validity of their results. **Objectives:** The objectives of this study were to assess the methodological quality of the CT published on keloid scars, mainly the outcomes used in them. **Methods:** Articles of CT for keloid scars were analyzed, their methodological quality was evaluated following the CONSORT guidelines and the risk of bias based on the Cochrane tool. All the clinical outcomes measured in CT were identified. **Results:** Fifty-two full-text articles were evaluated, of which only 3.84% of the CT mentioned important changes in the methodology after starting patient recruitment. Fifty-nine percentages of the CT were assessed as high risk of performance bias due to mistakes in blinding participants and personnel. The most frequent outcome was the height or thickness of the keloids. **Conclusions:** We recommend that participants with hypertrophic scars should be excluded from keloids' CT, and that the main outcome must be the scar height and flattening. The pain and quality of life of patients should also be measured.

Keywords: Scars. Clinical trials. Keloids. Risk of bias.

Resumen

Antecedentes: Existe heterogeneidad en el diseño de los ensayos clínicos (EC) para el tratamiento de cicatrices queloides que compromete la validez de sus resultados. **Objetivo:** Evaluar la calidad metodológica de los EC publicados sobre cicatrices queloides, principalmente las medidas de eficacia utilizadas. **Método:** Se analizaron los EC publicados sobre tratamientos para cicatrices queloides y se evaluó su calidad metodológica siguiendo los lineamientos CONSORT y el riesgo de sesgo según la herramienta de Cochrane. Se identificaron todas las medidas de eficacia utilizadas en los EC. **Resultados:** Se incluyeron 52 artículos, de los cuales solo en el 3.84% se mencionan los cambios importantes en la metodología después de iniciar el reclutamiento de los pacientes. El 59.6% de los EC fueron evaluados con alto riesgo de sesgo de realización por errores en el cegamiento del equipo de investigación. La variable de respuesta más frecuente fue la altura o grosor de las cicatrices. **Conclusiones:** Se recomienda que en los EC de cicatrices queloides se excluya a los participantes con cicatrices hipertróficas, y que la variable de resultado principal sea la altura de la cicatriz. También deben medirse el dolor y la calidad de vida de los pacientes.

Palabras clave: Cicatrices. Ensayos clínicos. Queloides. Riesgo de sesgo.

Correspondence:

*Martha A. Morales-Sánchez

José María Vértiz 464

Col. Buenos Aires, Alcaldía Cuauhtémoc

C.P. 06780, Ciudad de México, México

E-mail: marthamoralessanchez@gmail.com

Date of reception: 17-08-2019

Date of acceptance: 07-05-2020

DOI: 10.24875/CIRUE.M22000172

Cir Cir (Eng). 2020;88(5):578-585

Contents available at PubMed

www.cirugiaycirujanos.com

2444-0507/© 2020 Academia Mexicana de Cirugía. Published by Permanyer. This is an open access article under the terms of the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

A scar is a physiological response to a break in continuity of any tissue. On the skin, scars are classified as normotrophic, atrophic, hypertrophic, and keloid^{1,2}. Keloid scars are an alteration in wound repair in genetically susceptible individuals, characterized by connective tissue exaggerated synthesis³. Keloid scars diagnosis is easy for a doctor; however, the challenge is to find an effective treatment that also reduces the risk of recurrence. Randomized clinical trials (CTs) are the standard method for measuring the efficacy and safety of a treatment and are based on evaluation of the change in keloid scars after any pharmacological or surgical intervention.

To evaluate the change in keloid scars, several scales have been designed, with the most widely used being the Vancouver Scale. However, in the past three decades, other scales have been created to measure in detail the pigmentation, vascularization, thickness, and appearance of scars, such as the Patient and Observer Scar Assessment Scale (POSAS), the Matching Assessment using Photographs with Scars (MAPS) assessment tool, and the Manchester Scar Scale (MSS), among others. In addition, the visual analog scale has been used to measure pain and pruritus triggered by keloid scars. The existence of several ways to measure changes in keloid scars has made for CTs on the subject to be barely comparable only for measuring the efficacy of interventions; in addition, scar evaluation using a clinical scale or index is operator-dependent. The above is added to the findings reported by the three most recent meta-analyses on interventions for keloid scars, in which the authors found CTs with small sample sizes, heterogeneity in efficacy evaluation, and lack of information on the blinding of interventions and the research team member who carries out the measurement of scars⁴⁻⁶. Only in the meta-analysis by Bi et al.⁶ was high comparability of results found since most of the 14 included that studies had been carried out by members of the same research team.

