



# Assessing Patient Preferences in Rare Diseases: Direct Preference Elicitation in the Rare Chronic Kidney Disease, Immunoglobulin A Nephropathy

Kevin Marsh<sup>1</sup> · Kerrie-Anne Ho<sup>2</sup> · Rachel Lo<sup>1</sup> · Nancy Zaour<sup>3</sup> · Aneesh Thomas George<sup>4</sup> · Nigel S. Cook<sup>3</sup>

Accepted: 20 April 2021 / Published online: 19 May 2021  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

**Background** Patient preference information is increasingly being used to inform decision making; however, further work is required to support the collection of preference information in rare diseases. This study illustrates the use of direct preference elicitation methods to collect preference data from small samples in the context of early decision making to inform the development of a product for the treatment of immunoglobulin A nephropathy.

**Method** An interview-based swing weighting approach was used to elicit preferences from 40 patients in the US and China. Attributes were identified through a background review, expert engagement and patient focus groups. Participants completed a series of tasks that involved ranking, rating and scoring improvements in the attributes to obtain attribute swing weights and partial value functions. The preference results were then incorporated into a benefit-risk assessment simulation tool.

**Results** Participants placed the greatest value on avoiding end-stage renal/kidney disease. Similar weight was given to short-term quality-of-life improvements and avoiding infections. Treatment burden (number of vaccinations) received the least weight. Heterogeneity in preferences was also observed. Consistency tests did not identify statistically significant variation in preferences, and qualitative data suggested that the elicitation exercise was sensitive to participants' interpretation of attributes and that participants were able to express their preferences.

**Conclusion** Direct preference elicitation methods can be used to collect preference data from small samples. Further work should continue to test the validity of the estimate generated by such methods.

## Key Points for Decision Makers

Understanding patient preferences is important for health care decision making, but commonly used quantitative methods require sample sizes that often prohibit their use in rare diseases.

This study illustrates the use of direct rating methods to collect preference data from a small sample of rare disease patients (immunoglobulin A nephropathy) to inform a benefit–risk analysis.

## 1 Introduction

Patient preference information (PPI) elucidates how patients trade-off between, for instance, the benefit and risks of treatments [1, 2]. With efforts to increase patient centricity in healthcare, PPI is being used more by regulatory and health technology assessment agencies [3–6]. As a consequence, many sponsors are now incorporating preference data into benefit–risk assessments to support both regulatory and reimbursement submissions and to inform internal decision making, such as product development and evidence generation planning [7].

One of the challenges facing the collection of PPI is how to do so with small sample sizes, which is often the case with rare diseases. This is particularly important as patients with rare diseases are often more willing to accept treatment risks, given their limited treatment options and poor prognoses [8].

✉ Kevin Marsh  
Kevin.Marsh@evidera.com

<sup>1</sup> Evidera, London, UK

<sup>2</sup> UCB Pharma, Slough, UK

<sup>3</sup> Novartis Pharma AG, Basel, Switzerland

<sup>4</sup> Novartis Healthcare Pvt. Ltd, Hyderabad, India

PPI can help inform decision makers of the risk tolerance of rare disease patients, but recruiting the sample sizes required for patient preference studies is often thought to be prohibitively time-consuming and expensive. For example, the most commonly employed method for eliciting PPI is the discrete choice experiment (DCE) [9–13], while precise sample size requirements for DCEs will vary between studies [9]; as a rule of thumb, the minimum sample size is often thought to be 100–150 participants. When such sample sizes are not feasible within the scope of research studies in rare disease populations, other methods such as thresholding or best–worst scaling (BWS) have been proposed. By collecting more information per attribute, BWS type 2 is thought to require smaller sample sizes than DCEs, but because BWS type 2 asks participants to identify best and worst ranked attributes, there are concerns that it only provides insight into attribute ranking rather than benefit–risk trade-offs [14]. Thresholding is less data-demanding and thus requires smaller sample sizes; however, it is not designed to determine how patients simultaneously trade-off across many attributes.

The objective of this study was to illustrate direct rating methods for eliciting preferences with small samples; specifically, partial value function elicitation and swing weighting. Direct rating methods were most notably applied to support quantitative benefit–risk assessment by the European Medicines Agency (EMA) [15], which concluded that the method could support regulatory decision making where the benefit–risk balance was marginal. DCE, BWS and thresholding all involve participants making choices. These choices are then analysed to understand the relative importance of the attributes that define the choices. In contrast, direct rating methods elicit cardinal estimates of the value of attributes from participants [16]. As a consequence, they collect more data from each participant, facilitating its use with small samples, but at the expense of greater cognitive burden for participants [17]. The greater complexity of the elicitation tasks requires that direct rating is facilitated in an interview of focus group setting. This requires a greater time commitment from each participant and more research team resource per participant, but offers the benefit of simultaneously collecting qualitative and quantitative data. In studies with large sample sizes, this approach would be very time and resource costly, and likely impractical, but this is less of an obstacle in rare diseases with small sample sizes.

Direct rating is demonstrated in the context of early decision making to inform the development of a product for the treatment of primary immunoglobulin A nephropathy (IgAN), a rare chronic autoimmune kidney disease with no approved treatments and poor prognosis that can lead to end-stage renal/kidney disease (ESRD) [18]. An interview-based direct rating approach was used to collect PPI, which

was incorporated into a benefit–risk assessment to evaluate alternative product profiles.

## 2 Methods

### 2.1 Study Population

To be eligible, participants had to be  $\geq 18$  years of age at the time of consent, self-report having biopsy-verified primary IgAN, currently receiving angiotensin-converting enzyme inhibitors/angiotensin receptor blockers (ACEi/ARBs) for at least 6 months, and be fluent in English (in the US) or Mandarin (in China). Patients receiving dialysis or who had received an organ transplant were excluded from the study as they may have complex or differing treatment needs and experiences that may influence their treatment preferences.

Patients were recruited via patient support groups, online research panels, social media, physician referrals and recruiters' databases. All participants consented to participate prior to the focus groups/interviews and were compensated at fair market value for their participation in this study.

