

ORIGINAL RESEARCH

# Subgroup analyses in randomized phase III trials of systemic treatments in patients with advanced solid tumours: a systematic review of trials published between 2017 and 2020

C. Paratore<sup>1,2†</sup>, C. Zichi<sup>2†</sup>, M. Audisio<sup>1</sup>, M. Bungaro<sup>1</sup>, A. Caglio<sup>2</sup>, R. Di Liello<sup>3</sup>, T. Gamba<sup>2</sup>, P. Gargiulo<sup>3</sup>, A. Mariniello<sup>1,4</sup>, M. L. Reale<sup>1</sup>, F. Perrone<sup>3</sup> & M. Di Maio<sup>2\*</sup>

<sup>1</sup>Department of Oncology, University of Turin, San Luigi Gonzaga Hospital, Orbassano; <sup>2</sup>Department of Oncology, University of Turin, Ordine Mauriziano Hospital, Turin; <sup>3</sup>Clinical Trials Unit, Istituto Nazionale Tumori – IRCCS – Fondazione Giovanni Pascale, Naples, Italy; <sup>4</sup>Emory Vaccine Center, Emory University School of Medicine, Atlanta, USA



Available online 12 October 2022

**Background:** Subgroup analyses of randomized controlled trials are very common in oncology; nevertheless, the methodological approach has not been systematically evaluated. The present analysis was conducted with the aim of describing the prevalence and methodological characteristics of the subgroup analyses in randomized controlled trials in patients with advanced cancer.

**Methods:** A systematic literature search using PubMed was carried out to identify all phase III randomized controlled trials conducted in adult patients affected by locally advanced or metastatic solid tumours, published between 2017 and 2020.

**Results:** Overall, 253 publications were identified. Subgroup analyses were reported in 217 (86%) publications. A statistically significant association of presence of subgroup analysis with study sponsor was observed: subgroup analyses were reported in 157 (94%) for-profit trials compared with 60 (70%) non-profit trials ( $P < 0.001$ ). Description of the methodology of subgroup analysis was completely lacking in 82 trials (38%), only cited without methodological details in 100 trials (46%) and fully described in 35 trials (16%). Forest plot of subgroup analyses for the primary endpoint was available in 195 publications (77%). Among publications with reported forest plots, the median number of subgroups for primary endpoint was 19 (range 6–78). Out of the 217 publications with subgroup analyses, authors discuss the heterogeneity of treatment effect among different subgroups in 173 publications (80%), although a formal test for interaction for subgroup analysis of primary endpoint was reported for at least one variable only in 60 publications (28%). Correction for multiplicity was explicitly carried out only in nine trials (4%).

**Conclusions:** The very high prevalence of subgroup analyses in published papers, together with their methodological weaknesses, makes advisable an adequate education about their correct presentation and correct reading. More attention about proper planning and conduction of subgroup analysis should be paid not only by readers, but also by authors, journal editors and reviewers.

**Key words:** subgroup analyses, systematic review, advanced solid tumours, methodology, heterogeneity of treatment effect

## INTRODUCTION

Subgroup analyses of randomized controlled trials are very common in oncology.<sup>1,2</sup> No doubt that—especially in the era of personalized medicine—it appears legitimate to ask, in

addition to the main result obtained in the overall study population, whether the efficacy of the experimental treatment is influenced by some specific characteristics of the patient or of the disease. Within a positive study, this could help to better define the target population, avoiding toxicity (and costs) of treatment in subjects who would not derive benefit. In the context of a negative study, however, subgroup analyses could be useful in avoiding ‘throwing the baby out with the bath water’, by identifying certain groups of patients in whom the experimental treatment appears to work.

Due to power and multiplicity of statistical tests, however, subgroup analyses are inherently associated with a well-established risk of spurious effects, which means

\*Correspondence to: Prof. Massimo Di Maio, Department of Oncology, University of Turin, Division of Medical Oncology, Ordine Mauriziano Hospital, Via Magellano 1, Turin 10128, Italy. Tel: +39-011-5082032

E-mail: [massimo.dimaio@unito.it](mailto:massimo.dimaio@unito.it) (M. Di Maio).

Twitter handle: [@MassimoDiMaio75](https://twitter.com/MassimoDiMaio75)

<sup>†</sup>Equally contributing first authors.

2059-7029/© 2022 The Authors. Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

false-negative and/or false-positive results.<sup>3-5</sup> If the aim is to identify patients who do not benefit from a treatment which showed superiority in the whole study population, testing for statistical significance of treatment comparison within each subgroup with subgroup-specific *P* values can be misleading, because the limited number of subjects in each subgroup is obviously associated with a lower statistical power.<sup>6</sup> If the aim is to identify subpopulations of patients who seem to benefit from a treatment which did not meet the main study endpoint, however, subgroup analyses can be misleading, testing repeatedly, within every subgroup, the same null hypothesis unsuccessfully tested and rejected in the main analysis.<sup>7</sup> In the latter situation, a positive result in a subgroup within a negative trial should not support treatment adoption: at best, that result should be hypothesis-generating, representing the rationale for further research.

Despite these caveats, subgroup analyses are included in the presentation of many studies, often affecting the overall interpretation of the result.<sup>8</sup>

The present analysis was conducted with the aim of describing the prevalence and technical characteristics of the subgroup analyses in randomized studies recently published in oncology. Furthermore, we evaluated the emphasis given by the authors to the results observed in the subgroups, and the incidence of regulatory decisions based on the results of subgroup analyses.

## METHODS

### Selection of publications

A systematic literature search using PubMed was carried out in May 2021 to identify all randomized phase III trials conducted in adult patients affected by locally advanced or metastatic solid tumours published between 1<sup>st</sup> January 2017 and 31<sup>st</sup> December 2020.

