

# Development and Pilot Testing of an Encounter Tool for Shared Decision Making About the Treatment of Graves' Disease

Juan P. Brito,<sup>1</sup> Ana Castaneda-Guarderas,<sup>2</sup> Michael R. Gionfriddo,<sup>2</sup> Naykky Singh Ospina,<sup>1</sup> Spyridoula Maraka,<sup>1</sup> Diana S. Dean,<sup>1</sup> Regina M. Castro,<sup>1</sup> Vahab Fatourehchi,<sup>1</sup> Hossein Gharib,<sup>1</sup> Marius N. Stan,<sup>1</sup> Megan E. Branda,<sup>3</sup> Rebecca S. Bahn,<sup>1</sup> and Victor M. Montori<sup>1</sup>

**Background:** The best treatment option for patients with Graves' disease (GD) depends on each person's situation and how the differences between the treatment options matter to them in bringing resolution to their illness. The objective of this study was to develop and test an encounter decision tool (GD Choice) for patients and clinicians to engage in shared decision making about the treatment of GD.

**Methods:** GD Choice was developed using an iterative process based on the principles of interaction design and participatory action research. To evaluate the impact of the tool, a controlled before–after study was conducted, assessing the use of GD Choice versus usual care (UC).

**Results:** Sixty-eight patients were enrolled, 37 to UC and 31 to GD Choice. At baseline, the groups were similar. Treatment discussion length was similar in both arms. After their visit, patients in both groups had similar knowledge about the options, except for GD Choice patients knowing significantly more about the complications of treatment (correctly answered by 83% vs. 55%;  $p=0.04$ ). Compared with UC, patients in the GD Choice arm had greater involvement in decision making observed on video recordings of clinical encounters (mean OPTION scale score, 35% vs. 30%;  $p=0.02$ ), but reported similar levels of decisional comfort and participation in shared decision making.

**Conclusions:** GD Choice increases engagement in the decision-making process and knowledge regarding intervention complications without increasing the length of consultation. These promising results support the conduct of a randomized trial of GD Choice versus UC in a large multicenter trial.

## Introduction

GRAVES' DISEASE (GD) IS THE CAUSE of hyperthyroidism in approximately 0.3–0.5% of the U.S. population (1). The abnormally elevated thyroid hormone levels arising from GD can result in a variety of symptoms, including weight loss, insomnia, muscle weakness, menstrual irregularities, tachycardia, and sight-threatening eye disease (1). Over time, GD decreases quality of life (2) with impacts on physical and social function, mental health, and overall well-being. Without treatment, GD may result in thyroid storm or life-threatening cardiac complications.

There are three treatment options for GD: antithyroid drugs (ATDs), radioiodine ablation, and surgery (3). These differ in their mechanisms to induce resolution of hyperthyroidism, as well as in their safety, convenience, and cost. No

one option is considered superior for all patients. Indeed, patients with GD face a daunting task in deciding with their endocrinologist which treatment option is best for them in their situation. In fact, The Guidelines for Hyperthyroidism from the American Thyroid Association and American Association of Clinical Endocrinologists, published in 2011, stated that “Once the diagnosis has been made, the treating physician and patient should discuss each of the treatment options, including the logistics, benefits, expected speed of recovery, drawbacks, potential side effects, and cost” (3). Nevertheless, the Guidelines offered no further specifics on how to structure such a discussion and offered no tools to facilitate or support it.

Decision aids (DAs) are tools designed for use during the clinical encounter to create conversations between patients and clinicians. In these conversations, they share evidence-based

<sup>1</sup>Division of Diabetes, Endocrinology, Metabolism, and Nutrition, Department of Medicine; <sup>2</sup>Knowledge and Evaluation Research Unit; <sup>3</sup>Health Services Research; Mayo Clinic, Rochester, Minnesota.

information about the options and their relative favorable and unfavorable features, and collaboratively deliberate about how each of the options will affect the patient's life and prognosis (4). A systematic review of randomized trials concluded that use of DAs increase patient knowledge, accurate risk perceptions, satisfaction with the decision, and number of patients achieving decisions that were informed and consistent with their values (5).

In this study, an encounter decision tool, Graves' Disease Choice (GD Choice), was developed, and its impact on the conversations patients with GD have with their endocrinologist and on the quality of the decision-making process they followed was evaluated.

## Methods

The Mayo Clinic Institutional Review Board approved all study procedures. Patients and clinicians provided written informed consent before participation.