The quality of a randomized CT is based on its design, outcome variable measurement (the way treatment effect is measured) and statistical analysis of its results, which should be reflected on the published article. An article that does not include enough information to assess its quality does not allow the physician to formulate a judgment about its validity, both internal and external, and makes for its scientific contribution

to be questioned, its results not to be trusted and probably not to be used in routine clinical practice^{1,7}. Therefore, the Cochrane Collaboration and the Consolidated Standards of Reporting Trials (CONSORT) initiative have currently promoted critical evaluation of the methodological quality of randomized CTs to ensure that only effective interventions reach the patients⁴.

In keloid scars, given the absence of a highly effective treatment, coupled with the elevated cost of those that are available, it is necessary to analyze whether the evidence obtained in published CTs is real or overestimated due to methodological errors. Consequently, the purpose of this systematic review was to evaluate the methodological characteristics of published CTs on keloid scars, mainly the efficacy measurements used in them, and to determine their risk of bias to conclude with recommendations for clinical researchers involved in the design and execution of research projects on the subject.

Methods

A systematic review was carried out on the methodology of keloid scars CTs focused on the ways to measure efficacy or changes in scars, the protocol of which was registered in 2018 on the PROSPERO platform with registration number CRD42018107933⁸.

Search strategy

A search was carried out in the MEDLINE and Embase databases through PubMed and OVID SP, and in SciELO (Scientific Electronic Library Online), from 1970 to August 2018. “keloid” and “scar” were used as keywords, with study design being limited to “CT,” without language or age group restrictions.

Study selection

Two authors (M.L. Peralta-Pedrero and M.A. Morales-Sánchez) independently reviewed the bibliographic search to identify, through article titles and abstracts, those that met the selection criteria. Subsequently, selected articles full-texts were obtained to verify the selection criteria. Discrepancies on inclusion of any study were solved by consensus between both authors.

Selection criteria

CTs that had recruited subjects with keloid scars were included in the study. Quasi-experimental

studies and those that had mainly recruited individuals with acne (rolling, boxcar, and ice pick) and burn scars were excluded from the study.

Data extraction

Two authors (M.A. Morales-Sánchez and A. De Villafranca-Dugelby) evaluated included studies full-text to retrieve trial characteristics, recruited sample, variables for measuring efficacy and main outcomes, as well as to analyze the risk of bias.

Methodology and risk of bias evaluation

To evaluate the methodology, the CONSORT group standards were used as a guideline, and for risk of bias, the tool designed by the Cochrane Collaboration was used, which includes the randomization sequence generation, blinding, incomplete data, and selective reporting domains, among others^{9,10}. The CONSORT initiative was created in 1996 with the purpose to standardize different criteria for reporting randomized CTs and to facilitate their critical reading and interpretation. In 2001, the guidelines were revised and, in 2010, a third and final version was published. It is made up of 25 items grouped into six domains: title and abstract, introduction, methodology, results, discussion, and other information, such as original study protocol registration and the funding, whereby the study was carried out⁹. Two authors (M.A. Morales-Sánchez and A. De Villafranca-Dugelby) independently took care of the evaluation and two others (M.L. Peralta-Pedrero and F. Jurado-Santa Cruz) reviewed it for reaching a consensus.

Statistical analysis

Given that this is a systematic review, only study characteristics proportions and risk of bias domains are reported. No quantitative synthesis or meta-analysis of the analyzed CTs were carried out.

Results

One-thousand four-hundred and forty-six articles were found, out of which 52 that met the selection criteria were evaluated. The articles of included CTs are broken down by author and bibliographic reference in Appendix 1. The study inclusion flowchart is shown in figure 1.