### 2.2 Attribute Selection

Preferences were elicited for the attributes and levels summarized in Table 1. These were initially identified based on discussions with subject matter experts, a review of literature [19], social media listening [20], a patient online bulletin board with IgAN patients [21], and from observing the US FDA-led Patient-Focused Drug Development (EL-PFDD) meeting on IgAN in August 2019. These attributes were then evaluated with IgAN patients in focus groups in the US and China [22]. As this study was conducted early in the development of potential new target treatments, precise data on the clinical performance on target IgAN treatments were not available; therefore, the attribute performance levels were based on published studies of the current standard of care and discussion with medical experts.

Most of the attributes and levels selected were based on endpoints observed in clinical trials. The likelihood of developing ESRD is not captured in trials. Instead trials captured a measure of kidney function that is predictive of ESRD—estimated glomerular filtration rate (eGFR). As eGFR levels are not well understood by patients, they were modelled into the likelihood of developing ESRD at 10 years [23, 24].

### 2.3 Preference Elicitation Tasks

Three preference elicitation tasks were firstly piloted in two focus groups (one in the US and one in China). Main data

**Table 1** Attribute definitions and performance ranges

Attribute domain	Attribute	Definition/measure	Performance levels		
Clinical efficacy (benefit)	Likelihood of ESRD/dialysis	The likelihood of developing ESRD/needing dialysis in the 10 years following treatment	Level 1 10%	Level 2 20%	Level 3 30%
Adverse effects (risks)	Risk of infections	The likelihood of experiencing infections while receiving treatment	0%	10%	20%
	Risk of other adverse effects	The likelihood of experiencing other adverse effects, such as weight gain and joint pain, while receiving treatment	0%	20%	40%
Quality of life	Ability to perform usual activities	The ability to perform usual activities due to physical tiredness, exhaustion or weakness associated with IgAN experienced by patients	Not at all	Somewhat	Very much
	Emotional burden	The level of emotional burden associated with IgAN experienced by patients	Not at all	Somewhat	Very much
Treatment burden	Number of vaccinations	The number of vaccinations required before commencing treatment and at 5-year intervals after treatment	1 vaccination	2 vaccinations	3 vaccinations

ESRD end-stage renal/kidney disease, IgAN immunoglobulin A nephropathy

collection was conducted as individual interviews (2 h in duration) via computer-assisted phone interviews in the US and face-to-face computer-assisted interviews in China. The interviews in China were later conducted as computer-assisted phone interviews due to the outbreak of coronavirus disease 2019 (COVID-19).

The three preference elicitation tasks completed by participants are described below (these are illustrated in electronic supplementary material [ESM] 1):

- **Attribute ranking** Participants were asked to rank improvements in each attribute (from the worst to best level – the ‘swing’) in the order of importance.
- **Attribute weights** Participants were then asked to compare the attributes pairwise in the order in which they were ranked. The first and second attributes were compared, then the second and third, third and fourth, and so on. In each case, the swing in the highest ranked attribute in the pair was given a score of 100 points, and the participant was asked to give a corresponding score to the swing in the other attribute on a scale of 0–100. During the exercise, participants were shown an ‘exchange rate’, the marginal rate of substitution implied by their score. Participants could update their responses until an exchange rate that accurately reflected the trade-offs they were willing to make was achieved. To check for consistency of the participants’ responses, an additional task was included where par-

ticipants were asked to directly compare between the first- and third-ranked attribute (hereafter ‘validation task’). The weights derived from the validation task were compared with those derived from the results from the original pairwise comparison.

- **Partial value functions** Participants were asked to value improvements within each attribute. Participants were told that levels 1 and 2 on the attribute had a score of 0 and 1, respectively, and were asked to score level 3 on a scale of 0–10.

Participants were asked to provide a rationale for their responses throughout the above exercises. The interviews included an introduction to IgAN and patient preferences, warm-up exercises (simple preference elicitation exercises using a real-world example), introduction to the attributes and levels, followed by three preference elicitation exercises. During the warm-up exercises and introduction to attributes, participants were probed on their understanding and the interviewer only proceeded to the main preference elicitation tasks once they were confident participants understood the tasks.

Prior to the focus group or interview, participants also answered a brief online survey with questions about their sociodemographic and clinical characteristics, as well as two quality-of-life instruments (not reported in this methodological paper).

This study was reviewed and approved by an independent Ethical and Independent (E&I) Institutional Review Board.

## 2.4 Data Analysis

Participant-specific weights for each attribute were calculated based on the following formula (Eq. 1):

$$W_k^{\text{ranked } n} = \text{RATE}_k^{\text{ranked } n-1} \left( \frac{\text{RATE}_k^{\text{ranked } n}}{100} \right), \quad (1)$$

where  $W_k^{\text{ranked } n}$  is the weight for attribute  $k$  ranked  $n$ th,  $\text{RATE}_k^{\text{ranked } n-1}$  is the rate for attribute  $k$  ranked  $(n-1)$ th as reported by the participant, and  $\text{RATE}_k^{\text{ranked } n}$  is the rate for attribute  $k$  ranked  $n$ th as reported by the participant.

The weights were normalized to obtain the preference weight for each attribute,  $w_k$ , such that (Eq. 2):

$$\sum_{k=1}^K w_k = 1. \quad (2)$$

The partial value functions obtained from the scoring exercise were normalized to constrain to a 0–1 scale, where the worst and best attribute levels were anchored at 0 and 1, respectively. Participant-level mid-points on the partial value function (the ‘indifference point’) were summarized using mean and standard deviation (SD). A Dirichlet regression was employed to model the distribution of the attribute weights. Dirichlet regression was particularly suited for this study due to the compositional nature of the attribute weights [25]. The weighting exercise elicits a set of weight vectors  $(w_1, \dots, w_n)$ , one for each participant. These weight vectors were then used to fit a Dirichlet regression model by maximum likelihood to obtain the distribution of the weights,  $f_D(w)$ .

The effects of participants’ observable characteristics (such as age and time since diagnosis) on the attribute weights was assessed with a Dirichlet regression that controlled for the effect of these characteristics one at a time. The selection of observable characteristics to be tested was guided by a priori hypotheses on characteristics likely to impact the participants’ treatment preferences for IgAN. All analyses were conducted for the whole sample, pooled across the US and China and also on subgroups for country and selected sociodemographic and clinical characteristics.