### Development of GD Choice

A multidisciplinary team was formed comprised of a senior research designer, patients, and clinicians. A novel practice-based, patient-centered approach was used, based on design/participatory action research, developed and validated by the authors' team for decision aid development (6,7). This process involves: (i) review and synthesis of the available evidence; (ii) analysis of usual practice; (iii) development of an initial prototype; (iv) field testing (i.e., use in real clinical encounters with patients facing the decision of interest and their clinicians) of the prototype in clinical settings under the study team's supervision; and (v) successive iterations and field testing of the prototype (Fig. 1). The development work continues until the tool is able to support the creation of a conversation consistently between patients and clinicians in which patients verbalize "trying on" the different options, testing the hypotheses that the option considered will be a best fit for them. The goal is not to maximize support, offering comprehensive

information, but rather to build on what clinicians and patients usually attempt while providing the smallest, least intrusive support that will achieve the tool's goal.

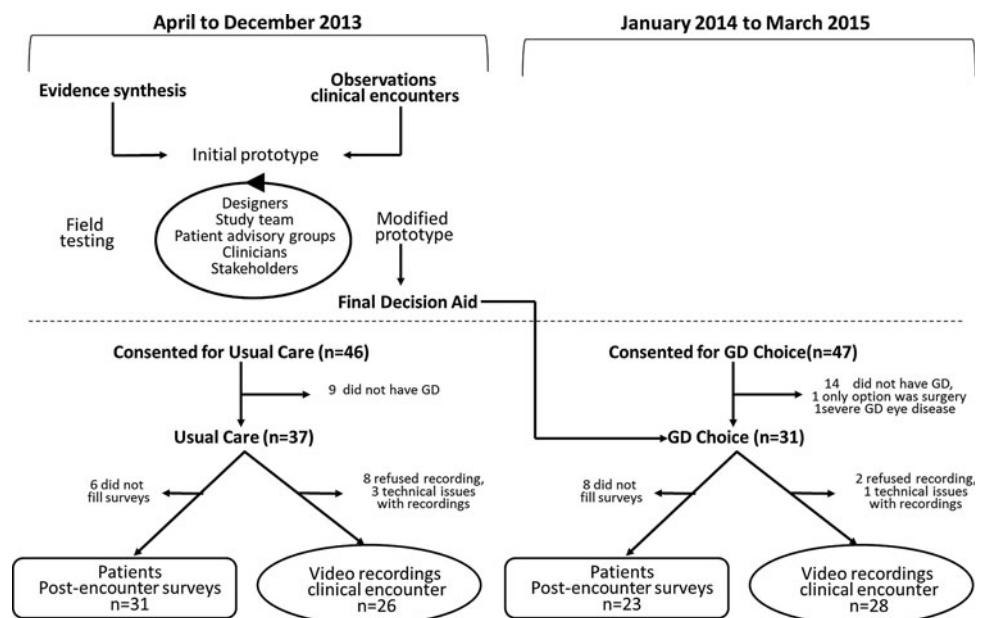
### Evaluation of tool's impact on patient-important outcomes

**Design, study setting, and participants.** This controlled before–after study (Fig. 1) took place at the Thyroid Clinic within the Division of Endocrinology at the Mayo Clinic (Rochester, MN). This group includes nine clinical endocrinologists with expertise in the management of thyroid disorders. For the first nine months (first cohort), all patients referred to the clinic for treatment of GD who gave consent to participate in this study experienced usual care (UC). For the following 15 months, a second cohort of patients with GD seen at the clinic who had not been seen in the previous nine months (and thus were not part of the UC arm) experienced care supplemented by the intervention (GD Choice).

Eligible patients were identified from the upcoming appointment schedules of the clinicians. Eligible patients were adults with GD who, as judged by their clinicians, needed treatment for this condition. Pregnant women were excluded because the tradeoff between benefits and harms for each option was different from the average patient with GD. Finally, clinicians were given the opportunity to exclude other patients where the risks and benefits of one or more treatment options were thought to be significantly different from that expected for the average GD patient (i.e., severe Graves' orbitopathy, large and compressive goiters). Eligible patients were invited to participate prior to their scheduled appointment, and were enrolled in the study after giving written informed consent for participation.

**Outcome measures.** Immediately after the visit, patients and clinicians completed a brief survey to ascertain patient knowledge about the options, and to document participants' perception of the quality of the decision-making process.