We considered only trials testing systemic anticancer treatments (chemotherapy, immunotherapy, target therapy and hormonal treatment), excluding trials testing supportive care drugs, non-pharmacological interventions and prevention strategies. Trials conducted in hematologic malignancies, in paediatric patients as well as trials conducted in early stages of disease (testing adjuvant/neoadjuvant treatment) were excluded. Publications in language other than English were excluded. Fields: *random\* AND cancer AND ("exten\*" OR "previously treated" OR "stage IV" OR "unresectable" OR advanced OR recurren\* OR metast\*) AND ("2017/01/01"[Date - Publication]: "3000"[Date - Publication])*; Filters applied: Article type (Clinical Trial); Publication date (2017-2020).

### Data collection

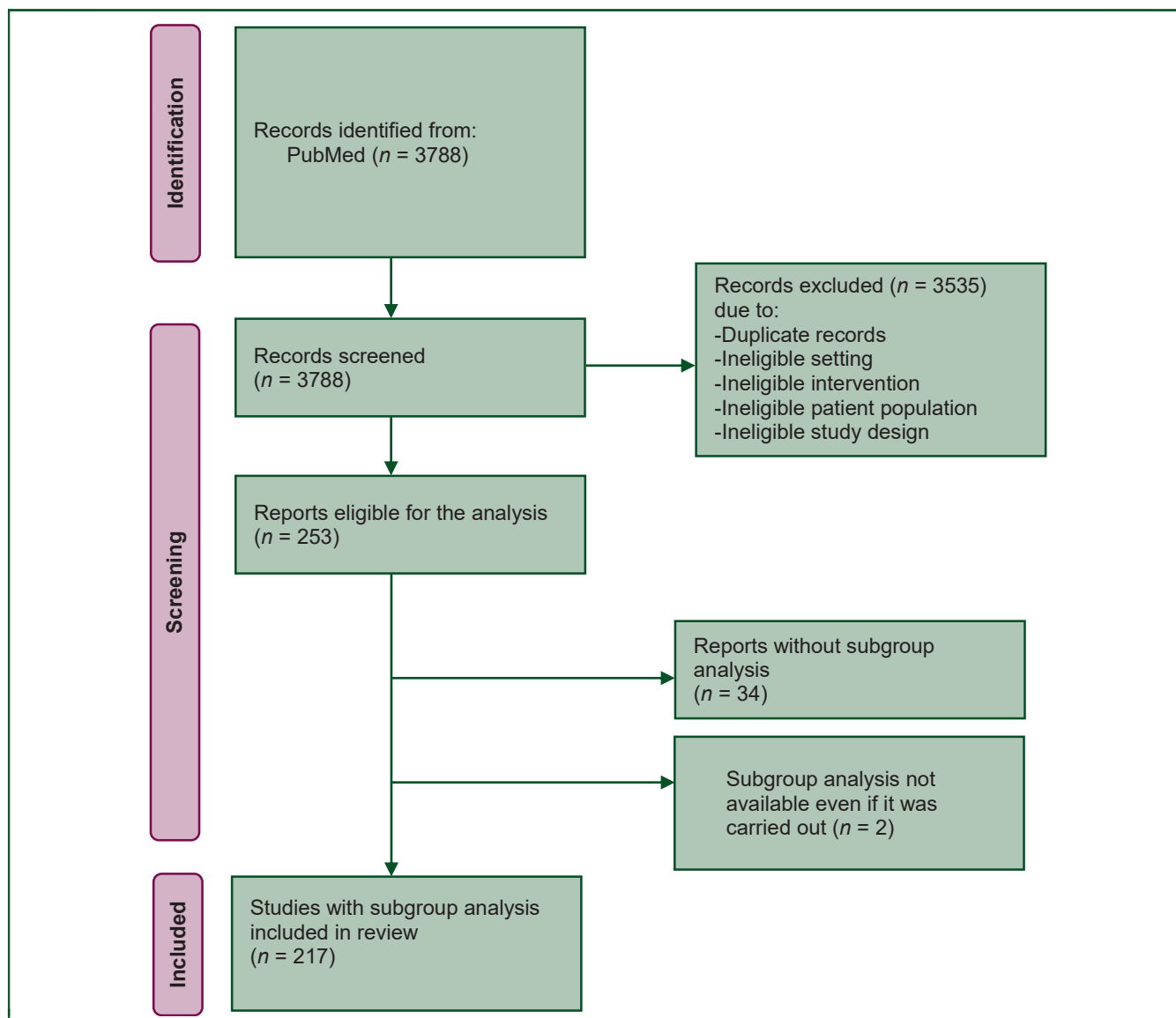
An electronic database was generated to collect data, with one record for each eligible paper. Each selected paper was reviewed by a young investigator, discussing doubts and controversies with a senior investigator.

For each study, information collected was about publication (date of publication, primary publication DOI, journals' impact factor (IF), availability of supplementary material and/or study protocol) and clinical trial, including disease setting (locally advanced; first-line for metastatic disease; second-line or further treatment of metastatic disease), type of primary tumour (breast; thoracic; gastrointestinal; urological; gynaecological; other cancers), and study sponsor (profit; non-profit). Namely, trials were considered as profit when sponsored by the drug company for commercial purposes and as non-profit when sponsored by an academic institution or a cooperative group, even when receiving drug supply and/or economic support from one or more drug companies. As for the type of experimental treatment, we classified them into one of four main groups, in the following conventional order of dominance in case of combination treatments: immunotherapy, targeted therapy, chemotherapy, hormonal treatment. According to the IF, the papers were divided into three categories, (low IF, intermediate IF and high IF), using as cut-off the 25<sup>o</sup> and 75<sup>o</sup> percentile.

