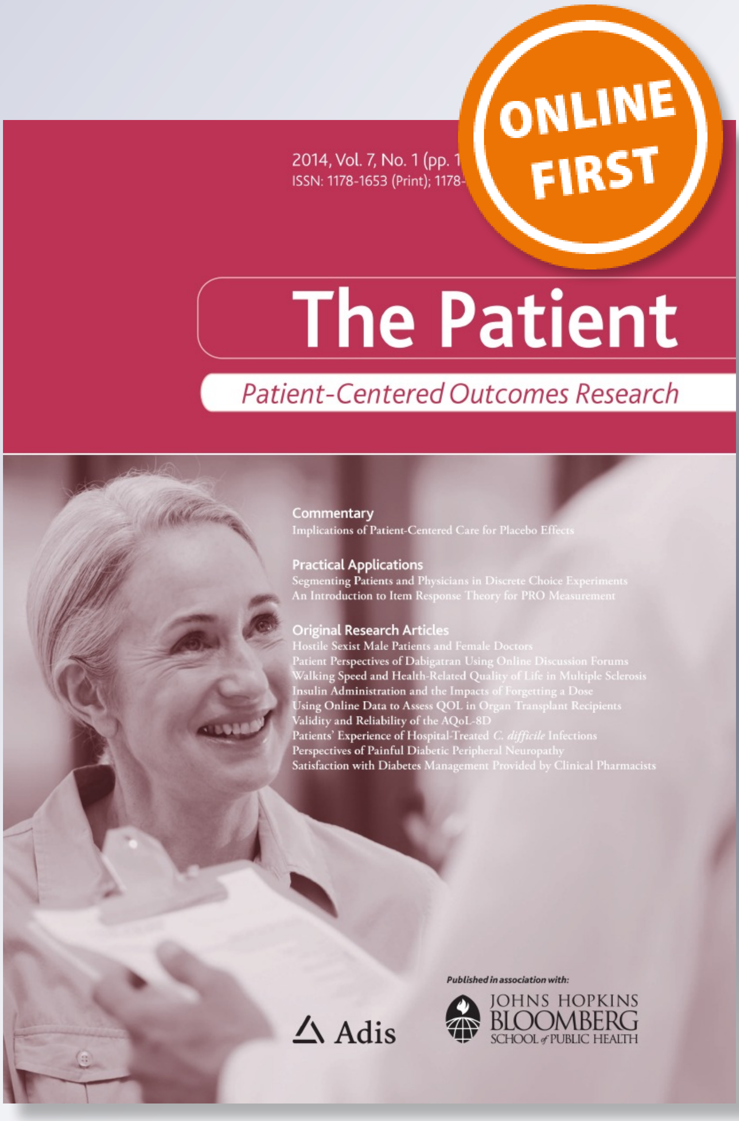
Group Decision Making with the Analytic Hierarchy Process in Benefit-Risk Assessment: A Tutorial

J. Marjan Hummel, John F. P. Bridges & Maarten J. IJzerman

The Patient - Patient-Centered Outcomes Research

ISSN 1178-1653

Patient
DOI 10.1007/s40271-014-0050-7



Patient
DOI 10.1007/s40271-014-0050-7

PRACTICAL APPLICATION

Group Decision Making with the Analytic Hierarchy Process in Benefit-Risk Assessment: A Tutorial

J. Marjan Hummel · John F. P. Bridges · Maarten J. IJzerman

Abstract The analytic hierarchy process (AHP) has been increasingly applied as a technique for multi-criteria decision analysis in healthcare. The AHP can aid decision makers in selecting the most valuable technology for patients, while taking into account multiple, and even conflicting, decision criteria. This tutorial illustrates the procedural steps of the AHP in supporting group decision making about new healthcare technology, including (1) identifying the decision goal, decision criteria, and alternative healthcare technologies to compare, (2) structuring the decision criteria, (3) judging the value of the alternative technologies on each decision criterion, (4) judging the importance of the decision criteria, (5) calculating group judgments, (6) analyzing the inconsistency in judgments, (7) calculating the overall value of the technologies, and (8) conducting sensitivity analyses. The AHP is illustrated via a hypothetical example, adapted from an empirical AHP analysis on the benefits and risks of tissue regeneration to repair small cartilage lesions in the knee.

Key Points for Decision Makers

In a step-by-step approach, it is illustrated how the analytic hierarchy process (AHP) can support groups making healthcare decisions.

The AHP facilitates the decision makers in discussing and valuing the multiple outcomes of alternative healthcare technologies.

The AHP can prioritize the healthcare technology to help decision makers in selecting the most valuable technology for patients.

1 Introduction

Multi-criteria decision analysis (MCDA) has been increasingly applied in healthcare [1]. One of the commonly applied MCDA techniques in healthcare is the analytic hierarchy process (AHP) [2]. It can support individual decision makers, as well as groups of decision makers [3]. The AHP aims to support shared decision making [4, 5], decisions on clinical guidelines [6, 7], decisions on the development of new technology [8, 9], organizational decisions [10, 11], and decisions on health policy [12–14], such as regulatory decisions, reimbursement decisions, or allocation of public research funding.

The purpose of this tutorial is to illustrate the use of the AHP to support group decision making. This tutorial provides information on the procedural steps of the AHP and provides recommendations on the organization of group panel sessions. A full illustration of the mathematical algorithms for AHP is beyond the scope of this paper and

J. M. Hummel (✉) · M. J. IJzerman
Department of Health Technology and Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands
e-mail: j.m.hummel@utwente.nl

J. F. P. Bridges
Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

J. F. P. Bridges
Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands

Published online: 13 March 2014

the reader may be referred for a detailed overview elsewhere [15, 16].

This tutorial is targeted at health outcomes researchers and policy makers interested in using the AHP, yet not experienced to do so. The tutorial illustrates each procedural step to undertake based on a hypothetical decision regarding the selection of a candidate to transfer from translation research to phase II clinical research. The potential candidate is tissue-engineered cartilage injected in the knee to repair small cartilage lesions. The expected benefits and risks of this treatment are compared with the benefits and risks of a currently applied treatment in clinical practice.

