



A Guide to Measuring and Interpreting Attribute Importance

Juan Marcos Gonzalez¹

Published online: 25 March 2019
© Springer Nature Switzerland AG 2019

Abstract

Stated-preference (SP) methods, such as discrete-choice experiments (DCE) and best–worst scaling (BWS), have increasingly been used to measure preferences for attributes of medical interventions. Preference information is commonly characterized using attribute importance. However, attribute importance measures can vary in value and interpretation depending on the method used to elicit preferences, the specific context of the questions, and the approach used to normalize attribute effects. This variation complicates the interpretation of preference results and the comparability of results across subgroups in a sample. This article highlights the potential consequences of ignoring variations in attribute importance measures, and makes the case for reporting more clearly how these measures are obtained and calculated. Transparency in the calculations can clarify what conclusions are supported by the results, and help make more accurate and meaningful comparisons across subsamples.

Key Points for Decision Makers

This article provides a guide to evaluate the concept of attribute importance, and how it is elicited in discrete-choice experiments and best–worst scaling studies.

The concept of importance varies across applications, partly because it can be estimated using different elicitation methods that result in subtly different versions of importance.

Even with the same elicitation method, different approaches to normalize importance weights can complicate comparisons of results.

Practitioners should clearly report how attribute importance is calculated to ensure that conclusions are properly supported by the study results.

hedonic principle that preferences for interventions can be defined as the aggregation of preferences for their fundamental characteristics (attributes) [4]. Attributes of medical interventions include health outcomes, convenience factors, and cost, among others [2].

DCEs and Case 3 BWS elicit attribute importance by asking respondents to choose or rank profiles of medical interventions. The profiles are defined in terms of attributes and the various manifestations of the attributes (attribute levels) achievable with each intervention [5]. An experimental design is used to systematically vary the attribute levels among profiles. Choices and rankings are assumed to respond to the differences in the attributes between profiles. Thus, by stating their preference between profiles, respondents provide an indirect way to determine attribute importance. Other versions of BWS ask respondents to directly rank the attributes (Case 1) or attribute levels (Case 2), where each question only elicits a partial ranking (i.e., best and worst) of the options [5, 6]. With Case 1 and Case 2 BWS, respondents provide direct information on the relative importance of attributes. Differences in the elicitation formats imply differences in the information obtained from respondents as well as the interpretation of the resulting measures of attribute importance. These differences across methods must be taken into account when comparing and interpreting preference results, but are seldom discussed explicitly in SP health applications.

This article provides a guide to evaluating the concept of attribute importance in the context of DCE and BWS methods. It also discusses the meaning of importance measures and

1 Introduction

Stated-preference (SP) methods, such as discrete-choice experiments (DCEs) and best–worst scaling (BWS), are increasingly used to measure preferences for attributes of medical interventions [1–3]. These methods rely on the

✉ Juan Marcos Gonzalez
jm.gonzalez@duke.edu

¹ Duke University, 200 Morris St, Durham, NC 27701, USA

the ways practitioners can normalize these measures to make them easier to interpret and to compare across subgroups in a sample. Finally, stylized DCE and BWS results are discussed as examples of options to calculate and interpret attribute importance.

2 The Concept of Utility

DCE and BWS methods rely on the concept of utility, which is assumed to be a latent measure of well-being that can only be quantified on a relative (ordinal) scale. For this reason, a single measure of utility conveys no meaningful information. Since preferences correspond with utility, the ordinal nature of utility also applies to preference measures. Thus, to interpret preference measures one must understand what relative relationship they convey. This, of course, extends to attribute importance.

3 Defining Attribute Importance

In the context of SP applications in health, I define attribute importance as the absolute change in utility associated with an attribute. This measure of importance represents the overall positive or negative effect that an attribute has on individuals' well-being relative to other attributes.

The proposed definition of attribute importance considers changes in utility between attributes, and changes in utility within attributes. Between-attribute importance captures the change in utility associated with changing one treatment attribute for another. For example, it would be the change in utility when a treatment that causes nausea is changed for a treatment that causes fatigue, all else being equal. To understand the importance of changes between attributes, preference information must provide a ranking of these attributes. Within-attribute importance captures the impact of changing an attribute between two relevant attribute levels. For example, it would be the change in utility when a treatment that causes severe nausea is changed for a treatment that causes mild nausea, all else being equal. Thus, to understand the importance of changes within an attribute, preference information must provide a ranking of attribute levels. Although the distinction between within- and between-attribute importance can seem trivial, it is significant in terms of how we interpret and how we can use importance information from SP methods.