For trials including subgroup analysis in the publication, we collected further details, namely the presence of forest plots (both for primary and secondary endpoints), the number of variables (e.g. sex) and the number of subgroups (e.g. men, women) reported in the plots, the presence of a test for interaction, the presence of a *P* value for each subgroup, the presence of correction for multiplicity of tests. Of note, we planned to describe the concordance between the analyses declared in the protocol and the subgroup analyses reported in the article, but we were not able to carry out this classification optimally because a full trial protocol was available only in slightly more than half of the publications. Further information was collected about the inclusion of details about subgroup analysis in the Abstract, in the Methods, Results and Discussion section of the publication. Studies were classified according to the description of subgroup analysis (concise or detailed) and the conclusions (balanced comments or excessive emphasis on subgroup analysis, according to the subjective impression of the reader). Finally, information about any drug approval by the US Food and Drug Administration (FDA) and/or European Medicines Agency (EMA) specifically based on subgroup analysis was collected.

### Statistical analysis

Analyses were mostly descriptive. The chi-square test was applied to determine the existence of a statistically significant association between the presence of subgroup analysis and main characteristics of study publication: year, study sponsor, type of primary tumour, disease setting and type of experimental treatment. The association of variables related to subgroup analysis and journal IF was tested by the chi-square test for linear trend (categorical variables) or the Jonckheere–Terpstra test (numerical variables). A *P* value <0.05 was considered statistically significant. Considering



**Figure 1. PRISMA flow diagram.** From: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>. For more information, visit: <http://www.prisma-statement.org/>. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

the descriptive and exploratory intent of the analysis, no adjustment for multiple testing was applied.

All analyses were carried out with IBM SPSS Statistics for Windows, version 27.0.

## RESULTS

### Characteristics of eligible trials

Overall, 253 publications of randomized phase III trials in patients with advanced solid tumours treated with systemic therapy were identified (Figure 1), with the highest number of publications in 2017 (87, 34.4%) and the lowest number in 2020 (44, 17.4%). Median IF was 32.956 (25<sup>o</sup> percentile 13.930; 75<sup>o</sup> percentile 44.544). The full trial protocol was available for slightly more than half of the publications (141, 55.7%). The main characteristics of the 253 eligible trials are summarized in Table 1. Study sponsor was a pharmaceutical

company in 167 (66.0%) trials and a non-profit organization in the remaining 86 (34.0%) trials. Most common types of tumours were gastrointestinal cancers (74 trials, 29.2%) followed by thoracic cancers (68, 26.9%), urological cancers (36, 14.2%) and breast cancer (32, 12.6%). Most trials were conducted in the first-line setting for metastatic disease (166 trials, 65.6%). Most frequent experimental treatments were targeted agents (119 trials, 47.0%) and chemotherapy (78 trials, 30.8%), followed by immunotherapy (45 trials, 17.8%) and hormonal treatment (11 trials, 4.3%). In nine trials, experimental treatment was a combination of target therapy plus hormonal treatment or chemotherapy, or a combination of immunotherapy plus chemotherapy.

### Characteristics of trials reporting subgroup analysis

Subgroup analyses were reported in 217 (85.8%) publications. The main characteristics of trials according to

Table 1. Characteristics of eligible trials and of trials with and without subgroup analysis				
	All eligible trials (n = 253)	Trials with subgroup analysis (n = 217)	Trials without subgroup analysis (n = 36)	P value (chi-square)
<b>Year of publication</b>				0.27
2017	87	72 (82.8%)	15 (17.2%)	
2018	62	55 (88.7%)	13 (11.3%)	
2019	60	49 (81.7%)	11 (18.3%)	
2020	44	41 (93.2%)	3 (6.8%)	
<b>Study sponsor</b>				<0.001
For profit	167	157 (94.0%)	10 (6.0%)	
Non profit	86	60 (69.8%)	26 (30.2%)	
<b>Disease</b>				0.41
Breast cancer	32	29 (90.6%)	3 (9.4%)	
Thoracic cancer	68	59 (86.8%)	9 (13.2%)	
GI cancers	74	63 (85.1%)	11 (14.9%)	
GU cancers	36	30 (83.3%)	6 (16.7%)	
Gyn cancers	13	13 (100%)	0	
Other cancers	30	23 (76.7%)	7 (23.3%)	
<b>Setting</b>				0.11
Locally advanced	7	6 (85.7%)	1 (14.3%)	
First line metastatic	166	137 (82.5%)	29 (17.5%)	
Second or subsequent lines	80	74 (92.5%)	6 (7.5%)	
<b>Experimental treatment<sup>a</sup></b>				0.007
Chemotherapy	78	60 (76.9%)	18 (23.1%)	
Hormonal treatment	11	11 (100%)	0	
Targeted agent	119	102 (85.7%)	17 (14.3%)	
Immunotherapy	45	44 (97.8%)	1 (2.2%)	
<b>Impact factor</b>				<0.001
Low IF	61	40 (65.6%)	21 (34.4%)	
Intermediate IF	122	111 (91.0%)	11 (9.0%)	
High IF	70	66 (94.3%)	4 (5.7%)	

Bold values correspond to a *P* value with statistically significant results.

GI, gastrointestinal; GU, genitourinary; Gyn, gynaecological; IF, impact factor.

<sup>a</sup>There were nine trials with combination experimental treatments. In these cases, we classified trials into one of four main groups, in the following conventional order of dominance: immunotherapy, targeted therapy, chemotherapy, hormonal treatment.

presence or absence of subgroup analysis are detailed in Table 1. There was no significant difference among the 4 years ( $P = 0.27$ ) in terms of presence of subgroup analyses, whereas a statistically significant association with study sponsor was observed: namely, subgroup analyses were reported in 157 (94.0%) for-profit trials compared with 60 (69.8%) non-profit trials ( $P < 0.001$ ). The proportion of trials including subgroup analysis was significantly lower in publications with lower IF ( $P < 0.001$ ).

