

Subgroup Analysis in Pulmonary Hypertension-specific Therapy Trials: a Systematic Review

H Rodríguez-Ramallo

Virgen del Rocio University Hospital

N Báez-Gutiérrez (✉ nerea.baez.sspa@juntadeandalucia.es)

Virgen del Rocio University Hospital

R Otero Candellera

Virgen del Rocio University Hospital

S Flores-Moreno

Virgen del Rocio University Hospital

L Abdel-kader Martín

Virgen del Rocio University Hospital

Research Article

Keywords: pulmonary hypertension, subgroup analyses, randomized controlled trials, methodological limitations

Posted Date: September 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-879986/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background. Pulmonary hypertension (PH) treatment decisions are driven by randomized controlled trials (RCTs) results. Subgroup analyses are often performed to assess whether the intervention effect will change due to the patient's characteristics. As subgroup claims may mislead clinician treatment decisions, there is a need for standards of such analyses.

Objective. To evaluate the appropriateness and interpretation of subgroup analysis performed in pulmonary hypertension-specific therapy RCTs.

Methods. A systematic review of the literature for pulmonary hypertension-specific therapy RCTs published between January 2000 and December 2020 was conducted. Claims of subgroup effects were evaluated with Sun X et al., 2012 criteria.

Results. 30 RCTs were included. Evaluated subgroup analyses presented: a high number of subgroup analyses reported, lack of prespecification, and interaction test. The trial protocol was not available for most RCTs; significant differences were found in those articles which published the protocol. Authors reported 13 claims of subgroup effect, with 12 claims meeting 4 or fewer Sun criteria.

Conclusion. Subgroup analyses in pulmonary hypertension-specific therapies are of poor quality. The lack of published protocols limited our capability to assess whether the published results correspond to the initially predefined analyses. Most claims of subgroup effect did not meet critical criteria.

Take Home Message

- Subgroup analyses performed in pulmonary hypertension-specific therapy Randomized Controlled Trials present severe methodological limitations, with publications delivering several potentially misleading claims of subgroup effect.

1. Introduction

Pulmonary hypertension (PH) is a relatively frequent complication of multiple clinical disorders [1]. Among other factors, the variety of aetiologies of PH makes it an extremely complex disease; for this reason, a clinical classification into 5 categories has been developed to group PH according to clinical presentation, findings, underlying conditions, and treatment [2]. As PH affects older patients disproportionately and may cause rapid deterioration and an increased risk of death, it is considered a major health issue, specifically in countries with older populations [3]. Several drugs with diverse pharmacological mechanisms have been developed for the treatment of PH. The choice of treatment for PH will vary according to the group of pulmonary hypertension to be treated, as therapies usually considered appropriate may even be harmful in a certain subgroup of patients [1].

Treatment decisions in PH are driven by results from randomized controlled trials (RCTs). Usually, only average results are reported in RCTs, and trial participants are often recruited from heterogeneous populations. However, clinicians ideally want more specific information to assist them in applying trial results to individual patients. Researchers conducting RCT usually perform subgroup analysis to assess whether the intervention effect will change due to the patient's baseline characteristics such as underlying pathologies, age, sex, or disease severity. Based on subgroup analysis results, researchers may report claims of subgroup effects. Nonetheless, subgroup claims should be interpreted cautiously since misstatements about subgroup effects may result in patients being denied beneficial treatments or even receiving treatments that may be harmful or ineffective [4–6].

The need for standards for the interpretation of subgroup analysis is crucial for treatment decisions in medical practice. Explicit criteria have been developed for this purpose [7–12]. Recent tools to evaluate subgroup credibility have been published, such as Gil-Sierra MD et al. 2020 [8] and Schandelmeier S et al. 2020 [7]. However, as far as we are concerned, the “10 criteria for assessing the credibility of a subgroup claim” [12] is the most reliable tool to assess confidence in subgroup analysis as they have been widely tested in several disciplines [13–16].

The central purpose of this study was to evaluate the appropriateness and interpretation of subgroup analysis performed in pulmonary hypertension-specific therapy RCTs. In order to achieve our goals, the following aspects have been studied:

- Description of subgroups analysis and claims of subgroup effects.
- Research characteristics of subgroup analysis.
- Analysis and interpretation of subgroup effects for primary outcomes.
- Assessment of subgroups claims credibility using the “10 criteria for assessing the credibility of a subgroup claim” [12].

2. Methods

2.1 Literature search.

This systematic review aims to summarize the available data to solve the following research questions, framed in the Population Intervention Comparator Outcome-Study (PICOS) design framework: Population, patients with pulmonary hypertension; Intervention, pulmonary hypertension-specific therapy; Comparison, studies with a comparator will be considered; Outcomes, subgroup analysis; Study design, randomized clinical trials.

As pulmonary hypertension-specific therapy was considered the following groups of drugs:

- Calcium channel blockers.
- Phosphodiesterase type 5 inhibitors.

- Endothelin receptor antagonists.
- Prostacyclin analogues and prostacyclin receptor agonists.
- Guanylate cyclase stimulators.

A systematic search was conducted according to the Preferred Reporting Items for a Systematic Review and Meta-analysis (PRISMA) guidelines [17]. The systematic review protocol was registered with the prospective register for systematic review protocols (PROSPERO), registration number: CRD42021242265.

The search was conducted between January 2000 and December 2020 using vocabulary and keywords controlled by Mesh terms in the MEDLINE database to identify RCTs assessing pulmonary hypertension-specific therapy for pulmonary hypertension patients.

The search was performed in March 2021. The full literature search strategy is available in Additional file 1.

The following criteria were used for the trial selection:

Eligibility criteria.