CTs design characteristics

Of total CTs, 80.8% (n = 42) had a parallel group design, and 19.2% (n = 10), a crossover design, in which each participant was his own control. Most CTs had two intervention groups; only one simultaneously compared four interventions. Allocation to interventions were random only in 63.5% (n = 33) of CTs; in 17.3% (n = 9) of articles, there is no mention on how said allocation performed. Blinding of interventions was carried out in 42.3% (n = 22) of CTs; in 40.4% (n = 21), it was not specified, while in 17.3% (n = 9), it was not possible due to the type of intervention. Among the 22 blinded CTs, the intervention was blinded for the efficacy measurement evaluator in 68.2% (n = 15), for the evaluator and participant (double-blind) in 22.7% (n = 5), and only for the participant in 9.1% (n = 2). The number of participants recruited per CT ranged from 5 to 176 (median, 31).

Participants' characteristics

In 50% (n = 26) of CTs, participants with keloid scars were recruited; in the rest, subjects with hypertrophic scars were also recruited. In just over half the CTs (51.9%, n = 27), study subjects were adults; 32.7% (n = 17) included children and adults, and 5.8% (n = 3) recruited only pediatric population. In 9.6% (n = 5) of CTs, the patient age was not reported in the selection criteria or in the results section. Scar topography was reported in 69.2% of CTs (n = 36), mainly in the head, trunk, and extremities; 50% of CTs (n = 26) reported the scar origin, with the majority being caused by surgical procedures, burns, and acne, while only 42.3% (n = 22) reported the time of evolution, which ranged from 1 to up to 280 days.

Interventions

The type of intervention that was tested in the CTs was a pharmacological treatment in 65.4%, with topical or intralesional administration, followed by silicone patches in 44.2%, cryosurgery in 30.8%, surgical removal in 27%, and laser in 19.2%. Seven studies had a placebo control group.

Efficacy measurement

The most widely used efficacy measurement was scar height or thickness in 75% of included studies, followed