There was no significant association with different types of tumours ( $P = 0.41$ ) or different treatment settings ( $P = 0.11$ ). There was a statistically significant association of presence of subgroup analysis with the type of treatment ( $P = 0.007$ ); namely, subgroup analysis was found in 100% of 11 trials testing hormonal treatments, in 97.8% of 45 trials testing immunotherapy, in 85.7% of 119 trials testing target therapy and in 76.9% of 78 trials testing chemotherapy.

### Statistical details

In 20 trials (8%) the primary analysis was planned by protocol to be done within a subgroup of the intention-to-treat population. Namely, the primary endpoint was assessed in subgroups defined by programmed death-ligand 1 (PD-L1) expression levels in 11 trials (55%), by molecular disease characteristics in 5 trials (25%), by histological features in 3

trials (15%) and by prognostic categories in 1 trial (5%). The details of the 20 trials are summarized in Table 2.

Description of the methodology of subgroup analysis was completely lacking in 82/217 papers (37.8%), only cited without methodological details in 100 (46.1%) and fully described in 35 (16.1%). Although, due to the unavailability of full protocol, in many cases we were not able to classify subgroup analyses in pre-planned, pre-specified and *post hoc* analyses, the vast majority of subgroup analyses included in the publications were not explicitly pre-planned.

As detailed in Table 3, a forest plot of subgroup analyses for the primary endpoint was available in 195/217 publications (89.9%), reported mostly in full article  $\pm$  supplementary material (81.0%), whereas in 19.0% of the cases it was reported in supplementary material only. A forest plot of secondary endpoints was found in 58 publications (26.7%), in the main article (62.1%) or in the supplementary material (38.9%). Among publications with a reported forest plot, we observed a median of nine variables (range three to nine) and a median of 19 subgroups (range 6-78) for primary endpoint, with similar data for secondary endpoints.

Out of the 217 publications with subgroup analyses, authors discuss the presence or absence of heterogeneity of treatment effect among different subgroups in 173 publications (79.7%). The test for interaction for subgroup analysis of primary endpoint was reported for at least one

**Table 2. Characteristics of trials where the primary analysis was planned by protocol to be done within a subgroup of the intention-to-treat population**

Author, year	Setting	Experimental treatment	Control treatment	Primary EP(s)	Subgroup considered for primary EP	Forest plot for primary EP	Interaction test for primary EP	P value for each subgroup for primary EP	Results of clinical trial (primary EP met)
Lee, 2017 <sup>9</sup>	NSCLC, first line	Paclitaxel + gemcitabine or pemetrexed	Cisplatin + gemcitabine or pemetrexed	OS	ERCC1+/-	No	Yes	Yes	Negative
Rittmeyer, 2017 <sup>10</sup>	NSCLC, second-third line	Atezolizumab	Docetaxel	OS	PD-L1-positive subgroups (TC1/2/3 or IC1/2/3)	Yes	Yes	No	Positive
Bellmunt, 2017 <sup>11</sup>	Urothelial carcinoma, second line	Pembrolizumab	Docetaxel or paclitaxel or vinflunine	PFS, OS	PD-L1 $\geq$ 10%	Yes	No	No	Positive
Shah, 2017 <sup>12</sup>	Gastroesophageal cancer, first line	mFOLFOX6 + onartuzumab	mFOLFOX6	OS	MET 2+/3+	Yes	No	No	Negative
Herbst, 2017 <sup>13</sup>	NSCLC, first line	Cetuximab, carboplatin, paclitaxel +/- bevacizumab	Carboplatin, paclitaxel +/- bevacizumab	PFS, OS	EGFR FISH+	Yes	No	Yes	Negative
Motzer, 2018 <sup>14</sup>	Renal cell carcinoma, first line	Nivolumab + ipilimumab	Sunitinib	PFS, OS, ORR	Intermediate and poor risk	Yes	No	No	Positive
Socinski, 2018 <sup>15</sup>	NSCLC, first line	Atezolizumab + paclitaxel + carboplatin +/- bevacizumab	Paclitaxel + carboplatin + bevacizumab	PFS, OS	High expression of an effector T-cell (Teff) gene signature	Yes	No	No	Positive
Hellmann, 2018 <sup>16</sup>	NSCLC, first line	Nivolumab + ipilimumab	CT based on tumour histologic type	PFS, OS	TMB $\geq$ 10 mutations per Mb, PD-L1 expression levels	Yes	No	No	Positive
Schmid, 2018 <sup>17</sup>	Triple-negative breast cancer, first line	Atezolizumab + nab-paclitaxel	Placebo + nab-paclitaxel	PFS, OS	PD-L1 $\geq$ 1%	Yes	No	No	Negative
Motzer, 2019 <sup>18</sup>	Renal cell carcinoma, first line	Avelumab + axitinib	Sunitinib	PFS, OS	PD-L1 $\geq$ 1%	Yes	No	No	Positive
Mok, 2019 <sup>19</sup>	NSCLC, first line	Pembrolizumab	Platinum-based CT	OS	PD-L1 $\geq$ 50%, 20%, 1%	Yes	No	No	Positive
Rini, 2019 <sup>20</sup>	Renal cell carcinoma, first line	Atezolizumab + bevacizumab	Sunitinib	PFS, OS	PD-L1 $\geq$ 1%	Yes	No	No	Positive
West, 2019 <sup>21</sup>	NSCLC, first line	Atezolizumab + carboplatin + nab-paclitaxel	Carboplatin + nab-paclitaxel	PFS, OS	EGFR WT and ALK NR	Yes	No	No	Positive
González-Martín, 2019 <sup>22</sup>	Ovarian cancer, maintenance after first line	Niraparib	Placebo	PFS	HRD	Yes	No	No	Positive