**FIG. 1.** Overview of study design. Patients consenting for usual care preceded patients consenting for GD Choice.



When patients and clinicians consented, the encounter was recorded using video or audio based on their preference.

To evaluate the decision-making process, patients and clinicians completed the 9-item Shared Decision Making Questionnaire (SDM-Q-9) (8,9). This instrument consists of nine statements, which can be rated on a six-point scale from 0 = "completely disagree" to 5 = "completely agree." Summing up all items leads to a raw total score between 0 and 45. When required, up to two missing items were imputed using the mean of the items that were filled out to calculate the raw score. No total score was calculated if more than two items were missing. Multiplication of the raw score by 20 and then dividing it by nine provided a new score that ranged from 0 to 100, where 0 indicates the lowest possible level of SDM and 100 the highest possible level (8).

To evaluate decisional conflict, the "uncertainty" subscale of the modified Decisional Conflict Scale (DCS) was used (10). This subscale consists of three questions on a 0–4 scale, where at least two responses were needed for analysis. The overall uncertainty subscale was converted to a 100-point scale, where a score of 0 means that a patient feels extremely certain about the best choice and 100 means that patient feels extremely uncertain about best choice.

A questionnaire was created to assess patients' knowledge about treatment options for GD. Responses were true, false, or I don't know. These items were fashioned after similar items that showed construct validity and responsiveness in previous decision aid trials (11,12).

In duplicate, video recordings were also evaluated for the extent to which clinicians engaged patients in the decision-making process using the Observer OPTION<sup>12</sup> scale (13). The Observer OPTION<sup>12</sup> scale is designed to evaluate the level of shared decision making occurring in clinical encounters, as captured through audio or video. The authors' group has used this tool extensively to assess video/audio recordings with excellent inter-rater reliability (intraclass correlation coefficient >0.7) (6,11). For this study, four observers conducted the review. Three of the observers had to come to agreement with a primary observer. The intraclass correlation coefficient ranged from 0.83 to 0.97.

To assess fidelity to the tool, a checklist was developed collaboratively by the study team that assessed the clinician's use of the DA as intended. The video/audio recording was used in both the intervention group and the control group. The items of the fidelity checklist were scored as present or not. The total score was calculated by adding and dividing the total number of items, with a range of 0% to 100%. Higher scores indicate higher fidelity to the intended use of the DA. In control group visits, this score reflects the extent to which elements of the tool were already present as part of the usual care of GD patients at Mayo Clinic. Assessing fidelity is important because the design of the DA can accommodate a range of clinician uses, some of which do not lead to shared decision making (14). Finally, the final treatment option for each patient was extracted, and the treatment decision was compared between the two arms.

**Analytic methods.** To estimate the sample size, a standard deviation difference was used in a continuous outcome (SDM-Q-9 scale) (9). Sample size estimates are based on having 80% power to detect a minimum difference of 10 points on a two-sided test ( $\alpha=0.05$ ). The sample size needed was 60 patients with 30 per arm. Baseline characteristics

were reported as means and standard deviations for continuous variables, and categorical values were reported as counts and frequencies and compared between study arms using *t*-tests and chi-square tests, respectively. As clinicians can have multiple patients per arm, any clustering impact the clinician might have had on the outcomes of interest was explored. The intra-cluster correlation was found to be 0 for all outcomes, except DCS uncertainty. The analysis comparing results across arms for this outcome is adjusted by the clustering effect. Any imbalances ( $p<0.05$ ) were accounted for in subsequent analyses. All data management and statistical analyses were conducted using SAS v9 (Cary, NC) and Stata v11 (College Station, TX).

## Results

### *Content and evidence*

To provide accurate and up-to-date evidence regarding the effectiveness of the three treatment options across critical outcomes in the DA, a systematic review of the literature was conducted, which is published elsewhere (15).

### *Prototype development*

Clinicians and patient representatives were convened to discuss the evidence summary, seeking to uncover blind spots in the evidence or in the review. The team also observed and recorded 10 encounters with patients with GD at the Endocrinology Clinic at Mayo Clinic-Rochester Campus, to identify how treatment decisions are usually made, how options are presented, and how deliberation takes place. From stakeholder meetings and encounter observations, a comprehensive set of salient issues for patients was identified: (i) effectiveness of therapy (when am I going to start feeling better?), (ii) side effects, (iii) the need for thyroid replacement therapy, (iv) cost of therapy, (v) frequency of follow-up during therapy and with thyroid replacement therapy, and (vi) the importance of concomitant Graves' ophthalmopathy. The evidence was then organized using this issue list (Table 1).