2 Setting the Stage

2.1 Administration of Group Judgments

If used in a group decision approach, the AHP can engage various stakeholders, including patients, care providers, researchers, and/or payers to value the multiple outcomes of healthcare technologies [8, 12]. Their judgments on the value of alternative technology can be administered through (online) questionnaires, electronic voting in a face-to-face group setting, or online voting in a dispersed group setting. In cases where judgments are collected by means of questionnaires, iterative Delphi rounds can be organized to reduce disagreements or inconsistent judgments. In a face-to-face group setting, or in a real-time dispersed group setting, the panel members can (online) share the arguments underpinning their judgments.

2.2 Group Facilitator

In a group setting, we recommend the group panel to be chaired by a facilitator who is able to understand the discussions about the new technology but who does not need to be an expert in this field. He or she should be competent in encouraging broad-based participation in the discussions, structuring and steering the communication processes, applying the AHP procedures, and have no personal interest in one of the decision alternatives to be selected [17].

2.3 Role of Software

Different software packages are available to provide interactive support to the group deliberations involved. These packages can enable the electronic submission of judgments on the value of healthcare technology, reveal the disagreements in judgments in the group, check the inconsistency in judgments, and present the (preliminary) results visualized in graphs. See the "Appendix" for a list of AHP-based software packages.

2.4 Informing the Group Members

To provide a common ground for sharing information, we recommend prior to the group session, sending an overview of the available evidence on the attributes of the alternative healthcare technology to compare. To be able to value the outcomes of the treatments, we advise the group members to be informed of the existing evidence on the outcomes of the new and existing treatments. In the case of group members having experience with one or more of the treatments, the experiences should be balanced over the treatments, so that during the group session information on all treatments can be shared. Moreover, it is informative to send in advance of the group session the program of the panel session, and information about the procedural steps of the AHP.

3 Procedural AHP Steps

The AHP distinguishes three stages in the decision-making process, i.e., (stage 1) structuring the decision problem to solve, (stage 2) evaluating the decision criteria and the decision alternatives, and (stage 3) categorizing, rank ordering, or prioritizing the decision alternatives. Figure 1 shows the three decision-making stages with the accompanying steps of the AHP. The eight procedural steps of the AHP will be explained and illustrated hereafter.

3.1 Problem Structuring

The first two AHP steps are to decompose and structure a complex decision problem. By breaking a decision problem into smaller sub-problems, the problem becomes more manageable.

3.1.1 Step 1: Defining the Decision Problem and Determining its Goal

As a first step, the decision problem and corresponding decision goal are defined. In general, the decision problem should be both relevant and complex enough to require a multi-criteria decision analysis. In defining the decision goal, assumptions need to be made explicit, such as from whose perspective the decision will be analyzed and who will, or should be affected by this decision.

Example A treatment with tissue-engineered cartilage has been developed to repair small cartilage lesions in the knee. Animal tests and phase I clinical trials have shown promising results. For instance, an expert panel with six

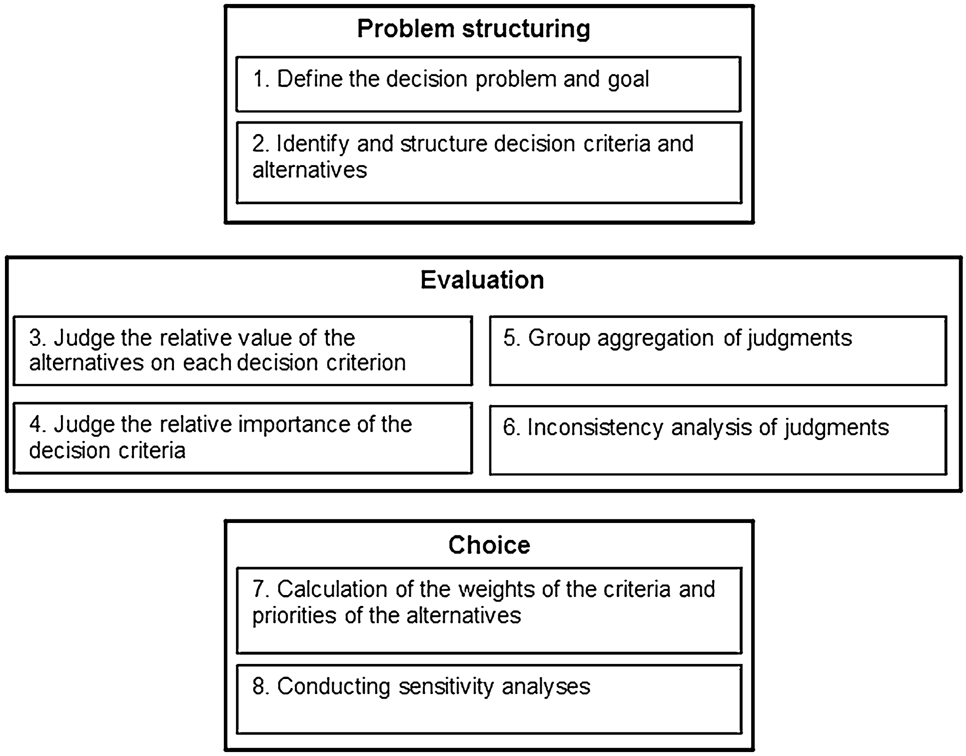


Fig. 1 The procedural steps of the analytic hierarchy process

orthopedic surgeons and two patient advocates is to advise regulators on the appropriateness of this tissue-engineered cartilage treatment as a candidate for phase II clinical trials. For the new treatment to be an appropriate candidate it should be preferred to the current treatment of cartilage damage; the creation of microfractures to stimulate cartilage growth. Accordingly, the goal of the expert panel's decision is to compare the benefits and risks of the two treatments of cartilage damage in the knee.

3.1.2 Step 2: Identifying and Structuring the Decision Alternatives and Criteria

When using the AHP, the decision problem to solve is represented as a hierarchical decision structure. In this decision structure, the goal of the decision is placed at the highest hierarchical level. The first intermediate level consists of the quantitative and/or qualitative criteria that are meaningful to the decision makers in comparing the alternatives. If required, each of these criteria can be sub-divided into a cluster of sub-criteria at the next intermediate level. For instance, a general criterion "health benefit" may be sub-divided into several specific benefits to the health of patients; the so-called sub-criteria. The lowest hierarchical level contains the decision alternatives. The decision alternatives are a finite set of alternatives that the decision makers aim to compare. These alternatives may, for instance, include treatment interventions still under development, treatments currently applied in clinical practice, and/or no treatment.