*Continued*

**Table 2. Continued**

Author, year	Setting	Experimental treatment	Control treatment	Primary EP(s)	Subgroup considered for primary EP	Forest plot for primary EP	Interaction test for primary EP	P value for each subgroup for primary EP	Results of clinical trial (primary EP met)
Tap, 2020 <sup>23</sup>	Soft tissue sarcoma	Doxorubicin + olaratumab	Doxorubicin + placebo	OS	Leiomyosarcoma	Yes	No	No	Negative
Galsky, 2020 <sup>24</sup>	Urothelial carcinoma, first line	Atezolizumab +/- platinum-based CT	Platinum based-CT	PFS, OS	PD-L1-positive subgroups (IC 2/3)	Yes	No	No	Positive
Powles, 2020 <sup>25</sup>	Urothelial carcinoma, maintenance after first line	Avelumab	Best supportive care	OS	PD-L1-positive subgroup	Yes	No	No	Positive
Shitara, 2020 <sup>26</sup>	Gastric cancer, first line	Pembrolizumab +/- standard CT	Placebo + standard CT	PFS, OS	PD-L1 ≥1%, 10%	Yes	No	No	Negative
Herbst, 2020 <sup>27</sup>	NSCLC, first line	Atezolizumab	Platinum-based CT	OS	PD-L1 ≥50%, 5%, 1%	Yes	No	No	Positive
Powles, 2020 <sup>28</sup>	Urothelial carcinoma, first line	Durvalumab +/- tremelimumab	Platinum-based CT	OS	High PD-L1 expression	Yes	No	No	Negative

ALK, anaplastic lymphoma kinase; CT, chemotherapy; EGFR, epidermal growth factor receptor; EP, endpoint; ERCC1, excision repair cross complementing group 1; HRD, homologous-recombination deficiency; IC, immune cells; NR, not rearranged; NSCLC, non-small-cell lung cancer; ORR, objective response rate; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; TC, tumour cells; TMB, tumour mutational burden; WT, wild type.

variable only in 60 publications (27.6%), however, whereas 21 publications (9.7%) reported the test for interaction for secondary endpoints. An interaction test for primary and secondary endpoints was reported in a very low proportion of papers even in journals with high IF (in 12.1% and 6.1% for primary and secondary endpoints, respectively). *P* value for each subgroup, however, was reported in 36 publications (16.6%) for primary endpoint and in 14 publications (6.5%) per secondary endpoints (Table 3), and this was more frequent in journals with lower IF. Correction for multiplicity was explicitly carried out only in nine trials (4.1%).

**Reporting subgroup analysis**

Subgroup analyses were mentioned in the Abstract in 41 publications (18.9%), in the Results section in 205 publications (94.5%) and in the Discussion or Conclusions in 174 publications (80.2%) (Table 3). In the section of Results or Discussion/Conclusion, according to our subjective judgement, authors focused excessively on treatment effect in different subgroups in 21 publications (9.7%). In detail, 9 of these 21 trials were positive for the primary endpoint analysis, whereas the remaining 12 trials failed to reach the primary endpoint. In 94 publications (43.3%), according to our subjective judgement, authors’ comments on subgroups were balanced and/or readers were invited to cautiously interpret the results of subgroup analysis and to explore their potential role in subsequent studies.

**Subgroup analyses and drug approvals by regulatory agencies**

Overall, out of the treatments tested in the eligible trials, we found eight drug approvals by the FDA and/or EMA based on the results of subgroup analyses. For instance, the FDA approved atezolizumab plus nab-paclitaxel in advanced triple-negative breast cancer with positive PD-L1, based on the results of the IMpassion130 trial.<sup>17</sup> In that case, the analysis of the subgroup with positive PD-L1 was formally pre-planned, although, according to the original study design, overall survival in the subgroup was to be tested hierarchically only in case of a statistically significant result in the intention-to-treat population. This was formally not the case, however, the statistically significant benefit in progression-free survival, the trend in overall survival improvement in the intention-to-treat population and the more convincing overall survival benefit in the PD-L1-positive subgroup led the regulatory agency to approve the experimental treatment in this subgroup. As an example of approval decision based on a *post hoc* subgroup analysis, durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer (NSCLC) was approved by the EMA only in patients with PD-L1 expression level ≥1%, despite the fact that this PD-L1 expression cut-off was not pre-planned.<sup>29</sup> A complete list of drug approvals taking into account subgroup analyses are detailed in Table 4.