Because the definition of attributes and levels often can be ambiguous, identifying changes as *within* or *between* attributes can be problematic. For example, a researcher might want to consider treatment-related tolerability issues in an SP study. One option would be to make each adverse effect outcome an attribute, and specify the attribute levels as the presence and absence of each outcome. A different

approach could be to create an attribute labeled 'treatment tolerability issues' and make each adverse effect outcome a level for the attribute. Essentially turning the attributes in the first approach into attribute levels.

With the first approach, the importance of eliminating a specific tolerability issue in a treatment can be evaluated within each adverse effect as preference information for the two attribute levels (i.e., presence and absence of a adverse effect) is available to calculate an absolute importance value for each outcome. This would be the difference between the utility for the presence of a adverse effect minus the utility value for the absence of the same adverse effect.

With the second approach, the importance of eliminating a adverse effect is only identifiable as respondents' change in utility when swapping one adverse effect for another. In other words, because the only utility values that can be estimated are those for the presence of each adverse effect relative to the presence of other adverse effects, the second approach does not produce an absolute measure of attribute importance.

It is crucial to note that although the calculation of within-attribute importance does not depend on the importance estimates for other attributes, the utility for attribute levels are estimated in the context of the other attributes and attribute levels presented to respondents. In that sense, within-attribute importance measures are only absolute for a specific SP application.

4 Differences in Importance Measures by Elicitation Method

The previous discussion highlights how the design of an experiment can determine whether within- or between-attribute importance can be evaluated. Next, I discuss how DCEs and the various BWS methods are inherently suited to collect within- or between-attribute importance. To accomplish this, however, one first needs to understand how the answers elicited with each SP method relate to utility.

The framework that is commonly used to estimate attribute importance with DCEs and BWS relies on random-utility theory [4]. The theory assumes that it is possible to describe the effect of medical interventions on people's well-being (utility) through a function. The utility function is typically represented as follows (Eq. 1):

$$U_j = V(\beta, X_j) + \varepsilon_j \quad (1)$$

where U_j represents the utility for medical intervention j , and $V(\beta, X_j)$ is a deterministic portion of utility defined by a vector of attributes that are specific to intervention j (X_j) and a set of attribute-specific model estimates (β) that determine the marginal effect of each attribute or attribute

level on utility. The deterministic portion of utility is typically assumed to be additively separable in the k attributes of medical interventions, as in Eq. 2:

$$V(\beta, X_j) = \sum_k \beta_k x_{jk} \quad (2)$$

In DCEs and Case 3 BWS, the choice or ranking of profiles is assumed to be determined by the utility for those profiles. More specifically, the probability of selecting an alternative over all available alternatives is equivalent to the probability that the utility of the preferred alternative is greater than the utility of all other alternatives. Thus, in a simple example with two alternatives (alternative j and alternative i), random-utility theory states that the probability that alternative j is chosen over—or ranked higher than—alternative i is determined by the difference in utility between the two options. More positive differences in the utility of the alternatives would be associated with greater likelihood of choosing or ranking j over i . Although different modeling approaches may modify the specific relationship between utility and choice, all methods used to estimate attribute importance from DCEs and Case 3 BWS data treat the probability of choice or ranking as a signal on utility changes [6, 7].

Exactly how differences in utility between alternatives j and i affect the probability of choosing or ranking options is defined by the probability density function assumed in the analysis of DCEs and Case 3 BWS data. The probability density function often assumed is that of a conditional logit model or some variation of it [4, 7]. Importantly, Case 3 BWS data are analyzed using a variation that specifies whether a respondent chooses best and worst simultaneously [8] or sequentially [5, 6]. Nevertheless, the basic structure of the conditional logit model is preserved with all commonly used variations and looks as follows (Eq. 3):

$$P(C = j|i, j) = \frac{e^{V(\beta, X_j)}}{e^{V(\beta, X_j)} + e^{V(\beta, X_i)}} = \frac{1}{1 + e^{V(\beta, X_i) - V(\beta, X_j)}} \quad (3)$$

where $P(C = j|i, j)$ is the probability that choice is—or the top rank is for—alternative j when both alternative i and j are available. Equation 3 shows how the probability of choosing alternative j —or ranking alternative j over alternative i —depends entirely on the utility difference between alternatives ($V(\beta, X_i) - V(\beta, X_j)$). Given Eq. 2, the utility differences in Eq. 3 can be redefined as $V(\beta, X_i) - V(\beta, X_j) = \sum_k \beta_k (x_{ik} - x_{jk})$ when the model variables are coded as continuous and linear, where $x_{ik} - x_{jk}$ represents the differences in the levels for attribute k across alternatives, and β_k is the estimated utility change with any change in the levels of the attribute. This simple example shows how estimates from the analysis of DCEs and Case 3

BWS data represent the change in utility associated with changes in attribute levels. The result conceptually is the same when using non-linear or categorical attribute variables. However, with non-linear and categorical coding, utility can change at different rates with the attribute levels considered.