The recommendations and standards for the construction of decision aids were followed from The International Patient Aid Standards (IPDAS) (16). The initial prototype of the decision aid was created through a series of discussions with members of the study team, clinicians, and patients. An initial prototype was piloted in three real-life clinical encounters, looking for patterns in the conversations, and the issues, problems, and challenges were documented. The study team then evaluated the quality of the conversations and the ability of the prototype to facilitate the decision-making process. This process was repeated 11 times until the entire team reached consensus that the prototype was successful in involving patients in decision making. After making the necessary revisions, the endocrinologists were introduced to the final prototype decision aid (see ref. [17]) and they were trained in its use, with its intended use being demonstrated in one-on-one sessions. Otherwise, using the DA was highly intuitive and required minimal training.

### *Impact of the decision aid, GD Choice*

Of the 103 patients approached, 93 expressed interest, of which 68 were eligible (25 patients had hyperthyroidism not due to GD). Of these, 37 received UC without the GD Choice

TABLE 1. SUMMARY OF EVIDENCE FOR PATIENT IMPORTANT OUTCOMES

	<i>Cure rate</i>	<i>Cost<sup>a</sup></i>	<i>Adverse effects</i>	<i>Need for lifelong thyroid replacement</i>	<i>Speed of recovery</i>	<i>Frequency of follow-up (first year)</i>	<i>Mild eye disease</i>
ATD	50–60%	\$300–400	One in four patients will drop treatment due to side effects (nausea, skin rash, pruritus)	0%	4–8 weeks	4–8 per year	Unchanged
RAI	90%	\$4000–5000	New eye disease (15%) Minor risk for malignancy Impaired male fertility (4–6 months) Avoid pregnancy (6 months)	90%	12–18 weeks	2–3 per year	May worsen
Surgery	95%	\$30,000–40,000	Permanent scar (1–2 inches) Temporal hoarseness (4%) Permanent hoarseness (<1%) Need for calcium pills for lifetime (1%)	100%	Immediately	1–2 year	Unchanged

<sup>a</sup>Cost does not include visits or tests.

ATD, antithyroid drugs; RAI, radioactive iodine therapy.

tool, and 31 received care with the GD Choice tool. Baseline characteristics were similar in both groups (Table 2).

#### *Fidelity and process*

Out of a maximum of 100 points (100%), encounters in the DA arm had an average fidelity score of 51% (range 5–90%). In contrast, encounters in the UC arm had an average score of 22% (range 5–40%), a mean difference of 29% [CI 18–41] (Table 3 and Supplementary Table S1; Supplementary Data are available at [www.liebertpub.com/thy](http://www.liebertpub.com/thy)).

Importantly, patients in the GD Choice arm received more information regarding considerations, adverse effects, and impact of the treatment options compared with patients in the UC arm (Fig. 2). For instance, patients in the GD Choice arm more frequently received information about the possibility of hoarseness as a complication of surgery than did patients in the UC arm (59% vs. 4%;  $p=0.01$ ).

Use of the tool had no material impact on encounter duration (information not available for nine encounters): median duration was 39 minutes (range 5–70 minutes) for UC encounters ( $n=30$ ), and 40 minutes (range 20–90 minutes) for GD Choice encounters ( $n=29$ ).

#### *Knowledge*

Patient knowledge scores were found to be similar between arms, with a median of four correct responses out of the five questions asked (Supplementary Table S2). More patients in the GD Choice arm answered correctly the only knowledge question regarding complications of treatment (people with GD who have surgery may have problems with their voice;  $p=0.04$ ; Table 3).

#### *Uncertainty, quality, and patient involvement in decision making*

Self-reported measures of the quality of shared decision making—DCS Uncertainty Subscale, and SDM-Q-9—failed to show a statistically significant difference between UC and GD Choice arms. Analysis of video recordings showed

greater patient involvement in decision making in the GD Choice arm as evaluated by the OPTION scale (Table 3).