Qualitative data were analysed to understand the reasons for participants’ preferences using semantic content analysis. It is not in the scope of the paper to report the qualitative findings in detail, but rather to illustrate the types of qualitative insights that were obtained. The results of the qualitative analysis of reasons for heterogeneity in preference weights and for the shape of two value functions—likelihood

of ESRD and ability to undertake usual activities—are reported.

The preference data were then incorporated into a benefit–risk assessment as follows (the calculation is illustrated in ESM 1) (Eq. 3):

$$U_j = \sum_{i=1}^n w_i v_i(x_{ij}), \quad (3)$$

where  $U_j$  is the overall value generated by treatment  $j$ ,  $w_i$  is the weight associated with attribute  $i$ ,  $v_i$  is the partial value function for attribute  $i$ , and  $x_{ij}$  is the performance of treatment  $j$  on attribute  $i$ .

To support early development decisions, a simulation tool was developed to apply the benefit–risk model to the evaluation of hypothetical treatment profiles. To illustrate the output from the simulation tool, this paper reports its application to estimate the proportion of patients who would prefer ‘treatment 1’ or ‘treatment 2’, which differ on some but not all attributes described in Table 2.

The study results were also presented to one patient advocate from the National Kidney Foundation and one patient advocate professional from the IgA Nephropathy Foundation America, and their feedback were incorporated in the interpretation of the study results.

## 3 Results

### 3.1 Participant Characteristics

Forty patients consented and participated in the study ( $n = 25$  in the US,  $n = 15$  in China) [see Table 3 for overall participant characteristics; further detail on the sample is provided in ESM 2]. No differences in the characteristics of the samples from the US and China were observed, except for age. Participants in China were younger (mean 35.7, SD 10.1 years) than those in the US (mean 45.2, SD 10.7 years).

**Table 2** Scenarios applied in the benefit–risk assessment-based simulation

Attribute	Treatment 1	Treatment 2
Likelihood of ESRD/dialysis	30%	10–30%
Risk of infections	10%	10–30%
Risk of other adverse effects	10%	10%
Ability to perform usual activities	Not able	Somewhat
Emotional burden	Very much emotionally burdened	Somewhat emotionally burdened
Number of vaccinations	1	3

ESRD end-stage renal disease

## 3.2 Preference Results

### 3.2.1 Preference Weights for Improvements in Treatment Attributes

Improving the likelihood of ESRD/dialysis (mean 0.314, standard error [SE] 0.021) was assigned a greater preference weight than improvements in all other attributes ( $p < 0.05$ ). There was no difference in the preference weights for reducing the risk of infection (mean 0.176, SE 0.016), reducing the risk of other adverse effects (mean 0.172, SE 0.016), or improving the ability to perform usual activities (mean 0.183, SE 0.016). Reducing the number of vaccinations (mean 0.06, SD 0.06) was assigned a lower preference weight than all other attributes ( $p < 0.05$ ) [Fig. 1].

Figure 2 demonstrates the heterogeneity observed in the attribute weights. Based on the ratio of SD to the mean, the extent of heterogeneity was more evident for the number of vaccinations (110.6%), emotional burden (63.8%) and likelihood of ESRD/dialysis (50.6%) than for other attributes (42.4–50.3%). This heterogeneity was also observed in the way that participants rank attribute swings. Reducing the likelihood of ESRD/dialysis was ranked most important by most participants, but 35% of participants ranked other

attribute swings first. Similarly, most participants ranked the number of vaccinations as their lowest priority, but 37.5% ranked other attributes as least important.

The relationship between participant characteristics and preference weights was modelled using a Dirichlet regression. Relative to the importance they attached to the likelihood of ESRD/dialysis, participants from China placed more importance on the risk of other adverse effects than participants from the US ( $p < 0.05$ ); those who were not in employment placed more importance on the ability to perform usual activity than those who were in employment ( $p < 0.05$ ); those who did not have a college/university degree placed more importance on the risk of other adverse effects than those who did have a college/university degree ( $p < 0.05$ ); and those with a kidney disease component summary (KDCS) score  $> 60$  placed less importance on the number of vaccinations than those with a KDCS score  $< 60$  ( $p < 0.05$ ) [see ESM 3 for full subgroup analyses].

Qualitative data provided insight into the rationale for preference heterogeneity. Most participants for which qualitative data were available were not concerned about the number of vaccinations ( $n = 5$ ), although others disliked injections ( $n = 1$ ) or associated injections with greater adverse effects ( $n = 1$ ). Participants who had lower weights