We considered all published pulmonary hypertension-specific therapy RCTs on pulmonary hypertension adults with subgroup analysis reported.

Exclusion criteria.

- Articles written in languages other than English, Spanish, and French.
- Post-hoc analyses of a previously published RCT.
- Articles that were not available.
- Trials in which subgroup analysis credibility was impossible to evaluate due to missing data.

2.2 Study screening and selection.

Two investigators independently checked the titles and abstracts of the search results using predefined inclusion criteria. The full text was accessed for all titles that seem to meet the inclusion criteria or have uncertainties. Two reviewers, HRR and NBG, assessed whether the article met the selection criteria. Any disagreements were resolved through discussion or arbitration with the third reviewer, LAM.

2.3 Data extraction.

For data extraction, other sources included in the study were used (i.e., trial registration, published protocols, and online supplements). Data were extracted and entered in a structured Microsoft Excel (Redmond, WA, USA) database.

Eligible RCTs were evaluated to determine whether a subgroup analysis was reported. A *subgroup analysis* was defined as a statistical analysis that explored whether or not the effects of the intervention differed according to the status of a subgroup variable. A *subgroup effect* was defined as a difference in the magnitude of a treatment effect across a group of a study population [12]. For each RCT reporting subgroup analysis and subgroup claims, the following information was collected:

Trial characteristics: Information on the funding source, year and journal of publication, journal impact factor, pulmonary hypertension classification according to Clinical classification of pulmonary hypertension [2], updated by the European Society of Cardiology and the European Respiratory Society (ESC/ERS) Guidelines [1], centre (multicentric or unicentric), trial design (parallel, cross-over, or factorial), trial type (superiority, noninferiority, or equivalence), allocation concealment, blinding of patients, the number of patients randomized. The primary endpoint was categorized according to whether the results were statistically significant and the type of outcome variable (time-to-event, binary, continuous, or count).

Reporting of subgroup analysis: Number of subgroup factors, type of subgroup factors (clinical factors or biomarkers), number of subgroup analysis and outcomes for subgroup analysis reported, forest plots used, prespecified or post hoc subgroup, the statistical method used to assess the heterogeneity of the treatment effect (descriptive only, subgroup P values and confidence interval or interaction test). When the trial protocol was available, the agreement on the number of subgroup factors, the number of subgroup analyses, and the pre-specification of such analyses between the journal publication and the trial protocol were measured.

A subgroup factor was defined as a study variable, by which the population may be categorized into different subgroups, i.e., sex, age, the presence of a mutation. A subgroup analysis was defined as a specific analysis performed to compare two categories within a subgroup factor. For example, within the age factor, the analysis that compares the subgroups: > 65 years vs. <65 years.

Claims of subgroup effects: Subgroup claims mode of presentation (abstract or text only), number of subgroup claims, subgroup variable (primary or secondary outcome), and number of outcomes for subgroup claims were recorded. A subgroup effect was considered to be claimed when the authors stated in the abstract or discussion that the intervention effect differed between the categories of the subgroup variable. The claims of subgroup effects were classified according to the strength of the claim into three categories: strong claim, a claim of a likely effect, or suggestion of a possible effect based on Sun et al. classification (Additional file 2). To evaluate the credibility of subgroup claims for primary outcomes, “the 10 criteria for assessing the credibility of a subgroup claim” were applied pair-wise (Additional file 3). If the subgroup claim met less than half the criteria, the credibility of this claim was considered low.

2.4 Assessment of risk of bias.

The risk of bias was assessed using the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials [18]. The risk of bias was assessed by two independent reviewers. Possible

disagreements between reviewers were resolved by discussion or arbitration by a third reviewer when consensus could not be reached.

2.5 Secondary analyses.

The quality of subgroup analysis reports during 4 time periods (2000–2004, 2005–2009, 2010–2014, and 2015–2019) were compared. This analysis aims to assess whether the methodology reported to perform subgroup analyses has improved over time.

2.6 Data analysis.

A descriptive analysis was developed. Continuous and categorical variables were presented as mean (range) and n (%), respectively.

For those RCTs that stated a subgroup effect without providing an interaction test, P interaction was calculated using the Joaquin Primo calculator [19] to verify that there was indeed statistical significance.

The inter-reviewer agreement for assessing the credibility of the subgroup claims was estimated by Cohen's kappa coefficient.

3. Results

The initial literature search identified 1837 studies. After the first review by title or abstract and the deletion of duplicates, 185 articles were selected for full-text review. Finally, 30 papers were included (Fig. 1). The excluded articles and the reasons for their exclusion are provided in the supplementary material (Additional file 4).

3.1 Trials characteristics.

The characteristics of included trials in this study are listed in Table 1. Included publications reported data on 7765 randomized patients (Median: 208; range: 52-1156).

Table 1
 Characteristics of trials included in the analysis (N = 30)

Variable		N° of Trials	%
Funding source	Industry	27	90
	Non- industry	1	3.3
	Not specified	2	6.7
Year of publication	2000–2004	3	10
	2005–2009	7	23.4
	2010–2014	10	33.3
	2015–2019	10	33.3
Journal	Chest	2	6.7
	Circulation	4	13.3
	European Heart Journal	2	6.7
	Journal of the American College of Cardiology	2	6.7
	The Lancet. Respiratory medicine	2	6.7
	The New England journal of medicine	8	26.7
	Others	10	33.3
Journal Impact factor	< 10	8	26.7
	> 10	22	73.3
Pulmonary Hypertension Group	Group 1 PH	20	66.6
	Group 2 PH	3	10
	Group 3 PH	2	6.7
	Group 4 PH	3	10
	Any	2	6.7
Centre	Multicentric	27	90
	Unicentric	2	6.7
	Not specified	1	3.3
Trial design	Parallel	30	100