<b>Table 3. Details of forest plots and subgroup analysis and frequency of reporting of subgroup analysis in the different sections of the publication and according to impact factor</b>					
	<b>Trials with subgroup analysis (n = 217)</b>	<b>Low IF</b>	<b>Intermediate IF</b>	<b>High IF</b>	<b>P value</b>
<b>Forest plot for the primary endpoint</b>	195/217 (89.9%)	28/40 (70.0%)	102/111 (91.9%)	65/66 (98.5%)	<b>P &lt; 0.001</b>
In the main article	158/195 (81.0%)	23/28 (82.1%)	84/102 (82.4%)	51/65 (78.5%)	
In the supplementary material only	37/195 (19.0%)	5/28 (17.9%)	18/102 (17.6%)	14/65 (21.5%)	
<b>Forest plot for the secondary endpoint</b>	58/217 (26.7%)	13/40 (32.5%)	30/111 (27.0%)	15/66 (22.7%)	<b>P = 0.27</b>
In the main article	36/58 (62.1%)	10/13 (76.9%)	18/30 (60.0%)	8/15 (53.3%)	
In the supplementary material only	22/58 (38.9%)	3/13 (23.1%)	12/30 (40.0%)	7/15 (46.7%)	
<b>Number of variables</b>					
Primary endpoint: median (range)	9 (3-19)	7 (3-14)	9 (3-19)	9 (4-19)	<b>P = 0.19</b>
Secondary endpoint: median (range)	8.5 (1-19)	8 (1-14)	8.50 (3-19)	9 (1-19)	<b>P = 0.98</b>
<b>Number of subgroups</b>					
Primary endpoint: median (range)	19 (6-78)	15.5 (6-30)	20 (6-78)	19 (8-38)	<b>P = 0.21</b>
Secondary endpoint: median (range)	20 (2-43)	20 (2-29)	20 (6-43)	21 (2-31)	<b>P = 0.79</b>
<b>Test for interaction</b>					
Primary endpoint, all trials	60/217 (27.6%)	9/40 (22.5%)	43/111 (38.7%)	8/66 (12.1%)	<b>P = 0.07</b>
Primary endpoint, only trials with forest plot	52/195 (26.7%)	6/28 (21.4%)	39/102 (38.2%)	7/65 (10.8%)	<b>P = 0.03</b>
Secondary endpoint, all trials	21/217 (9.7%)	3/40 (7.5%)	14/111 (12.6%)	4/66 (6.1%)	<b>P = 0.61</b>
Secondary endpoint, only trials with forest plot	17/58 (29.3%)	3/13 (23.1%)	12/30 (40.0%)	2/15 (13.3%)	<b>P = 0.51</b>
<b>P value for each subgroup</b>					
Primary endpoint, all trials	36/217 (16.6%)	12/40 (30.0%)	20/111 (18.0%)	4/66 (6.1%)	<b>P = 0.001</b>
Primary endpoint, only trials with forest plot	29/195 (14.9%)	8/28 (28.6%)	17/102 (16.7%)	4/65 (6.2%)	<b>P = 0.004</b>
Secondary endpoint, all trials	14/217 (6.5%)	4/40 (10.0%)	7/111 (6.3%)	3/66 (4.5%)	<b>P = 0.28</b>
Secondary endpoint, only trials with forest plot	6/58 (10.3%)	1/13 (7.7%)	3/30 (10.0%)	2/15 (13.3%)	<b>P = 0.62</b>
<b>Reporting of subgroup analysis in the different sections of the publications</b>					
Abstract	41/217 (18.9%)	11/40 (27.5%)	17/111 (15.3%)	13/66 (19.7%)	<b>P = 0.46</b>
Results	205/217 (94.5%)	39/40 (97.5%)	106/111 (95.5%)	60/66 (90.9%)	<b>P = 0.13</b>
Discussion/conclusions	174/217 (80.2%)	34/40 (85.0%)	88/111 (79.3%)	52/66 (78.8%)	<b>P = 0.48</b>

Bold values correspond to a P value with statistically significant results.  
IF, impact factor.

## DISCUSSION

This systematic review of randomized controlled trials recently published in oncology showed a very high prevalence of subgroup analyses. Namely, 86% of the eligible publications included some analysis in one or more subgroups, with a particularly high prevalence in trials testing immune checkpoint inhibitors and targeted agents compared with trials testing chemotherapy and in papers with higher IF. From a methodological point of view, there is room for improvement in the conduction and reporting of subgroup analysis: (i) correction of statistical testing for multiplicity is rarely considered; (ii) test for interaction is applied (or at least is reported) only in a minority of cases; (iii) in some cases there is the wrong approach of testing the statistical significance of the difference between treatments (with a P value) within each specific subgroup; (iv) the vast majority of subgroup analysis seems to be conducted *post hoc*, although in most cases it is difficult to understand whether the analyses were pre-planned or at least pre-specified.

Subgroup analyses are unavoidably associated with some increased risk of false-positive and/or false-negative results. Of course, these risks increase with the multiplicity of tests carried out. Interestingly, among publications with a reported forest plot, we counted a median of 19 subgroups (range 6-78) for the primary endpoint, and similar data (median of 20 subgroups, range 2-43) for secondary endpoints. This means that the risk of falsely declaring and

discussing some heterogeneity in treatment effect among different subgroups is more than concrete, mainly because we found that, in the majority of cases, no formal test for interaction is presented, even in journals with high IF. Nevertheless, even when the test for interaction is included, readers should be aware of the risk of a false-negative result (due to the limited statistical power of the test, if the study was not sized to test the interaction) and, however, of the risk of a false-positive result (due to the absence of correction for multiplicity). Furthermore, the widespread use of subgroup analysis in high IF journals may have a serious influence on the scientific community, that requires strict methodological skills to critically evaluate the results. The issue of subgroup analysis in randomized trials published in oncology has been already studied by other authors. In 2015, Zhang and colleagues<sup>1</sup> described subgroup analyses in trials conducted in solid tumours, published between 2011 and 2013, showing that the reporting of subgroup analyses was neither uniform nor complete, with testing of a large number of subgroups, reporting of subgroups without pre-specifications and inadequate use of interaction tests. When commenting on those results, the authors themselves emphasized that an improvement was needed to ensure consistency and to provide critical information for guiding patient care, and Altman<sup>2</sup> suggested that journal editors should implement policies to reduce the risk of publishing misleading results. The indirect comparison of their results with our analysis, however, shows that a