#### *Treatment choice*

Of the patients who responded to the question about treatment choice at the end of the encounter ( $n=55$ ; 32 UC, 23 DA), more patients in the GD Choice arm (70% vs. 44%;  $p=0.058$ ) reported having reached a treatment decision (Table 3). Medical record review showed more patients choosing radioactive iodine therapy (RAI; 42%) in the GD Choice arm than in the UC arm (31%;  $p=0.66$ ).

#### **Discussion**

Compared with UC, patients participating in encounters using GD Choice were more involved in the decision-making process, received and understood more information about adverse consequences of treatment, and were more likely to report making a decision at the visit. GD Choice use did not have an effect on self-reported shared decision making or on treatment choice. These effects were achieved with suboptimal fidelity, but with minimal training and without lengthening duration of the encounter.

#### *Limitations*

The design of the study poses the potential risk of selection bias. However, randomization at the clinician level, in a small trial, could have led to imbalanced groups in terms of number of participants and in terms of the communication skills of participating clinicians. Likewise, randomization at patient level had the potential for contamination, as the small group of clinicians could have potentially transferred elements of the decision aid to the control encounters, biasing the trial results toward the null. Therefore, the controlled before–after design was the most feasible approach for this study.

The main limitation of the study relates to its sample size, producing imprecise results and limited robustness of statistically significant results (few different responses on the surveys would have been needed to change a statistically

TABLE 2. BASELINE CHARACTERISTICS

	Control (n=37)	GD Choice (n=31)	Total (n=68)	p-Value <sup>a</sup>
Sex				0.82
Female	29 (78%)	25 (81%)	54 (79%)	
Male	8 (22%)	6 (19%)	14 (21%)	
Age (years)				0.61
Mean (SD)	41.8 (14)	44.1 (16)	42.8 (15)	
Range	(18–72)	(19–76)	(18–76)	
Missing	0	1		
Current smoker				0.08
No	28 (87.5%)	20 (69.0%)	48 (78.7%)	
Yes	4 (12.5%)	9 (31.0%)	13 (21.3%)	
Missing	5	2		
Thyroid gland (g)				0.57
Mean (SD)	35.7 (26)	35.7 (19)	35.7 (23)	
Range	(10–150)	(10–100)	(10–150)	
Missing	4	0		
Graves' orbitopathy				0.71
No	30 (81%)	28 (90%)	58 (85%)	
Yes	5 (13%)	3 (10%)	8 (12%)	
Missing	2 (6%)	0 (0%)	2 (3%)	
Total number of prescribed medications patient is currently taking				0.46
Mean (SD)	4.8 (4)	6 (5)	5 (4.7)	
Range	(0–16)	(0–21)	(0–21)	
Missing	1	1		
Race/ethnicity				0.68
Asian	1 (3%)	1 (4%)	2 (4%)	
African American	0 (0%)	1 (4%)	1 (2%)	
White	28 (90%)	20 (87%)	48 (90%)	
Other	2 (7%)	1 (4%)	3 (6%)	
Missing	6	8		
Education				0.58
High school graduate or less	9 (30%)	5 (22%)	14 (26%)	
Some college or associates degree	9 (30%)	9 (40%)	18 (34%)	
Four year college graduate	3 (10%)	5 (22%)	8 (15%)	
Graduate or professional school degree	9 (30%)	4 (17%)	13 (24%)	
Missing	7	8		
Marital status				0.28
Married/life partner	22 (73%)	13 (56%)	35 (66%)	
Separated	1 (3%)	0 (0%)	1 (2%)	
Divorced	2 (7%)	6 (26%)	8 (15%)	
Widowed	0 (0%)	1 (4%)	1 (2%)	
Single	5 (17%)	3 (13%)	8 (15%)	
Missing	7	8		
Income				0.20
<\$40k	5 (17%)	6 (35%)	11 (24%)	
\$40k to <\$80k	8 (27%)	6 (35%)	14 (30%)	
\$80k+	16 (55%)	5 (29%)	21 (46%)	
Missing	8	14		

<sup>a</sup>Continuous comparisons conducted using *t*-tests or Wilcoxon's rank-sum tests, and categorical comparisons conducted using chi-square or Fisher's exact tests, as appropriate.

significant to nonsignificant results). Additionally, it was not possible to obtain video recording and survey data for all patients; missing data impair the ability to draw any robust conclusions and further the need for a larger study. Moreover, this study enrolled many participants who had already had an initial discussion regarding GD before the clinical encounters at Mayo Clinic, likely an unavoidable effect of conducting this work in a referral endocrine practice. Patients with less knowledge could derive greater benefit from knowledge

transfer and involvement in decision making (18). In addition, given that the clinicians used the DA only a few times, it is likely that repeated use may improve its impact.