PH: Pulmonary Hypertension

Variable		N° of Trials	%
Type of trial	Superiority	30	100
Allocation concealment	Yes	14	46.7
	No	1	3.3
	Unclear	15	50
Blinding	Open label	1	3.3
	Double blinded	28	93.3
	Not specified	1	3.3
Protocol was freely available	Yes	7	23.3
	No	23	66.7
N° patients randomized*	Total	7765	
	Median (Range)	208 (52-1156)	
N° arms	Median (Range)	2 (2–5)	
Type of primary endpoint	Time-to-event	5	16.7
	Binary	2	6.66
	Continuous	23	76.67
Trial met primary endpoint *	Yes	19	63.3
	No	8	26.7
PH: Pulmonary Hypertension			

Most studies were funded by industry (90%, n = 27) and were published principally during 2013 (20%; n = 6). The most selected journals for publication were The New England Journal of Medicine (26.7%; n = 8) and Circulation journal (13.3%; n = 4). 73.3% (n = 22) of the studies were published in high impact journals (impact factor > 10).

The most common pulmonary hypertension types explored were type 1 (66.6%; n = 20), 2 (10%; n = 3) and 4 (10%; n = 3). Stated primary endpoint was statistically significant in 63.3% (n = 19) of trials.

3.2 Subgroup analyses.

Characteristics of reported subgroup analysis are listed in Table 2. Subgroup analyses were mostly mentioned in the result (90%; n = 27) and the discussion (63.3%; n = 19) sections.

Table 2
 Characteristics of subgroup analysis reporting (N = 30)

Reporting of subgroup analysis		N° of Trials	%
Mode of presentation	Abstract	3	10
	Methods	11	36.7
	Results	27	90
	Discussion	19	63.3
	Supplementary material	8	26.7
N° subgroup factors	2-4	2	6.7
	5-10	10	33.3
	>10	1	3.3
	Unclear	17	56.7
	Median (range)	6 (2-17)	
N° subgroup analysis reported	2-4	1	3.33
	5-10	11	36.7
	>10	1	3.3
	Unclear	17	56.7
	Median (Range)	7 (2-36)	
N° subgroup outcomes	1	21	
	2-5	2	
	>5	3	
	Unclear	4	
	Median (Range)	1 (1-12)	
Forest plot	Yes	16	53.3
	No	14	46.7
Prespecified or post hoc	Prespecified	14	46.7
	Post hoc	5	16.7
	Unclear	9	30
	Prespecified and post hoc	2	6.66
CI: Confidence interval			

Reporting of subgroup analysis		N° of Trials	%
Statistical method	Descriptive	10	33.3
	Subgroups P or CI	6	20
	Interaction test	11	36.7
	Unclear	3	10
Subgroup claim	Yes	8	26.7
	No	22	73.3
CI: Confidence interval			

Most trials, 56.7% (n = 17), did not clearly report the number of subgroup factors or subgroup analysis carried out. The remaining trials reported at least 5 subgroup factors or subgroup analyses in 36.7% (n = 11) and 40% (n = 12) of the trials, respectively. Subgroup analysis for more than one outcome was reported in 16.7% (n = 5) of trials. Forest plots used to report subgroup analyses data in 53.3% (n = 16) of the trials.

For 30% (n = 9) of trials, it was unclear whether subgroup analysis was pre-planned or post hoc, in 46.7% (n = 14) of trials were prespecified and 16.7% (n = 5) were post hoc.

Only 36.7% (n = 11) of trials used an interaction test to assess heterogeneity of the treatment effect; 33.3% (n = 10) reported subgroup analysis without any statistical analysis.

The clinical trial protocol was available for 8 of the 30 RCTs included. Relevant differences were found for all 8 of the RCTs when comparing the trial protocol and the published manuscript:

- Subgroup analyses: 6 RCTs reported a fewer number of subgroup analyses than prespecified in the protocol, the two RCTs remaining reported subgroup analyses that were not prespecified in the protocol; in both cases, these analyses were characterized as prespecified in the published manuscript.
- Subgroup factors: The number of subgroup factors reported differed between the protocol and the published manuscript in 7 cases: 5 RCTs reported fewer factors than those specified in the protocol, the remaining two added several subgroup factors that were not previously defined.
- Selective reports of subgroup analyses by outcome: There were differences in the number of subgroup analyses reported for the primary outcome in 7 RCTs. In addition, in 4 protocols, authors specified that subgroup analysis would be carried out for primary and secondary endpoints; however, the published manuscript only reported the subgroup analyses for the primary endpoint on three of these RCTs.

3.3 Claims of subgroup effects.

Table 3 lists the characteristics of subgroup claims identified. In 11 RCTs [20–28], authors claim heterogeneity of treatment effect of at least one subject subgroup. Two RCTs made two claims of subgroup differences [29, 30]. Of the 11 RCTs with claims of subgroup effect: 4 reached the primary endpoint, 5 did not reach it, and for the rest, a clear primary endpoint was not defined. Only three (27.7%) RCTs provided interaction test results to prove a subgroup difference.