3.1.2.1 Organizational Setting The process of identifying and structuring the decision hierarchy can be managed in three distinct approaches:

- *Brainstorming session in a group setting* [18] The group members list all (sub)criteria and alternatives they deem important. The group facilitator clusters similar sub-criteria and relates each cluster of sub-criteria to a covering criterion. The proposed decision hierarchy is discussed and modified in the group until each level is composed of (sub)criteria that are mutually exclusive, clear, comprehensive, and are of importance within the same order of magnitude.

In cases where the most relevant decision criteria are known from the literature, we recommend one of the alternative options:

- *Preparation of the decision structure before the group session* [12] The criteria and alternatives to include in the analysis are derived from the literature, and/or interviews with experts.
- *Combination of the above* The decision structure is prepared in advance of the group session. During the group session, the prepared decision hierarchy is discussed. Appropriate relevant criteria and alternatives

that were neglected in the literature can be added to the decision structure. Criteria and alternatives that are inappropriate or irrelevant according to the latest insights of the experts can be deleted.

3.1.2.2 Structuring the Decision Hierarchy As in any approach for the MCDA, an adequate decision hierarchy identifying, specifying, and structuring the criteria is essential. The following recommendations are important to keep in mind during the problem-structuring stage:

- At each hierarchical level, the criteria and the sub-criteria need to be mutually exclusive. Furthermore, the criteria and alternatives need to be clearly defined to avoid misunderstandings.
- The decision structure needs to include as many as possible decision criteria that have a relevant impact on the preferences for the alternatives. In the case where a criterion is of negligible importance in comparison with the other criteria in its cluster, Saaty recommends the adaptation of the decision structure [15]. The criterion of relatively low importance could become a criterion on a lower hierarchical level, so, for instance, a criterion could become a sub-criterion of another covering criterion, or be removed from the decision structure.
- To create a well-managed decision structure, we recommend the number of criteria and the number of sub-criteria within each cluster to be between three and five [19].

Example The goal of the decision was to analyze if the new treatment with tissue-engineered cartilage was likely to become preferred to the dominant treatment in clinical practice to repair small cartilage lesions. Consequently, the decision alternatives were the new treatment (tissue-engineered cartilage), and the treatment generally accepted in clinical practice (microfracture technique). By means of a literature study, the decision criteria were identified. The decision criteria, used in this example, are the benefits and risks of the treatments: e.g., pain relief, initial improvement of knee function, prevention of wear, adverse events such as infection, bone disruption, swelling, and rehabilitation load as caused by the duration of rehabilitation and the restrictions imposed upon the patients during rehabilitation. A hierarchical decision structure was proposed by the facilitator of the panel. After discussion, criteria related to the efficiency and user friendliness of the surgical procedures were added. The orthopedic surgeons considered these criteria to be important to the clinical acceptance of a new surgical treatment. Subsequently, the group panel approved of the decision hierarchy. The structure is depicted in Fig. 2.

3.2 Evaluation

In the evaluation stage, each group member judges the relative value of the alternatives on the decision criteria,

Fig. 2 The hierarchical decision structure

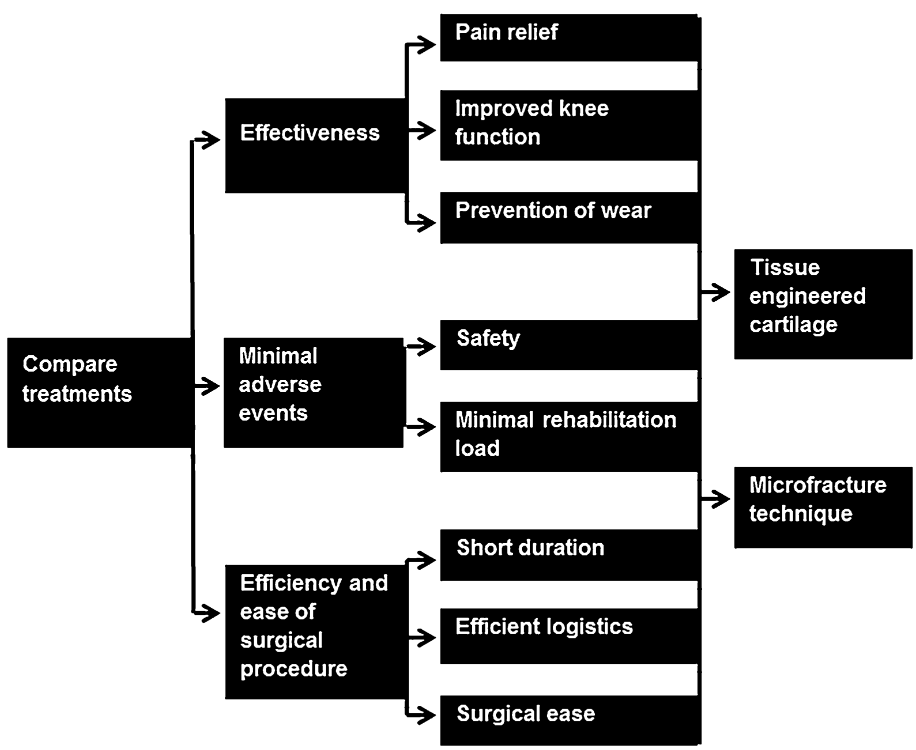


Table 1 Original analytic hierarchy process scale

Numerical rating	Verbal judgments
9	Extremely more important or preferred
8	Very strongly to extremely more important or preferred
7	Very strongly preferred more important or preferred
6	Strongly to very strongly more important or preferred
5	Strongly preferred more important or preferred
4	Moderately to strongly more important or preferred
3	Moderately preferred more important or preferred
2	Equally to moderately more important or preferred
1	Equally preferred more important or preferred

and judges the relevance of the criteria and sub-criteria. The individual judgments are aggregated into group judgments, and feedback is provided on the consistency in judgments. These four AHP steps that belong to the evaluation stage are explained and illustrated hereafter.

