

Assessment

Cite this article: Zozaya N, Arrizubieta Basterrechea MI, Bollo E, Castellví I, Espín J, Ortego N, Poveda-Andrés JL, Rodríguez Portal JA, Rivero A, Marcos-Rodríguez JA, Verde L (2022). A multi-criteria decision analysis on the value of nintedanib for interstitial lung diseases. *International Journal of Technology Assessment in Health Care*, **38**(1), e64, 1–10 <https://doi.org/10.1017/S0266462322000459>

Received: 21 January 2022

Revised: 17 May 2022

Accepted: 21 June 2022

Key words:

Interstitial lung diseases; Systemic sclerosis; Pulmonary fibrosis; Multi-criteria decision analysis


Author for correspondence:

*Néboa Zozaya,

E-mail: neboa.zozaya@weber.org.es

We would like to thank Alfonso Sánchez Lorite and Jorge Cuadrelli (in memorial) for their valuable contributions to this project, by participating as patient's representatives in the advisory committee.

A multi-criteria decision analysis on the value of nintedanib for interstitial lung diseases

Néboa Zozaya^{1,2*} , Maria Iciar Arrizubieta Basterrechea³, Elena Bollo⁴, Iván Castellví⁵, Jaime Espín⁶, Norberto Ortego⁷, José Luis Poveda-Andrés⁸, José Antonio Rodríguez Portal^{9,10}, Agustín Rivero¹¹, José Antonio Marcos-Rodríguez¹² and Luis Verde¹³

¹Department of Health Economics, WEBER, Madrid, Spain; ²Department of Quantitative Methods in Economics and Management, University of Las Palmas de Gran Canaria, Las Palmas, Spain; ³Department of Pneumology, Hospital Universitario de Galdakao-Usansolo, Bizkaia, Spain; ⁴Department of Pneumology, Complejo Asistencial Universitario de León, León, Spain; ⁵Department of Rheumatology, Hospital Universitari de la Santa Creu i Sant Pau, Barcelona, Spain; ⁶Andalusian School of Public Health, Granada, Spain; ⁷Unit of Systemic Autoimmune Diseases, Hospital Universitario Clínico San Cecilio, Granada, Spain; ⁸Department of Pharmacy, Unit of Clinical Medications, Hospital Universitari i Politècnic La Fe, Valencia, Spain; ⁹Department of Pneumology, Hospital Universitario Virgen del Rocío, Seville, Spain; ¹⁰Centro de Investigación Biomédica en Red – Enfermedades Respiratorias (CIBERES), Madrid, Spain; ¹¹Bioregión de Salud y Bienestar (BioMad), Madrid, Spain; ¹²Department of Pharmacy, Hospital Universitario Virgen Macarena, Seville, Spain and ¹³Management Department, Integrated Health Area, A Coruña, Spain

Abstract

Objectives: Our aim was to assess the value of nintedanib for non-idiopathic progressive fibrosing interstitial lung disease (non-IPF PF-ILD) and systemic sclerosis-associated ILD (SSc-ILD) in the Spanish context, using a multi-criteria decision analysis (MCDA).

Methods: Following an adaptation of the Evidence and Value: Impact on Decision Making (EVIDEM) MCDA methodology, the estimated value of nintedanib was obtained by means of an additive linear model that combined individual weights (100-points distribution) of criteria with the individual scoring of nintedanib in each criterion for every indication, assigned by a multidisciplinary committee of twelve clinicians, patients, pharmacists, and decision-makers. To assess the reproducibility, an alternative weighting method was applied, as well as a re-test of weights and scores at a different moment of time.

Results: The experts committee recognized nintedanib as an intervention with a positive value contribution in comparison to placebo for the treatment of non-IPF PF-ILD (0.50 ± 0.16 , on a scale from -1 to 1) and SSc-ILD (0.40 ± 0.12), diseases which were considered as very severe and with high unmet needs. The drug was perceived as a treatment that provides an added therapeutic benefit for patients (0.06 – 0.07), given its proven clinical efficacy (0.05 – 0.06), slight improvements in patient-reported outcomes (0.01 – 0.02), and similar safety profile than placebo (-0.04 – 0.00), which will likely be positioned as a recommended therapy in the next clinical practice guidelines updates.

Conclusions: Under this increasingly used methodology, nintedanib has shown to provide a positive value estimate for non-IPF PF-ILD and SSc-ILD when compared to placebo in Spain.

Decisions on the allocation of public resources in health care are complex, given that they involve trade-offs between multiple and often conflicting objectives, such as the high demand for access to drugs, devices, and services, and the protection of the financial sustainability of the system, in an environment of increasing demographic, technological, social, and budgetary challenges.

Based on this scenario, the multi-criteria decision analysis (MCDA) is a framework that helps to inform and make the preferences inherent to decisions explicit, in a consistent and transparent way (1). Its use in the health field is relatively recent, with some examples of pilots and applications in real practice in a few countries, including Spain, with a special focus on rare diseases (2).

Interstitial lung diseases (ILDs), also referred to as diffuse parenchymal lung disease, encompass a large and diverse group of restrictive lung diseases, many of which are formally classified as rare. The major abnormality in ILD is the disruption of the distal lung parenchyma which is comprised of thin-walled alveoli through which gas exchange occurs. Current clinical understanding of ILD posits that all ILDs are activated by repetitive chronic epithelial or vascular injuries or by granulomatous inflammation, both of which activate pathological pathways in the lung tissue with varying consequences including cell destruction (3).

A proportion of patients with certain types of ILD develop a progressive fibrosing (PF) phenotype. As a result, the fibrosis becomes progressive, self-sustaining, and independent of the original clinical association or disease trigger. PF-ILD is a defining feature of idiopathic

pulmonary fibrosis (IPF) in which all patients exhibit the phenotype; however, the phenotype also occurs in a smaller proportion of patients with other non-IPF ILDs (4).

On the other hand, systemic sclerosis (SSc) is a serious, heterogeneous, and chronic autoimmune disease that is characterized by microvascular damage and progressive fibrosis of skin and internal organs. Most patients with SSc have lung involvement, with ILD featuring as the principal complication. SSc-associated ILD (SSc-ILD) is a potentially fatal complication with symptoms that range from subclinical lung involvement to major pulmonary disease progressing to respiratory failure and death (5).

Nintedanib (Ofev, Boehringer Ingelheim, Ingelheim am Rhein, Germany), an intracellular inhibitor of tyrosine kinases, is the only drug approved by the European Medicines Agency (EMA) for the treatment of non-IPF PF-ILD and SSc-ILD, with both marketing authorizations granted in July and April, 2020, respectively (6). Other treatments used for the management of the disease are mainly immunomodulatory off-label medications (7).

The main objective of this study was to assess the value estimate of nintedanib for non-IPF PF-ILD and SSc-ILD in the Spanish context, using a MCDA. The comparison was made versus placebo, as there are no approved treatments for any of the indications for which nintedanib is being evaluated. The term value should be interpreted throughout this report as a value judgment elicited from an expert committee, based on the application of a value measurement approach (in this MCDA: a compositional, direct ranking approach), which is the most common technique used in health care (8).