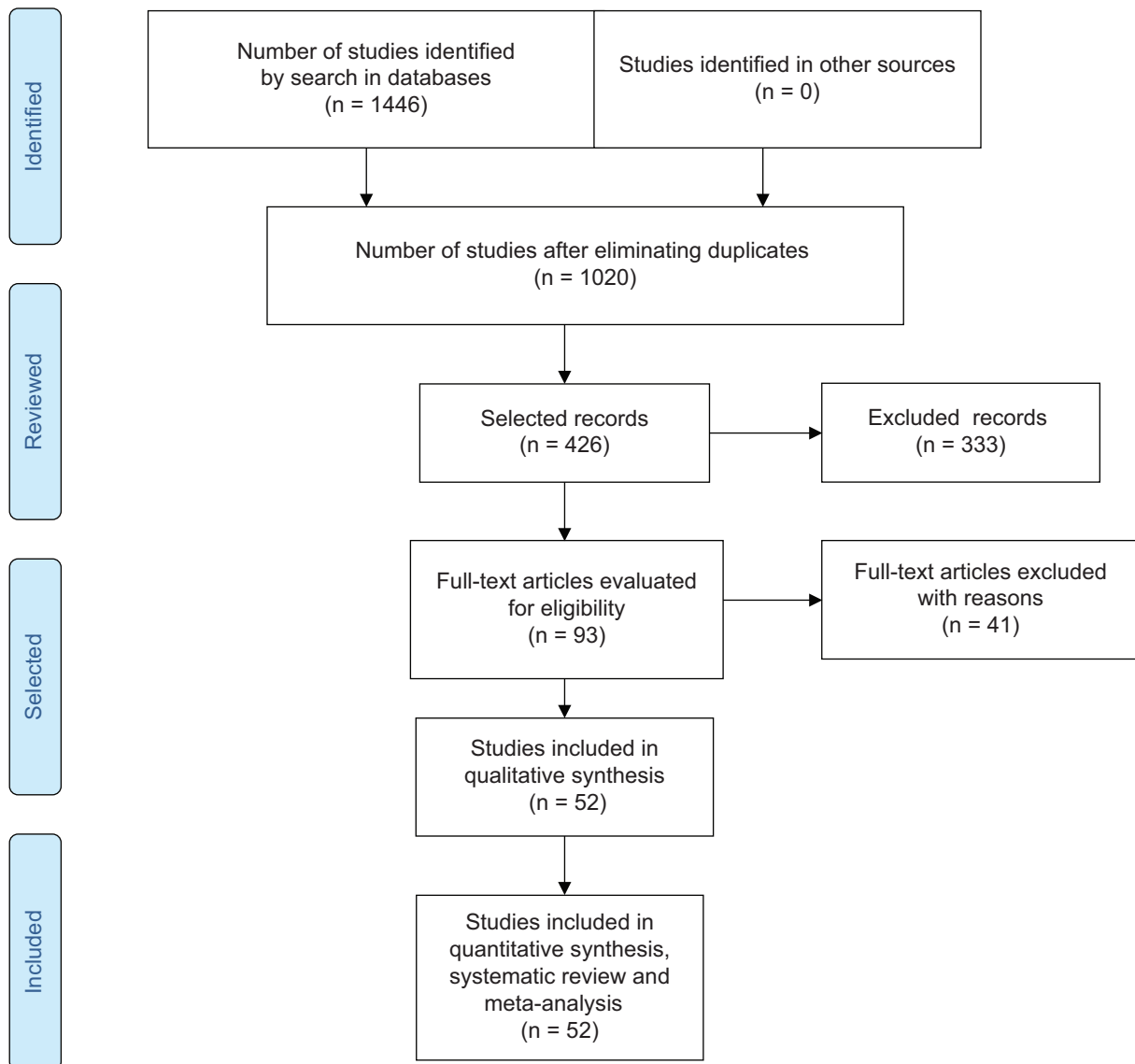


Figure 1. PRISMA flowchart.

by the use of a scale or clinical index in 71.2%, and elasticity and erythema or vascularity, with 55.8% each. To evaluate scar height or thickness, only 13 articles refer having used a graduated ruler. The symptoms most often considered as response variables were pruritus in 50% of CTs, and pain in 34.6%. In 21.2% ($n = 11$) of CTs, patient-perceived pain was measured at the time of intervention administration. Scar flattening or zero height was only considered as a response or outcome variable in 28.8% of CTs. The rest of efficacy measures included in the CTs are shown in table 1. It is important mentioning that only 25% of CTs ($n = 13$) evaluated keloid scar recurrence after the interventions, and the time at which it was evaluated was from 6 months after the start of the intervention to

5 years, with a median of 12 months. The authors reported the presence of adverse events in 28 articles (53.8%), with the most common being erythema, pain, hypopigmentation, atrophy, dermatitis, and telangiectasia.

As for the clinical scales that were used to measure the efficacy of interventions, the Vancouver scale was applied in 21.2% ($n = 11$) of CTs, followed by subjective improvement scales applied by the researcher in 19.2% ($n = 10$), patient-directed subjective improvement scales in 17.3% ($n = 9$), and the POSAS scale in 7.7% ($n = 4$). In 19.2% ($n = 10$) of CTs, scales or indices specifically designed for these studies were used, all focused on making a subjective evaluation of scar appearance and symptoms.

Table 1. Efficacy measurements of clinical trials on keloid scars

Variable	% (n = 52)
Scar height	75 (39)
Clinical scale or index	71.2 (37)
Elasticity	55.8 (29)
Erythema or vascularity	55.8 (29)
Pruritus	50 (26)
Pigmentation	40.4 (21)
Flattening	28.8 (15)
Pain	34.6 (18)
Scar area	26.9 (14)
Surface irregularity	13.5 (7)
Volume	1.9 (1)
Mobility limitation	1.9 (1)

As regards efficacy evaluation or outcome variable measurement, only in 25% (n = 13) of CTs did the authors mention who performed such measurement. Among the 13 CTs, in nine evaluation was carried out by a dermatologist, while in the rest, evaluators were plastic surgeons and pathologists (one study, in which histological changes were the response variable).

Manuscript content evaluation

According to CONSORT guidelines, in 100% of CTs were the results consistent in terms of benefits and harms considering relevant scientific evidence, while in 98% (n = 51), interventions for each group were detailed, allowing their replication. On the other hand, 3.84% of CTs mentioned relevant changes in methodology after patient recruitment was started, and only in one study was the research protocol full-version found. Only for five CTs was their registration found on any database, and changes in variables were documented after participants' recruitment was started. A breakdown of compliance with the manuscripts' content is shown in table 2.