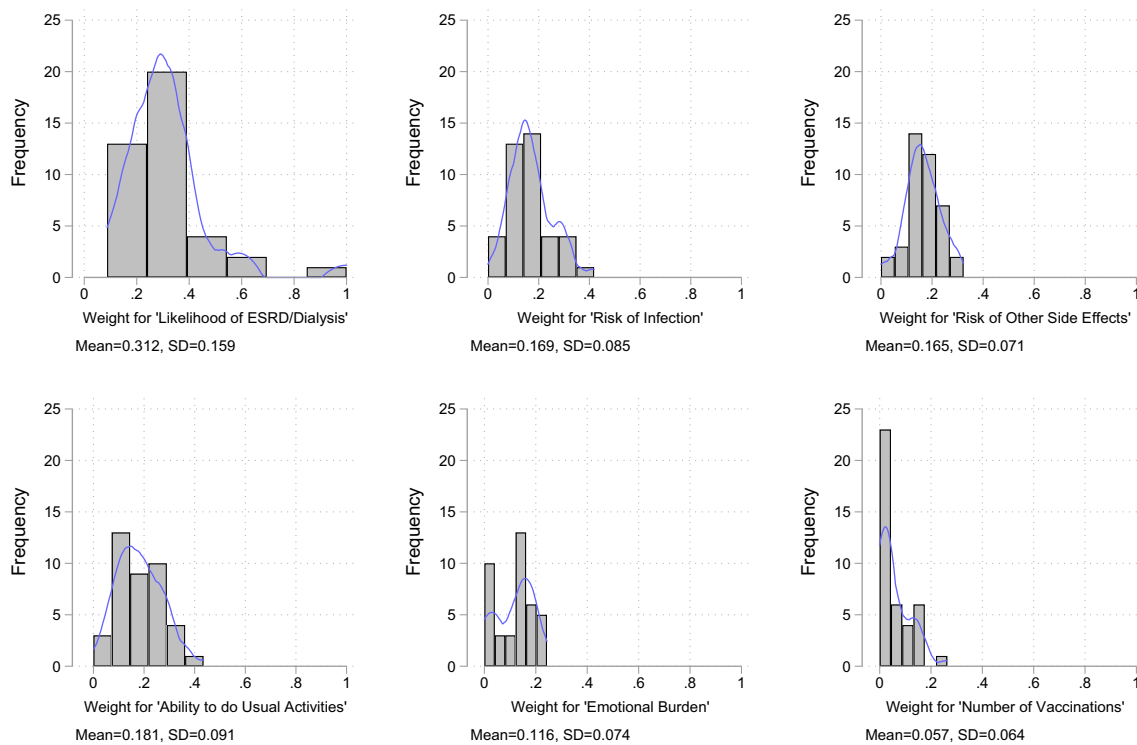
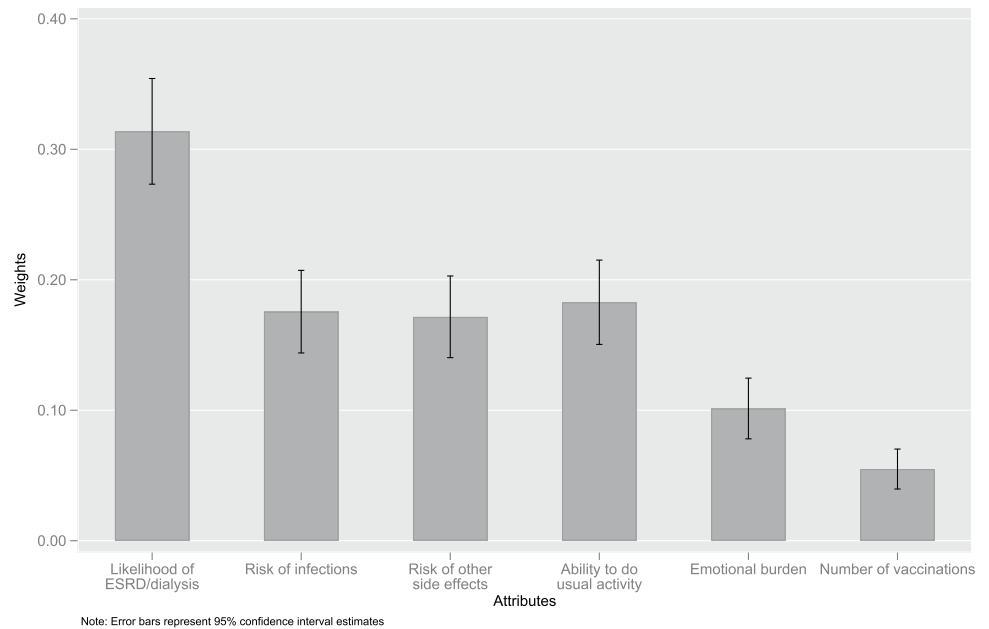
**Table 3** Participant characteristics

	Overall sample [ $N = 40$ ]
Sex	
Female	27 (67.5)
Male	13 (32.5)
Age, years	
Min, max	20, 63
Median (Q1, Q3)	41.0 (35.0, 47.8)
Mean (SD)	41.6 (11.3)
Time since diagnosis	
6–12 months	3 (7.5)
1–2 years ago	8 (20.0)
2–5 years ago	8 (20.0)
5–10 years ago	6 (15.0)
10–20 years ago	11 (27.5)
20+ years ago	4 (10.0)
Ability to perform usual activity	
I am not at all able to perform my usual activities	0 (0.0)
I am somewhat able to perform my usual activities	14 (35.0)
I am very much able to perform my usual activities	26 (65.0)
Level of emotional burden	
I am not at all emotionally burdened by my disease	4 (10.0)
I am somewhat emotionally burdened by my disease	30 (75.0)
I am very much emotionally burdened by my disease	4 (10.0)

Data are expressed as  $n$  (%) unless otherwise specified

$N$  sample size,  $Q1$  interquartile range, first quartile,  $Q3$  interquartile range, quartile 3,  $SD$  standard deviation,  $Min$  minimum,  $Max$  maximum

**Fig. 1** Attribute preference weights. Error bars represent the standard deviation. *ESRD* end-stage renal/kidney disease



**Fig. 2** Heterogeneity in attribute weights. The *blue line* represents the density plot of the weight distribution. *ESRD* end-stage renal/kidney disease, *SD* standard deviation

for improving the level of emotional burden thought that emotions could be controlled or self-regulated ( $n = 4$ ), that they had the support system to manage emotional burdens ( $n = 1$ ), and that emotional burden was a consequence of other factors that needed to be addressed ( $n = 1$ ). Those

who weighted emotional burden highly perceived emotional health to impact other aspects of their life and health/treatment outcomes ( $n = 4$ ), they did not want to burden others ( $n = 1$ ) or they perceived emotional burden as an ongoing issue they had to manage ( $n = 1$ ). Participants who placed

high weights on improving the likelihood of ESRD/dialysis cited its impact on life expectancy ( $n = 4$ ) and fear of undergoing/not wanting to undergo dialysis ( $n = 2$ ). Participants who placed lower weights on ESRD did so as they did not consider it worth improving the likelihood of ESRD/dialysis at the expense of other treatment attributes ( $n = 2$ ) or they placed higher priority on short-term quality-of-life impacts than longer-term ESRD risks ( $n = 2$ ).

All attributes except the number of vaccinations were ranked third by at least one participant and were thus included in the consistency test. Paired  $t$ -tests showed that the weights elicited from the original tasks and the validation tasks were not statistically significant at both the  $p < 0.05$  and  $p < 0.1$  levels (Fig. 3).