Method

This study is built on an adaptation of the Evidence and Value: Impact on DEcision Making (EVIDEM) framework (10th edition) for MCDA, a methodology based on the application and evaluation of a set of criteria used to estimate an overall value for an intervention (9). The framework also includes a contextual tool for evaluating qualitative criteria.

The steps taken to perform this theoretical MCDA exercise (Figure 1) were grounded on previously published good methodological practices (8;10). First, a committee of twelve experts was formed, including four clinicians (two pneumologists, one rheumatologist, and one internal medicine physician), one nurse, two patients (one with non-IPF PF-ILD and one with SSc-ILD), two healthcare managers, two hospital pharmacists, and one health economist. The number of experts, as well as its composition (i.e., different profiles and perspectives included) is in line with previously published MCDA in the context of estimating the value of a drug in comparison to an alternative treatment (11–17).

Members were selected based on their expertise and with the aim of achieving a balanced geographical representation, to incorporate, as much as possible, views that could reflect differences which may exist between the autonomous regions in the country. The consultancy firm WEBER (Madrid, Spain) established the contact with the panelists, except for the patients, who were selected by their physicians—also members of the committee. WEBER oversaw the training of the experts committee on MCDA, reviewing and sharing the evidence with the panelists prior to the meetings, coordinating the activities held by the committee, and analyzing and presenting the results.

Secondly, the experts committee validated the quantitative and qualitative criteria to be included in the analysis, in a first online meeting, making sure the criteria fulfilled the requirements related to

completeness, nonredundancy, nonoverlapping, and preference independence (18). Based on the EVIDEM framework, the experts committee selected eleven quantitative and seven qualitative criteria as relevant for the exercise. This was done through a voting, followed by a debate in case consensus was not reached (≥ 75 percent of votes). The two excluded criteria from the standard EVIDEM set were type of preventive benefit (quantitative) and environmental impact (qualitative). The criterion size of the population was transferred from the quantitative to the qualitative criteria set.

The third step, also performed during the first online meeting, was the assignation of weights, through the 100-points distribution method. Each expert allocated a weight to each quantitative criterion according to its perceived importance, independently from the interventions to be evaluated, and provided that the sum of all weights resulted in 100. Under this weighting method, the more points a criterion receives, the greater is its relative importance in the view of the experts. If an expert, for example, believes all criteria have the same relevance, he/she would assign a weight of approximately 9.1 to each criterion (9.1 times eleven criteria = 100 points).

The fourth step consisted of carrying out two additional online meetings. Prior to that, a comprehensive literature review was conducted to collect the available evidence regarding the criteria and drug included. The information was assembled in two evidence matrixes, which were reviewed and validated by the clinicians (except for the information related to the qualitative criteria). The search was performed using major biomedical databases, such as PubMed and Medline, clinical trials registries, clinical practice guidelines, official European and Spanish healthcare evaluation bodies webpages, namely EMA, Spanish Medicines and Healthcare Products Agency (AEMPS) and Spanish regional and hospital evaluations, as well as grey literature. No date or language restrictions were applied (Supplementary Files 5 and 6).

Based on the available evidence, on individual experiences and perceptions and on the debate generated during the meeting, each expert assigned a score to each quantitative criterion for the two indications under evaluation. Both the drug and the indications were evaluated. The scores for the absolute criteria (those not involving comparisons with placebo) could range from 0 to 5, 0 being the lowest value and 5 the highest. For relative criteria (those involving comparisons with placebo), the scale ranged from -5 to 5 to reflect the full range of comparative effects. The impact of the seven selected qualitative criteria, which were rated as negative, neutral, or positive, was discussed in the first online meeting.

To obtain the estimated value for each intervention, a linear additive model was used, combining the relative weighting of each criterion with the score for each intervention in each criterion. Each estimated value was transformed into a 0–1 scale to facilitate its interpretation. In particular, the following formula was used ((1) (19):

$$V = \sum_{x=1}^n V_x = \sum_{x=1}^n \left(\frac{W_x}{\sum W_n} S_x \right), \quad (1)$$

where V is the total estimated value, n is the number of experts in the experts committee, V_x is the weight contribution of criterion x , W_x is the weighting of criterion x , $\sum W_n$ is the sum of all weights, and S_x is the scoring of criterion x .

To check the degree of consistency and replicability of the analysis, three steps were taken: (i) a re-test of weights (using the 100-points distribution) and scores was performed, where the experts assigned weights and scores to each quantitative criterion. The degree of agreement between the responses made at the two