**Table 4. Treatments approved by FDA and/or EMA based on subgroup analysis**

Drug	Pivotal clinical trial	Setting	Disease	Primary endpoint(s)	Pivotal subgroup analysis	Subgroup	Agency
Atezolizumab	Impower 110 <sup>27</sup>	First line	Advanced NSCLC	OS in preplanned subgroup	Preplanned subgroup	PD-L1 $\geq$ 50% of TC or IC $\geq$ 10%	FDA, EMA
Atezolizumab and nab-paclitaxel	Impassion 130 <sup>17</sup>	First line	Advanced TNBC	OS and PFS preplanned subgroup and ITT	Preplanned subgroup	PD-L1 $\geq$ 1%.	FDA, EMA
Durvalumab	PACIFIC trial <sup>29</sup>	Consolidation therapy after CT-RT	Locally advanced NSCLC	OS and PFS in ITT	<i>Post hoc</i> analysis	PD-L1 $\geq$ 1%	EMA
Nivolumab plus ipilimumab	CheckMate 227 <sup>16</sup>	First line	Advanced NSCLC	OS and PFS in preplanned subgroup	Preplanned subgroup	PD-L1 $\geq$ 1%	FDA
Olaparib plus bevacizumab	PAOLA-1 trial <sup>30</sup>	First line maintenance	Advanced ovarian cancer	PFS in ITT	Prespecified subgroups	HRD-positive status (BRCA mutation, and/or genomic instability)	FDA, EMA
Pembrolizumab	KEYNOTE-042 <sup>19</sup>	First line	Advanced NSCLC	OS in preplanned subgroups and ITT	Preplanned subgroups	PD-L1 TPS $\geq$ 50%	FDA, EMA
Pembrolizumab single-agent <sup>a</sup>	KEYNOTE-048 <sup>31</sup>	First line	Advanced HNSCCs	PFS and OS in preplanned subgroups and ITT	Preplanned subgroups	PD-L1 CPS $\geq$ 1 %	FDA
Pembrolizumab with or without platinum and fluorouracil	KEYNOTE-048 <sup>31</sup>	First line	Advanced HNSCCs	PFS and OS in preplanned subgroups and ITT	Preplanned subgroups	PD-L1 CPS $\geq$ 1 %	EMA

CPS, combined positive score; CT-RT, chemo-radiotherapy; IC, infiltrating immune cells; EMA, European Medicines Agency; FDA, Food and Drug Administration; HRD, homologous recombination deficiency; HNSCC, head and neck squamous cell carcinoma; ITT, intention to treat; NSCLC, non-small-cell lung cancer; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; TC, tumour cells; TNBC, triple-negative breast cancer; TPS, tumour proportion score.

<sup>a</sup> Differently from EMA, the combination with platinum and fluorouracil (FU) was approved by FDA for all patients with metastatic head and neck tumours, regardless of PD-L1 level.

definitive improvement in the methodology of subgroup analysis is largely yet to come, at least in terms of clarity in pre-specification and pre-planning of subgroup analyses and in terms of test for interaction.

Our analysis has several limitations. First, it included trials published in a limited time interval (4 years between 2017 and 2020) and this interval is probably too short to capture time trends, if any, in the presence of subgroup analysis and/or in the methodology applied. The period analysed includes very recent trials, however, so our results can be considered a timely picture of this methodological issue. Second, at least in principle, subgroup analyses could be subject to selective reporting bias, and the number of subgroups tested could be even higher than those reported in the publications. Unfortunately, our analysis was based exclusively on the papers and on the study protocol when available, so we had no way of verifying the coherence between the analysis actually carried out and the results presented in the publication. Third, the judgement about the excessive emphasis or the presence of balanced comments on subgroup analyses is a subjective measure, not based on objective parameters. Although the same description could be judged differently by another reader, however, this is a rough measure of how some readers could be misled by some reports of subgroup analyses.

Subgroup analyses should be considered hypothesis-generating more than a definitive demonstration of heterogeneity of treatment effect. One of the trials included in our systematic review can be considered a good example of

this principle.<sup>32</sup> The REACH-2 trial tested the efficacy of the anti-angiogenic ramucirumab compared with placebo as second-line treatment in patients with advanced hepatocellular carcinoma and high levels of alpha-fetoprotein, based on the hypothesis generated by the subgroup analysis of the previous randomized trial.<sup>33</sup> In the first trial, ramucirumab did not meet the primary endpoint in the intention-to-treat population, but subgroup analysis suggested a significant heterogeneity of treatment efficacy according to levels of alpha-fetoprotein. Following this finding, a second trial was carried out, which confirmed the hypothesis and led to regulatory approval in that specific subgroup. We are perfectly aware that in many cases it is not easy to carry out another trial, but at least in the case of a subgroup suggesting a positive treatment effect within a negative trial in the overall population, this approach should be recommended. The debated decision by the EMA of restricting the approval of durvalumab in locally advanced NSCLC to the treatment of cases with positive PD-L1 expression, however, although that subgroup analysis was not pre-planned, is a clear example that, in some cases, even regulatory agencies could decide to assume important decisions on the basis of subgroup analyses.<sup>34</sup> The exception does not invalidate the rule: when evaluating the results of subgroup analyses, caution should be utmost; results should be considered hypothesis-generating; statistical tests should be corrected for multiplicity; tests for interaction, although with limited statistical power, should be reported; consistency of results among different trials, if



available, should be analysed; biological and clinical plausibility of results should matter.

### Conclusion

In conclusion, particularly in the era of precision medicine, subgroup analyses are a legitimate attempt at better tailoring treatment choices. The very high prevalence of these analyses in published papers, together with their methodological weaknesses, however, makes advisable an adequate education about their correct presentation and correct reading. More attention about methodological issues of subgroup analysis should be paid not only by readers, but starting by authors, by journal editors and by reviewers.

### FUNDING

None declared.