3.2.1 Step 3 and 4: Judging the Relative Value of Alternatives and Criteria

3.2.1.1 The Verbal Rating Scale In pairwise comparisons between criteria, the group members compare two criteria on their importance. Most commonly, these comparisons are judged on a verbal nine-point rating scale. If criteria are judged to be equally important, both criteria are assigned a score of one. If one of the criteria is judged to be more important than the other one, the more important criterion is assigned a score from 2 up to 9. A 2 represents a value between equally to moderately more important, and 9 represents extremely more important (see Table 1).

Likewise, the importance of each pair of sub-criteria stemming from the same cluster of sub-criteria is compared. Sub-criteria in different clusters are not compared directly. For instance, the importance of pain relief is directly compared with the improvement of the knee function. Both sub-criteria are related to the effectiveness of the treatments. Pain relief is not pairwise compared with one of the adverse events. On a similar nine-point scale, the preferences for the alternatives are compared in pairs with regard to each sub-criterion. In this case, 1 reflects equal preference, and 9 reflects extremely higher preference. In our example, the preferences for the tissue-engineered cartilage treatment and the microfracture technique are compared regarding, for example, these treatments' impact on pain relief.

3.2.1.2 The Original and Alternative Numerical Scales The nine-point AHP scale has the properties of a ratio scale. In the original AHP scale, the verbal judgments are converted into the associated numerical ratings in Table 1. Accordingly, an extremely higher importance is, for instance, assumed to have a nine-times higher importance.

The validity of these numerical ratings has been extensively discussed in the literature. These ratings may not accurately reflect the value judgments on the pairwise comparisons [20]. In a response to this discussion, alternative numerical AHP scales have been developed. For example, Saaty suggested a nine-point scale that has a range between 1.1 and 1.9 to compare alternatives that differ only slightly [21]. Other investigators have proposed alternative linear, geometric, or logarithmic scales [22–25]. A possible solution is to use a continuous graphic mode of judging the pairwise comparisons, which is offered in some software packages. This continuous scale offers the possibility of small incremental steps in changing relative priorities.

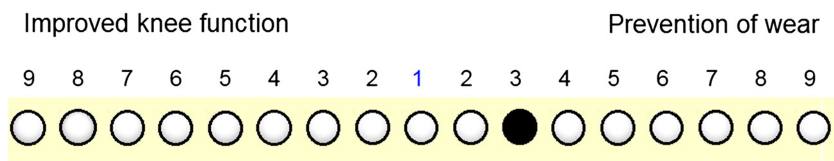
3.2.1.3 Framing the Pairwise Comparisons The importance of each pair of criteria is compared with respect to the goal of the decision problem; in this case, the question is which criterion is more important in comparing the value of the alternative technology. The importance of the sub-criteria is compared with respect to the criterion at the higher hierarchical level in the decision hierarchy. In this case, the question is which sub-criterion is more important in fulfilling the covering criterion (see example). To increase comparability, all (sub)criteria are framed as positive measures of value. Adverse events, for example, can be framed as the minimal adverse events.

Example Which sub-criterion do you consider to be more important in valuing the effectiveness of the cartilage treatments, and to what extent it is more important?

In this example, the sub-criterion: the prevention of wear was rated to be *moderately* more important than the sub-criterion: improvement of the knee function. At the long term, the prevention of wear was considered to be more important to the effectiveness of the cartilage treatment than the initial improvement of the knee function.

3.2.1.4 Amount of Pairwise Comparisons To compare n criteria, a cluster of n sub-criteria, or to compare n alternatives with respect to a criterion, one needs to make $n(n-1)/2$ pairwise comparisons [15]. For example, in comparing three criteria, three pairwise comparisons need to be made: criterion 1 is compared with criterion 2, criterion 2 is compared with criterion 3, and criterion 1 is compared with criterion 3. Reciprocity in judgments is assumed. This means, for example, that after having

Fig. 3 A pairwise comparison of sub-criteria



compared criterion 1 with criterion 2, one does not need to compare criterion 2 with criterion 1.

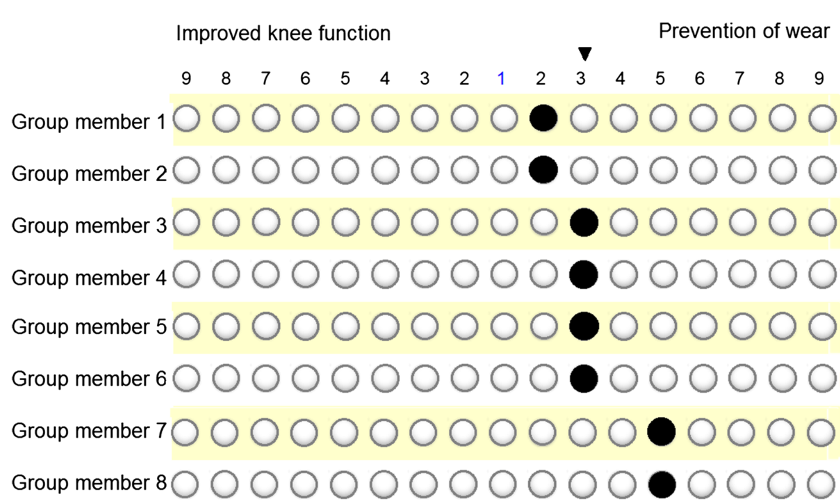
Example To weight all (sub-)criteria in our example, ten pairwise comparisons were judged: three pairwise comparisons to weight the three criteria, three pairwise comparisons to weight the three sub-criteria of effectiveness, one pairwise comparison to weight the two sub-criteria of adverse events, and three pairwise comparisons to weight the three sub-criteria related to the surgical procedure. In addition, eight pairwise comparisons were made to prioritize the two treatment alternatives on all eight sub-criteria. The total of pairwise comparisons in the full AHP analysis was 18 (Fig. 3).

3.2.1.5 Relative or Direct Rating of Alternatives The pairwise comparison approach is used to judge the relative value of a limited amount of alternatives. In the case where large numbers of alternatives are to be prioritized, it is possible to directly value the alternatives on qualitative or quantitative intensity scales [14, 26]. An intensity scale can, for example, have the intensity levels: excellent, above average, average, below average, and poor. The highest intensity level receives the priority 1 and the other levels are proportionally smaller [26]. These priorities can be established by means of pairwise comparisons. The alternatives are rated by selecting the appropriate intensity levels. The advantage of this direct rating technique is that once the intensity levels have been prioritized, each alternative can directly be rated with the priority of the corresponding intensity level. In the case of large amounts of alternatives, this approach will avoid a laborious set of pairwise comparisons among the alternatives themselves. Relative rating with the pairwise comparisons approach and direct rating on intensity scales are particularly suitable where insufficient quantitative evidence is available on the value of the alternatives. When sufficient quantitative evidence exists, it is also possible to directly convert absolute data on the value of the alternatives into priorities [26]. For an example, see Hummel et al. [12].