As with Case 3 BWS, the analysis of Case 1 and Case 2 BWS data requires an assumption about whether the selection of best and worst was done simultaneously or sequentially. Nevertheless, these assumptions do not change the basic structure of the conditional logit model, which is still often used to relate the probability of ranking an attribute above others with the utility derived from that attribute. Since the ranking in Case 1 and Case 2 BWS does not happen between profiles, but directly between attributes, the parameter estimates in the analysis of these data reflect utility differences between attributes.

For Case 2 BWS, however, respondents are asked to evaluate between-attribute importance as attribute levels change. Changes in the ranking of attributes at different levels provide an indirect way to estimate within-attribute importance values. Note that, because the utility levels within attributes are inferred from changes in attribute rankings, Case 2 BWS does not directly provide the impact of changes in one attribute versus changes in other attributes.

With all of this in mind, we can say that DCEs, Case 3 BWS, and Case 2 BWS all provide within-attribute importance. On the other hand, Case 1 BWS provides between-attribute importance.

5 Interpreting Attribute Importance Across Elicitation Methods

In DCEs, respondents choose between profiles that simultaneously provide some constructed desirable improvements in certain attributes and undesirable changes in other attributes. As respondents choose between profiles, they are expected to evaluate the combination of desirable and undesirable attribute changes across alternatives [9]. Thus, the measure of importance with DCEs stems from the degree to which respondents are willing to accept one attribute change for another. Also, because DCEs elicit importance from choices between profiles, the importance measures are directly related to the impact an attribute has on the decision to choose a medical intervention over others.

For BWS, the frequency with which a profile or an attribute is selected as best or worst indicates how far in one direction or another the attribute is on a common scale of interest [6, 10]. The specific interpretation of the measures depends on the context of the ranking, as BWS exercises can ask respondents to partially rank items based on a variety of

contexts such as how important attributes are in a decision to choose treatment, or how burdensome the items are expected to be. For this reason, results from BWS methods cannot necessarily be interpreted in the context of choices between treatments or as tradeoff information between attributes.

Since Case 3 and Case 2 BWS elicit attribute importance in the context of treatment profiles, importance measures can be more directly related to the desirability of medical interventions. With a BWS scale that relates to the likelihood of choosing a treatment, importance measures from Case 3 and Case 2 can be equivalent to those obtained with DCEs. However, interpreting results from Case 2 BWS as those of DCEs requires assuming that the elicited attribute level rankings appropriately capture the mechanism by which respondents would make treatment choices.

6 Normalizing Importance: Comparing Importance Measures

Importance measures are often normalized to aid in their interpretation or to compare attribute importance across subgroups within a sample. There are two main types of normalizations used with attribute importance values: (1) attribute-based normalization—which presents attribute importance as a proportion of a reference attribute importance; and (2) profile-based normalization—which presents attribute importance as a proportion of the total utility induced by exchanging treatment profiles.

Attribute-based normalization can be described with Eq. 4:

$$I_{kj} = S \times \frac{(V_{k1} - V_{k2})}{(V_{j1} - V_{j2})} \quad (4)$$

where I_{kj} is the normalized importance value for attribute k , $(V_{k1} - V_{k2})$ is the estimated importance for attribute k , and $(V_{j1} - V_{j2})$ is the estimated importance for attribute j , given any two levels of interest in the attribute. For the purpose of the normalization, $(V_{j1} - V_{j2})$ is considered a reference attribute importance. The reference attribute and the attribute levels chosen to define importance could be based on clinical or practical considerations. For example, this reference attribute importance could be a clinically meaningful risk increase with treatment or the greatest difference in utility among all study attributes. The reference attribute importance could also be the importance of a 1-unit change in another attribute (e.g., treatment risk or out-of-pocket cost), which makes the attribute-based normalization a measure of the marginal rate of substitution between attributes [10–12]. Finally, S is an arbitrary scaling factor that

sets the normalized importance of the reference attribute at a particular value. Some authors set S to be equal to 10 or 100 as intuitive anchors for the normalized importance measures [13, 14].

The attribute-based normalization implies that attribute importance values are no longer in utility units, but rather in the units of the reference attribute change. For example, if the reference attribute importance corresponds to the change in utility associated with an increase of 10 percentage points in the risk of an adverse event, the normalized importance measures represent how much more or less important each attribute is relative to that 10 percentage-point increase. Thus, if the normalized importance for the attribute of interest is 2, that attribute is two times as important as a 10 percentage point risk of an adverse event. When the reference attribute importance is a \$1 change in cost, the normalized importance measures become a representation of the number of dollars in cost that are equivalent to each attribute¹ [11].