#### Comparison with prior research

The present findings are consistent with prior studies assessing the impact of DAs on patient involvement, but this study stands alone as the first effort to promote shared

TABLE 3. IMPACT OF THE GD CHOICE ON TREATMENT CHOICE, DECISION-MAKING PROCESS, AND KNOWLEDGE

Source of data	Outcomes	Control	GD Choice	Mean diff GD Choice – UC [CI]	p-Value
Chart review	Treatment decision	n = 36	n = 31		
	ATDs	16 (44%)	10 (32%)		0.66
	RAI	11 (31%)	13 (42%)		
	Surgery	6 (17%)	4 (13%)		
	No decision	0 (0%)	1 (3%)		
Surveys	Other	3 (8%)	3 (10%)		
	Level of uncertainty (DCS scale)	n = 31	n = 23		
	Mean [CI]	30 [12–47]	27 [7–47]	–2.8 [–27–22]	0.2 <sup>a</sup>
	Median (IQR)	25 (8–50)	25 (0–50)		
	Level of shared decision making (SDMQ-Patient)	n = 28	n = 23		
	Mean [CI]	19 [17–21]	20 [19–21]	0.99 [–0.98–3.0]	0.47
	Median (IQR)	19 (17–22)	21 (18–22)		
	Level of shared decision making (SDMQ-Clinicians)	n = 32	n = 30		
	Mean [CI]	19 [17–21]	20 [18–23]	1.4 [–1.5–4]	0.18 <sup>a</sup>
	Median (IQR)	21 (16–22)	20 (18–23)		
Video recordings	Knowledge transfer	n = 31	n = 23		
	Overall median (IQR)	4 (3–5)	4 (3–5)		0.33
	Question side effects, n (%)	17 (55%)	19 (83%)		0.043
	Fidelity score	n = 26	n = 28		
	Mean (SD)	22 (9)	51 (28)	29 (18,41)	0.01
Video recordings	Median IQR	23 (10–30)	45 (30–80)		
	Patient engagement (OPTION score)	n = 26	n = 28		
	Mean (SD)	30 (8)	35 (9)	5 (0.2–10)	0.02
Video recordings	Median (IQR)	30 (26–36)	34 (31–41)		

Results shown as proportions with Fisher’s exact test statistics for categorical and mean [CI], median (IQR) with Wilcoxon’s rank-sum for continuous outcomes.

<sup>a</sup>Clustering of clinician accounted for in estimates and p-value; p-value (ICC) listed.

decision making for treatment decisions in patients with GD. The Diabetes Medication Choice study (11) found a significant increase in involvement, with an average of 21.8/100 points on the OPTION scale. Compared with this trial and others, in the present study, the observed gain in patient

involvement with the use of a DA was smaller. This may represent the limited use of the tool by any individual clinician (suboptimal fidelity), and gains in this area may be seen with continued use or further training. Other causes of the small effect size may relate to issues with the design of the