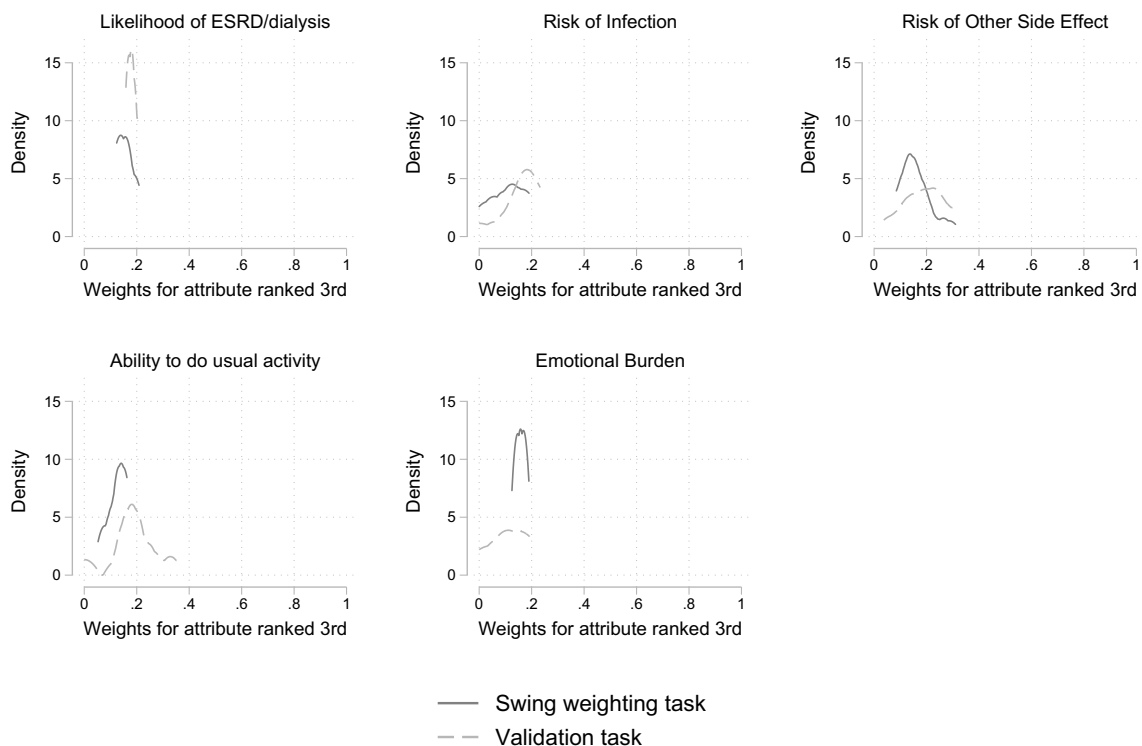
### 3.2.2 Partial Value Functions

Participants had a linear partial value function for reducing the number of vaccinations, i.e. the marginal returns were linear. For all other attributes, participants had a non-linear partial value function, demonstrating increasing marginal returns, i.e. for one unit of improvement for an attribute, participants placed greater value on further improvements on the attribute (e.g. participants placed greater weight on reducing the likelihood of ESRD/dialysis from 20 to 10% than reducing it from 30% to 20%) [Fig. 4]. For the number of vaccinations, participants valued every unit of

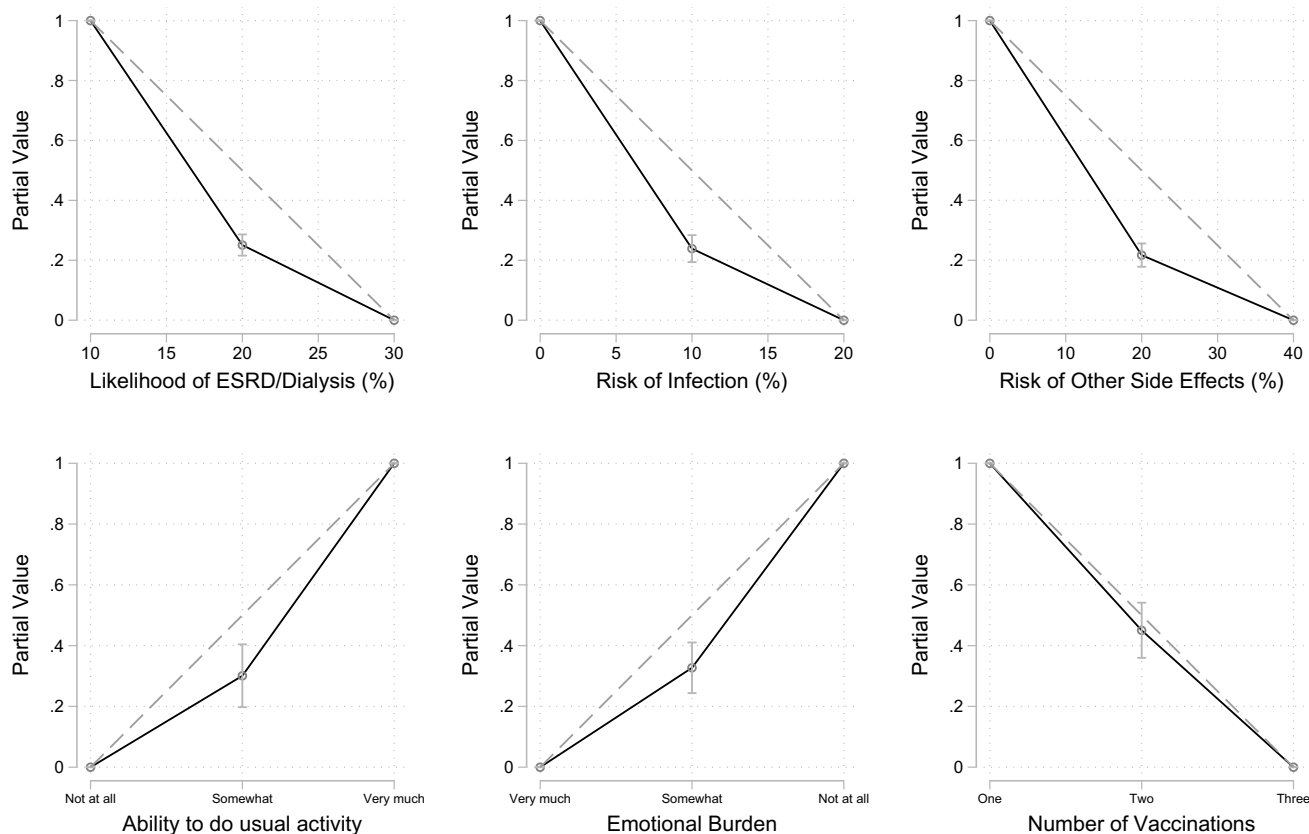
improvement similarly. There was little variation in the shape of participant's value functions for likelihood of ESRD/dialysis, risk of infections and risk of other adverse effects (indifference point: SD range 0.111–0.141). There was more variation in value functions for the ability to perform usual activity, the level of emotional burden and the number of vaccinations (indifference point: SD range 0.260–0.323).