Attribute-based normalizations also are commonly applied by setting the reference attribute importance to be the greatest attribute importance in a study (i.e., the greatest difference in utility between the most and least preferred level for any given attribute). With this reference every normalized attribute importance value lies between 0 and S , and can be interpreted as a fraction of the most important attribute in the exercise. Note that the same reference attribute importance must be used to compare normalized attribute importance across subgroups in a sample—even if the attribute is not the most important in all subgroups. Otherwise, the normalized importance values represent different fractions and cannot be compared directly.

The profile-based normalization calculates importance values relative to the aggregate utility differences between two treatment profiles. This can be represented as follows:

$$I_k = S \times \frac{(V_{k1} - V_{k2})}{\sum_j (V_{j1} - V_{j2})} \quad (5)$$

where $\sum_j (V_{j1} - V_{j2})$ is the sum of all attribute differences between profiles 1 and 2, including the differences in attribute k .

In this context, normalized importance values are interpreted as the proportion of the value of a profile change captured by a specific attribute, or how much of the value of one intervention relative to another is due to a specific attribute change. Note that with this normalization, the sum of all normalized importance values is also equal to S , as shown in Eq. 6:

¹ Assuming the marginal utility of cost is constant.

$$\sum_k^K I_k = \sum_k^K S \times \frac{(V_{k1} - V_{k2})}{\sum_j^J (V_{j1} - V_{j2})} = S \times \frac{\sum_k^K (V_{k1} - V_{k2})}{\sum_j^J (V_{j1} - V_{j2})} = S \quad (6)$$

Since $\sum_k^K (V_{k1} - V_{k2}) = \sum_j^J (V_{j1} - V_{j2})$ by the definition of this normalization. If the scaling factor S is set to 1, the normalized relative importance values can be interpreted as the percentage of the difference between profiles attributable to attribute k [15, 16].

The interpretation of relative importance with the profile-based normalization is consistent with the concept of attribute importance in a decision to choose a specific medical intervention. Also, with this normalization the relative importance value for an attribute depends on all the other attribute changes considered in the treatment profiles. Thus, although the ratios of normalized importance values are always the same, the specific attribute importance values change with the profiles considered. In the absence of clinically relevant treatment profiles to use with the profile-based normalization, practitioners could consider profiles that represent the most extreme attribute changes in the experiment.

When using profile-based normalizations, the comparison of importance values across subgroups must be done using the same profile differences for all subgroups. Otherwise, the normalized values have different meanings as fractions of different profile changes.

It is also worth noting that the two normalizations shown here do not apply as presented to *between-attribute* importance measures obtained with Case 1 BWS. With *between-attribute* importance, some authors have used information on choice probabilities to recover absolute importance measures that can be normalized [17, 18]. The steps for such transformations are discussed in Sect. 7.

7 Normalizing Between-Attribute Importance

It is possible to transform between-attribute importance values so they can be normalized with attribute- or profile-based normalizations. To do this, however, more information is needed. One way to incorporate the additional information is to use what we know about the relationship between utility and choice through the conditional logit probability density function. Specifically, one can set the assumed value for the best [$V(b)$] and worst [$V(w)$] attributes in the conditional logit probability density function to be the regression estimate for an attribute of interest (δ_k) and zero for the omitted effect in the model, respectively. If the attribute of interest and the omitted effect were the only ones included in a Case 1 BWS question, we could calculate the probability that the attribute of interest is selected as best using Eq. 7:

$$A_k = \frac{e^{\delta_k}}{e^{\delta_k} + \sum_l^{L-1} e^0} = \frac{e^{\delta_k}}{e^{\delta_k} + L - 1} \quad (7)$$

where A_k is considered the adjusted importance estimate for attribute k , and L is the number of items in each of the BWS questions. Conceptually, this adjustment uses the probability of selecting an attribute against the omitted category in the model specification to generate an absolute measure of attribute importance [19]. With adjusted importance values for each attribute, it is straightforward to apply the normalizations in Eqs. 4 and 5 by substituting the utility differences in each normalization for the adjusted importance values. The final form of the normalizations for between-attribute importance measures is then as follows (Eqs. 8 and 9):

$$I_{kj} = S \times \frac{A_k}{A_j} (\text{attribute - based normalization}) \quad (8)$$

$$I_k = S \times \frac{A_k}{\sum_j^J A_j} (\text{profile - based normalization}) \quad (9)$$

where A_j is the adjusted importance of the reference attribute, and $\sum_j^J A_j$ represents the sum of the adjusted importance measures for attributes that would make up a profile for a medical intervention.

8 Stylized Example

To illustrate the interpretation and normalization of attribute importance with SP methods, consider the following stylized examples of a DCE and Case 1 BWS.