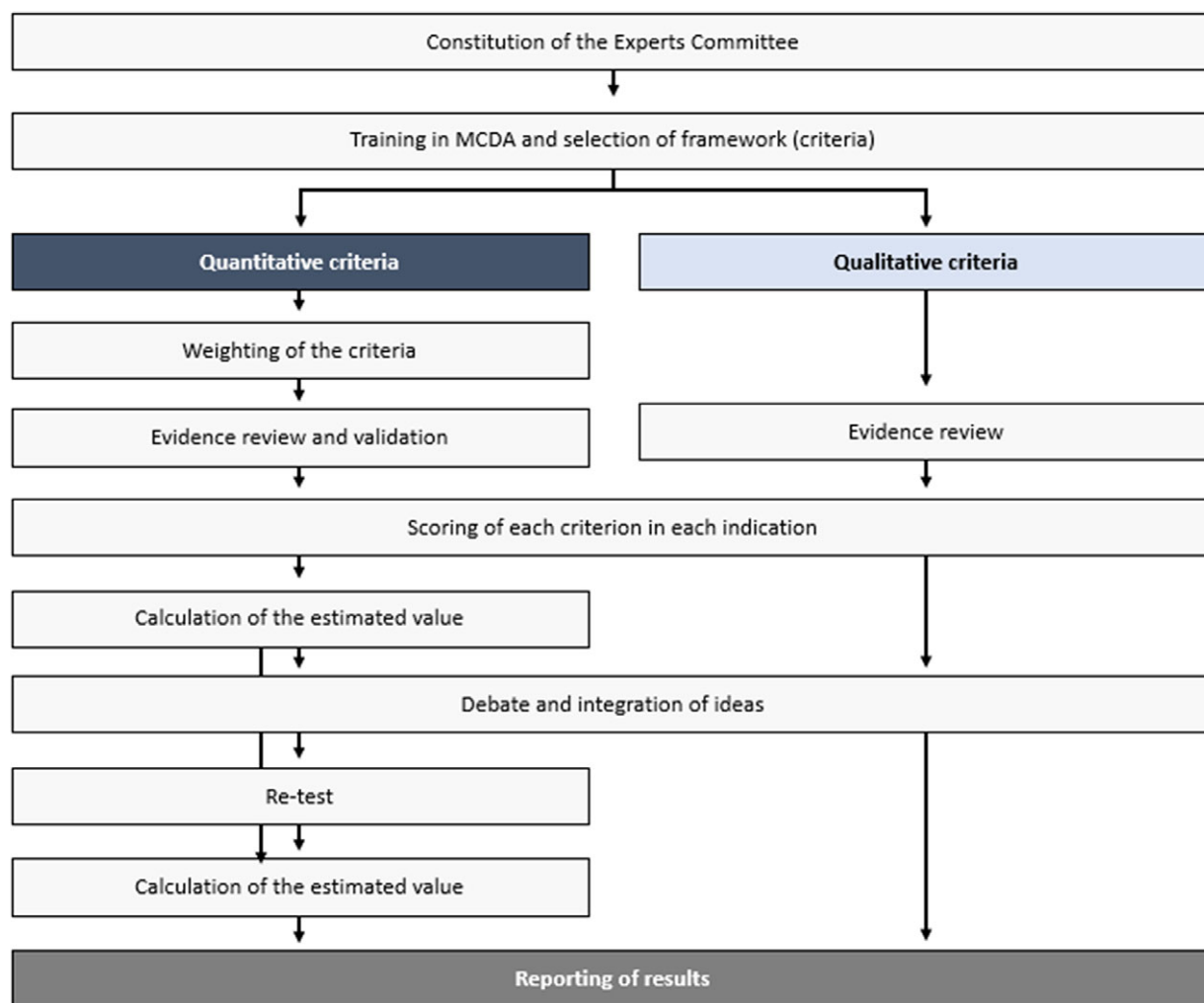


Figure 1. Study design.

timepoints was evaluated through the intrarater correlation coefficients (ICC 3.1) using STATA version 14 (STATA Corp., LP, College Station, TX). The re-test was done online, 2 weeks after the third meeting was performed; (ii) additionally, an alternative weighting method (scales from 1 to 5) was applied as part of the exercise (in the test phase, as well as in the re-test phase), where each expert assigned a weight between 1 and 5 to each quantitative criterion according to its importance, where 1 was the lowest relevant and 5, the most; (iii) a sensitivity analysis was carried out to test what influence would the weights assigned have on the value resulted from the analysis, as well as whether the weights assigned have external validity. For that, we have replaced weightings assigned in this MCDA by the weightings resulting from the work of Badia et al. (20), based on preferences of ninety-eight Spanish evaluators and decision makers (twenty-five at national level and seventy-three at regional level), with a perspective of appraising orphan drugs (Supplementary File 4).

Results

Weights

Under this method, the experts assigned a weight to each criterion, given that the sum of all weights would result in 100 points.

The three criteria considered as with the highest relevance by the experts committee were disease severity (18.4 ± 6.13), unmet needs (14.0 ± 4.59), and type of therapeutic benefit (13.2 ± 3.09), whilst the three lowest ranked criteria in terms of relevance were non-medical costs (2.5 ± 2.76), expert consensus/clinical practice guidelines (4.4 ± 2.37), and other medical costs (4.7 ± 2.95). The criteria with the highest variability in responses were disease severity (standard deviation [SD] 6.13), quality of evidence (SD 5.56), and cost of intervention (SD 5.46) (Figure 2).

Clinicians granted the highest weights for disease severity, unmet needs, and type of therapeutic benefit, whilst considering nonmedical costs and other medical costs as of low relevance. For patients, the most important criterion was patient-reported outcomes, and the least, cost of intervention. Managers and pharmacists weighted the highest on efficacy and cost of intervention (Supplementary File 1).

Scores for non-IPF PF-ILD

Regarding the mean scores for nintedanib versus placebo for the treatment of non-IPF PF-ILD (Figure 3 and Table 1), the experts committee gave the highest scores, with the lowest variability amongst participants, to disease severity (4.4 ± 0.7 out of 5.0) and unmet needs (4.3 ± 0.6). This fact shows that non-IPF PF-

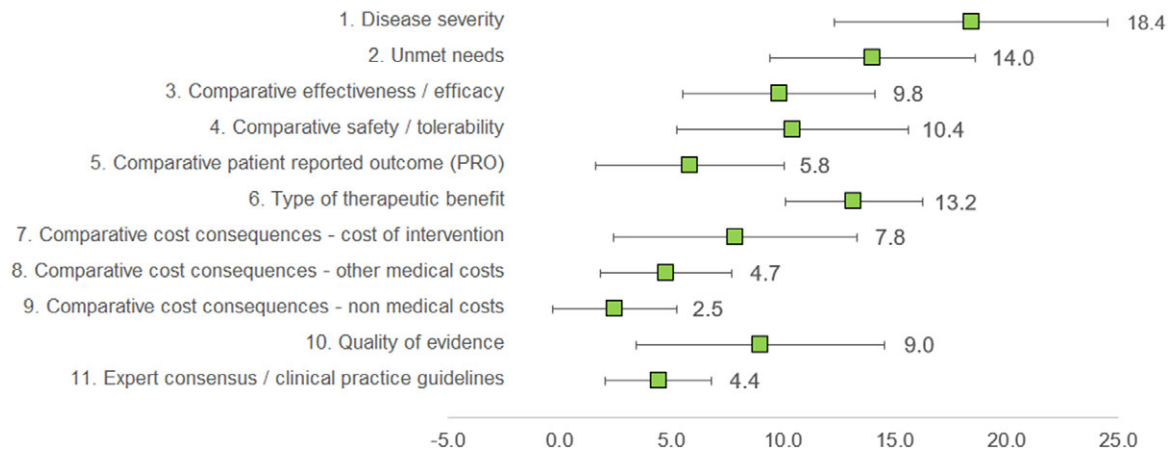


Figure 2. Mean weights for the decision criteria by the experts committee. A 100-points distribution was used, where the experts assigned a weight to each criterion, given the sum of the weights resulted in 100.

ILD is perceived as a very serious disease, with a very low survival rate and high morbidity, with important unmet needs, as off-label treatments are associated with serious adverse events, undemonstrated efficacy level, and that usually only target the symptoms and do not modify disease progression (21).

The next two highest-scoring criteria were quality of evidence (3.7 ± 0.9) and comparative efficacy (3.2 ± 0.6), indicating the superiority of nintedanib over placebo in slowing the progression of the disease and on the other efficacy criteria included in a trial that was considered as relevant and valid by the experts committee (INBUILD trial) (22).

The only criterion with a negative score was cost of intervention (-3.2 ± 1.5), reflecting the notable difference in costs between nintedanib and placebo. On the other hand, the other two cost criteria were granted positive scores (1.2 ± 1.9 and 2.0 ± 1.2 for nonmedical costs and other medical costs, respectively), representing a belief by the experts committee that the introduction of nintedanib would generate an offset effect on other cost lines, mainly on direct costs (other medical costs), as participants perceived that nintedanib could decrease the number of hospitalizations, medical visits, and other medical costs in general.

The criteria type of therapeutic benefit (2.7 ± 0.7), expert consensus/clinical practice guidelines (2.7 ± 1.1), and patient-reported outcomes (1.6 ± 1.1) also received moderate–high scores. Some of the assumptions made by the experts for scoring those criteria were that they believe that nintedanib will receive positive recommendation in the next update of the guidelines and, when it comes to patient-reported outcomes, the fact that the drug maintains quality of life levels can be seen as positive, given the nature of the disease.