Table 3
Articles with claims of subgroup differences (N = 11)

Claim of subgroup difference		N° of Trials	%
Mode of presentation	Abstract	4	36.4
	Text only	7	63.6
N° subgroup claims	1	9	63.6
	2	2	18.2
Subgroup variable	Primary endpoint	11	100
Forest plot	Yes	2	18.2
	No	9	63.6
N° subgroup analysis	1–4	0	9.1
	5–10	2	18.2
	> 10	1	9.1
	Unclear	8	72.2
	Median (Range)	7 (7–12)	
N° of outcomes for subgroup claims	1	8	72.7
	2–5	1	9.1
	> 5	1	9.1
	Unclear	1	9.1
	Median (Range)	1 (1–12)	
Statistical method	Descriptive	3	27.3
	Subgroups P or CI	5	45.5
	Interaction test	3	27.7
Prespecified or post hoc	Prespecified	3	27.3
	Post hoc	4	36.4
	Prespecified and post hoc	1	9.1
	Unclear	3	27.3
Protocol was freely available	Yes	1	9.1
	No	10	90.1

A total of 13 subgroup differences were claimed in 11 trials. These claims were classified as: three (23.1%) strong claims, one (7.7%) claim of a likely effect, and 9 (69.2%) suggestions of a possible effect.

Concerning the 10 criteria to assess the credibility of subgroups claims (Table 4): authors included subgroup variables for the primary outcome measured at baseline for all 13 claims, used subgroup variable as stratification factor at randomization for three (23.1%) claims, clearly prespecified their hypothesis for three (23.1%) claims, tested a small number of hypothesis for one (7.7%) claims, carried out a test of interaction that provides statistically significant for 4 (30.8%) claims, correctly prespecify direction for one (7.7%) claim, documented replication of a subgroup effect with previously related studies for 8 (61.5%) claims, and provide a biological rationale for the effect for 6 (46.2%) claims. Of the 13 claims, 12 (92.3%) met 4 or fewer of the 10 criteria. For strong claims, only one (33.3%) met 5 criteria.

Table 4
Claims meeting subgroup criteria for primary outcomes

Criteria	Strong claim (n = 3)	Claim of likely effect (n = 1)	Suggestion of effect (n = 9)	Total (n = 13)
Subgroup variable as a baseline characteristic *	3 (100%)	1 (100%)	9 (100%)	13 (100%)
Subgroup variable a stratification factor at randomization	0 (0%)	1 (100%)	2 (22.2%)	3 (23.1%)
Subgroup hypothesis specified a priori	0 (0%)	0 (0%)	3 (33.3%)	3 (23.1%)
A small number of hypothesised effects tested (≤ 5)	0 (0%)	0 (0%)	0 (0%)	1 (7.7%)
Significant interaction test ($P < 0.05$) ¹	0 (0%)	0 (0%)	4 (44.5%)	4 (30.8%)
Independence of interaction *	-	-	-	-
Direction of the subgroup effect correctly prespecified?	1 (33.3%)	0 (0%)	0 (0%)	1 (7.7%)
Subgroup effect consistency across studies	2 (66.7%)	0 (0%)	6 (66.7%)	8 (61.5%)
Subgroup effect consistent across related outcomes	-	-	-	-
Compelling indirect evidence	1 (33.3%)	0 (0%)	5 (55.6%)	6 (46.2%)
* Two trials claimed two subgroup claims each.				
¹ For those RCTs that stated a subgroup effect without providing an interaction test, P interaction was calculated using the Joaquin Primo calculator [19] to verify that there was indeed statistical significance.				

The inter-reviewer agreement for the assessment of the credibility of the subgroup claims was 0.88 (95% CI: 0.77–0.98), representing substantial to almost perfect agreement.

Risk of Bias Graphs Within Studies and across studies are available at supplemental material (Additional file 5).

3.4 Secondary analyses.

Figure 2 shows the evolution of the quality of the subgroup analyses reported over 4 periods of time.

An improvement was observed for most methodological characteristics of pulmonary hypertension-specific therapy RCTs over time, except for the use of subgroup variables as a stratification factor at randomization.

4. Discussion

Subgroup analyses have the potential to generate investigation hypotheses, discover new treatments, and identify baseline factors that may influence treatment efficacy or toxicity. However, when subgroup analyses are misused may also lead to spurious findings and misleading interpretations [31–33]. The most frequent methodological limitations of subgroup analyses in RCTs have been reported extensively; multiple testing of hypotheses, inadequate statistical power, inappropriate a priori specification, and lacking biological rationale [4, 5, 33–35].

As a result of this review, we can observe that, generally, the subgroup analyses carried out in RCTs of pulmonary hypertension-specific therapy are of low quality, despite being published primarily in high-impact factor journals. It highlights the lack of clarity in the allocation concealment. For most clinical trials, the study protocol is not available; therefore, it is challenging to verify aspects such as the pre-specification of the subgroup analyses. Furthermore, of the 11 RCTs with subgroup effect claims, only one has a publicly available protocol. For those studies whose protocol was available, subgroup analyses reported in the manuscript lacked description and were significantly different from those planned in the protocol.

Other factors that stand out the methodological errors when performing subgroup analyses in this study were identified; A high number of subgroup analyses reported, the high number of post hoc analyses, and the lack of interaction test to confirm the existence of subgroup effects.

When multiple subgroup analyses are carried out, the results obtained should be interpreted with caution since the probability of obtaining a false positive can be significantly augmented [5]. This risk may be increased, especially if, in addition, the hypothesis of the subgroup analyses has not been pre-specified [5, 13, 33]. The approximately calculated risk for a false positive result for 5 subgroup analyses is 25%; however, it may increase as the number of subgroup analyses arises. We identified a median of 6 subgroup analyses reported among the RCTs evaluated in this review.

The pre-specification of subgroup analysis is a frequent parameter measured in order to estimate methodological quality. For a subgroup analysis to be prespecified, it must be planned and documented before any examination of the data; this is based on the premise that a prespecified analysis usually follows a biological rationale. However, pre-specification alone may not lead to solid subgroup analyses as prespecified analysis may be based on unlikely and poorly formulated hypotheses [36]. In pulmonary-specific therapy RCT, 46.7% (14) of subgroup analyses were prespecified.