8.1 Example Discrete-Choice Experiment

Assume a DCE evaluated three attributes with three levels each for two different subgroups in a sample. These attributes include treatment efficacy, safety, and convenience. Table 1 presents the attributes and levels for this example.

If data for subgroups are not pooled, traditional analysis of DCEs would produce a set of estimates representing utility values (or preference weights) for each subgroup [7].

If estimates look like those in Table 1, absolute importance values can be obtained by simply taking the difference between any two levels of an attribute. For example, the importance of changing efficacy from level 2 to level 1 is 0.1 ($0.1 = 0.2 - 0.1$). Overall attribute importance would be calculated as the difference between the most and least preferred level in each attribute, or 0.5 ($0.5 = 0.2 - (-0.3)$) for

Table 1 Estimates from discrete-choice experiments

| Attribute | Estimates | |
|--------------------|------------|------------|
| | Subgroup 1 | Subgroup 2 |
| <i>Efficacy</i> | | |
| Efficacy 1 | 0.2 | 0.5 |
| Efficacy 2 | 0.1 | - 0.1 |
| Efficacy 3 | - 0.3 | - 0.4 |
| <i>Safety</i> | | |
| Safety 1 | 0.9 | 0.4 |
| Safety 2 | - 0.2 | - 0.1 |
| Safety 3 | - 0.7 | - 0.3 |
| <i>Convenience</i> | | |
| Convenience 1 | 0.4 | 0.3 |
| Convenience 2 | 0.3 | 0.2 |
| Convenience 3 | - 0.7 | - 0.5 |

efficacy, 1.6 (1.6=0.9 to - 0.7) for safety, and 1.1 (1.1=0.4 to - 0.7) for convenience. Although these values are absolute importance measures because they represent within-attribute importance, they must be interpreted in the context of all the other attributes presented to respondents, and are not directly comparable across subgroups. To interpret and compare these absolute importance values across subgroups, results need to be normalized using one of the two approaches described earlier.

The normalized overall importance values are presented in Fig. 1. To the left, the importance values for each subgroup are normalized using the attribute-based normalization based on the overall importance of the safety attribute. That is, the overall importance of efficacy and convenience are presented as a proportion of the overall importance of safety for each subgroup. As expected with this normalization, the relative importance value for safety is 1 in both subgroups. Figure 1 also shows the normalized importance values for each subgroup as a proportion

of overall changes in medication profiles (profile-based normalization considering a change from the treatment with the lowest utility to the treatment with the highest utility). As expected, the sum of all importance values with the profile-based normalization is equal to 1 in each subgroup.

Although all ratios of relative importance values are preserved with both normalizations, one can see how the two approaches seem to provide different information about attribute importance across subgroups. For example, with the attribute-based normalization, safety is assumed to have the same importance in both subgroups. This, of course, is an assumption in the normalization. However, the importance of safety seems to be dramatically different across subgroups with the profile-based normalization. The difference between the two sets of normalized results can be explained by the fact that the two normalizations are providing different representations of attribute importance. Thus, differences across subgroups would also support different conclusions.

Table 2 Best-worst scaling items and results

| Items | Estimates | |
|------------------------------|------------|------------|
| | Subgroup 1 | Subgroup 2 |
| Improved efficacy | 0.7 | 1.0 |
| Safety problems | 0.6 | 0.6 |
| Lack of convenience | 0.2 | 0.3 |
| Tolerability problems | 0.1 | 0.1 |
| Quality-of-life problems | 0.1 | 0 |
| Administration at home | - 0.1 | - 0.1 |
| Delayed onset of action | - 0.3 | - 0.3 |
| Prolonged duration of effect | - 0.4 | - 0.9 |
| Cost | - 0.9 | - 0.7 |