The safety criterion received almost a null score with great variability amongst groups (0.1 ± 2.8) which in average represents no difference in safety and tolerability versus placebo. Some of the positive scores were given by the consideration of the fact that the patients on the placebo group had more serious events, meaning that nintedanib provided a better control on patients. Some of the lowest scores were assigned based on the frequency of diarrhea, dose reduction, and treatment discontinuation (22).

All subgroups of experts have agreed on the highest score, granted to disease severity (4.5 for all groups, except for patients, who scored 4.0), as well as on the lowest score for cost of intervention (-2.0 for patients and others; -3.0 for pharmacists; -3.8 for clinicians, and -4.5 for managers). Safety was the criterion with

the greatest variability, as patients assigned 4.0; managers: 1.5; others: 1.0; pharmacists: -1.0 , and clinicians: -2.5 . For expert consensus/clinical practice guidelines, managers and clinicians scores (3.5 and 3.3, respectively) were higher than among other groups (others: 2.5; pharmacists: 2.0 and patients: 1.5) (Supplementary File 1).

Scores for SSc-ILD

The highest mean scores for nintedanib versus placebo for the treatment of SSc-ILD (Figure 3 and Table 1) granted by the experts were disease severity (4.1 ± 0.5) and unmet needs (3.9 ± 0.5), with low variability between groups. Relatively to non-IPF PF-ILD, SSc-ILD was considered as a slightly less severe disease with fewer unmet needs. This could be attributed to, among other factors, the fact that the decline in lung function is slower in SSc-ILD and, in some cases, it stabilizes (23).

The next highest-scoring criteria were type of therapeutic benefit (2.3 ± 0.8), expert consensus/clinical practice guidelines (2.3 ± 1.1), other medical costs (1.8 ± 1.4), and nonmedical costs (1.5 ± 1.2). Similarly, the magnitude of those scores was relatively lower in SSc-ILD in comparison to non-IPF PF-ILD. Some members of the experts committee believe that, even in the absence of data, the therapeutic benefit with regard to SSc-ILD could be higher in the long term, resulting in an improvement in survival. They also added that nintedanib was already included in a clinical practice recommendation paper (24), reason for which the drug will be well positioned in future clinical guidelines.

The criterion patient-reported outcomes received an almost null mean score (0.3 ± 1.1), meaning that there were no differences between the study drug and placebo on the quality of life reported by the patients.

The two criteria with negative scores were safety/tolerability (-2.1 ± 0.8) and cost of intervention (-3.3 ± 0.8), with a low variability between groups. Differently from the scores provided for non-IPF PF-ILD, nintedanib safety profile for SSc-ILD was considered as relatively worse than placebo.

In the subgroup analysis, disease severity and unmet needs were the criteria with the highest scores granted by all groups. For disease severity, patients and managers assigned a score of 4.5, while clinicians and pharmacists granted 4.0, and others scored 3.5. For unmet needs, results were highly similar amongst groups (all subgroups scored 4.0, except managers: 3.8).

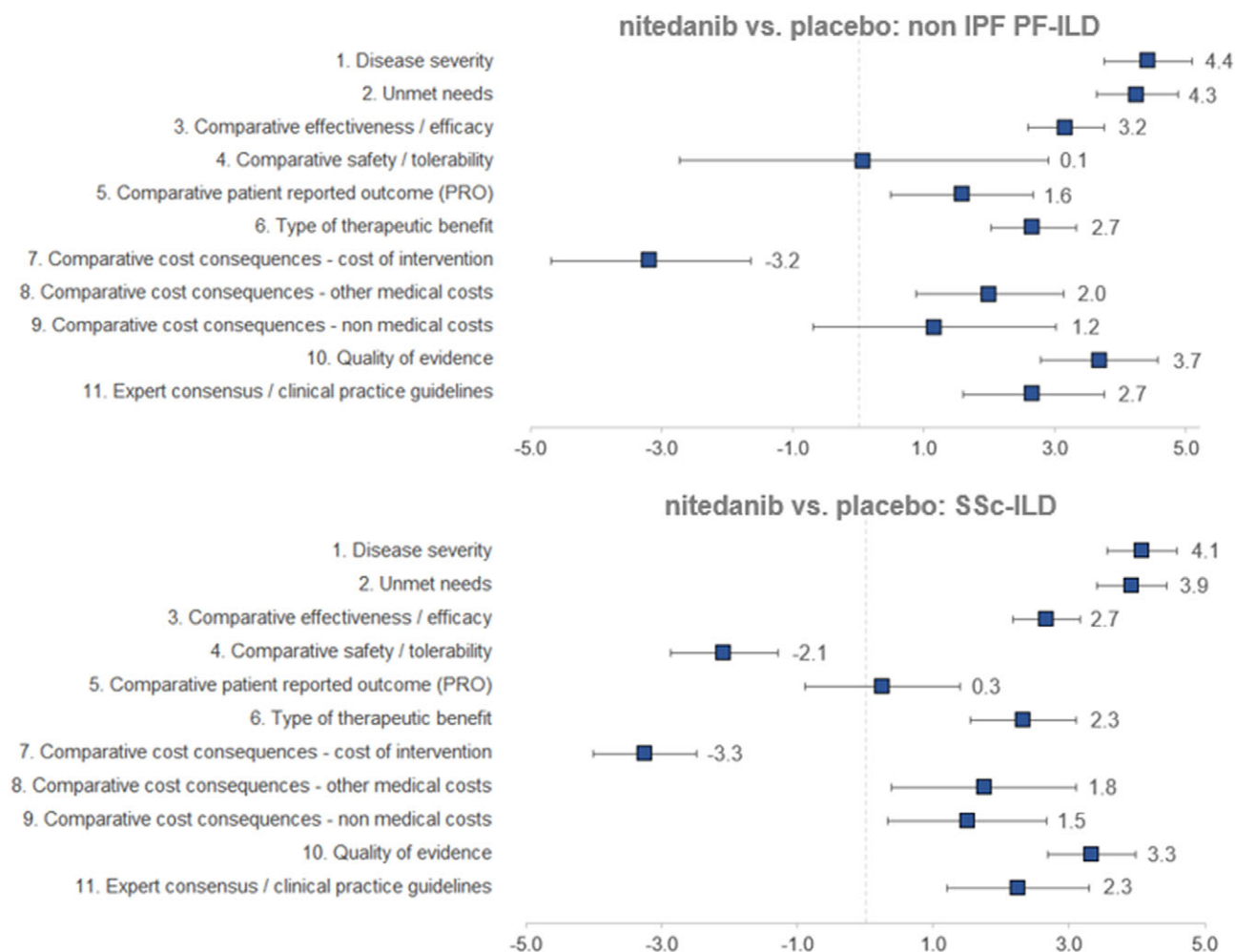


Figure 3. Mean scores for the decision criteria on the appraisal of nintedanib versus placebo for non-IPF PF-ILD and SSc-ILD. Criteria evaluated in absolute terms (with a score between 0 and 5) were disease severity, unmet needs, type of therapeutic benefit, quality of evidence, and expert consensus/clinical practice guidelines, whilst the rest were evaluated in relative terms (with a score between -5 and 5). ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; PF, progressive fibrosis; SSc, systemic sclerosis.