In addition to the pre-specification of the subgroup analysis, the correct direction of subgroup hypotheses must also be specified. For those claims in which the direction of the effect has not been or has been

wrongly identified, their credibility could be reduced.

A common mistaken belief among authors is to claim a subgroup difference when a statistically significant effect is found in one subgroup but not in the other. One of the essential criteria to appropriately establish a claim of subgroup effect is performing an interaction test [37]. The p-value of an interaction test provides information about the probability that the existence of a subgroup difference is due to an accidental finding or chance rather than an actual subgroup effect. In this review, we observed that only 37.7% of the RCTs performed an interaction test to confirm the existence of a subgroup claim. Of the 9 claims of subgroup difference identified in this study, 44.4% (n = 4) were based on a significant interaction test. When comparing our results with others carried out in other areas, we found mixed results. Wallach et al. identified that among a sample of articles that made at least one claim in the abstract, 40% of the subgroups' claims were based on the result of an interaction test [38]. On the other hand, Khan et al. evaluate the quality of subgroup analyses in heart failure RCTs, reporting 70% of claims based on significant interaction tests [39].

Most of the studies included in this review were industry-funded (90%), which could have influenced our results. The source of funding of clinical trials may play a role in the quality of the reports of subgroup analyses; industry-funded RCTs are more likely to report subgroup analyses [40–42], even when an overall treatment effect for a primary outcome could not be proved [40]. Industry funding was also correlated with suboptimal reporting of subgroup effects; often, the subgroup hypotheses were not pre-specified, and the use of an interaction test was rare [40, 42]. This is consistent with our findings in this primarily industry-funded sample of RCTs as, among the articles that claimed difference of subgroup effect, only 4 (36.4%) RCTs reached the primary endpoint.

Previous studies have found that the methodological quality reported on the methods sections of published articles is lacking compared to study protocols [43, 44], finding high-quality studies being poorly reported. Protocols provide a complete insight into the analysis methods utilized in RCT. It is recommended to publish trial protocols all together with the publication of the RCT and its publication in clinical trial registries, thus providing the reader a transparent and complete description of the prespecified methods. However, several studies have found that RCT protocols are often not freely available [41, 45]; this is consistent with our findings, as only 7 out of 30 RCTs provided the study protocol, and discrete growth in protocol publishing was observed during the studied period.

The fact that protocols are not systematically accessible is alarming; even when voluntarily published, discrepancies with journal publications are relatively frequent when reporting study outcomes [46–54]. Similarly, high inconsistency between protocols and publications has been described in several methodological characteristics of subgroup analysis: Omitted prespecified analyses [54], interaction test, pre-specification of subgroup analyses, and minor differences for the anticipated direction of the effect [41]. Due to these prevalent discrepancies, the credibility of subgroup methods may be questionable if the study protocol is not accessible.

Our findings coincide with previous reports; few studies (23.3%) published the protocol either in the journal publication or clinical trial registries. 46.7% (n = 14) of studies reported a prespecified subgroup analysis, with only half publishing the study's protocol. Furthermore, 30% (n = 9) of studies did not report clearly whether the subgroup analysis was prespecified or post-hoc; in none of these cases, the protocol was freely available.

Despite subgroup analysis methodological limitations in RCTs are increasingly recognized, a review of 437 randomly selected RCTs published in high-impact journals found a decrease in the appropriateness of reporting subgroup analyses from 2007 to 2014 [42].

In contrast with these results, we observed an improvement of most methodological characteristics of pulmonary hypertension-specific therapy RCTs: a priori specification, forest plot utilization, and interaction test improved from 2002 to 2019. However, a decline of subgroup variables set as stratification factors during randomization was observed. This decrease adds to the hypothesis that most subgroup analyses, even when prespecified, are exploratory. When a particular characteristic is known to influence the trial outcome, it should be used as a stratification factor at randomization.

Claims of subgroup effect are common in RCT reports. Several systematic reviews and analyses have shown that authors believe and report a difference in treatment effects between patient subgroups in 40–60% of all RCT reporting subgroup analyses [13, 36, 55]. Few systematic reviews have described a relatively low number of subgroup claims [14, 39]. Our results were in line with the latest, as we found that pulmonary hypertension-specific therapy RCTs reported claims of subgroup effect on 26.7% (n = 9) of RCTs reporting subgroup analyses. Fewer subgroup claims may indicate that authors are cautious in their reporting, as these claims may result in changes in clinical practices.

4.1 Strengths.

To our knowledge, this is the first systematic review of the credibility of subgroup analysis and subgroup effect claims reported on pulmonary hypertension-specific therapy RCTs. A rigorous systematic method was employed. Standardized criteria were used in order to assess the credibility of subgroup claims.

4.2 Limitations.

This study has some limitations: First, although we use a scale to determine the credibility of the claims, the sun criteria were not designed to provide a score; therefore, the later interpretation of its results is not without subjectivity.

Secondly, when assessing the strength of a claim, there is an undeniable subjective value in interpreting what the authors state. However, the pair-wise work and the high agreement in the results of both researchers suggest that the limitation in this sense was not significant.

Third, in most of the studies, we were unable to find the study protocols. In many cases, we could not know whether the published results correspond to the initially defined objectives; this limits our capability

to judge the credibility of subgroup claims. For this purpose, authors must provide detailed information about the conduct and results of subgroups analysis.