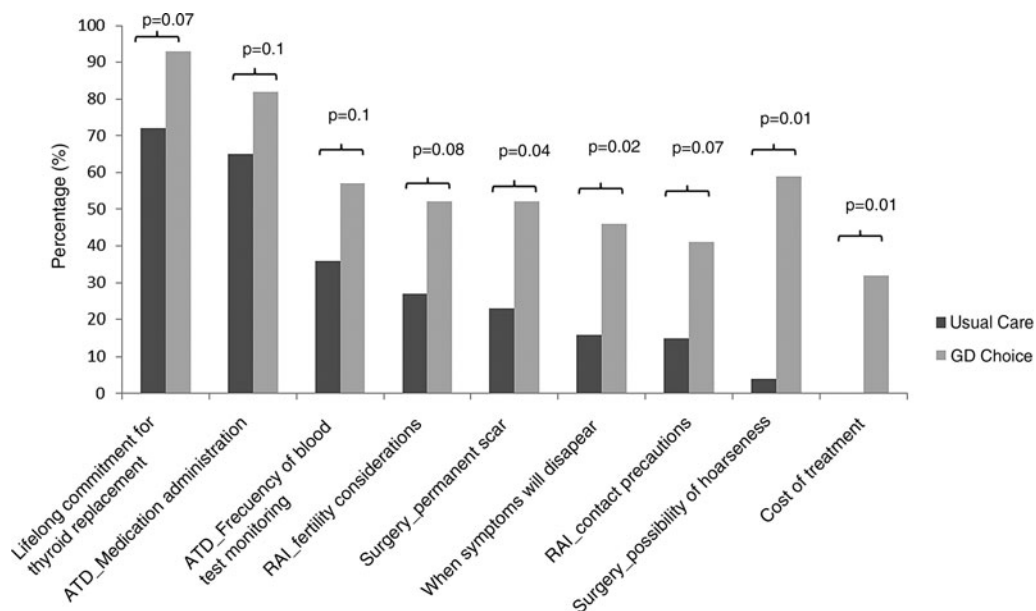


FIG. 2. Frequency of topics discussed during the encounters.

tool that did not promote conversation as expected and high shared decision-making skills of the clinicians at baseline.

While not statistically significant, an improvement in decisional uncertainty was found among patients in the DA arm; similar findings were observed in a randomized trial of a DA for statin treatment (12). In the latter, patients in the statin decision aid group had greater decisional comfort (10.6 points higher on a 100-point scale) than did patients randomized to UC. In contrast to these two and other trials using decision aids, improvement in overall knowledge transfer could not be demonstrated in the present study (18).

The present findings also demonstrate disagreement between two validated tools to assess shared decision making: SDM-Q-9 and OPTION. A prior cross-sectional study compared these scales among patients with chronic diseases facing medical decisions (19). This study found that the association between the total scores was weak ( $r=0.19$ ). The OPTION scale assesses decision making from the perspective of an external observer through video or audio recordings, while SDM-Q-9 assesses shared decision making from the patient and clinician perspective (20). Thus, it is possible that each scale measures a different construct of the shared decision-making process. Because patients in the UC group were not exposed to the shared decision-making experience, they may report the UC experience as meeting their expectations for involvement, unaware that their participation could be greater. Similarly, clinicians reported SDM-Q-9 in the UC arm before they had the experience of using GD Choice.

It was found that patients with GD using the GD Choice had similar treatment choices to patients in UC. Nevertheless, a nonstatistically significant increase was observed, in the number of patients in the GD Choice arm choosing RAI and the number of patients in the UC arm choosing ATDs. A systematic review of DAs found that the effect of DAs on choices was variable and inconsistent (5). This effect on choice of treatment could be a spurious effect resulting from the small sample size, and may not represent a true effect of the decision aid. Additionally, other studies using DAs have found the length on consultation may vary from eight minutes shorter to 23 minutes longer (5). A statistically significant difference in the length of the consultation with or without the DA was not detected. This finding strengthens the argument that the use of a DA, besides being an ethical imperative for patient autonomy and patient-centered care, can be feasibly integrated into the specialty care of patients with GD.

## Conclusions

GD Choice—a feasible, acceptable, efficient, and effective encounter decision aid—was developed to promote shared decision making about treatment options for GD in the specialty setting. GD Choice increased engagement in the decision-making process and knowledge regarding side effects of the interventions without increasing the length of consultation. The results of this preliminary evaluation support plans to conduct a larger multicenter trial to evaluate its impact on the care of patients with GD facing treatment decisions.

## Acknowledgments

This publication was made possible by CTSA Grant Number UL1 TR000135 from the National Center for Ad-

vancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH), and the Mayo Clinic Robert D and Patricia E. Kern Center for the Science of Health Care Delivery. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH or the Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery.

Clinical Trial Registration: URL: <https://clinicaltrials.gov>. Unique identifier: NCT02107794.

## Author Disclosure Statement

The authors have nothing to disclose.