The efficacy criterion received similar scores by all groups (3.0 for patients, managers, and pharmacists; 2.5 for clinicians, and 2.0 for others), while safety was rated similarly by clinicians and patients (-2.8 and -2.5 , respectively), with the remaining groups granting a slightly better score (-1.5). On the other hand, patients granted a higher score for patient-reported outcomes (1.5) in comparison to the other groups (between -0.5 and 1.0).

Variability between subgroups was also observed in the following criteria: expert consensus/clinical practice guidelines (between 1.5 and 3.0), nonmedical costs (0.0–3.0), and other medical costs (0.5–3.0) (Supplementary File 1).

Value Estimates

The overall value estimated is a result of the integration of weights and scores of each panelist on a scale between 0 and 1. Nintedanib provided an added value in both diseases analyzed, against placebo (Figure 4). Specifically, the estimated value for nintedanib versus placebo for the treatment of non-IPF PF-ILD was 0.50 ± 0.16 , and the value for nintedanib versus placebo for the treatment of SSc-ILD was 0.40 ± 0.12 .

The five criteria with the highest positive contribution to the overall value estimates of nintedanib were disease severity

(0.17 ± 0.07 and 0.15 ± 0.06 for non-IPF PF-ILD and SSc-ILD, respectively), unmet needs (0.12 ± 0.05 and 0.11 ± 0.04), type of therapeutic benefit (0.07 ± 0.03 and 0.06 ± 0.03), quality of evidence (0.07 ± 0.05 and 0.06 ± 0.04), and compared efficacy (0.06 ± 0.03 and 0.05 ± 0.02).

On the other hand, cost of intervention contributed negatively to the overall value estimates for nintedanib versus placebo for both indications analyzed (-0.06 ± 0.05 and -0.05 ± 0.04 for non-IPF PF-ILD and SSc-ILD, respectively) and safety/tolerability had a negative contribution only for SSc-ILD (-0.04 ± 0.02).

In the subgroups analyzed, the highest estimated values were observed in the patient group, followed by clinicians. Managers estimated highest values than pharmacists (Supplementary File 1).

Replicability and Consistency

The results showed good reproducibility, with a relatively high degree of agreement and consistency of the results. On the re-test of the 100-points distribution weighting method, five out of the eleven criteria (45.5 percent) were weighted identically as on the test, whilst two (18.2 percent) differed by 1 point, three (27.3 percent) by 2 points, and one (9.1 percent) by 3 points. The ICC for weights was high (0.84).

Table 1. Mean Scores and Key Comments from the Experts Committee, for Each Criterion, Non-IPF PF-ILD and SSc-ILD

Criteria	Non-IPF PF-ILD		SSc-ILD	
	MS ± SD	Key comments from the experts committee	MS ± SD	Key comments from the experts committee
Disease severity	4.4 ± 0.7	“Unfortunately, any additional lifetime we can provide the patient with is not symptom free (suffocation and physical limitation)”. “Very low survival rates (3–5 yrs.)”	4.1 ± 0.5	“Progression is usually slower than in non-IPF PF-ILD patients. In some cases, it stabilizes.” “ILD is the main cause of death in SSc patients”
Unmet needs	4.3 ± 0.6	“Some medications target ILDs, but for the progressive ILDs, those treatments are not effective.” “There is no approved treatment alternative”	3.9 ± 0.5	“Trials with immunomodulators, immunosuppressors and biologics are of very low quality, and results in real practice are very poor.” “There is no approved treatment option”
Comparative effectiveness/ efficacy	3.2 ± 0.6	“Very effective. It slows disease progression by 57% compared to placebo. Its maximum effect suggests an FVC decrease similar to the observed in healthy people”	2.7 ± 0.5	“Effective drug, as it slows progression by 44% versus placebo”. “Extension of the SENSICIS trial shows efficacy is being kept after 52 weeks”
Comparative safety/ tolerability	0.1 ± 2.8	“Serious AE were more common in the placebo group, reflecting those patients with nintedanib are well controlled”	−2.1 ± 0.8	“A more extensive list of usual medications was allowed for patients with placebo in the non-IPF PF-ILD trial, vs. the SSc-ILD trial.” “Most AE are manageable”
Comparative patient-reported outcome	1.6 ± 1.1	“It is positive that quality of life is maintained in a very serious disease like this.” “Instruments used in the trial are not the correct ones to capture PRO improvements”	0.3 ± 1.1	“Instruments used, patients’ related comorbidities and trial duration could be associated with the non-statistically significant results”
Type of therapeutic benefit	2.7 ± 0.7	“Nintedanib alleviates symptoms and slows disease progression. This could potentially translate in higher survival, although trial results were not significant”	2.3 ± 0.8	“A change of the disease course and the alleviation of symptoms could improve mortality in a 4–5 years’ horizon, especially in patients with rapid progression”
Cost of the intervention	−3.2 ± 1.5	“Official costs vs. placebo are high. Nevertheless: (1) costs in real practice are lower; (2) placebo was combined with other drugs, in some cases”	−3.3 ± 0.8	Similar comments as to non-IPF PF-ILD, with one addition: “being an additional indication to an approved drug, prices could be lower than the ones currently practiced”
Other medical costs	2.0 ± 1.1	“Although hospitalizations were not measured in the trial, they are correlated with acute exacerbations, which decreased by >30% with the use of nintedanib”	1.8 ± 1.4	“Alleviation of dyspnea and cough can decrease number of hospitalizations, medical visits and other medical costs in general”
Nonmedical costs	1.2 ± 1.9	“Average age of onset is high (60’s), hence productivity is less affected in this patient population.” “Patients with nintedanib noted more willingness to work”	1.5 ± 1.2	“Slightly higher effect in SSc-ILD patients in comparison with non-IPF PF-ILD, as onset of disease occurs in patients with working age (50’s)”
Quality of evidence	3.7 ± 0.9	“High quality trial, especially considering the challenges of recruiting and running trials in rare diseases. 663 patients are a very representative number”	3.3 ± 0.7	“High quality trial, with a very representative number of patients (576), despite challenges of being a rare disease with a heterogeneous population”
Expert consensus/clinical practice guidelines	2.7 ± 1.1	“The results of the trial and current expert opinions clearly indicate that this drug will be well positioned in the future updates of clinical practice guidelines”	2.3 ± 1.1	“There are no specific CPG for SSc-ILD patients, as they are included in SSc guidelines. Current publications with similar methodology as the CPGs recommend Nintedanib”

AE, adverse events; CPG, clinical practice guidelines; FVC, forced vital capacity; ILD, interstitial lung diseases; MS, mean score; Non-IPF PF-ILD, non-idiopathic progressive fibrosing interstitial lung disease; PRO, patient-reported outcome; SD, standard deviation; SSc-ILD, systemic sclerosis-associated interstitial lung disease.