4.3 Improvement on the reporting of subgroup analyses proposals.

Although the methodological limitations of subgroup analyses are consistently reported in the literature, similar mistakes are carried when conducting and reporting subgroup analyses in recent RCTs. As improvement measures to change the current state of subgroup analyses, we propose the following:

Firstly, subgroup analysis should be prespecified and documented in trial registries. Secondly, scientific journals should request authors to make the study protocol accessible to reviewers and readers as a requirement for publishing the results of RCTs. Thirdly the use of guidelines or tools for the correct publication of subgroup analyses should be enforced. Fourthly, researchers should be cautious when claiming subgroup differences, even when a robust methodology for subgroup analyses was followed.

5. Conclusions

Subgroup analysis in pulmonary hypertension-specific therapies is of poor quality; flaws identified in previous studies were common. Although the fulfilment of several criteria improved over time, most studies did not set subgroup variables as stratification factors at randomization, prespecified the subgroup analyses, or published the study protocol.

Subgroup claims credibility was low. Most claims did not meet critical criteria; therefore, clinicians should be sceptical of claims of subgroup effects if these differences are not confirmed in later RCTs.

6. List Of Abbreviations

ESC: European Society of Cardiology.

ERS: European Respiratory Society.

PH: Pulmonary hypertension.

PICOS: Population Intervention Comparator Outcome-Study.

PRISMA: Preferred Reporting Items for a Systematic Review and Meta-analysis.

PROSPERO: Prospective register for systematic review protocols.

RCT: Randomized controlled trials.

7. Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests

Funding

The authors received no specific funding for this work.

Authors' contributions

LAM and ROC conceived the study. H.R.-R. and N.B.-G. collected the data and wrote the manuscript. LAM, ROC and SFM contributed to the interpretation of the results and reviewed the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

8. References

1. Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J*. 2016;37(1):67-119.
2. Simonneau G, Galiè N, Rubin LJ, Langleben D, Seeger W, Domenighetti G, et al. Clinical classification of pulmonary hypertension. *J Am Coll Cardiol* 2004;43(Suppl 1):S5–S12
3. Hoeper MM, Humbert M, Souza R, Idrees M, Kawut SM, Sliwa-Hahnle K, et al. A global view of pulmonary hypertension. *Lancet Respir Med*. 2016;4(4):306-22
4. Wittes J. On looking at subgroups. *Circulation*. 2009;119(7):912-5.

5. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357(21):2189-94.
6. Koch A, Framke T. Reliably basing conclusions on subgroups of randomized clinical trials. *J Biopharm Stat*. 2014;24(1):42-57.
7. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192(32):E901-6.
8. Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, Fraga-Fuentes MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, et al. Checklist for Clinical Applicability of Subgroup Analysis. *J Clin Pharm Ther*. 2020 Jun;45(3):530-8
9. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
10. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.
11. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311(4):405-11.
12. Sun X, Briel M, Busse JW, Akl EA, You JJ, Mejza F, et al. Subgroup Analysis of Trials Is Rarely Easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. *Trials*. 2009;10:101.
13. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ*. 2012;344:e1553.
14. Báez-Gutiérrez N, Rodríguez-Ramallo H, Flores-Moreno S, Abdel-Kader Martín L., et al. Subgroup analysis in haematologic malignancies phase III clinical trials: A systematic review. *Br J Clin Pharmacol*. 2021;87(7):2635-44.
15. Saragiotto BT, Maher CG, Moseley AM, Yamato TP, Koes BW, Sun X, et al. A systematic review reveals that the credibility of subgroup claims in low back pain trials was low. *J Clin Epidemiol*. 2016;79:3-9.
16. Paquette M, Alotaibi AM, Nieuwlaat R, Santesso N, Mbuagbaw L. A meta-epidemiological study of subgroup analyses in cochrane systematic reviews of atrial fibrillation. *Syst Rev*. 2019;8(1):241
17. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg*. 2010;8:336–41.
18. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
19. Primo J, Escrig J. MetaSurv: Excel calculator for survival meta-analyses. 2008. Available in: <http://www.redcaspe.org/herramientas/descargas/MetaSurv.xls>
20. Galiè N, Humbert M, Vachiéry JL, Vizza CD, Kneussl M, Manes A, et al. Effects of beraprost sodium, an oral prostacyclin analogue, in patients with pulmonary arterial hypertension: a randomized, double-blind, placebo-controlled trial. *J Am Coll Cardiol*. 2002 May 1;39(9):1496-502.