Qualitative data provided insights into the reasons for the shape of participants' value functions. For instance, the greater variation in shape of the value functions for the quality-of-life attributes may be due to variation in the interpretation of the attribute levels. For instance, participants who interpreted 'somewhat able' to perform usual activities to include greater abilities, such as being able to take care of themselves or being able to work, had a more linear partial value function ( $n = 6$ , average indifferent point = 0.315), while those who interpreted 'somewhat able' to still include restrictions on their life had a less linear partial value function ( $n = 8$ , average indifferent point = 0.205).

The qualitative data also provided insight into the reason participants had increasing returns to improvements in attributes. Using the likelihood of ESRD/dialysis as an example, some participants were unable to articulate a reason for the shape of their value function. However, others offered reasons that suggest the shape of the function may be both due to substantive value concepts (value of hope,



**Fig. 3** Consistency test comparing the weights for the third-ranked attribute. *ESRD* end-stage renal/kidney disease



**Fig. 4** Partial value functions. The *dotted line* represents the linear partial value function. *ESRD* end-stage renal/kidney disease

i.e. providing hope that technology can advance further ( $n = 5$ ); rewarding the technological difficulty associated with achieving better performance ( $n = 4$ ); reaching their expectation of a ‘good treatment’ ( $n = 2$ ) and artefacts of the study design (an anchoring of 10 points at 0% risk ( $n = 6$ ); and the application of relative risk (for instance, 10% is half of 20%, while 20% is two-thirds of 30%).

### 3.2.3 Benefit–Risk Assessment

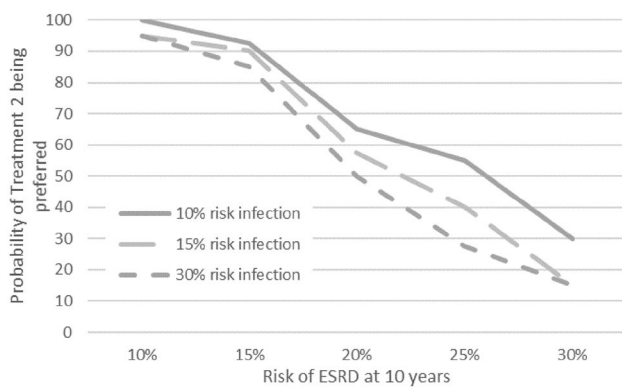
Various scenarios were run to explore how treatments could be assessed across the different attributes. As an example, Fig. 5 shows the output generated by the ‘simulation tool’ for the illustrative scenario defined in Table 2, which outlines the attribute levels of two hypothetical treatments, i.e. ‘Treatment 1’ and ‘Treatment 2’. It demonstrates how the tool can inform the evaluation of alternative treatment profiles early in the product development process by estimating the proportion of participants in the preference study who would prefer a treatment profile (‘treatment 2’) and how this proportion of participants changes as the performance of ‘treatment 2’ varies; for example, if both the likelihood of ESRD at 10 years and the risk of infection increase or

decrease. Specifically, the uptake of ‘treatment 2’ is particularly sensitive to the likelihood of ESRD at 10 years, varying from 94% to 100% when the likelihood of ESRD at 10 years is only 10%, to 15–30% when the likelihood of ESRD at 10 years increases to 30%.

## 4 Discussion

As health care decision makers are increasingly using patient preference data to inform approval and reimbursement, it is important that methods to capture such data for patients with rare diseases are also explored. Commonly used preference methods, such as DCE or thresholding, either require too large a sample size to be applied in many rare diseases or do not capture preferences for multiple attributes. As direct rating methods, such as swing weighting, capture a complete set of preference data from each participant, they require fewer participants than choice-based methods, which capture a partial set of preference data from each participant. However, the benefit of eliciting a complete set of preference data from each participant is accompanied by the concern that the cardinal value elicitation places a high cognitive burden on participants, which needs to be considered carefully in the





**Fig. 5** Using the simulation tool to understand the impact of treatment profiles on patient preferences. *ESRD* end-stage renal/kidney disease

design of the study. One-to-one interaction with participants, which can be conducted virtually, may help to guide them through the exercise, and additionally allows for collection of qualitative information to understand participants' choices and preferences.

This study illustrates the use of direct rating, specifically swing weighting and partial value functions, to elicit patient preferences in the context of a small sample ( $n = 40$ ) of patients with a rare disease. There are no existing studies on the magnitude of patients' preferences against which the results of the study can be compared. Nevertheless, the consistent tests and qualitative data provide some reassurance as to the validity of the elicited preferences, as well as the rationale behind heterogeneity in the preference weights placed on improvements in the different attributes. The consistency tests did not identify statistically significant variation in preferences across elicitation tests, although this may partly be a function of the small sample sizes. The qualitative data suggested that the elicitation exercise was sensitive to variations in the participants' interpretation of the attributes and that patients were able to use the exercise to share their preferences. However, the study was not designed to formally test the validity of direct rating methods. Further research could usefully test the validity of the results of direct rating exercises, such as by comparing the results from swing weighting and choice-based methods, providing a sufficient sample size could be reached.