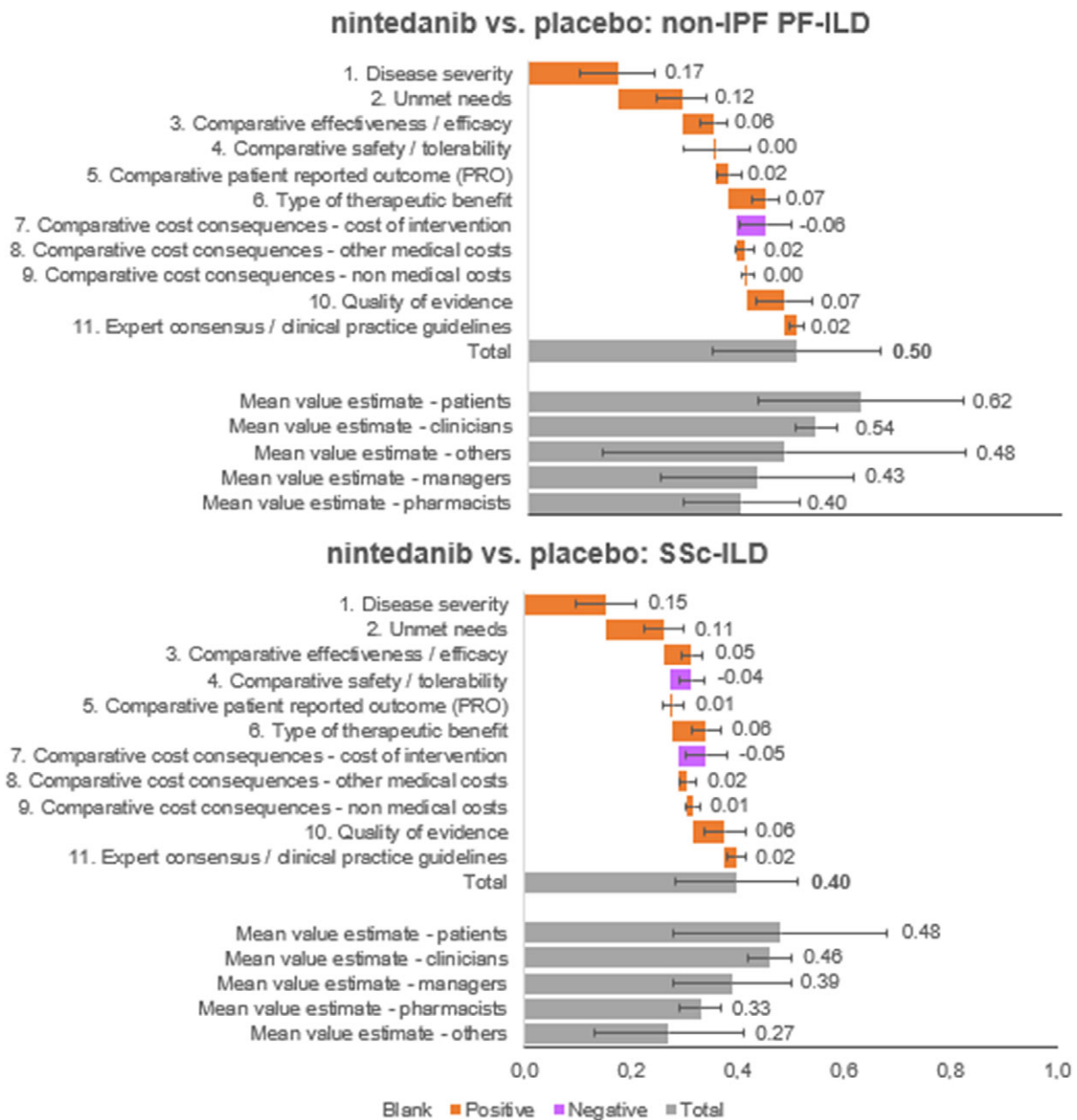


Figure 4. Mean value contributions of each quantitative criterion and overall value estimates for nintedanib versus placebo for the treatment of non-IPF PF-ILD and SSc-ILD. Average of the experts committee and for each subgroup (patients, clinicians, managers, pharmacists, and others). ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; PF, progressive fibrosis; SSc, systemic sclerosis.

The scores for the re-test were highly similar to the ones in the test, for both indications. The major deviations observed were of low magnitude: -0.75 for patient-reported outcomes in non-IPF PF-ILD, -0.50 on other medical costs for non-IPF PF-ILD, -0.50 on nonmedical costs for SSc-ILD, and 0.42 on safety/tolerability for SSc-ILD. The analysis of the ICC showed the high concordance in the scores given to the criteria at both moments of time (0.94 for non-IPF PF-ILD and 0.96 for SSc-ILD).

The value estimate obtained in the re-test (100-points distribution) for the two indications was relatively consistent with that of the primary analysis (0.50 vs. 0.46 for non-IPF PF-ILD and 0.40 vs. 0.40 for SSc-ILD), with ICCs of 0.75 for non-IPF PF-ILD and 0.67 for SSc-ILD (Supplementary File 2).

The values estimated with the alternative weighting method (scales from 1 to 5) were between 11.5 percent (non-IPF PF-ILD)

and 13.4 percent (SSc-ILD) lower when compared with the 100-distribution method, and the key difference related to disease severity.

The results of the sensitivity analysis show that value estimates would vary from 0.50 to 0.42 in non-IPF PF-ILD and from 0.40 to 0.31 in SSc-ILD when replacing the original weights assigned to the ones found by the study from Badia et al. (20). Moreover, the correlation between weights assigned by the experts and the ones resulting from Badia et al. (20) were high (0.8209) (Supplementary File 4).

Qualitative Criteria

The panelists also assessed the impact of each qualitative criterion on both interventions (Supplementary File 3). Panelists considered

that the introduction of nintedanib would have a positive impact on all evaluated qualitative criteria, meaning that the drug would be aligned with actual evaluation and decision-making context in Spain. There was a high degree of agreement among participants, as five out of seven criteria were considered positive by more than 75 percent of participants.

Discussion

The MCDA methodology allows for the consideration of different perspectives and criteria when evaluating a new healthcare technology (1). Although many of the criteria included in this MCDA are already implicitly considered in decision making, this exercise identifies, in a holistic, explicit, and quantifiable way, all the value drivers involved in it. In addition, it allows for a multidisciplinary reflection on the value provided by the evaluated therapy, helping to clarify, quantify, and assess the most relevant elements for the different types of agents.

In recent years, there has been an increase in the use of the MCDA in health care, both internationally and nationally, and one very representative example of this is the pilot carried out in Spain, by the Catalan Health Service for the evaluation of orphan drugs (25;26). There are multiple publications in which a specific drug has been considered or evaluated through this methodology, especially in the dermatological (12;27), oncological (14;28) and, closer to the present case, in rare diseases fields (16;29).