Risk of bias assessment

Of total CTs, 26.9% (n = 14) were classified with high risk of selection bias for not having a suitable

Table 2. Methodological evaluation according to CONSORT

Domain	Item	% (n = 52)
Methods	Identified as a randomized trial in the title	13.4 (7)
	Structured abstract	84.6 (44)
	Scientific background and rationale	94.23 (49)
	Specific objectives or hypothesis	90.38 (47)
	Description of trial design, including the reason for allocation	94.23 (49)
	Important changes in methods after starting the trial	3.84 (2)
	Participants' selection criteria	92.3 (48)
	Origin (centers and institutions) of recorded data	65.3 (34)
	Detailed description of the intervention	98 (51)
	A priori specification of response variables	94.23 (49)
	Any change in response variables after trial initiation	9.6 (5)
	Sample size calculation	21.15 (11)
	If applicable, explain any interim analysis and interruption rules	19.23 (10)
	Method used for random allocation sequence to be generated	44.23 (23)
	Randomization type; details on any restrictions	42.3 (22)

method for generating intervention groups' allocation sequence. Seventeen CTs (32.7%) did not use a good method for concealing intervention groups allocation, regardless of the generation of a random sequence, which is why they were also classified with high risk of selection bias. More than half the CTs (59.6%, n = 31) were evaluated with high risk of performance bias due to blinding errors, both of participants and investigators, which might have influenced the response variable. As for detection bias, 38.5% (n = 20) of the studies had high risk, because the researchers who evaluated treatment response knew the intervention assigned to the participants that they evaluated, given that proper blinding was not possible. In 21.1% of CTs (n = 11), a statistical analysis that included all recruited patients was not performed and only the results of participants who completed the interventions were reported; some did not report all the variables they measured according to the methodology of the article or those that should have been measured with the clinical scales or indices they used. No CT was classified with high risk of reporting bias, but 76%

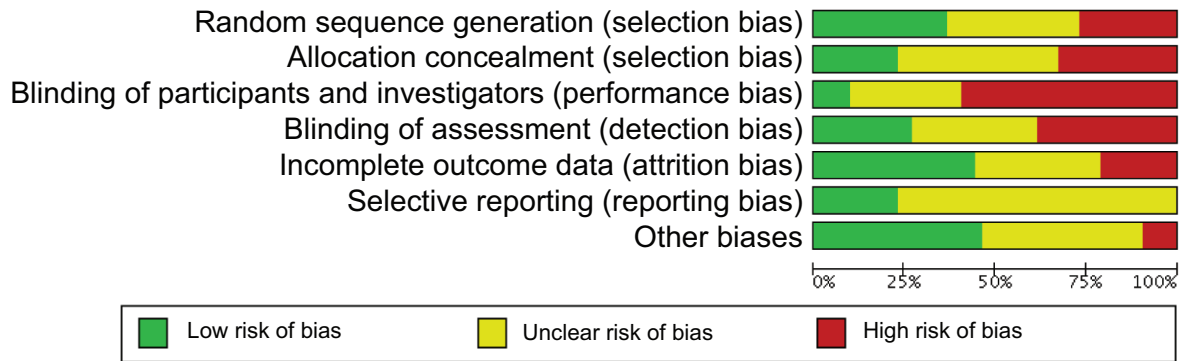


Figure 2. Risk of bias graph.

($n = 40$) were classified with unclear risk of reporting bias, because the protocol registration was not found, and interventions' efficacy was not measured with appropriate instruments. It should be noted that few CTs were classified with low risk of bias in all domains; in fact, only 5 (9.6%) were classified with low risk of performance bias. A breakdown of risk of bias assessments is shown in figure 2. In the "other biases" domain, 9.6% ($n = 5$) of CTs were found to have between-group differences at baseline or did not provide data to verify groups' comparability according to their baseline profile.

Discussion

In our systematic review, we found that the efficacy measurement that is most commonly used in CTs on keloid scars treatment is height or thickness of the scars, since it was the response variable in 75% of CTs, followed by the use of any clinical scale or index in 71.2%. However, although most studies use a scale to assess scar characteristics, the most widely used is the Vancouver scale, in 21.2% of CTs. In other words, there is heterogeneity in the way keloid scars are evaluated when using a scale, despite the existence of systematic reviews on the properties of these instruments, which indicate that those with the best validation process are POSAS and Patient-Reported Impact of Scars Measure (PRISM), since both include items that measure the impact on patients¹¹. However, these scales are long, and both have poor correlation with objective measurements such as height, erythema, pigmentation, and elasticity, and PRISM, which is the most robust, does not measure the appearance of scars¹². One of the advantages of using the Vancouver scale is that it is focused on evaluating

scars clinical characteristics, such as height, elasticity, vascularity, and pigmentation, in addition to being easy to apply for clinical researchers despite not having good inter-observer reproducibility, which is why, in a CT, it should always be applied by the same evaluator to avoid measurement errors¹¹. In turn, the POSAS scale has no sensitivity to scar characteristics changes during a CT, despite having many domains for their evaluation¹².