As has been observed in previous studies [26, 27], the swing weighting exercise identified substantial preference heterogeneity, particularly for the reduction in the number of vaccinations, improving the level of emotional burden and improving the likelihood of ESRD/dialysis. This may suggest that methods which elicit preferences at an individual level, such as swing weighting, capture more heterogeneity than population-level methods, such as DCE. However,

further comparative work would be required to demonstrate this, and to also test that the heterogeneity reflects substantive variation in participants preferences and is not an artefact of the study design. One published study identified a comparable large amount of heterogeneity in both an adaptive swing weighting exercise and a thresholding exercise [27]. This may suggest that large amounts of preference heterogeneity may be the result of the use of methods that elicit individual-level patient-preference data, rather than being an artefact of swing weighting specifically. However, further research could usefully compare classic swing weighting and population-level choice methods such as DCE.

The study provided insights into patient preferences that are informative to the development of drugs in IgAN. It was expected that patients placed a high value on avoiding ESRD, which is consistent with findings from existing research that have identified kidney function and avoiding the need for dialysis or transplant as important outcomes for patients with glomerular disease [28]. However, the study generated two sets of valuable insight. First, it estimated the rate at which patients are willing to trade such benefits to avoid possible treatment adverse effects. Such data are being increasingly relied on by regulators to inform approval decisions [2]. Second, it demonstrated that patients were willing to trade-off some of the longer-term reduction in the likelihood of ESRD in exchange for short-term improvements in quality of life, particularly improvements in their ability to perform usual activities. This is consistent with the findings of studies in other diseases [29].

These insights can support the prioritization of endpoints for inclusion in clinical studies. Clinical trials in IgAN typically focus on surrogate endpoints [30–33] that physicians and clinical researchers deem important, but may not necessarily reflect outcomes or aspects of treatment that patients' value. Understanding patient preferences can help to ensure a patient-centric trial design and that new drugs are measured and evaluated against outcomes that are important to patients.

The study also demonstrates how insights into patient preferences can support the assessment of alternative product profiles early in the development process. A simulation tool was built that facilitated engagement with the preference data to flexibly evaluate treatment profiles.

This study is subject to some limitations. While the objective of the study was to illustrate the use of swing weighting with a small sample of patients, such small sample sizes may also limit the ability to explore variation in preferences between patients. This is of particular concern given the heterogeneity in patient preferences. One implication of this is that it might not be possible to prospectively identify the subgroups of patients who would benefit from a treatment. In such instances, it may be necessary to adopt shared decision making and individual patient preference elicitation at

the point of prescribing drugs. Furthermore, given the challenges in recruiting patients with rare diseases, the resulting sample may not be representative of the broader patient population, thereby limiting the generalizability of the results.

## 5 Conclusion

Direct rating methods can be used to collect preference data from small sample groups of patients, such as in rare diseases. Further work should continue to test the validity of the preference data captured using direct rating methods.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40271-021-00521-3>.

## Declarations

**Funding** This study was supported by Novartis Pharma AG.

**Conflict of Interest** NZ and ATG are full-time employees of Novartis, while NC was a full-time employee of Novartis at the time the study was conducted. KM and RL are employees of Evidera, contracted to implement this study. KAH was an employee of Evidera at the time of writing this manuscript.

**Ethics Approval** This study was approved by the Ethical and Independent (E&I) Institutional Review Board (E&I study number: 19147-01).

**Consent to Participate** All participants provided informed consent prior to study participation.

**Consent for Publication** Not applicable.

**Availability of Data and Material** Not applicable.

**Code Availability** Not applicable.

**Author Contributions** All authors were involved in the design and conception of the study; critically revised and approved the final version for publication, and meet the following four criteria for authorship recommended by the International Committee of Medical Journal Editors (ICMJE): (1) substantial contributions to the conception and design of the work; (2) drafting the work or revising it critically for important intellectual content; (3) final approval of the version to be published; and (4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References

- Craig BM, Lancsar E, Muhlbacher AC, Brown DS, Ostermann J. Health preference research: an overview. *Patient*. 2017;10(4):507–10. <https://doi.org/10.1007/s40271-017-0253-9>.
- Food and Drug Administration (FDA). Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and De Novo Classifications. Guidance for Industry and Food and Drug Administration Staff 2019. 2019. <https://www.fda.gov/media/99769/download>. Accessed 28 Sep 2020.
- Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of patient preference studies in HTA decision making: A NICE perspective. *Patient*. 2020;13(2):145–9. <https://doi.org/10.1007/s40271-019-00408-4>.
- Ho M, Saha A, McCleary KK, Levitan B, Christopher S, Zandlo K, et al. A framework for incorporating patient preferences regarding benefits and risks into regulatory assessment of medical technologies. *Value Health*. 2016;19(6):746–50. <https://doi.org/10.1016/j.jval.2016.02.019>.
- Marsh K, van Til JA, Molsen-David E, Juhnke C, Hawken N, Oehrlein EM, et al. Health preference research in Europe: a review of its use in marketing authorization, reimbursement, and pricing decisions—report of the ISPOR stated preference research special interest group. *Value Health*. 2020;23(7):831–41. <https://doi.org/10.1016/j.jval.2019.11.009>.
- van Overbeeke E, Whichello C, Janssens R, Veldwijk J, Cleemput I, Simoens S, et al. Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(1):57–68. <https://doi.org/10.1016/j.drudis.2018.09.015>.
- Marsh K, Hauber B, DiSantostefano R. Quantitative Benefit-Risk Assessment: What Methods are Being Used? How Far has Industry Come? DIA 2020 Virtual Global Annual Meeting; 2020.
- Morel T, Ayme S, Cassiman D, Simoens S, Morgan M, Vandebroek M. Quantifying benefit-risk preferences for new medicines in rare disease patients and caregivers. *Orphanet J Rare Dis*. 2016;11(1):70. <https://doi.org/10.1186/s13023-016-0444-9>.
- Hauber AB, Fairchild AO, Reed JF. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11(4):319–29. <https://doi.org/10.1007/s40258-013-0028-y>.
- Jackson Y, Flood E, Rhoten S, Janssen EM, Lundie M. Acro-Voice: eliciting the patients' perspective on acromegaly disease activity. *Pituitary*. 2019;22(1):62–9. <https://doi.org/10.1007/s11102-018-00933-9>.
- Muhlbacher A, Johnson FR. Choice experiments to quantify preferences for health and healthcare: state of the practice. *Appl Health Econ Health Policy*. 2016;14(3):253–66. <https://doi.org/10.1007/s40258-016-0232-7>.
- Ryan M, Bate A, Eastmond C, Ludbrook A. Use of discrete choice experiments to elicit preferences. *Qual Health Care*. 2001;10(Suppl 1):i55–60.
- Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324–31. <https://doi.org/10.1016/j.drudis.2019.05.001>.
- Krucien N, Watson V, Ryan M. Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Econ*. 2017;26(12):e1–16. <https://doi.org/10.1002/hec.3459>.
- European Medicines Agency (EMA). Benefit-risk methodology. 2020. <https://www.ema.europa.eu/en/about-us/support-research/benefit-risk-methodology>. Accessed 23 Sep 2020.
- Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making—an introduction: Report 1 of the ISPOR MCDA emerging good practices task force. *Value Health*. 2016;19(1):1–13. <https://doi.org/10.1016/j.jval.2015.12.003>.
- Tervonen T, Gelhorn H, Sri Bhashyam S, Poon JL, Gries KS, Rentz A, et al. MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiol Drug Saf*. 2017;26(12):1483–91. <https://doi.org/10.1002/pds.4255>.