Rare diseases have a series of differential characteristics, in which the implementation of a MCDA takes on greater importance. Being pathologies that, in general, affect a small population, lack therapeutic alternatives, and pose a serious risk to the lives of patients who suffer from them and a great impact on their quality of life, it seems appropriate to consider a wide spectrum of elements of value, as well as having a more holistic view of the elements to consider in its evaluation. In this sense, it is necessary to take into account the different voices involved, both that of clinicians, who are on the front line of healthcare, and other healthcare professionals, managers, and patient representatives.

This MCDA allowed for the adoption of a clear and transparent methodological approach where, based on the EVIDEM methodology, the eleven quantitative and seven qualitative selected criteria helped to assess the same drug authorized for two indications: non-IPF PF-ILD and SSc-ILD.

As with all MCDA, the results of the final evaluation of the therapy are conditioned by the relative importance (or weighting) that each one of the experts gives to each element of value, and that differs depending on the expert subgroup. Thus, even though the experts agreed on the high relevance of the disease severity and unmet needs, and on the low relevance of the criteria for nonmedical costs and expert consensus/clinical practice guidelines, they diverged as to the relative importance of other criteria, such as patient-reported outcomes, other medical costs, and quality of evidence. For example, patients give greater importance to patient-reported outcomes, while clinicians consider that the quality of evidence is relatively more relevant, and healthcare managers, hospital pharmacists and others, the cost of intervention.

Moreover, the final value estimate is based on the aggregation of weights and scores assigned by experts. The scores assigned are based on the evidence provided and their own perception. For this reason, most of this study focuses on reporting the value judgment done by the experts (i.e., weights and scores) rather than on the evidence available. Overall, when compared to placebo, nintedanib

was perceived as a drug which adds value in relation to therapeutic benefits. In the exercise performed, the estimated final values of nintedanib versus placebo were 0.50 ± 0.16 for non-IPF PF-ILD and 0.40 ± 0.12 for SSc-ILD. This means that nintedanib provides added value in both indications studied, compared to placebo.

This study is methodologically aligned with other MCDA, including the ones performed in Spain for rare diseases, such as the cases of the assessment of orphan drugs by the Catalan Health Services and the protocol for the assessment of rare diseases developed by the Working Group for Rare Diseases and Orphan Drugs from the Spanish Society for Hospital Pharmacy (Orphar-SEFH) (26;30). The importance of those studies relies in understanding which elements of the new drug are of highest added value. In this sense, the multidisciplinary debate generated around each of the elements was key to understand the strengths and weaknesses of nintedanib for each of them in the two indications.

This study has certain limitations, inherent to all MCDA, which should be noted. The first comes from the composition of the expert committee. Although it has different points of view, the limited number of experts may not represent the opinion of all the actors involved. Moreover, the weights and scores assigned in this MCDA reflect the perception of the experts committee on the current exercise, and although they can serve as a reference in other ILD (or rare diseases) evaluation processes, they may change under different decision-making settings, that is, other countries, different composition of a decision-making body, and so forth. For this reason, the external validity of the results will not be evident, and often the final assessment obtained will not be directly generalizable to other comparators, nor will it be durable over time. In the same way, it must be considered that the final assessment of the MCDA is done based on the experience, knowledge, and value judgments of the committee members; hence, the analysis contains certain subjectivity. Nevertheless, this subjectivity is intrinsic to most of the decision processes in health care, and the MCDA was not developed to provide an objective ratio or a unique answer to a decision problem, as it should be used as a complementary tool to the current existing set of frameworks applied.

Likewise, the analysis entails a certain cognitive complexity, especially considering that this MCDA was carried out in two indications. Therefore, the exercise requires concentration from the evaluator and a clear understanding of how the tool works. It is therefore important that the scoring is carried out during the online meeting, preceded by an exhaustive explanation of the MCDA methodology and the assumptions made, giving the evaluator enough time to carry out the evaluations.

Finally, another limitation could be the one related to the existing debate around whether the criterion quality of evidence should be used as a penalty or to modulate other scores, such as the clinical ones (efficacy, safety, PRO), rather than being used as a standalone criterion (31). In the case of this MCDA, we have used EVIDEM as the reference framework; hence, the criterion quality of evidence was placed as a separate one.

Conclusions

Nintedanib, an intracellular inhibitor of tyrosine kinases, is the only drug approved by EMA for the treatment of non-IPF PF-ILD and SSc-ILD. Its value was successfully assessed by a multidisciplinary experts committee formed by relevant stakeholders involved in clinical management, drug evaluation, and healthcare decision making in Spain.

The experts committee recognized nintedanib as an intervention with a positive value contribution to the treatment of non-IPF PF-ILD and SSc-ILD, diseases which are considered as very severe and with high unmet needs, given the lack of alternative therapies which are either efficient or approved in the indications.

Nintedanib, when compared to placebo, was perceived as a treatment that provides an added therapeutic benefit for patients, given its proven clinical efficacy and slight improvements in patient-reported outcomes, with a good safety profile which, based on high quality evidence, will likely be positioned as a recommended therapy in the upcoming updates of the clinical practice guidelines.

These types of exercises allow us to understand where the value of health interventions lie from the perspective of the different agents, serving as a communication link between stakeholders, as well as a reference on the decision-making process regarding evaluation, financing, and reimbursement. In the future, it would be desirable to continue advancing in the development of the MCDA methodology and to extend its use, so that health decision making is carried out within a framework of greater transparency, consistency, and efficiency.

Supplementary Material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0266462322000459>.

Funding statement. This project was funded by Boehringer–Ingelheim Spain. Their participation was limited to sponsoring the project. They had no involvement in the online meetings held with the experts committee nor in the writing of this manuscript.

Conflicts of interest. The authors declare no conflicts of interest that are relevant to the content of this study. N.Z. is employed by the consultancy firm WEBER, which received funding for the development of this project. M.I.A.B., E.B., I.C., J.E., N.O., J.L.P., J.A.R.P., A.R., J.A.M.R., and L.V. received fees from the consultancy firm WEBER for the participation in the project.

