

# High Level of Agreement between Electronic and Paper Mode of Administration of a Thyroid-Specific Patient-Reported Outcome, ThyPRO

Sofie Larsen Rasmussen<sup>a</sup> Lars Rejnmark<sup>b</sup> Eva Ebbelhøj<sup>b</sup>  
Ulla Feldt-Rasmussen<sup>a</sup> Åse Krogh Rasmussen<sup>a</sup> Jakob Bue Bjorner<sup>c</sup>  
Torquil Watt<sup>a</sup>

<sup>a</sup>Department of Medical Endocrinology, Copenhagen University Hospital Rigshospitalet, Copenhagen, and

<sup>b</sup>Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark;

<sup>c</sup>QualityMetric (an Optum Company), Lincoln, R.I., USA

## Key Words

Thyroid diseases · Health-related quality of life · Patient-reported outcome measure · Agreement · Mode of administration

## Abstract

**Introduction and Purpose:** Use of electronic questionnaires to collect health-related quality-of-life data has evolved as an alternative to paper questionnaires. For the electronic questionnaire to be used interchangeably with the validated paper questionnaire, measurement properties similar to the original must be demonstrated. The aim of the present study was to assess the equivalence between the paper version and the electronic version of the thyroid-related quality-of-life questionnaire ThyPRO. **Methods:** Patients with Graves' hyperthyroidism or autoimmune hypothyroidism in a clinically stable phase were included. The patients were recruited from two endocrine outpatient centers. All patients completed both versions in a randomized test-retest set-up. Scores were compared using intraclass correlation coefficients (ICCs), paired t tests and Bland-Altman plots. Limits of agreement were compared with data from a previous paper-paper test-retest study. **Results:** 104 patients were included. ICCs were generally high for the 13 scales, ranging from 0.76

to 0.95. There was a small but significant difference in the scale score between paper and electronic administration for the Cosmetic complaints scale, but no differences were found for any other scale. Bland-Altman plots showed similar limits of agreement compared to the earlier test-retest study of the paper version of ThyPRO. **Conclusion:** Based on our analyses using ICCs, paired t tests and Bland-Altman plots, we found adequate agreement between the paper and electronic questionnaires. The statistically significant difference in score found in the Cosmetic complaints scale is small and probably clinically insignificant.

© 2016 European Thyroid Association  
Published by S. Karger AG, Basel

## Introduction

In clinical trials, questionnaires about health-related quality of life (HRQL) are increasingly used as outcomes for evaluating impact of health on patients' lives and effect of treatment [1, 2]. Traditionally, HRQL data has been captured by paper questionnaires. In recent years, electronic collection of HRQL data has evolved as an alternative, and many established paper questionnaires have been converted into electronic equivalents [2], thus reducing missing and ambiguous data as well as the time-con-

suming effort of manual typing [2]. Electronic data collection can be made accessible through the Internet, and might thus be more convenient for the respondent [3].

ThyPRO is a patient-reported outcomes questionnaire measuring HRQL in patients with benign thyroid disorders. The paper version of the questionnaire has been shown to be valid and reliable and is recommended for use in clinical trials [4, 5]. An electronic version of the ThyPRO has recently been developed. For the electronic questionnaire to be used interchangeably with the validated paper questionnaire, it must apply to the same standards as the original questionnaire in terms of measurement properties. A general concern has been that the outcome might not be equivalent to the paper version [1]. Lack of equivalence might arise from insufficient experience with electronic devices. Also, the electronic version might appear graphically different (e.g. font size, radio buttons), which could increase measurement error. Researchers can aim at reducing these differences, but can never fully control the graphical appearance because it depends on respondent screen size, screen, and browser settings [6]. There is also concern about a more direct effect on outcomes, which might be caused by systematic differences in response patterns [3]. Also, response bias might occur because characteristics of (non)responders may differ between the two modes, leading to indirect effects on score outcome. For example, Mayr et al. [7] found that participants responding electronically were younger, better educated, and more often male, compared with participants preferring the paper-and-pencil version, but after adjusting for this bias there was no direct effect on outcome scales.