18. Chouaid C, Germain N, De Pourville G, Aballea S, Korchagina D, Baldwin M, et al. Patient preference for chronic obstructive pulmonary disease (COPD) treatment inhalers: a discrete choice experiment in France. *Curr Med Res Opin.* 2019;35(5):785–92. <https://doi.org/10.1080/03007995.2019.1574507>.
19. George AT, Zaour N, Nic Lochlainn EM. PUK31—the burden associated with immunoglobulin A nephropathy (IgAN). *Value Health.* 2018;21:S480. <https://doi.org/10.1016/j.jval.2018.09.2829>.
20. Tyagi N, Aasaithambi S, Chauhan J, George A, Zaour N. PUK32 patient insights for immunoglobulin A nephropathy (IgAN) using social media listening. *Value Health.* 2019;22:S919. <https://doi.org/10.1016/j.jval.2019.09.2716>.
21. Zaour N, Mayländer M, Walda S, George AT. PUB159: Patient Journey, Perceptions, and Burden Associated with Immunoglobulin A Nephropathy (IgAN): A Qualitative Study. *J Am Soc Nephrol.* 2020;31(abstract supplement).
22. Hollin IL, Craig BM, Coast J, Beusterien K, Vass C, DiSantostefano R, et al. Reporting formative qualitative research to support the development of quantitative preference study protocols and corresponding survey instruments: guidelines for authors and reviewers. *Patient.* 2020;13(1):121–36. <https://doi.org/10.1007/s40271-019-00401-x>.
23. Reich HN, Troyanov S, Scholey JW, Cattran DC, Toronto GR. Remission of proteinuria improves prognosis in IgA nephropathy. *J Am Soc Nephrol.* 2007;18(12):3177–83. <https://doi.org/10.1681/ASN.2007050526>.
24. Nam KH, Kie JH, Lee MJ, Chang TI, Kang EW, Kim DW, et al. Optimal proteinuria target for renoprotection in patients with IgA nephropathy. *PLoS ONE.* 2014;9(7):e101935. <https://doi.org/10.1371/journal.pone.0101935>.
25. Tervonen T, Pignatti F, Postmus D. From individual to population preferences: comparison of discrete choice and Dirichlet models for treatment benefit-risk tradeoffs. *Med Decis Making.* 2019;39(7):879–85. <https://doi.org/10.1177/0272989X19873630>.
26. Postmus D, Richard S, Bere N, van Valkenhoef G, Galinsky J, Low E, et al. Individual trade-offs between possible benefits and risks of cancer treatments: results from a stated preference study with patients with multiple myeloma. *Oncologist.* 2018;23(1):44–51. <https://doi.org/10.1634/theoncologist.2017-0257>.
27. SriBhashyam S, Marsh K, Quartel A, Weng HH, Gershman A, Longo N, et al. A benefit-risk analysis of pegvaliase for the treatment of phenylketonuria: a study of patients' preferences. *Mol Genet Metab Rep.* 2019;21:100507. <https://doi.org/10.1016/j.ymgmr.2019.100507>.
28. Carter SA, Gutman T, Logeman C, Cattran D, Lightstone L, Bagga A, et al. Identifying Outcomes Important to Patients with Glomerular Disease and Their Caregivers. *Clin J Am Soc Nephrol.* 2020;15(5):673–84. <https://doi.org/10.2215/CJN.13101019>.
29. Postmus D, Mavris M, Hillege HL, Salmonson T, Ryll B, Plate A, et al. Incorporating patient preferences into drug development and regulatory decision making: results from a quantitative pilot study with cancer patients, careers, and regulators. *Clin Pharmacol Ther.* 2016;99(5):548–54. <https://doi.org/10.1002/cpt.332>.
30. Lv J, Zhang H, Wong MG, Jardine MJ, Hladunewich M, Jha V, et al. Effect of Oral Methylprednisolone on Clinical Outcomes in Patients With IgA nephropathy: the TESTING Randomized Clinical Trial. *JAMA.* 2017;318(5):432–42. <https://doi.org/10.1001/jama.2017.9362>.
31. Rauen T, Eitner F, Fitzner C, Sommerer C, Zeier M, Otte B, et al. Intensive supportive care plus immunosuppression in IgA nephropathy. *N Engl J Med.* 2015;373(23):2225–36. <https://doi.org/10.1056/NEJMoa1415463>.
32. Lafayette RA, Canetta PA, Rovin BH, Appel GB, Novak J, Nath KA, et al. A randomized, controlled trial of rituximab in IgA nephropathy with proteinuria and renal dysfunction. *J Am Soc Nephrol.* 2017;28(4):1306–13. <https://doi.org/10.1681/ASN.2016060640>.
33. Shima Y, Nakanishi K, Kaku Y, Ishikura K, Hataya H, Matsuyama T, et al. Combination therapy with or without warfarin and dipyridamole for severe childhood IgA nephropathy: an RCT. *Pediatr Nephrol.* 2018;33(11):2103–12. <https://doi.org/10.1007/s00467-018-4011-6>.