21. Olschewski H, Simonneau G, Galiè N, Higenbottam T, Naeije R, Rubin LJ, Nikkho S, et al. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med.* 2002 ;347(5):322-9
22. Simonneau G, Rubin LJ, Galiè N, Barst RJ, Fleming TR, Frost AE, et al. Addition of sildenafil to long-term intravenous epoprostenol therapy in patients with pulmonary arterial hypertension: a randomized trial. *Ann Intern Med.* 2008;149(8):521-30.
23. Benza RL, Barst RJ, Galie N, Frost A, Girgis RE, Highland KB, et al Sitaxsentan for the treatment of pulmonary arterial hypertension: a 1-year, prospective, open-label observation of outcome and survival. *Chest.* 2008;134(4):775-82.
24. Barst RJ, Oudiz RJ, Beardsworth A, Brundage BH, Simonneau G, Ghofrani HA, et al. Pulmonary Arterial Hypertension and Response to Tadalafil (PHIRST) Study Group. Tadalafil monotherapy and as add-on to background bosentan in patients with pulmonary arterial hypertension. *J Heart Lung Transplant.* 2011;30(6):632-43.
25. Ghofrani HA, Galiè N, Grimminger F, Grünig E, Humbert M, Jing ZC, et al. Riociguat for the treatment of pulmonary arterial hypertension. *N Engl J Med.* 2013;369(4):330-40.
26. Tapson VF, Jing ZC, Xu KF, Pan L, Feldman J, Kiely DG, et al. Oral treprostinil for the treatment of pulmonary arterial hypertension in patients receiving background endothelin receptor antagonist and phosphodiesterase type 5 inhibitor therapy (the FREEDOM-C2 study): a randomized controlled trial. *Chest.* 2013;144(3):952-8.
27. Hoendermis ES, Liu LC, Hummel YM, van der Meer P, de Boer RA, Berger RM, van Veldhuisen DJ, et al. Effects of sildenafil on invasive haemodynamics and exercise capacity in heart failure patients with preserved ejection fraction and pulmonary hypertension: a randomized controlled trial. *Eur Heart J.* 2015;36(38):2565-73.
28. Chang HJ, Song S, Chang SA, Kim HK, Jung HO, Choi JH, et al. Efficacy and Safety of Udenafil for the Treatment of Pulmonary Arterial Hypertension: a Placebo-controlled, Double-blind, Phase IIb Clinical Trial. *Clin Ther.* 2019;41(8):1499-507.
29. McLaughlin V, Channick RN, Ghofrani HA, Lemarié JC, Naeije R, Packer M, et al. Bosentan added to sildenafil therapy in patients with pulmonary arterial hypertension. *Eur Respir J.* 2015;46(2):405-13.
30. Vizza CD, Jansa P, Teal S, Dombi T, Zhou D. Sildenafil dosed concomitantly with bosentan for adult pulmonary arterial hypertension in a randomized controlled trial. *BMC Cardiovasc Disord.* 2017;17(1):239.
31. Izem R, Liao J, Hu M, Wei Y, Akhtar S, Wernecke M, et al. Comparison of propensity score methods for pre-specified subgroup analysis with survival data. *J Biopharm Stat.* 2020;30(4):734-51.
32. Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med.* 2020;172(1):35-45.
33. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis) uses of baseline data in clinical trials. *Lancet.* 2000;355 (9209):1064-9.

34. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol.* 2004;57(3):229-36.
35. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy. *J Clin Epidemiol.* 1998;51(4):297-9.
36. Pharoah P. Response to "Credibility of claims of subgroup effects in randomised controlled trials: systematic review". *BMJ* 2012;344:e1553.
37. Sainani K. Misleading comparisons: the fallacy of comparing statistical significance. *PM&R.* 2010;2(6):559-62.
38. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials. *JAMA Intern Med.* 2017 Apr 1;177(4):554-60.
39. Khan MS, Khan MAA, Irfan S, Siddiqi TJ, Greene SJ, Anker SD, et al. Reporting and interpretation of subgroup analyses in heart failure randomized controlled trials. *ESC Heart Fail.* 2021;8(1):26-36.
40. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomized controlled trials: systematic review. *BMJ.* 2011;342:d1569.
41. Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blümle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. *BMJ.* 2014;349:4921.
42. Gabler NB, Duan N, Ranases E, Suttner L, Ciarametaro M, Cooney E, et al. No improvement in the reporting of clinical trial subgroup effects in high-impact general medical journals. *Trials.* 2016;17(1):320.
43. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol.* 2012;65(6):602-9.
44. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the radiation therapy oncology group. *BMJ.* 2004;328(7430):22-4.
45. Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ.* 2008;337:a2299.
46. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
47. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-9

48. Chan AW, Krleza-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
49. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252-60.
50. Von Elm E, Röllin A, Blümle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly*. 2008;138(13-14):197-203.
51. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977-84.
52. Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by The Lancet. *Lancet* 2008;372:201.
53. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.gov: a cross-sectional analysis. *PLoS Med* 2009;6: e1000144.
54. Li G, Abbade LPF, Nwosu I, Jin Y, Leenus A, Maaz M, et al. A systematic review of comparisons between protocols or registrations and full reports in primary biomedical research. *BMC Med Res Methodol*. 2018;18(1):9.
55. Vidic A, Chibnall JT, Goparaju N, Hauptman PJ. Subgroup analyses of randomized clinical trials in heart failure: facts and numbers. *SC Heart Fail*. 2016;3(3):152-7.