A wide variety of studies have previously shown good equivalence between electronic and paper questionnaires [2, 8–17], but it is still generally recommended to evaluate measurement properties of the electronic version separately, rather than expecting validations of the paper version to a priori be extrapolated to the electronic version [1, 3]. Thus, the aim of the present study was to assess the equivalence between the paper version and the electronic version of the ThyPRO questionnaire [1].

## Methods

### *Study Population*

We included patients with a diagnosis of Graves' hyperthyroidism or autoimmune hypothyroidism who were in a clinically stable phase of their disease and who had a valid e-mail address and access to the Internet at home. Clinically stable phase was based on an overall clinical judgment, including symptomatology (no

symptoms or sign of overt thyroid dysfunction), biochemistry (euthyroid or judged to remain at the current state of thyroid function), and treatment (e.g. no newly instituted thyroid treatment). Thus, patients were excluded if they were newly diagnosed, pregnant, with significant comorbidity, less than 18 years of age, or not fluent in Danish.

Patients were recruited from the endocrine outpatient clinics at Copenhagen University Hospital Rigshospitalet (RH; May 2012 to October 2013) and Aarhus University Hospital (AUH; May 2012 to March 2013). Patients with autoimmune hypothyroidism were only included between May 2012 and October 2012. Patients were recruited in two ways. Clinicians at AUH included eligible patients when they were visiting the outpatient clinic. At RH, eligible patients were identified in electronic patient files and approached by telephone.

All patients gave their oral informed consent. According to Danish law, patient-reported outcomes research does not require and thus cannot obtain approval by ethical committees, and a completed questionnaire is regarded as consent. The study was approved by the Danish Data Protection Agency (case file 30-0679) and conducted in accordance with the Declarations of Helsinki.

### *Study Procedure*

The study applied a randomized cross-over test-retest design. At the time of inclusion, patients were assigned a serial number. The serial numbers were beforehand randomly coupled to one of two groups, determining the order of questionnaire completion. The clinical staff was not blinded to the result of the randomization. Included patients received the paper questionnaire and a letter with instructions as to which questionnaire to complete first. These items were handed over to the patients at AUH and sent by postal mail at RH.

Patients assigned to group 1 first completed the paper questionnaire. After 14 days they received an email with a link to the electronic questionnaire, along with a completion instruction. Patients assigned to group 2 first completed the electronic questionnaire, as described for group 1. After 14 days they received an email with instructions to complete and return the paper questionnaire. Fourteen days after completing the second questionnaire, the patients in both groups completed an electronic questionnaire about sociodemographic data.

### *Patient-Reported Outcome*

ThyPRO is self-administered and measures HRQL with 13 scales, covering physical and mental symptoms, well-being, and function, as well as impact of thyroid disease on participation (i.e. social and daily life) and overall HRQL [4, 5]. It consists of 85 items and, on average, takes 14 min to complete. Each scale ranges from 0 to 100 with increasing scores indicating more symptoms or greater impact of disease.

### *Survey Conduction*

The electronic questionnaire was developed in and administered via the software system SurveyXact [18]. It was developed to resemble the original paper version as closely as possible, but one change was made – items concerning physical symptoms spanned three pages, in contrast to only two pages in the paper version, to avoid need for scrolling.

Patients were only allowed to choose one response option in the electronic version. When one or more items were unanswered,

the respondent was prompted to go back and complete all items. However, the respondent was also given the option to continue to the next page, in order to keep it equivalent to the paper version, where patients may choose not to complete particular items. If two adjacent responses were marked in a paper questionnaire, one of the two was randomly picked. Data from all paper questionnaires were manually entered twice, to reduce typing errors.

Patients who did not answer the electronic questionnaire were sent an electronic reminder 3 days after they had received the questionnaire, and again after 6 days if they still had not responded. Similarly, patients who did not answer the paper questionnaire were sent an electronic reminder after 7 and 12 days, respectively.

Clinical, descriptive data were stored in a SAS database and all data were stored on a secured server.

### Statistical Analyses

Data management and analyses were conducted with SAS 9.3.