3.2.1.6 Bottom-Up or Top-Down Valuation Because the weights of the criteria may be dependent on how well the set of alternatives fulfill these criteria [27], we recommend a bottom-up approach of conducting the pairwise comparisons. This means that the relative priorities of the alternatives on the criteria have to be evaluated first, after which the weights for the criteria can be judged. Conversely, in a top-down approach, first the weights for the criteria are evaluated, and subsequently the priorities of the alternatives. For an example, see the work by Steele et al. [27] for an explanation and discussion of dependency between criterion weights and priority values of the alternatives.

Example The panel session was conducted in a face-to-face setting. The chairman informed the panel members about the background of the decision problem, and illustrated the procedures of the AHP. The first clinical evidence on the treatments was presented. Using hand-held scoring keypads, the panel members compared their preferences for the two treatments and subsequently the importance of the criteria and sub-criteria using the original nine-point scale. Individual judgments on the pairwise

Fig. 4 Judgments in the group on the pairwise comparison



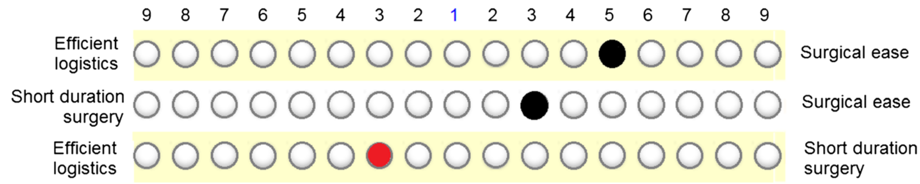


Fig. 5 Inconsistent judgments on the pairwise comparisons

comparisons were projected on a screen, allowing the members of the panel to discuss the rationales behind their individual scores (see Fig. 4). During the discussions, the panel members could alter their judgments.

3.2.2 Step 5: Group Aggregation

Group judgments can be set by means of a consensus vote on the pairwise comparisons. If a group consensus is unwanted, individual judgments can be aggregated [28]. Aggregation of individual judgments can take place during the evaluation stage, or during the later choice stage of the decision-making process. During the evaluation stage, differences in judgments on the pairwise comparisons can be reduced by discussing previously unshared information about the properties of the criteria or alternatives that were inconsistently compared.

The group average of the final scores on each pairwise comparison is calculated to reflect the opinion of the group as a whole. As the pairwise comparisons are rated on a ratio scale, the geometric mean is used to calculate the average score on each pairwise comparison. In the choice stage, weights and priorities are calculated using these group averages. Alternatively, in a setting that could be better described as negotiated decision making, the group members individually make the pairwise comparisons. Only in the last stage of the decision analysis; the choice stage, the individual outcomes are aggregated. When averaging the weights and priorities of the individual group members, the arithmetic mean is used. In this setting, the group members only need to agree upon the final choice for one of the alternatives, irrespective of the differences in rationale behind this choice.

Example As depicted in Fig. 4, four group members considered the sub-criterion "prevention of wear" to be moderately more important (score 3) than the sub-criterion "improved knee function". Two other group members stated the prevention of wear to be only slightly more important (score 2), and the last two group members were convinced that it is strongly more important (score 5) than the improvement of the knee function. After the group deliberations, the group score is to reflect the opinion of the group as a whole. Accordingly, the group score is calculated with the geometric mean of the scores on the pairwise comparisons, which is in this example: $(2 \times 2 \times 3 \times 3 \times 3 \times 3 \times 5 \times 5)^{(1/8)} = 3.08$. If the aim had been to aggregate the judgments of individual decision makers, weights and priorities would, in the later-choice stage, be calculated for each of the decision makers. These weights and priorities would, subsequently, be averaged to attain the group results.

3.2.3 Step 6: Inconsistency Analysis

After each set of pairwise comparisons (comparison of criteria; pairwise comparisons of sub-criteria within each cluster; or comparison between alternatives regarding each sub-criterion), a consistency ratio (CR) is calculated [15, 16]. The CR shows if each pairwise comparison is logically sound with regard to the remainder of the comparisons. It indicates the degree to which the pairwise judgments resemble a purely random set of pairwise comparisons. Judgments that have a CR lower than 0.1 are reasonable, lower than 0.2 is tolerable, and higher than 0.2 should be revised or discarded [15]. In the case of higher inconsistency, the decision makers are advised to check for accidental mistakes and to reconsider their pairwise comparisons, until the consistency measure is below the threshold indicated (Fig. 5) [15].

Example The first two comparisons indicated that "surgical ease" is *strongly* more important than "efficient logistics", and *moderately* more important than the "short duration of surgery". This suggests that the "short duration of surgery" is more important than the "efficient logistics". Conversely, in the third pairwise comparison, "efficient logistics" was rated to be *moderately* more important than "short duration of surgery". The resulting consistency ratio is 0.28, suggesting the need to revise.

Revision of the last score into a *slightly* higher importance of "short duration of surgery" resulted in the acceptable consistency ratio of 0.00.

3.3 Prioritizing Alternatives

In the last stage of the decision-making process, overall priorities are calculated for the alternative technologies. Alternatives with a higher priority are assumed to be more valuable, or more preferred. The overall priorities can be used to select the most preferred alternative; to rank order the alternatives from most preferred to least preferred; or to determine the relative value of these alternatives. Subsequently, in a sensitivity analysis, the robustness of the preferences for the alternatives can be analyzed. The results can be used to underpin a decision about one of the healthcare technologies. This final decision does not need to be made by the group panel. It can be made by another formal decision-making body, being informed by the results of the AHP analysis.