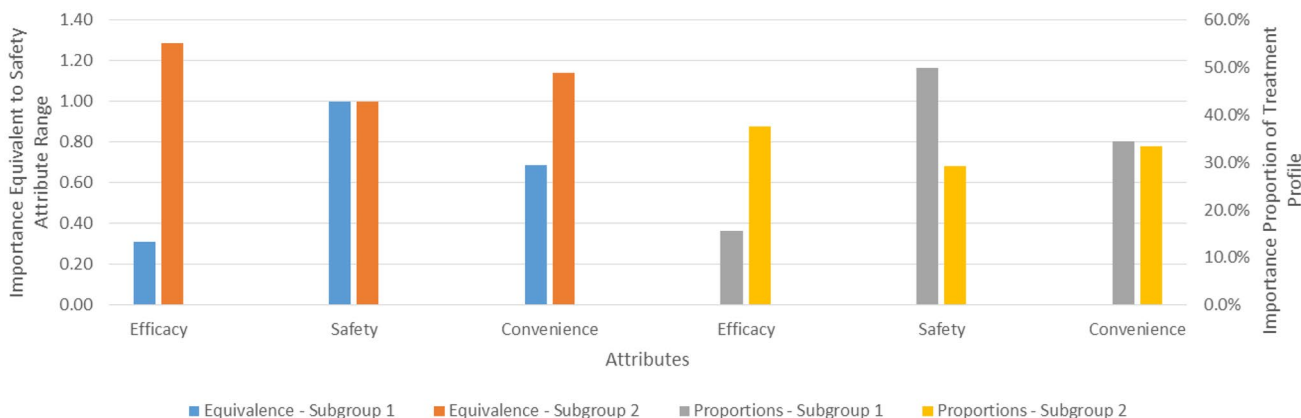


Fig. 1 Normalized importance values from discrete-choice experiments

Table 3 Example importance adjustment and normalization of best–worst scaling results

| Items | Adjusted estimates (mean effect = 1) | | Adjusted estimates (mean effect = 10) | | Normalized adjusted estimates (mean effect = 1) | | Normalized adjusted estimates (mean effect = 10) | |
|-------------------------|--------------------------------------|------------|---------------------------------------|------------|---|------------|--|------------|
| | Subgroup 1 | Subgroup 2 | Subgroup 1 | Subgroup 2 | Subgroup 1 | Subgroup 2 | Subgroup 1 | Subgroup 2 |
| Efficacy | 1.7 | 2 | 10.7 | 11 | 17.00 | 6.67 | 1.18 | 1.18 |
| Safety | 1.6 | 1.6 | 10.6 | 10.6 | 16.00 | 5.33 | 1.16 | 1.14 |
| Convenience | 1.2 | 1.3 | 10.2 | 10.3 | 12.00 | 4.33 | 1.12 | 1.11 |
| Tolerability | 1.1 | 1.1 | 10.1 | 10.1 | 11.00 | 3.67 | 1.11 | 1.09 |
| Quality of life | 1.1 | 1 | 10.1 | 10 | 11.00 | 3.33 | 1.11 | 1.08 |
| Route of administration | 0.9 | 0.9 | 9.9 | 9.9 | 9.00 | 3.00 | 1.09 | 1.06 |
| Onset of action | 0.7 | 0.7 | 9.7 | 9.7 | 7.00 | 2.33 | 1.07 | 1.04 |
| Duration of effect | 0.6 | 0.1 | 9.6 | 9.1 | 6.00 | 0.33 | 1.05 | 0.98 |
| Cost | 0.1 | 0.3 | 9.1 | 9.3 | 1.00 | 1.00 | 1.00 | 1.00 |

While the attribute-based normalization focuses strictly on how the attributes relate to each other, the profile-based normalization evaluates the relationship between the attributes and profile choice. Finding differences across subgroups when importance values are normalized with an attribute-based approach implies that the average relationship between attributes varies by subgroup. Differences in importance values that are normalized with a profile-based approach imply that the attributes contributed to treatment choices at different rates across subgroups.

Information from each normalization can be used to address different problems. For example, practitioners may want to evaluate marginal rates of substitution, in which case information from the attribute-based normalization would be necessary. Alternatively, practitioners may want to use attribute importance information to support multi-criteria decision analysis. In such a case, profile-based normalizations can inform how attributes influence a relevant decision and facilitate consensus among stakeholders.

8.2 Case 1 Best–Worst Scaling Example

Assume a Case 1 BWS exercise with three items in each question that considered nine items in total. These items are summarized in Table 2. The items cover attributes that could be used to build the profile of a medical intervention, although the questions will only ask respondents to rank the most and least important attribute in determining the desirability of a medical intervention. Thus, the importance of these attributes would not strictly correspond to the effect of choosing specific medical interventions.

Responses to BWS tasks can be analyzed following standard regression tools [6]. Results of such analysis show the importance of the attributes and could look like those presented in Table 2. These results correspond to an effect-coded specification—note that the result for cost is the negative sum of all other items in both subgroups—but could

easily be changed to represent deviations from any item in the list (dummy coding).

With effect-coded attributes, the results from Case 1 BWS identify differences between the importance of each attribute and the mean importance of all attributes. In other words, these differences represent how much more or less important attributes are relative to an unobserved mean effect for all attributes.² An absolute measure of attribute importance then requires adding each estimate to a mean importance measure that is not available. This is a problem because the absolute measure of attribute importance will change with the value of the unobserved mean. A simple example can help us show this.