Level of agreement was evaluated by intraclass correlation coefficients (ICCs), ranging from 0 to 1. When assessing agreement, an ICC of 0.61 and above is considered 'substantial', and a value of 0.81 and above is considered 'almost perfect' [19]. ICCs take into account random error, between-subject variability, and within-subject variability. It also takes systematic error into account, which makes it the most suitable measure of agreement [2].

To assess the magnitude of a possible systematic error, we also compared the mean difference of scale scores between paper and electronic responses (paper minus electronic) with a paired t test. Order effects were tested by an independent t test comparing the difference in scale score between time 1 and time 2 in the two groups.

Data were evaluated graphically by Bland-Altman plots of the scale scores [20]. This was done by plotting the score difference (paper minus electronic) against the average paper and electronic score for each individual, including 95% limits of agreement (LOA) calculated by  $1.96 SD_{diff}$ . LOA then indicates how well the two methods agree. It allows further analysis of the data because it separates the systematic bias from the random error. Furthermore, it is possible to see the direction of any possible bias as well as outliers. We compared the Bland-Altman plots with plots of the data from a previous test-retest study of the paper version of ThyPRO [5], which applied a similar design.

## Results

A total of 191 eligible patients were invited to participate. Of these, 176 accepted the invitation and 120 completed both questionnaires, corresponding to a response rate of 68%. Some lacking responses were caused by an initial breakdown of the email delivery system, which caused some mails to be sent out too late. The completion rates were similar in both groups: 69% for the electronic-first group versus 67% for the paper-first group. Fifteen of the 120 patients who completed both questionnaires were excluded because they did not respond to the two questionnaires within the predefined time interval of 10–

25 days. One patient was excluded because she became pregnant. Thus, the final sample comprised 104 patients (fig. 1). No differences in scale score levels between the compliant sample and the dropouts were found (unpaired t test).

Table 1 shows the characteristics of the study population. Seventy-three of the included patients were diagnosed with Graves' disease, and 31 were diagnosed with autoimmune hypothyroidism. As shown in table 2, all ICCs were above 0.70, and all but 3 were above 0.81.