### DISCLOSURE

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: MDM reports personal fees from AstraZeneca, Pfizer, Novartis, Roche, Takeda, Janssen, Eisai, Astellas, Merck Sharp & Dohme (MSD), Boehringer Ingelheim, grants from Tesaro-GlaxoSmithKline, outside the submitted work. FP reports grants and personal fees from Bayer, AstraZeneca, Pierre Fabre, Roche, Incyte, MSD, Janssen Cilag, personal fees from Daichii Sankyo, Clovis, Bristol Myers Squibb, Astellas, Ipsen, Seagen, Eli Lilly, GlaxoSmithKline, grants from Tesaro, Pfizer, Exelixis, Aileron, outside the submitted work. MLR had a role as a consultant for Eli Lilly, AstraZeneca and MSD outside the submitted work. All other authors have declared no conflicts of interest.

### REFERENCES

- Zhang S, Liang F, Li W, Hu X. Subgroup analyses in reporting of phase III clinical trials in solid tumors. *J Clin Oncol*. 2015;33(15):1697-1702.
- Altman DG. Clinical trials: subgroup analyses in randomized trials—more rigour needed. *Nat Rev Clin Oncol*. 2015;12(9):506-507.
- Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266(1):93-98.
- Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess*. 2001;5(33):1-56.
- Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ*. 2015;351:h5651.
- Pocock SJ, Assmann SF, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917-2930.
- Sormani MP, Bruzzi P. Reporting of subgroup analyses from clinical trials. *Lancet Neurol*. 2012;11(9):747;author reply 747-8.
- Di Maio M, Audisio M, Cardone C, et al. The use of not-negative conclusions to describe results of formally negative trials presented at oncology meetings. *JAMA Oncol*. 2020;6(6):926-927.
- Lee SM, Falzon M, Blackhall F, et al. Randomized prospective biomarker trial of ERCC1 for comparing platinum and nonplatinum therapy in advanced non-small-cell lung cancer: ERCC1 trial (ET). *J Clin Oncol*. 2017;35(4):402-411.
- Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389(10066):255-265.
- Bellmunt J, de Wit R, Vaughn DJ, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017;376(11):1015-1026.
- Shah MA, Bang YJ, Lordick F, et al. Effect of fluorouracil, leucovorin, and oxaliplatin with or without onartuzumab in HER2-negative, MET-positive gastroesophageal adenocarcinoma: The METGastric randomized clinical trial. *JAMA Oncol*. 2017;3(5):620-627.
- Herbst RS, Redman MW, Kim ES, et al. Cetuximab plus carboplatin and paclitaxel with or without bevacizumab versus carboplatin and paclitaxel with or without bevacizumab in advanced NSCLC (SWOG S0819): a randomised, phase 3 study. *Lancet Oncol*. 2018;19(1):101-114.
- Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med*. 2018;378(14):1277-1290.
- Socinski MA, Jotte RM, Cappuzzo F, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378(24):2288-2301.
- Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378(22):2093-2104.
- Schmid P, Adams S, Rugo HS, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N Engl J Med*. 2018;379(22):2108-2121.
- Motzer RJ, Penkov K, Haanen J, et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med*. 2019;380(12):1103-1115.
- Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819-1830.
- Rini BI, Powles T, Atkins MB, et al. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet*. 2019;393(10189):2404-2415.
- West H, McCleod M, Hussein M, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase3 trial. *Lancet Oncol*. 2019;20(7):924-937.
- González-Martín A, Pothuri B, Vergote I, et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med*. 2019;381(25):2391-2402.
- Tap WD, Wagner AJ, Schöffski P, et al. Effect of doxorubicin plus olaratumab vs doxorubicin plus placebo on survival in patients with advanced soft tissue sarcomas: The ANNOUNCE randomized clinical trial. *JAMA*. 2020;323(13):1266-1276.
- Galsky MD, Arija JÁA, Bamias A, et al. Atezolizumab with or without chemotherapy in metastatic urothelial cancer (IMvigor130): a multicentre, randomised, placebo-controlled phase 3 trial. *Lancet*. 2020;395(10236):1547-1557.
- Powles T, Park SH, Voog E, et al. Avelumab maintenance therapy for advanced or metastatic urothelial carcinoma. *N Engl J Med*. 2020;383(13):1218-1230.
- Shitara K, Van Cutsem E, Bang YJ, et al. Efficacy and safety of pembrolizumab or pembrolizumab plus chemotherapy vs chemotherapy alone for patients with first-line, advanced gastric cancer: the KEYNOTE-062 phase 3 randomized clinical trial. *JAMA Oncol*. 2020;6(10):1571-1580.
- Herbst RS, Giaccone G, de Marinis F, et al. Atezolizumab for first-line treatment of PD-L1-selected patients with NSCLC. *N Engl J Med*. 2020;383(14):1328-1339.
- Powles T, van der Heijden MS, Castellano D, et al. Durvalumab alone and durvalumab plus tremelimumab versus chemotherapy in

- previously untreated patients with unresectable, locally advanced or metastatic urothelial carcinoma (DANUBE): a randomised, open-label, multicentre, phase 3 trial. *Lancet Oncol.* 2020;21(12):1574-1588.
29. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med.* 2017;377(20):1919-1929.
  30. Ray-Coquard I, Pautier P, Pignata S, et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *N Engl J Med.* 2019;381(25):2416-2428.
  31. Burtneß B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet.* 2019;394(10212):1915-1928.
  32. Zhu AX, Kang YK, Yen CJ, et al. Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased  $\alpha$ -fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2019;20(2):282-296.
  33. Zhu AX, Park JO, Ryoo BY, et al. Ramucirumab versus placebo as second-line treatment in patients with advanced hepatocellular carcinoma following first-line therapy with sorafenib (REACH): a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol.* 2015;16(7):859-870.
  34. Faehling M, Schumann C, Christopoulos P, et al. Durvalumab after definitive chemoradiotherapy in locally advanced unresectable non-small cell lung cancer (NSCLC): Real-world data on survival and safety from the German expanded-access program (EAP). *Lung Cancer.* 2020;150:114-122.