3.3.1 Step 7: Calculation of Weights and Priorities

3.3.1.1 Calculation of Weights for the Criteria When inconsistency is reduced to an acceptable degree, Saaty recommends the calculation of weighting factors and performance priorities by using the principal right eigenvector approach [15, 16]. This eigenvector method can be interpreted as a simple averaging process by which the final weights are the average of all possible ways of comparing the scores on the pairwise comparisons. A higher weight assigned to one of the criteria reflects a higher importance of this criterion (see Table 2, discussed later in the example). Alternative approaches exist to calculate weights and priorities, among others, the frequently cited geometric means approach [29].

3.3.1.2 Local Weights and Global Weights of Sub-Criteria When pairwise comparing sub-criteria, local weights are calculated for the sub-criteria. The local weights of the sub-criteria in any cluster add up to 1. Global weights of the sub-criteria are calculated by multiplying the local weights of the sub-criteria with the weight of the covering criterion. Consequently, the global weights of the sub-criteria within the same cluster sum to the weight of the covering criterion.

Example As calculated from the revised pairwise comparisons in the previous example, the local weights of sub-criteria duration, logistics, and surgical ease were respectively, 0.23, 0.12, and 0.65. The local weights of these sub-criteria sum up to 1. To calculate the global weights of these three sub-criteria, their local weights are multiplied with the weight of the covering criterion "efficiency and ease of surgical procedure" (0.14), resulting in the global weights of duration, logistics, and surgical ease of respectively 0.03, 0.02, and 0.09. These latter weights can be found in Table 2.

3.3.1.3 Calculating the Priorities for the Alternatives In a similar manner, the priorities of the alternatives are calculated regarding each of the criteria. A higher priority reflects a stronger preference for the corresponding alternative. After knowing the priorities of the alternatives on all sub-criteria, the AHP software uses an additive value function to calculate the overall priorities for the alternatives. The overall priority is the weighted average of all priorities: the sum of the priority of this alternative on each criterion multiplied by the weight of the corresponding criterion [15, 16].

3.3.1.4 The Overall Prioritization of Alternatives: Ideal vs. Distributive Mode In calculating the priorities of the

Table 2 Global weights, priorities, and overall priorities

Criteria	Sub-criteria	Tissue-engineered cartilage	Microfracture technique
Effectiveness (0.67)	Pain relief (0.27)	0.73	0.27
	Improvement knee function (0.10)	0.69	0.31
	Prevention of wear (0.30)	0.77	0.23
Minimal adverse events (0.19)	Safety (0.13)	0.54	0.46
	Minimal rehabilitation load (0.06)	0.75	0.25
Efficiency and ease of surgical procedure (0.14)	Short duration (0.03)	0.51	0.49
	Efficient logistics (0.02)	0.15	0.85
	Surgical ease (0.09)	0.45	0.55
Overall	(1.00)	0.67	0.33

Notes: The weights of the (sub-)criteria are within brackets under the corresponding (sub-)criteria; in the cells are the priorities of the treatment alternatives. In the last row, the overall priorities of the treatments are given alternatives, it is possible to choose from two different modes of synthesis [30]. The distributive mode should be used if the performance of an alternative is dependent on the performance of all other alternatives. The distributive mode can be appropriate if, for example, the decision problem includes only two decision alternatives that are relevant to take into consideration. This mode normalizes the priorities of the alternatives so that the priorities of all alternatives together sum up to 1. The ideal synthesis mode should be used if the decision maker is concerned with how well each alternative performs only relative to one benchmark alternative. In the ideal mode, the priorities are normalized by dividing the priority of the alternative under consideration by the score of the benchmark alternative. In this manner, the priority of an alternative under consideration is only dependent on the priority of the benchmark alternative and not on the priorities of the other alternatives. In a clinical setting, the fixed benchmark could be the gold standard of treatment. If comparing (new) technologies with the gold standard in a clinical setting, we recommend applying the ideal synthesis mode with the gold standard as the fixed benchmark. In the case where no benchmark alternative is available, the most preferable alternative under each criterion or sub-criterion is in the ideal mode of synthesis assigned the full priority of the (sub)criterion. The other alternatives receive a priority proportional to their preferences relative to the most preferred alternative [31].

A point of criticism on the AHP focuses on the possibility that the rank order of the prioritized decision alternatives can change [20]. When using the distributive mode of synthesis, the rank order can change when adding new decision alternatives to the analysis. Particularly when similar alternatives are added to the decision analysis, the rank order of the original alternatives might change. This counterintuitive rank reversal is caused by the use of relative priorities in combination with the additive value function of the original AHP. Namely, the overall priority of an alternative depends on how all other alternatives perform. One solution is to apply the ideal mode of synthesis of the weights and priorities using a fixed benchmark alternative. Accordingly, the rank order of the alternatives is preserved when adding or deleting other alternatives besides the benchmark alternative [16]. In the case where the distributive mode is used, synthesis in the ideal mode can be applied in the sensitivity analysis to explore the possibility of a rank reversal of alternatives. A more fundamental solution to avoid rank reversals has been proposed by Lootsma. He suggested the use of a multiplicative value function instead of the additive value function in the multiplicative AHP [32, 33].

3.3.1.5 Discussion and Approval of the Results After showing the weights for the (sub-)criteria, the priorities of the alternatives regarding each criterion, and the overall priorities of the alternatives, the validity of these outcomes is discussed. If desired after these discussions, the judgments on the pairwise comparisons can be adapted.

3.3.2 Step 8: Conducting Sensitivity and Heterogeneity Analyses

The alternative with the highest overall priority is considered to be the preferred option in the decision tree, logically followed by rank 2 and further. By gradually changing the weight of each criterion and the priorities of each technology, one can check if the initial rank order of technologies is likely to reverse. In this simple deterministic sensitivity analysis, the robustness of the decision outcomes is studied. A procedure for sensitivity analysis on the weights of criteria has been suggested by Mareschal [34]. A procedure for sensitivity analysis that also includes altering the priorities of the alternatives has been suggested by Triantaphyllou and Sanchez [35].

In addition, it can be relevant to study the heterogeneity in priorities among subgroups in the panel. In our example, it could be relevant to examine the priorities of the clinicians vs. the priorities of the patient advocates. See Dolan et al. [4] for an example of the analysis of patient heterogeneity.

Example Judgments were revised after group discussions and warnings of excessive inconsistencies in pairwise comparisons. Based on the final group judgments, priorities of the two treatments were calculated with the principal right eigenvector approach. The following table shows the global weights of the sub-criteria and the priorities of the treatments. By presenting the global weights, a comparison is allowed between the importance of the sub-criteria over all clusters of sub-criteria.