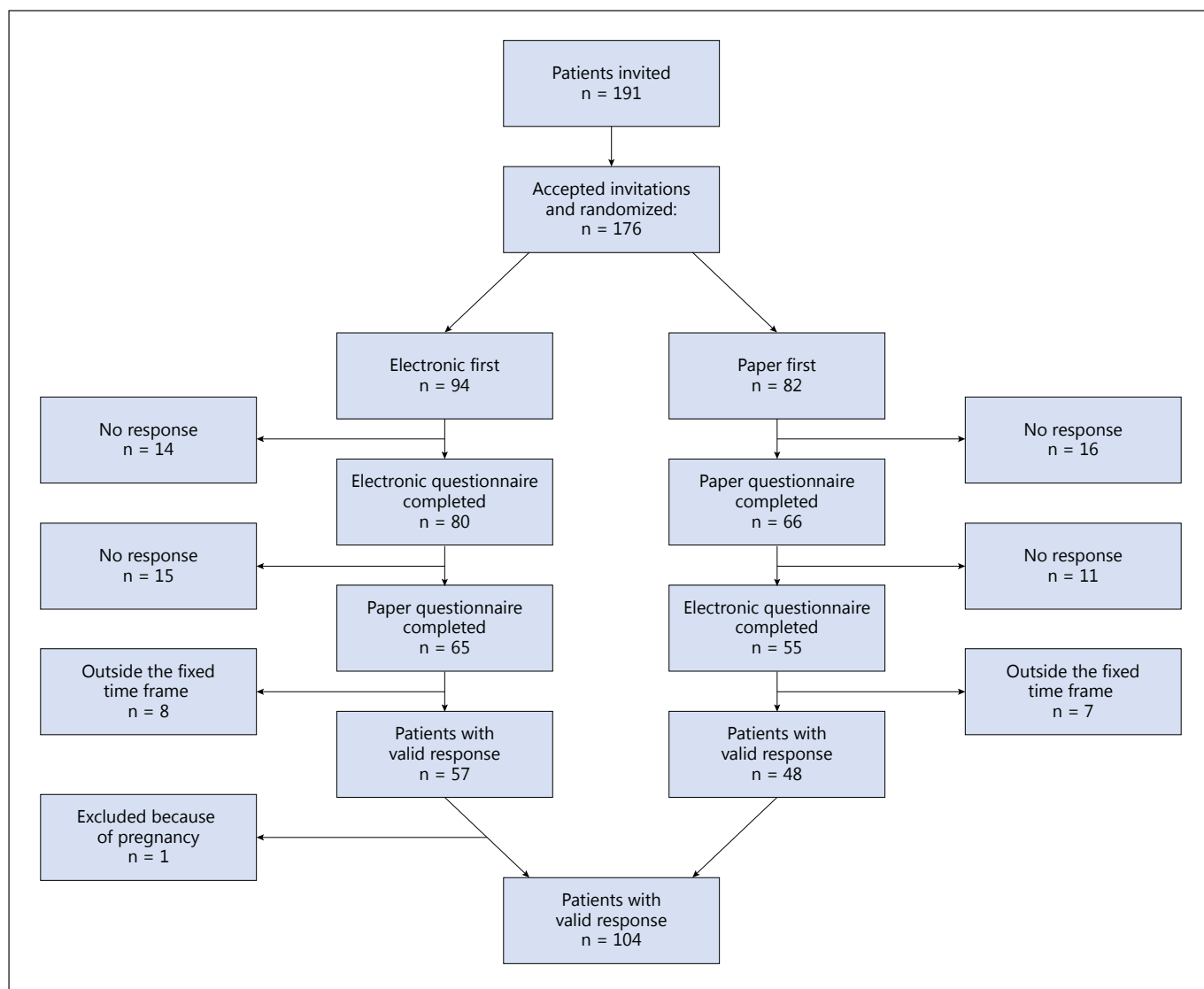
There was no difference in score level between the two administration modes, except for the Cosmetic complaints scale, where the mean score with the electronic mode was 2.5 points (on a scale of 0–100) higher than the paper mode. For the same scale, a significant order effect was identified, in that the score difference was 4.7 points higher in group 2 versus group 1 ( $p = 0.0021$ ).

Figure 2 shows the Bland-Altman plots with LOA from both the present study and the previous test-retest study of the paper version of ThyPRO. As shown, no patterns indicating nonuniformity of disagreement was found; the LOAs were similar to those of the paper-paper study and were around  $\pm 20$ –30 points in magnitude.

## Discussion

In this study, our goal was to evaluate the equivalence of the electronic and paper versions of ThyPRO. The results showed generally good ICCs – all but 3 (out of 13) were above 0.81. In concordance with our results, a previous study reported generally high ICCs when it came to the paper-paper test-retest of the ThyPRO questionnaire [5]. The previous study was conducted in a different population of patients with a larger number of benign thyroid diagnoses and with different inclusion criteria. The results are thus not directly comparable since the ICC is dependent on the between-subjects variability of the sample. Despite this, the similarity of the study results supports that the two modes can be used interchangeably [2]. Variation between the electronic and paper scores thus most likely reflect random variation due to the test-retest setup, and not attribution of differences between the two modes. We chose the same time interval (10–25 days) as in the previous study, so any time-related difference should be similar.

The differences in score between the electronic and paper administration were insignificant, except for the cosmetic complaints scale. However, the difference in mean score was small (2.5 on a scale of 0–100) and most likely



**Fig. 1.** Patient flow chart.

clinically insignificant, and the methods might agree very well, even if there is a small bias [21].

The Bland-Altman plots provide a visual impression of the agreement in each subject (fig. 2). Two basic assumptions must be fulfilled to correctly analyze the Bland-Altman plots: difference between the methods must be constant, i.e. zero slope, and constant variance of measures across the entire range of measures, i.e. homoscedasticity [20]. Based on a visual impression of the plots, these assumptions are most probably fulfilled on all scales. Levels of agreement from the current study were also similar to those from paper-paper agreement tests,

suggesting that the identified range represents the inherent variability of ThyPRO and over time, and not systematic errors attributable to different modes. There might be some true change incorporated in the LOA caused by nonstable clinical conditions within the patients. Thus, some of the constructs under investigation might have truly changed during the 2 weeks. This might especially be the case of the mental state scales since the patients were primarily evaluated by their clinical stability, not their mental state.