Methodologically, all CTs should include measurements of patient-reported interventions efficacy (patient-reported outcomes), among which pruritus and pain were measured in keloid scars in 50% and 34.6%, respectively. In 17.3% of CTs, patients evaluated their treatment response using a visual analog scale of improvement. However, a systematic review of instruments or scales designed to measure changes in keloid scars documented that there is no correlation between what the patient and the researcher report on treatment responses¹³. Despite the difference in said evaluations, in a CT on keloid scars, it is essential for patient perception and quality of life to be measured since it is highly likely that the treatment will not accomplish a clinically significant improvement, but symptomatology modification could be a benefit for the patient. In this sense, it is important to point out that in none of the reviewed CTs was quality of life measured.

When we evaluated the CTs according to CONSORT guidelines, only 13.4% were identified as randomized CTs in the title; in two, methodology changes after patient recruitment and at the start of interventions were not mentioned, although changes were actually made, and only in one was the study protocol full-version found. This shows the lack of methodological rigor in the preparation of CTs reports, and that in the

journals, where they were published, adherence to CONSORT is not considered among the requirements for publication. At present, most indexed international journals request the CONSORT checklist as a submission requirement for manuscripts that report CTs. Interestingly, only half the reviewed articles (51.9%) were published in dermatology journals; the rest appeared in journals of surgery and medicine in general.

On the other hand, when we evaluated the risk of bias, we found important flaws that directly affect CTs results, casting doubts on both their internal and external validity. The most commonly identified error was failure to blind the research team and participants to the interventions, which impact on the results when the evaluator of treatment response or outcome variable is the same doctor who administers the treatment and the patient is aware of the therapy he/she receives^{4,5}. In addition, most scales that were used, such as Vancouver and POSAS, are observer-dependent and involve intra-observer and inter-observer agreement and, therefore, the fact that the authors do not report who measured the intervention may be a factor for questioning the results.

Finally, 26.9% of CTs had a high risk of selection bias due to flaws in the generation of the sequence of intervention assignment, which compromises the quality of the results of any CT, and in other areas of medicine, it is unusual for this to occur. In addition, most manuscripts did not report essential information for allowing CTs risk of bias to be assessed.

As recommendations derived from this systematic review, we propose, as a rule to be followed in CTs on keloid scars, to exclude patients with hypertrophic scars from recruitment, since in half the analyzed CTs, both populations were included and were not separately analyzed. Due to their natural evolution and pathophysiology, hypertrophic scars tend to improve over time and easily respond to compression treatment, with silicone patches and others, unlike keloid scars, which should be controlled right from CTs design to avoid confusion in treatment effect¹⁴. The primary outcome variable should be keloid scar height or thickness, and the percentage of participants who achieve scar flattening or zero height should be reported since this is the clinically significant outcome for both patients and investigators. Secondary response variables should include the measurement of pain and pruritus with visual analog scales, and quality of life using generic questionnaires or with the PRISM scale, which measures said impact on patients. Keloid scar recurrence should be another

variable to be measured, but at least 12 months after the intervention; therefore, we recommend that any CT on scars should have at least 12 weeks of treatment, ideally 6 months, and 1-year follow-up. Given the difficulty for interventions with different administration routes to be blinded, which is the rule in treatments for keloid scars, our suggestion is that the evaluator of treatment response should be independent and blinded to interventions, that is, he/she should not be the main investigator, who usually looks after patients and can easily recognize the type of intervention that the patient is receiving and be influenced. All these recommendations are aimed at seeking a consensus, on which the outcome variables to be measured in CTs on keloid scars should be, as it has already been done in other pathologies¹⁵.