The performance of the new treatment with tissue-engineered cartilage was valued relative to the performance of the microfracture technique. Only in the case where the tissue-engineered cartilage treatment was expected to perform better than the microfracture technique performs, did the tissue-engineered cartilage treatment receive a priority higher than 0.50. This means that the priority of the tissue-engineered cartilage treatment was made dependent on the priority of all other alternatives; in this case, solely the microfracture technique. Accordingly, it was appropriate to use the distributive mode of prioritizing the alternatives. If there had been more alternatives available, and the priority of the tissue-engineered cartilage treatment was only to be dependent on the priority of the microfracture technique and not on the priorities of the other alternatives, the distributive mode would not have been appropriate. Then, the ideal mode of synthesis should have been used, with the microfracture technique as a benchmark alternative.

The weights show that the group panel considered the prevention of wear (weight 0.30) and the relief of pain (weight 0.27) to be the most important decision criteria. Particularly because of the high priority of the tissue-engineered cartilage treatment on these criteria, this alternative has the highest overall priority as well. The advantages related to its effectiveness and its minimal adverse events weigh up against the disadvantages related to its inefficient surgical procedure. The tissue-engineered cartilage's overall priority is $(0.27 \times 0.73) + (0.10 \times 0.69) + (0.30 \times 0.77) + (0.13 \times 0.54) + (0.06 \times 0.75) + (0.03 \times 0.51) + (0.02 \times 0.15) + (0.09 \times 0.45) = 0.67$.

In a sensitivity analysis, the impact of safety on the overall preferences was examined. The group varied strongly in opinion on the potential safety of the tissue-engineered cartilage relative to the microfracture technique. When lowering the priority of tissue-engineered cartilage on safety to the lowest possible priority (priority = 0.10), the overall priority of tissue-engineered cartilage would reduce from 0.67 to 0.61. Accordingly, a different value assigned to the safety of tissue-engineered cartilage was impossible to evoke a rank reversal of alternatives.

On the basis of these outcomes, the expert panel concluded that the tissue-engineered cartilage treatment was an appropriate candidate for clinical trials in the treatment of small cardiac lesions in the knee. Nevertheless, considering the inefficiency and complexity of the surgical procedure of the tissue-engineered cartilage treatment, the panel recommended a careful selection of the appropriate hospitals to involve in the clinical trials. Sufficient skills and resources should be available to the surgeon and the surgical team.

4 Discussion

The AHP can support the decision-making process to arrive at a decision that the panel members trust and are able to rationalize. High inconsistencies in judgments can indicate the need to further clarify the definitions of the criteria, and to discuss counterintuitive, or uncertain priorities of the technologies. Disagreements in judgments can show the need to share more information. The judgments can be adapted until the group members are satisfied with the decision outcomes. The impact of remaining differences in judgments on the decision outcomes can be analyzed in the sensitivity analysis.

However, as in all group settings, group dynamics may also negatively impact the decision outcomes. Peer pressure may evoke group members to revise their judgments, or group members may deliberately attempt to steer the results by submitting too extreme judgments. See for instance, Hummel et al. on the impact of the AHP on group dynamics in sociodynamic processes in group decision making [36].

This tutorial illustrates the basic procedures of the AHP. More advanced analyses are possible as supported by the approaches of fuzzy AHP [37] and the analytic network process [38]. Instead of judging the pairwise comparisons in one deterministic number, fuzzy AHP uses a fuzzy scale covering multiple numbers. The analytic network process explicitly takes into account interdependencies among criteria. New advancements are biannually discussed by academics and practitioners at the International Symposium on the AHP (see http://www.ISAHP.org). Besides the AHP, other methods for MCDA can be suitable to support the comparison of healthcare technology [39–41]. Alternatives to the AHP that emphasize the deliberative support to the decision-making process as well are, for example, the Simple Multi-Attribute Rating Technique (SMART) [42, 43], Swing weighting procedures[44], or Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) [45].

5 Conclusion

The AHP supports the processes of deliberation in making group decisions. When taking into account the lessons learnt from the ongoing scientific discussions on the methodology, it can be appropriately applied to gain an overview of the mean advantages and disadvantages of new healthcare technology in comparison with a benchmark alternative. This result helps to underpin the selection of candidates for further development, clinical trials, or full health economic analyses. Moreover, the AHP can support decisions that cannot be based on considerations of cost effectiveness alone.

Acknowledgments The authors declare they have no conflicts of interest. Marjan Hummel contributed to conceptualizing the paper, developing the procedural steps of the AHP and the hypothetical case study, and to writing the draft manuscript and revising the final manuscript. John Bridges contributed to conceptualizing the paper, and reviewing and revising the drafts and final manuscript. Maarten IJzerman contributed to the conceptualization of the paper, the development of the hypothetical case study, and reviewing and revising the drafts and final manuscript. Marjan Hummel acts as a guarantor for the content of the article.

Appendix

See Table 3

Table 3 Examples of analytic hierarchy process-based software packages

Software package	Internet source
Team Expert Choice	http://www.expertchoice.com
Decision Lens	http://www.decisionlens.com
HIPRE 3+	http://sal.aalto.fi/en/resources/downloadables/hipre3
SuperDecisions	http://www.superdecisions.com
SelectPro Decision Support Software	http://www.selectprosoftware.com
EasyMind	http://www.community.easymind.info
MakeItRational	http://www.makeitrational.com/analytic-hierarchy-process/ahp-software
TransparentChoice	http://www.transparentchoice.com
MindDecider Team	http://www.minddecider.com