Other studies investigating agreement between electronic and paper questionnaires have chosen a wide vari-

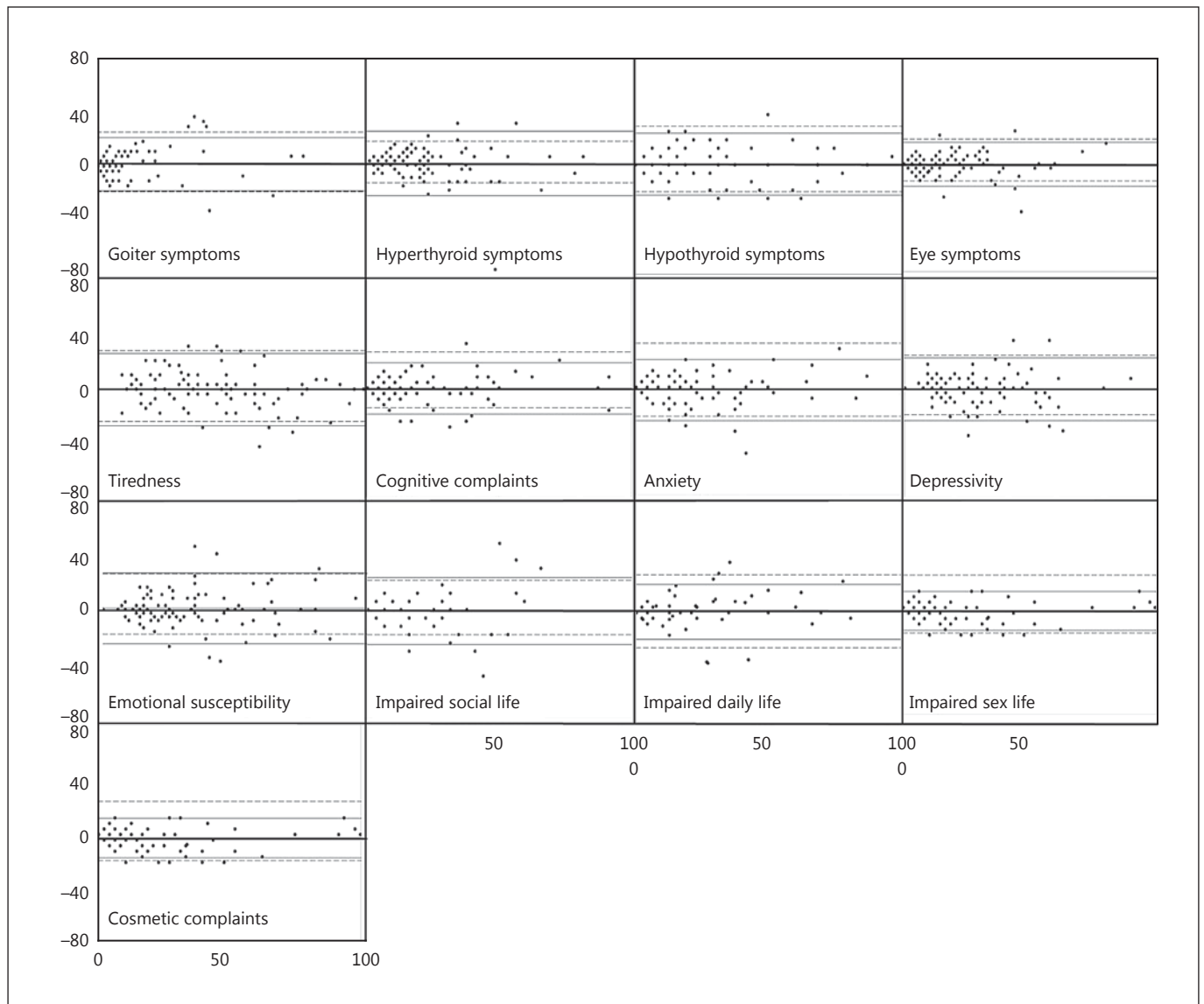
**Table 1.** Sociodemographic and clinical characteristics of the study populations