Figures

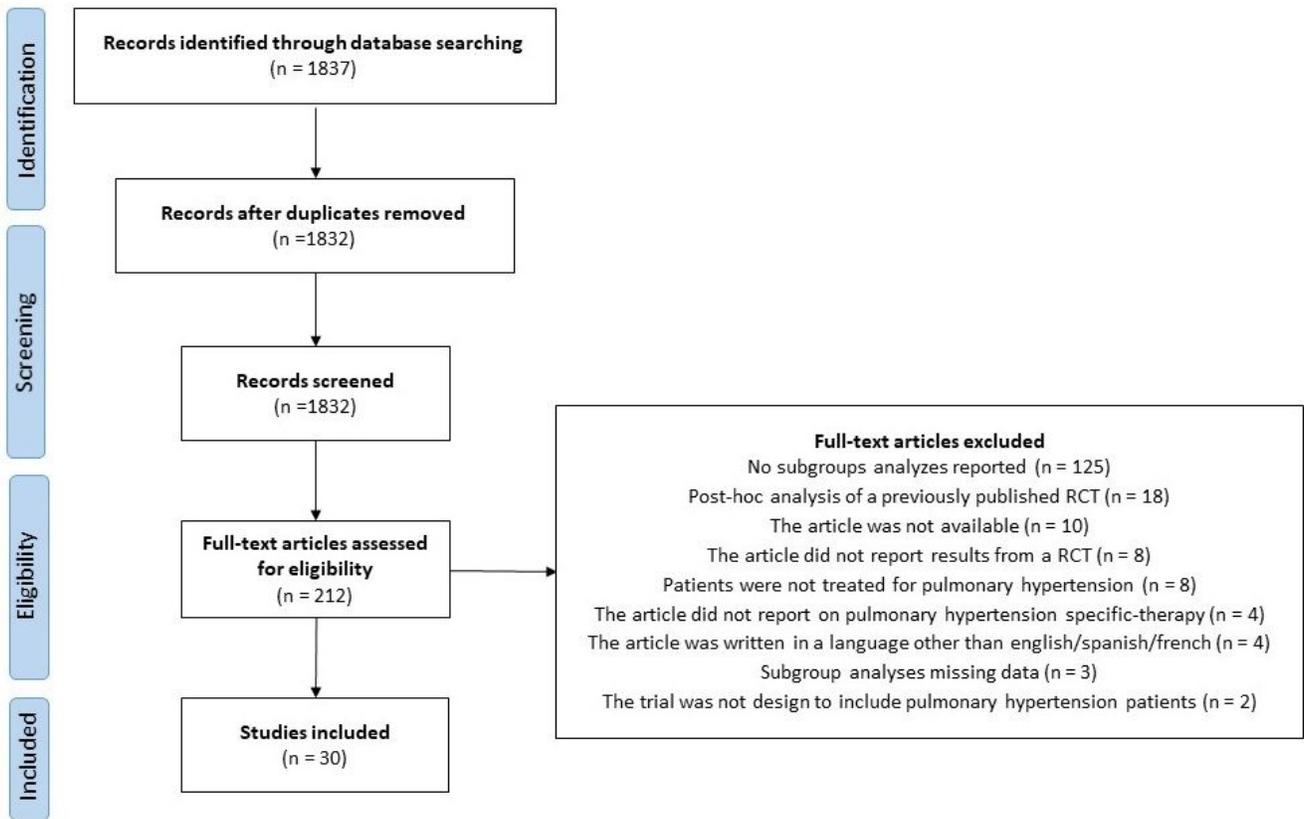


Figure 1

Flowchart of screening of randomized clinical trials included in this analysis.

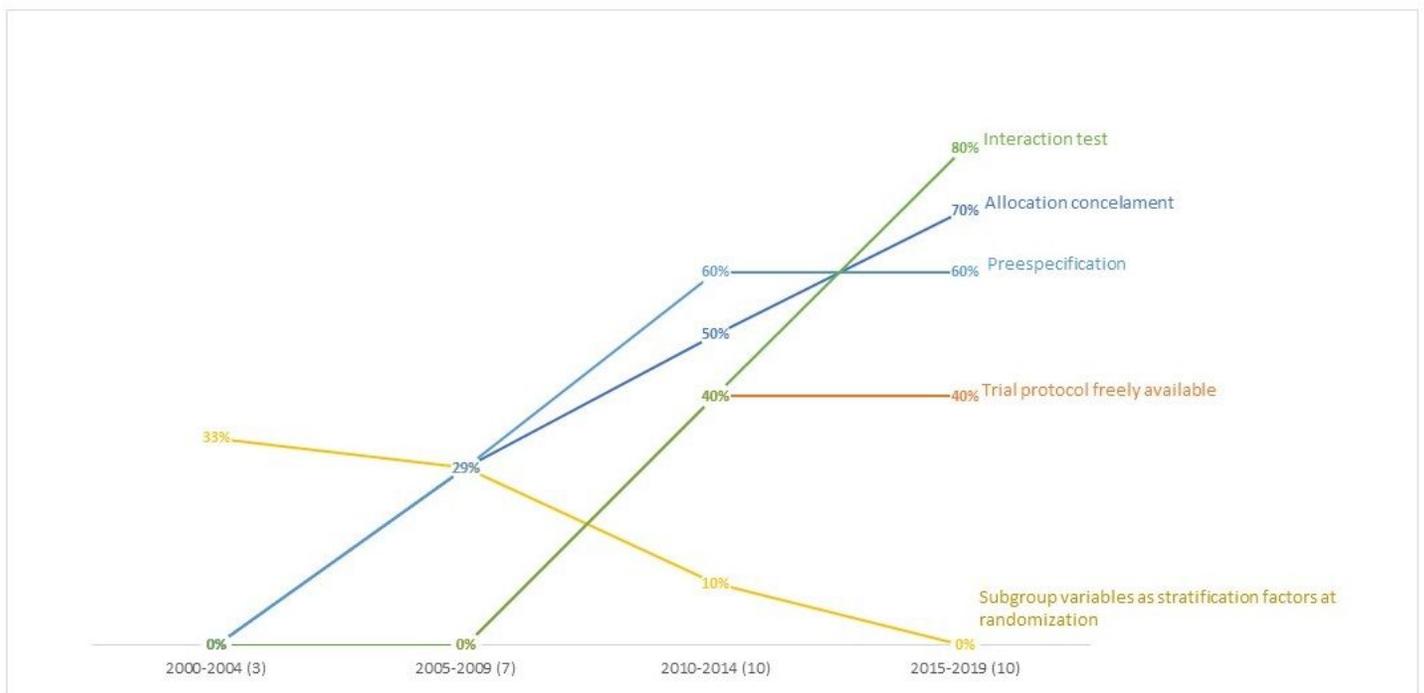


Figure 2

Evolution of subgroup analyses reporting.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf.pdf](#)
- [Additionalfile2.pdf.pdf](#)
- [Additionalfile3.pdf.pdf](#)
- [Additionalfile4.pdf.pdf](#)
- [Additionalfile5.pdf.pdf](#)