References

- Zozaya González N, Oliva Moreno J, Hidalgo Vega A, et al (2018) *El análisis de decisión multi-criterio en el ámbito sanitario*. Fundación Weber, Madrid, Spain.
- Drake JI, de Hart JCT, Monleón C, Toro W, Valentim J (2017) Utilization of multiple-criteria decision analysis (MCDA) to support healthcare decision-making. *J Mark Access Health Policy*. 5, 1360545. <https://doi.org/10.1080/20016689.2017.1360545>.
- Travis WD, Costabel U, Hansell DM, et al (2013) An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 188, 733–748. <https://doi.org/10.1164/rccm.201308-1483ST>.
- Cottin V, Hirani NA, Hotchkiss DL, et al (2018) Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*. 27, 180076. <https://doi.org/10.1183/16000617.0076-2018>.
- Perelas A, Silver RM, Arrossi AV, Highland KB (2020) Systemic sclerosis-associated interstitial lung disease. *Lancet Respir Med*. 8, 304–320. [https://doi.org/10.1016/S2213-2600\(19\)30480-1](https://doi.org/10.1016/S2213-2600(19)30480-1).
- European Medicines Agency (EMA). *Ofev® (nintedanib)*. European Medicines Agency. Disponible en: <https://www.ema.europa.eu/en/medicines/human/EPAR/ofev>. Accessed 2021.
- Wijsenbeek M, Cottin V (2020) Spectrum of fibrotic lung diseases. *New Eng J Med*. 383, 958–968. <https://doi.org/10.1056/NEJMra2005230>.
- Thokala P, Devlin N, Marsh K, et al (2016) Multiple criteria decision analysis for health care decision making—An introduction: Report 1 of the ISPOR MCDA emerging good practices task force. *Value Health*. 19, 1–13. <https://doi.org/10.1016/j.jval.2015.12.003>.
- Goetghebeur MM, Cellier MS (2018) Can reflective multicriteria be the new paradigm for healthcare decision-making? The EVIDEM journey. *Cost Eff Resour Alloc*. 16, 54. <https://doi.org/10.1186/s12962-018-0116-9>.
- Marsh K, Ijzerman M, Thokala P, et al (2016) Multiple criteria decision analysis for health care decision making—Emerging good practices: Report 2 of the ISPOR MCDA emerging good practices task force. *Value Health*. 19, 125–137. <https://doi.org/10.1016/j.jval.2015.12.016>.
- Alerany. *Determining the value contribution of Ocrevus® (Ocrelizumab) for the treatment of multiple sclerosis in Spain by multi-criteria decision analysis (MCDA) ISPOR Scientific Presentations Database - Presentation details*. Disponible en: <https://tools.ispor.org/ScientificPresentationsDatabase/Presentation/87528>. Accessed 2019.
- Badia X, P de la Cueva, Moreda FR, Trillo JL, Guarga I (2017) Application of multi-criteria decision analysis (MCDA) to determine the value of treatments for the moderate to severe plaque psoriasis in Spain. *Value Health*. 20, A564. <https://doi.org/10.1016/j.jval.2017.08.937>.
- de Andrés-Nogales F, Casado MÁ, Trillo JL, et al (2020) A multiple stakeholder multicriteria decision analysis in diabetic macular edema management: The MULTIDEX-EMD study. *Pharmacoecoon Open*. 4, 615–624. <https://doi.org/10.1007/s41669-020-00201-2>.
- Hsu JC, Lin J-Y, Lin P-C, Lee Y-C (2019) Comprehensive value assessment of drugs using a multi-criteria decision analysis: An example of targeted therapies for metastatic colorectal cancer treatment. *PLoS One*. 14, e0225938. <https://doi.org/10.1371/journal.pone.0225938>.
- Goetghebeur MM, Wagner M, Houry H, et al (2010) Combining multi-criteria decision analysis, ethics and health technology assessment: Applying the EVIDEM decisionmaking framework to growth hormone for Turner syndrome patients. *Cost Eff Resour Alloc*. 8, 4.
- Wagner M, Houry H, Bennetts L, et al (2017) Appraising the holistic value of Lenvatinib for radio-iodine refractory differentiated thyroid cancer: A multi-country study applying pragmatic MCDA. *BMC Cancer*. 17, 272. <https://doi.org/10.1186/s12885-017-3258-9>.
- Zozaya N, Abdalla F, Alfonso Zamora S, et al (2022) Assessing the value contribution of bimekizumab for the treatment of moderate-to-severe psoriasis using a multidisciplinary reflective multi-criteria decision analysis. *Exp Rev Pharmacoecon Outcomes Res*. 1–13. <https://doi.org/10.1080/14737167.2022.2063842>.
- Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J (2014) Assessing the value of healthcare interventions using multi-criteria decision analysis: A review of the literature. *PharmacoEcon*. 32, 345–365. <https://doi.org/10.1007/s40273-014-0135-0>.
- Tony M, Wagner M, Houry H, et al (2011) Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): Field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. *BMC Health Serv Res*. 11, 329.
- Badia X, Gil A, Shepherd J (2018) PHP169 - MCDA EVIDEM reference value framework for drug evaluation and decision making in Spain. *Value Health*. 21, S179. <https://doi.org/10.1016/j.jval.2018.09.1063>.
- Wijsenbeek M, Kreuter M, Fischer A, et al (2018) Non-IPF progressive fibrosing interstitial lung disease (PF-ILD): The patient journey. *Am Thorac Soc*. 197, A1678.
- Flaherty KR, Wells AU, Cottin V, et al (2019) Nintedanib in progressive fibrosing interstitial lung diseases. *New Engl J Med*. 381, 1718–1727. <https://doi.org/10.1056/NEJMoa1908681>.
- Guler SA, Winstone TA, Murphy D, et al (2018) Does systemic sclerosis-associated interstitial lung disease burn out? Specific phenotypes of disease progression. *Ann Am Thorac Soc*. 15, 1427–1433. <https://doi.org/10.1513/AnnalsATS.201806-362OC>.
- Hoffmann-Vold A-M, Maher TM, Philpot EE, et al (2020) The identification and management of interstitial lung disease in systemic sclerosis: Evidence-based European consensus statements. *Lancet Rheumatol*. 2, e71–e83. [https://doi.org/10.1016/S2665-9913\(19\)30144-4](https://doi.org/10.1016/S2665-9913(19)30144-4).
- Gilbert-Perramon A, Torrent-Farnell J, Catalan A, et al (2017) Drug evaluation and decision making in Catalonia: Development and validation of a methodological framework based on multi-criteria decision analysis (MCDA) for orphan drugs. *Int J Technol Assess Health Care*. 33, 111–120. <https://doi.org/10.1017/S0266462317000149>.

26. **Guarga L, Badia X, Obach M**, et al (2019) Implementing reflective multi-criteria decision analysis (MCDA) to assess orphan drugs value in the Catalan Health Service (CatSalut). *Orphanet J Rare Dis.* **14**, 157. <https://doi.org/10.1186/s13023-019-1121-6>.
27. **Zozaya N, Martínez-Galdeano L, Alcalá B**, et al (2018) Determining the value of two biologic drugs for chronic inflammatory skin diseases: Results of a multi-criteria decision analysis. *BioDrugs.* **32**, 281–291. <https://doi.org/10.1007/s40259-018-0284-3>.
28. **Camps C, Badia X, García-Campelo R**, et al (2020) Development of a multicriteria decision analysis framework for evaluating and positioning oncologic treatments in clinical practice. *J Oncol Pract.* **16**, e298–e305. <https://doi.org/10.1200/JOP.19.00487>.
29. **Schey C, Krabbe PFM, Postma MJ, Connolly MP** (2017) Multi-criteria decision analysis (MCDA): Testing a proposed MCDA framework for orphan drugs. *Orphanet J Rare Dis.* **12**, 10. <https://doi.org/10.1186/s13023-016-0555-3>.
30. **ORPHAR-SEFH** (2020) *Manual para el desarrollo de un informe de evaluación de medicamentos huérfanos por parte del grupo ORPHAR-SEFH usando metodología de Análisis de Decisión Multicriterio.*
31. **Angelis A, Kanavos P** (2017) Multiple criteria decision analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The advance value framework. *Soc Sci Med.* **188**, 137–156. <https://doi.org/10.1016/j.socscimed.2017.06.024>.