Variables	Subsample		Whole sample (n = 104)
	electronic first (n = 56)	paper first (n = 48)	
Women/men	52 (93)/4	41 (85)/7	93 (89)/11
Age	44±12	43±11	43±12
Diagnosis			
Graves' disease	39 (70)	34 (71)	73 (70)
Autoimmune hypothyroidism	17 (30)	14 (29)	31 (30)
Months since diagnosis	55 (2–197)	43 (3–381)	46 (2–381)
Thyroid treatment			
No current thyroid treatment	9 (16)	7 (15)	16 (15)
Current antithyroid medication	12 (22)	18 (38)	30 (29)
Current L-thyroxine	28 (50)	17 (35)	45 (43)
Radioiodine	3 (5)	1 (2)	4 (4)
Thyroidectomy	4 (7)	5 (10)	9 (9)
Current thyroid function			
Euthyroid	40 (71)	33 (69)	73 (70)
Mildly hypothyroid	1 (2)	1 (2)	2 (2)
Overtly hypothyroid	0	0	0
Mildly hyperthyroid	6 (11)	10 (21)	16 (15)
Overtly hyperthyroid	5 (9)	0	5 (5)
Education			
No vocational training	5 (9)	2 (4)	7 (7)
Currently studying	2 (4)	2 (4)	4 (4)
Short vocational training	1 (2)	0	1 (1)
Apprenticeship	2 (4)	2 (4)	4 (4)
Short academic education (1–3 years)	11 (20)	10 (21)	21 (20)
Medium academic education (3–5 years)	14 (25)	12 (25)	26 (25)
Long academic education (>5 years)	13 (23)	15 (31)	28 (27)

Values are given as n (%), means ± SD, or medians (range).

**Table 2.** Differences in electronic and paper questionnaires

ThyPRO scale	Mean ± SD scale score		Score difference (95% CI)	p (score difference ≠ 0)	ICC (95% CI)
	paper	electronic			
Goiter symptoms	10±16	9±15	0.9 (–0.4 to 2.0)	0.17	0.93 (0.86 to 0.96)
Hyperthyroid symptoms	20±18	19±19	1.1 (–1.3 to 3.38)	0.38	0.79 (0.60 to 0.91)
Hypothyroid symptoms	21±23	21±23	0.5 (–1.29 to 3.38)	0.67	0.88 (0.80 to 0.92)
Eye symptoms	15±18	16±18	–0.3 (–1.89 to 1.33)	0.73	0.89 (0.82 to 0.94)
Tiredness	42±24	43±26	–0.5 (–3.16 to 2.10)	0.69	0.86 (0.79 to 0.90)
Cognitive complaints	19±22	19±21	–0.4 (–2.32 to 1.44)	0.64	0.90 (0.84 to 0.94)
Anxiety	17±21	20±20	–2.2 (–4.4 to 0.09)	0.06	0.84 (0.74 to 0.90)
Depressivity	25±19	26±19	–1.0 (–3.30 to 1.30)	0.39	0.80 (0.71 to 0.87)
Emotional susceptibility	31±23	31±23	0.5 (–2.11 to 3.07)	0.71	0.83 (0.75 to 0.89)
Impaired social life	11±18	12±19	–1.1 (–3.52 to 1.36)	0.38	0.76 (0.67 to 0.84)
Impaired daily life	13±21	13±19	0.6 (–1.37 to 2.58)	0.55	0.87 (0.78 to 0.93)
Impaired sex life	25±35	28±33	–2.6 (–6.04 to 0.30)	0.08	0.89 (0.82 to 0.93)
Cosmetic complaints	15±23	18±23	2.5 (–4.22 to –1.32)	0.003	0.95 (0.90 to 0.97)



**Fig. 2.** Bland-Altman plots of the 13 ThyPRO scales. The horizontal axis is the mean score of the paper (P) and electronic (E) questionnaires ( $(P + E)/2$ ). The vertical axis is the difference in score between the P and E questionnaires ( $P - E$ ). The solid horizontal gray lines are the LOA of the present study, whereas the punctuated gray lines represent the LOA from a previous P-P test-retest study of ThyPRO.

ety of time intervals since there are no established standards. We chose a time interval of 2 weeks because this made it possible to compare our results with a previous paper-paper reliability study of ThyPRO, but a shorter interval might have produced higher agreement due to less change.