## References

1. Brent GA 2008 Clinical practice. Graves' disease. *N Engl J Med* **358**:2594–2605.
2. Abraham-Nordling M, Torring O, Hamberger B, Lundell G, Tallstedt L, Calissendorff J, Wallin G 2005 Graves' disease: a long-term quality-of-life follow up of patients randomized to treatment with antithyroid drugs, radioiodine, or surgery. *Thyroid* **15**:1279–1286.
3. Bahn RS, Burch HB, Cooper DS, Garber JR, Greenlee MC, Klein I, Laurberg P, McDougall IR, Montori VM, Rivkees SA, Ross DS, Sosa JA, Stan MN 2011 Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid* **21**:593–646.
4. The 2012 IPDAS Background Document. Available at: <http://ipdas.ohri.ca/resources.html> (accessed July 2014).
5. Stacey D, Legare F, Col NF, Bennett CL, Barry MJ, Eden KB, Holmes-Rovner M, Llewellyn-Thomas H, Lyddiatt A, Thomson R, Trevena L, Wu JH 2014 Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* **1**:CD001431.
6. Breslin M, Mullan RJ, Montori VM 2008 The design of a decision aid about diabetes medications for use during the consultation with patients with type 2 diabetes. *Patient Educ Couns* **73**:465–472.
7. Montori VM, Breslin M, Maleska M, Weymiller AJ 2007 Creating a conversation: insights from the development of a decision aid. *PLoS Med* **4**:e233.
8. Scholl I, Kriston L, Dirmaier J, Buchholz A, Harter M 2012 Development and psychometric properties of the Shared Decision Making Questionnaire—physician version (SDM-Q-Doc). *Patient Educ Couns* **88**:284–290.
9. Kriston L, Scholl I, Holzel L, Simon D, Loh A, Harter M 2010 The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Educ Couns* **80**:94–99.
10. O'Connor AM 1995 Validation of a decisional conflict scale. *Med Decis Making* **15**:25–30.
11. Mullan RJ, Montori VM, Shah ND, Christianson TJ, Bryant SC, Guyatt GH, Perestelo-Perez LI, Stroebel RJ, Yawn BP, Yapuncich V, Breslin MA, Pencille L, Smith SA 2009 The diabetes mellitus medication choice decision aid: a randomized trial. *Arch Intern Med* **169**:1560–1568.
12. Weymiller AJ, Montori VM, Jones LA, Gafni A, Guyatt GH, Bryant SC, Christianson TJ, Mullan RJ, Smith SA 2007 Helping patients with type 2 diabetes mellitus make treatment decisions: statin choice randomized trial. *Arch Intern Med* **167**:1076–1082.
13. Elwyn G, Hutchings H, Edwards A, Rapport F, Wensing M, Cheung WY, Grol R 2005 The OPTION scale: measuring

- the extent that clinicians involve patients in decision-making tasks. *Health Expect* **8**:34–42.
14. Wyatt KD, Branda ME, Anderson RT, Pencille LJ, Montori VM, Hess EP, Ting HH, LeBlanc A 2014 Peering into the black box: a meta-analysis of how clinicians use decision aids during clinical encounters. *Implement Sci* **9**:26.
  15. Sundaresh V, Brito JP, Wang Z, Prokop LJ, Stan MN, Murad MH, Bahn RS 2013 Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network meta-analysis. *J Clin Endocrinol Metab* **98**:3671–3677.
  16. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, Thomson R, Barratt A, Barry M, Bernstein S, Butow P, Clarke A, Entwistle V, Feldman-Stewart D, Holmes-Rovner M, Llewellyn-Thomas H, Moumjid N, Mulley A, Ruland C, Sepucha K, Sykes A, Whelan T 2006 Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* **333**:417.
  17. Mayo Clinic Shared Decision Making National Resource Center 2015 Graves' Disease Treatment Choice. Available at: <http://shareddecisions.mayoclinic.org/decision-aid-information/graves-disease-decision-aid/> (accessed October 9, 2015).
  18. Coylewright M, Branda M, Inselman JW, Shah N, Hess E, LeBlanc A, Montori VM, Ting HH 2014 Impact of socio-demographic patient characteristics on the efficacy of decision AIDS: a patient-level meta-analysis of 7 randomized trials. *Circ Cardiovasc Qual Outcomes* **7**:360–367.
  19. Scholl I, Kriston L, Dirmaier J, Harter M 2015 Comparing the nine-item Shared Decision-Making Questionnaire to the OPTION Scale—an attempt to establish convergent validity. *Health Expect* **18**:137–150.
  20. Sepucha KR, Scholl I 2014 Measuring shared decision making: a review of constructs, measures, and opportunities for cardiovascular care. *Circ Cardiovasc Qual Outcomes* **7**:620–626.

Address correspondence to:

*Victor M. Montori, MD*

*Knowledge and Evaluation Research Unit*

*Division of Diabetes, Metabolism, and Nutrition*

*Mayo Clinic*

*200 First Street SW*

*Rochester, MN 55905*

*E-mail: montori.victor@mayo.edu*