We can first assume the unknown mean effect of all attributes in the two populations is 1. If so, one can adjust the results for the BWS data by adding 1 to each importance estimate and apply the normalizations proposed before. This is shown in Table 3 where attribute-based normalizations were applied. In this example, the reference attribute importance was set to be the adjusted importance for cost, but any other attribute of interest could be used. One can follow the same steps after assuming the mean effect value was 10, instead of 1. Table 3 also shows the adjusted estimates and the resulting normalized importance values if the mean effect value was 10.

Note that the adjusted estimates and normalized importance values vary greatly with the assumed value of the mean effect.³ Thus, identifying an absolute importance level requires more information. This information can be obtained from the probability density function that generated the BWS estimates. A more general form of Eq. 7 is

² This would be the omitted attribute in the case of dummy-coded variables.

³ Or the value of the omitted attribute in the case where dummy coding is used.

Table 4 Best–worst scaling results adjusted using probability density function

| Items | Adjusted estimates (mean effect = 1) | | Adjusted estimates (mean effect = 10) | | Normalized adjusted estimates (mean effect = 1 and mean effect = 10) | |
|-------------------------|--------------------------------------|------------|---------------------------------------|------------|--|------------|
| | Subgroup 1 | Subgroup 2 | Subgroup 1 | Subgroup 2 | Subgroup 1 | Subgroup 2 |
| Efficacy | 0.50 | 0.58 | 0.50 | 0.58 | 2.97 | 2.90 |
| Safety | 0.48 | 0.48 | 0.48 | 0.48 | 2.82 | 2.40 |
| Convenience | 0.38 | 0.40 | 0.38 | 0.40 | 2.24 | 2.03 |
| Tolerability | 0.36 | 0.36 | 0.36 | 0.36 | 2.11 | 1.79 |
| Quality of life | 0.36 | 0.33 | 0.36 | 0.33 | 2.11 | 1.68 |
| Route of administration | 0.31 | 0.31 | 0.31 | 0.31 | 1.84 | 1.57 |
| Onset of action | 0.27 | 0.27 | 0.27 | 0.27 | 1.60 | 1.36 |
| Duration of effect | 0.25 | 0.17 | 0.25 | 0.17 | 1.49 | 0.85 |
| Cost | 0.17 | 0.20 | 0.17 | 0.20 | 1.00 | 1.00 |

used to generate the adjusted importance values with the various assumed means. This variation would simply update the assumed value of the omitted mean from zero to 1 or 10.

$$A_k = \frac{e^{\delta_k+m}}{e^{\delta_k+m} + \sum_l^{L-1} e^m} \quad (10)$$

where m is the assumed mean importance value (i.e., 1 or 10 in this example) and L is still the number of items in the BWS question. It is hopefully apparent that when $m=0$, Eq. 10 is exactly the same as Eq. 7. What is perhaps less apparent is that Eq. 10 is actually indistinguishable from Eq. 7 with any assumed mean value as the new terms in Eq. 10 cancel each other. This shows that these adjustments do not depend on the value of the unknown mean importance.

After the adjustment, all importance values can be normalized using any of the two options described here (i.e., attribute- or profile-based normalizations). Table 4 shows the adjusted values and the normalized adjusted values following the attribute-based approach. As anticipated, the adjusted importance values are now invariant to changes in the mean attribute importance.

9 Other Sources of Variation in Importance Measures

Importance measures can vary both in value and interpretation depending on the method used to elicit preferences (e.g., DCEs versus Case 1 BWS), the specific context of the questions (e.g., whether a BWS question asks respondents to rank attributes based on how important they are when choosing a treatment versus how important they are when reducing treatment burden), and the approach used to normalize attribute effects (i.e., attribute-based versus profile-based normalization). When evaluating attribute importance,

at least two other factors can influence the interpretation of attribute importance: (1) the specific changes considered within an attribute; and (2) the context under which the importance was elicited.

9.1 Attribute Change

Within-attribute importance measures depend entirely on the attribute change characterized by the measure. This is because the attribute change will determine the expected change in utility. Larger attribute changes can be expected to be more important than smaller attribute changes. For example, the importance of cost will depend upon the change in cost considered in a study. Changing the cost of a profile by \$1000 is expected to be more important than changing the cost by \$10.

Even Case 2 BWS, which in principle elicits between-attribute importance, may be influenced by the ranges of the levels in the attributes. An attribute with an extreme level may be more likely to be selected best or worst in the exercise, which would affect the implied within-attribute importance measures derived for an attribute.

The selection of levels to determine attribute importance is not generally an issue with Case 1 BWS applications because these applications do not ask respondents to rank attribute levels. That said, and given that study designs can blur the distinction between attributes and attribute levels, certain attribute labels or descriptions could lead respondents to think that an attribute implies greater changes in the status of a patient. For example, the need for insulin to treat patients with type 2 diabetes mellitus (T2DM) may have a greater impact than a broader label about the need for injectable treatments, because patients with T2DM can be aware that they would not need insulin unless their condition has deteriorated significantly. In such an example, the attribute label can be understood to imply a different intensity that would affect its importance.