All things considered, the results of our study suggest that the paper and electronic version of ThyPRO can be used interchangeably and produce identical results. The

results of this study are consistent with most other studies of comparisons between the two administration formats [2, 6, 9–15, 17, 22–28].

The strength of our study is a well-characterized study population and a relevant study design with an adequate, similarly designed study for comparison. The randomized cross-over design minimizes potential practice effects in a test-retest study. Also, the statistical analyses are among those recommended to establish agreement [1].

The clinical stability of the patients was evaluated according to the most recent evaluation of patients in our outpatient clinics. However, some of these were assessed several months prior to study inclusion. Thus, we cannot exclude that some patients may have been less stable than assumed, which may have contributed to a higher variability. However, as our study showed a good concordance between the two time points of assessments, we believe that such unrecognized lack of stability in disease activity of the patients is only of minor importance to our findings. We only included two subpopulations of thyroid patients, and the sample might be younger and more educated than the general population. All included patients reported themselves as being skilled or highly skilled computer and Internet users, and they all reported using a computer every day or nearly every day. This may have limited our power to detect any differences. However, previous studies have not found any differences between response patterns based on age, education, or income [3, 13, 22]. There was considerable attrition between the first and second questionnaire assessments, but our attrition analysis did not show any significant differences in responses between the two groups (data not shown). Finally, an important limitation to be noted is that our study did not provide information on whether differences exist between populations with *access* to or *no access* to a computer (and the Internet). If so, selection

bias may apply if a questionnaire is administered using only an electronic version.

To conclude, we found adequate agreement between the paper and electronic version of ThyPRO, which suggests that the electronic version of ThyPRO can be used without compromising the psychometric quality of the data and that data from the two modes can be combined without taking mode of administration into account. When appropriate, data collection can offer both modes of administration and in this setting the paper and electronic questionnaire can be used interchangeably according to individual patient visits, and thus facilitate participation and compliance.

### Acknowledgments

We thank our colleague Per Cramon for technical help. The study was supported by grants from Agnes and Knut Mørk's Foundation, the Danish Agency for Science, Technology and Innovation. UFRs research salary was provided by an unrestricted research grant from Novo Nordisk Foundation.

### Disclosure Statement

The authors have nothing to disclose.

### References

- 1 Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al: Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health* 2009;12:419–429.
- 2 Gwaltney CJ, Shields AL, Shiffman S: Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11:322–333.
- 3 van Gelder MM, Bretveld RW, Roeleveld N: Web-based questionnaires: the future in epidemiology? *Am J Epidemiol* 2010;172:1292–1298.
- 4 Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, et al: Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. *Qual Life Res* 2009;18:483–496.
- 5 Watt T, Hegedus L, Groenvold M, Bjorner JB, Rasmussen AK, Bonnema SJ, et al: Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *Eur J Endocrinol* 2010;162:161–167.
- 6 Brock RL, Barry RA, Lawrence E, Dey J, Rolffs J: Internet administration of paper-and-pencil questionnaires used in couple research: assessing psychometric equivalence. *Assessment* 2012;19:226–242.
- 7 Mayr A, Gefeller O, Prokosch HU, Pirkl A, Fröhlich A, de Zwaan M: Web-based data collection yielded an additional response bias – but had no direct effect on outcome scales. *J Clin Epidemiol* 2012;65:970–977.
- 8 Bishop FL, Lewis G, Harris S, McKay N, Prentice P, Thiel H, et al: A within-subjects trial to test the equivalence of online and paper outcome measures: the Roland Morris disability questionnaire. *BMC Musculoskelet Disord* 2010;11:113.
- 9 Bot AG, Menendez ME, Neuhaus V, Mudgal CS, Ring D: The comparison of paper- and web-based questionnaires in patients with hand and upper extremity illness. *Hand (NY)* 2013;8:210–214.
- 10 Bushnell DM, Reilly MC, Galani C, Martin ML, Ricci JF, Patrick DL, et al: Validation of electronic data capture of the Irritable Bowel Syndrome – Quality of Life Measure, the Work Productivity and Activity Impairment Questionnaire for Irritable Bowel Syndrome and the EuroQol. *Value Health* 2006;9:98–105.
- 11 Cook AJ, Roberts DA, Henderson MD, Van Winkle LC, Chastain DC, Hamill-Ruth RJ: Electronic pain questionnaires: a randomized, crossover comparison with paper questionnaires for chronic pain assessment. *Pain* 2004;110:310–317.
- 12 Egger MJ, Lukacz ES, Newhouse M, Wang J, Nygaard I: Web versus paper-based completion of the epidemiology of prolapse and incontinence questionnaire. *Female Pelvic Med Reconstr Surg* 2013;19:17–22.