The limitations of this systematic review are those inherent to the lack of information found in the manuscripts, a fact that reinforces the hypothesis that, in CTs for keloid scars, an expert consensus should be reached to define the treatment response variables that need to be measured, CT procedures and adherence to international guidelines on the publication of scientific articles.

Conclusions

This systematic review allowed us to know the flaws in the design and reporting of CTs on the treatment of keloid scars and to assume that the heterogeneity in the obtained results and the difficulty for performing meta-analyses is a result of their methodological deficiencies. Based on this analysis, guidelines are proposed for the design of CTs on the treatment of keloid scars that can guide the clinical researcher in his everyday tasks. Finally, in view of the emergence of new genetic and epigenetic therapies for keloid scars, it is necessary for CTs with good methodological quality to be carried out to determine if said treatments are efficacious from the clinical point of view.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical disclosures

Protection of human and animal subjects. The authors declare that no experiments were performed on humans or animals for this research.

Confidentiality of data. The authors declare that they have followed the protocols of their work center on the publication of patient data.

Right to privacy and informed consent. The authors declare that no patient data appear in this article.

Appendix

Additional material can be consulted in the appendix at <https://doi.org/10.24875/CIRU.20001506>.

References

1. Sanclemente G. Evidencias en dermatología: ensayos clínicos, revisiones sistemáticas y guías de práctica clínica. Tesis doctoral. Departamento de Pediatría, Obstetricia, Ginecología y Medicina Preventiva, Universidad Autónoma de Barcelona, España; 2016.
2. Herranz P, Santos-Heredero X. Cicatrices. Guía de valoración y tratamiento. Meda Pharma. Publicidad Just in Time S.L., España; 2012.
3. Ogawa R. Keloid and hypertrophic scars are the result of chronic inflammation in the reticular dermis. *Int J Mol Sci*. 2017;18. pii: E606.
4. Ren Y, Zhou X, Wei Z, Lin W, Fan B, Feng S. Efficacy and safety of triamcinolone acetonide alone and in combination with 5-fluorouracil for treating hypertrophic scars and keloids: a systematic review and meta-analysis. *Int Wound J*. 2017;14:480-7.
5. Wong T-S, Li JZ-H, Chen S, Chan JY-W, Gao W. The efficacy of triamcinolone acetonide in keloid treatment: a systematic review and meta-analysis. *Front Med*. 2016;3:71.
6. Bi M, Sun P, Li D, Dong Z, Chen Z. Intralesional injection of botulinum toxin type A compared with intralesional injection of corticosteroid for the treatment of hypertrophic scar and keloid: a systematic review and meta-analysis. *Med Sci Monit*. 2019;25:2950-8.
7. Moraga J, Cartes-Velásquez R. Pautas de chequeo, parte I: Consort y Trend. *Rev Chil Cir*. 2015;67:225-32.
8. Morales-Sánchez MA, Peralta-Pedrero ML, Jurado-Santa Cruz F. Core outcomes of clinical trials for the treatment of keloid scars. Citado 08 abril 2019. Disponible en: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018107933
9. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010;340:c869.
10. Higgins J, Green S. *Cochrane Handbook for systematic reviews of interventions*. Version 5.1.0, 2011. Disponible en: <https://handbook-5-1.cochrane.org/>
11. Vercelli S, Ferriero G, Sartorio F, Cisari C, Bravani E. Clinimetric properties and clinical utility in rehabilitation of postsurgical scar rating scales: a systematic review. *Int J Rehabil Res*. 2015;38:279-86.
12. Mundy LR, Miller HC, Klassen AF, Cano SJ, Pusic AL. Patient-reported outcome instruments for surgical and traumatic scars: a systematic review of their development, content, and psychometric validation. *Aesthetic Plast Surg*. 2016;40:792-800.
13. Tan A, Glass DA 2nd. Patient-reported outcomes for keloids: a systematic review. *G Ital Dermatol Venereol*. 2019;154:148-65.
14. Slemper AE, Kirschner RE. Keloids and scars: a review of keloids and scars, their pathogenesis, risk factors, and management. *Curr Opin Pediatr*. 2006;18:396-402.
15. Schmitt J, Apfelbacher C, Spuls PI, Thomas KS, Simpson EL, Furue M, et al. The harmonizing outcome measures for eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol*. 2015;135:24-30.