9.2 The Options in the Questions

High or low attribute importance values are only meaningful given the options and the attributes evaluated. For example, convenience associated with route of administration may be perceived to be unimportant when compared against very salient health outcomes such as longevity or quality of life, even if the changes in the health outcomes are minor. On the other hand, convenience could be perceived to be very important when compared with asymptomatic clinical outcomes such as changes in surrogate markers. Hence, low importance for an attribute cannot be considered an unconditional indicator that something is unimportant.

This issue can be appreciated when using profile-based normalizations. For example, in such a normalization the profile differences can include large changes in an efficacy attribute which could minimize the normalized importance of other attributes like convenience. In contrast, the profile changes could effectively exclude treatment efficacy by assuming no change in that attribute across profiles. This in turn could augment the normalized importance of convenience.

10 Concluding Remarks

Although attribute importance is reported in many of the publications that use SP methods, this article shows that not all importance values are the same. The measures may not communicate the same construct, or represent the same relative assessment of value. Practitioners should take these differences into account to ensure that the conclusions in studies are consistent with the way attribute importance was calculated. Practitioners should also report more clearly how attribute importance is calculated in their analysis so people using their results can understand what exactly is supported by the data, and make more accurate and meaningful comparisons across samples and potentially across studies.

Acknowledgements I would like to thank the anonymous reviewers for their many insightful comments and suggestions.

Compliance with Ethical Standards

Conflict of interest Juan Marcos Gonzalez has no conflicts of interest that need to be disclosed.

Funding The author received no specific funding for this work.

References

- de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* 2012;21(2):145–72.
- Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics.* 2014;32(9):883–902.
- Marshall D, Bridges JFP, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint analysis applications in health—how are studies being designed and reported? *Patient.* 2010;3(4):249–56.
- McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. *Frontiers in econometrics.* New York: Academic; 1973.
- Lancsar E, Louviere J, Donaldson C, Currie G, Burgess L. Best worst discrete choice experiments in health: methods and an application. *Soc Sci Med.* 2013;76:74–82.
- Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. *J Health Econ.* 2007;26(1):171–89.
- Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health.* 2016;19(4):300–15.
- Louviere JJ, Flynn TN, Marley AA. *Best-worst scaling: theory, methods and applications.* Cambridge: Cambridge University Press; 2015.
- Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health.* 2013;16(1):3–13.
- Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. *Soc Sci Med.* 2007;64(8):1738–53.
- Mühlbacher A, Johnson FR, Yang J-C, Happich M, Belger M. Do you want to hear the bad news? The value of diagnostic tests for Alzheimer’s disease. *Value Health.* 2016;19(1):66–74.
- Okumura K, Inoue H, Yasaka M, Gonzalez JM, Hauber AB, Levitan B, et al. Comparing patient and physician risk tolerance for bleeding events associated with anticoagulants in atrial fibrillation—evidence from the United States and Japan. *Value Health Reg Issues.* 2015;6:65–72.
- Hauber A, Arden N, Mohamed A, Johnson FR, Peloso P, Watson D, et al. A discrete-choice experiment of United Kingdom patients’ willingness to risk adverse events for improved function and pain control in osteoarthritis. *Osteoarthritis Cartil.* 2013;21(2):289–97.
- Luyten J, Kessels R, Goos P, Beutels P. Public preferences for prioritizing preventive and curative health care interventions: a discrete choice experiment. *Value Health.* 2015;18(2):224–33.
- Beusterien K, Middleton M, Feng Wang P, Rao S, Kotapati S, Sabater J, et al. Patient and physician preferences for treating adjuvant melanoma: a discrete choice experiment. *J Cancer Ther.* 2017;8:37–50.
- Regier DA, Diorio C, Ethier M-C, Alli A, Alexander S, Boydell KM, et al. Discrete choice experiment to evaluate factors that influence preferences for antibiotic prophylaxis in pediatric oncology. *PLoS One.* 2012;7(10):e47470.
- Yuan Z, Levitan B, Burton P, Poulos C, Brett Hauber A, Berlin JA. Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives. *Curr Med Res Opin.* 2014;30(9):1733–41.
- Mamounas E, Poulos C, Goertz H-P, González JM, Pugh A, Antao V. Neoadjuvant systemic therapy for breast cancer: factors influencing surgeons’ referrals. *Ann Surg Oncol.* 2016;23(11):3510–7.
- Software Sawtooth. The MaxDiff system technical paper. Orem: Sawtooth Software, Inc.; 2013.