- 13 Handa VL, Barber MD, Young SB, Aronson MP, Morse A, Cundiff GW: Paper versus web-based administration of the Pelvic Floor Distress Inventory 20 and Pelvic Floor Impact Questionnaire 7. *Int Urogynecol J Pelvic Floor Dysfunct* 2008;19:1331–1335.
- 14 MacKenzie H, Thavaneswaran A, Chandran V, Gladman DD: Patient-reported outcome in psoriatic arthritis: a comparison of Web-based versus paper-completed questionnaires. *J Rheumatol* 2011;38:2619–2624.
- 15 Sjöström M, Stenlund H, Johansson S, Umefjord G, Samuelsson E: Stress urinary incontinence and quality of life: a reliability study of a condition-specific instrument in paper and web-based versions. *NeuroUrol Urodyn* 2012; 31:1242–1246.
- 16 Wu RC, Thorpe K, Ross H, Micevski V, Marquez C, Straus SE: Comparing administration of questionnaires via the Internet to pen-and-paper in patients with heart failure: randomized controlled trial. *J Med Internet Res* 2009; 11:e3.
- 17 Marsh JD, Bryant DM, Macdonald SJ, Naudie DD: Patients respond similarly to paper and electronic versions of the WOMAC and SF-12 following total joint arthroplasty. *J Arthroplasty* 2014;29:670–673.
- 18 Cramon P, Rasmussen AK, Bonnema SJ, Bjorner JB, Feldt-Rasmussen U, Groenvold M, et al: Development and implementation of PROgmatic: a clinical trial management system for pragmatic multi-centre trials, optimised for electronic data capture and patient-reported outcomes. *Clin Trials* 2014;11:344–354.
- 19 Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- 20 Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
- 21 Bartko JJ: Measures of agreement: a single procedure. *Stat Med* 1994;13:737–745.
- 22 Bernstein AN, Levinson AW, Hobbs AR, Lavery HJ, Samadi DB: Validation of online administration of the sexual health inventory for men. *J Urol* 2013;189:1456–1461.
- 23 Bushnell DM, Martin ML, Parasuraman B: Electronic versus paper questionnaires: a further comparison in persons with asthma. *J Asthma* 2003;40:751–762.
- 24 Kleinman L, Leidy NK, Crawley J, Bonomi A, Schoenfeld P: A comparative trial of paper-and-pencil versus computer administration of the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire. *Med Care* 2001; 39:181–189.
- 25 Raat H, Mangunkusumo RT, Landgraf JM, Kloek G, Brug J: Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): Internet administration compared with the standard paper version. *Qual Life Res* 2007;16:675–685.
- 26 Salaffi F, Gasparini S, Ciapetti A, Gutierrez M, Grassi W: Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. *Rheumatology (Oxford)* 2013;52:2062–2070.
- 27 Thoren ES, Andersson G, Lunner T: The use of research questionnaires with hearing impaired adults: online versus paper-and-pencil administration. *BMC Ear Nose Throat Disord* 2012;12:12.
- 28 Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, et al: Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. *Eur J Epidemiol* 2010;25:287–296.