References

1. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: a bibliometric analysis. Oper Res Health Care. 2013;2:20–4.
2. Liberatore MJ, Nydick RL. The analytic hierarchy process in medical and health care decision making: a literature review. EJOR. 2008;189(1):294–307.
3. Dyer RF, Forman EH. Group decision support with the analytic hierarchy process. Decis Support Syst. 1992;8:99–124.
4. Dolan JG, Boohaker E, Allison J, Imperiale TF. Patients' preferences and priorities regarding colorectal cancer screening. Med Decis Making. 2013;33(1):59–70.
5. Kitamura Y. Decision-making process of patients with gynecological cancer regarding their cancer treatment choices using the analytic hierarchy process. Japan J Nurs Sci. 2010;7(2):148–57.
6. Singh S, Dolan JG, et al. Optimal management of adults with pharyngitis: a multi-criteria decision analysis. BMC Med Inform Decis Mak. 2006;6:14.
7. Van Til JA, Renzenbrink GJ, Dolan JG, IJzerman MJ. The use of the analytic hierarchy process to aid decision making in acquired equinovarus deformity. Arch Phys Med Rehabil. 2008;89(3):457–62.
8. Hilgerink MP, Hummel JM, Manohar S, et al. Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis. Med Devices (Auckl). 2011;4:107–15.
9. Kim K, Kyung T, Kim W, et al. Efficient management design for swimming exercise treatment. Korean J Physiol Pharmacol. 2009;13(6):497–502.
10. Li XJ, Bin GF, Dhillon BS. Model to evaluate the state of mechanical equipment based on health value. Mech Mach Theory. 2011;46(3):305–11.
11. Baykasoglu A, Dereli T, et al. Application of cost/benefit analysis for surgical gown and drape selection: a case study. Am J Infect Control. 2009;37(3):215–26.
12. Hummel JM, Volz F, van Manen JG, Danner M, et al. Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment. Patient. 2012;5(4):225–37.
13. Kim W, Han SK, Oh KJ, et al. The dual analytic hierarchy process to prioritize emerging technologies. Technol Forecast Social Change. 2010;77(4):566–77.
14. Smith J, Cook A, Packer C. Evaluation criteria to assess the value of identification sources for horizon scanning. Int J Technol Assess Health Care. 2010;26(3):348–53.
15. Saaty TL. The analytic hierarchy process: planning, priority setting, resource allocation. New York: McGraw-Hill; 1980.
16. Saaty TL. Highlights and critical points in the theory and application of the analytic hierarchy process. Eur J Oper Res. 1994;74:426–47.
17. DeSanctis G, Gallupe RB. A foundation for the study of group decision support systems. Manag Sci. 1987;33:589–609.
18. Hummel JM, van Rossum W, Verkerke GJ, Rakhorst G. Product design planning with the analytic hierarchy process in inter-organizational networks. R&D Manag. 2002;32(5):451–8.
19. Murphy CK. Limits of the analytical hierarchy process from its consistency index. Eur J Oper Res. 1993;65:138–9.
20. Holder RD. Some comments on the analytic hierarchy process. J Opl Res Soc. 1990;41(11):1073–6.
21. Saaty TL. Decision making with the analytic hierarchy process. Int J Serv Sci. 2008;1(8):83–98.
22. Lootsma FA. Conflict resolution via pairwise comparison of concessions. Eur J Opl Res. 1989;40(1):109–16.
23. Beynon M. An analysis of distributions of priority values from alternative comparison scales within AHP. Eur J Oper Res. 2002;140(1):104–17.
24. Salo AA, Hämäläinen RP. On the measurement of preference in the analytic hierarchy process. J Multi Crit Decis Anal. 1997;6(6):309–19.
25. Ishizaka A, Balkenborg D, Kaplan T. Influence of aggregation and measurement scale on ranking a compromise alternative in AHP. J Oper Res Soc. 2011;62(4):700–10.
26. Saaty TL. Rank from comparisons and from ratings in the analytic hierarchy/network processes. Eur J Oper Res. 2006;168:557–70.
27. Steele K, Carmel Y, Cross J, Wilcox C. Uses and misuses of multicriteria decision analysis (MCDA) in environmental decision making. Risk Anal. 2009;29(1):26–33.
28. Forman E, Peniwati K. Aggregating individual judgments and priorities with the analytic hierarchy process. Eur J Oper Res. 1998;108:165–9.
29. Ishizaka A, Labib A. Review of the main developments in the analytic hierarchy process. Expert Syst Appl. 2011;38(11):14336–45.
30. Millet I, Saaty TL. On the relativity of relative measures: accommodating both rank preservation and rank reversals in the AHP. Eur J Oper Res. 2000;121(1):205–12.
31. Forman EH, Gass SI. The analytic hierarchy process: an exposition. Oper Res. 2001;49(4):469–86.
32. Lootsma FA. Scale sensitivity in a multiplicative variant of the AHP and SMART. J Multi Crit Decis Anal. 1993;2:87–110.
33. Stam A, Duarte Silva AP. On multiplicative priority rating methods for the AHP. Eur J Oper Res. 2003;145(1):92–108.
34. Mareschal B. Weight stability intervals in multicriteria decision aid. Eur J Oper Res. 1988;33:54–64.
35. Triantaphyllou E, Sanchez A. A sensitivity analysis approach for some deterministic multi-criteria decision making methods. Decis Sci. 1997;28:151–94.
36. Hummel JM, Van Rossum W, Verkerke GJ, Rakhorst G. The effects of Team Expert Choice on group-decision making in collaborative new product development, a pilot study. J Multi Crit Decis Anal. 2000;9(1–3):90–8.
37. Kahraman C, Cebeci U, Ulukan Z. Multi-criteria supplier selection using fuzzy AHP. Logist Inf Manag. 2003;16(6):382–94.
38. Saaty TL, Vargas LG. Decision making with the analytic network process: economic, political, social and technological applications with benefits, opportunities, costs and risks. New York: Springer Science and Business Media; 2006.

39. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value Health. 2012;15:1172–81.
40. Valerie Belton V, Theodor J, Stewart TJ. Multiple criteria decision analysis an integrated approach. Dordrecht: Kluwer Academic Publishers; 2002.
41. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge: Cambridge University Press; 1993.
42. Lootsma FA, Schuijt H. The multiplicative AHP, SMART and ELECTRE in a common context. J Multi Crit Decis Anal. 1997;6:185–96.
43. Edwards W, Barton FH. Smarts and smarter: improved simple methods for multi attribute utility measurement. Organ Behav Human Decis Process. 1994;60:306–25.
44. von Winterfeldt D, Edwards W. Decision analysis and behavioral research. New York: Cambridge University Press; 1986.
45. Bana e Costa CA, Chagas MP. A career choice problem: an example of how to use MACBETH to build a quantitative value model based on qualitative value judgments. Eur J Oper Res. 2004;153(2):323–31.