

Penskalaan *Thurstone* pada Aitem *Thyroid Dysfunction Questionnaire* (TDQ) Berbasis Gejala Biopsikosial

*Mutalazimah*¹

Program Studi Gizi Fakultas Ilmu Kesehatan Universitas Muhammadiyah Surakarta

Saifuddin Azwar
Fakultas Psikologi UGM

Bhisma Murti
Fakultas Kedokteran UNS

Budi Mulyono
Fakultas Kedokteran UGM

Abstract. Thyroid function disorders can affect the biological, psychological and social conditions of the sufferer. This study aimed at developing a thyroid dysfunction questionnaire (TDQ) based on a biopsychosocial phenomenon, which is the impact of hypothyroidism and hyperthyroidism. The development of this instrument began with identification of the measurement objectives, domains and items. The procedures of *Thurstone* Scaling were done with analysis of content validity using the *Content Validity Ratio* (CVR) and *Content Validity Index* (CVI), which were formulated by three experts of thyroid and endocrinology. Every item in the TDQ was determined through median analysis based on the valuation results done by 44 internists from hospitals and 44 doctors from *Puskesmas* (*Public Health Center*) with 11 categories of response (1-11). The results of CVR analysis ranged from 0.33 to 1.00 and the value of CVI was 78%. It showed that the experts had adequate rate of homogeneity to judge the relevance of the 33 items in TDQ. The median analysis showed that the 33 items of TDQ had scores ranging from 4 to 10. Thus, it was expected that the TDQ scores on this interval scale would facilitate observer to make an interpretation of the examination of thyroid function of the subjects.

Keywords: *Thurstone scaling, thyroid dysfunction questionnaire, experts' judgment, biopsychosocial*

Abstrak. Gangguan fungsi tiroid berpengaruh terhadap kondisi biologis, psikologis dan sosial penderitanya. Penelitian ini bertujuan mengembangkan kuesioner disfungsi tiroid yang disebut *thyroid dysfunction questionnaire* (TDQ), berdasarkan fenomena biopsikosial yang merupakan dampak dari kondisi hipotiroidisme dan hipertiroidisme. Pengembangan kuesioner diawali dengan mengidentifikasi tujuan pengukuran, domain dan aitem. Selanjutnya, prosedur penskalaan *Thurstone* diawali dengan menganalisis validitas isi, menggunakan *content validity ratio* (CVR) dan *content validity index* (CVI) dari tiga orang ahli tiroid dan endokrinologi. Skor setiap aitem dalam TDQ ditentukan menggunakan analisis median, berdasarkan hasil penilaian 44 internis dari rumah sakit dan 44 dokter dari Puskesmas dengan 11 kategori respons (1 - 11). Hasil analisis CVR berkisar 0,33-1,00 dan CVI sebesar 78%, menunjukkan bahwa para ahli memiliki tingkat kesepakatan yang memadai untuk menilai relevansi dari 33 aitem dalam TDQ. Selanjutnya analisis median menunjukkan 33 aitem TDQ mempunyai skor mulai dari 4 – 10. Diharapkan skor TDQ pada skala interval ini memudahkan observer melakukan interpretasi hasil pemeriksaan fungsi tiroid pada subjek.

Kata kunci: *penskalaan Thurstone, thyroid dysfunction questionnaire, penilaian ahli, biopsikosial*

¹Korespondensi mengenai isi artikel ini dapat dilakukan melalui: mutalazimah@gmail.com

Hasil survei di seluruh Indonesia, menunjukkan adanya peningkatan prevalensi *total goitre rate* (TGR), dari 9,8% pada tahun 1998, menjadi sebesar 11,1% pada tahun 2003 (Tim GAKY Pusat, 2005). Suatu wilayah yang mempunyai karakteristik yang menyebabkan berkurangnya kandungan yodium dalam tanah ini disebut sebagai daerah endemis gangguan akibat kekurangan yodium (GAKY) (Djokomoeljanto, 2008; Bayram, Beyazyildiz, Gökçe, Budak, Erdoğan, Kurtoğlu, Kula, Ünlühızarıcı & Keleştimur, 2009). Dampak dari defisiensi yodium mencakup spektrum yang sangat luas, seperti: *abortus*, lahir mati, cacat bawaan, kematian perinatal, kematian bayi, kretin, gondok, *hypothyroidism*, penurunan IQ, gangguan fungsi mental, gangguan fungsi otot, pertumbuhan terhambat dan *iodine induced hyperthyroidism* atau IIH (Sebotsa, Dannhauser, Mollentze, Oosthuizen, Mahomed & Jooste, 2009).

Pengukuran defisiensi yodium bisa dilakukan melalui beberapa metode, yakni melalui cakupan konsumsi garam beryodium, asupan yodium dalam makanan, pembesaran kelenjar tiroid dan kadar ekskresi yodium dalam urin. Idealnya semakin banyak metode yang digunakan, hasilnya akan semakin baik. Baku standar pengukuran defisiensi yodium adalah pemeriksaan ekskresi yodium dalam urin (EYU) (Bogale, Abebe, Stoecker, Abuye, Ketema, & Hambidge, 2009; Vejbjerg, Knudsen, Perrild, Laurberg, Andersen, Rasmussen, Ovesen, & Jorgensen, 2009; Singh, Marwal, & Lakshminarayana, 2010; Medani, Elnour, & Saeed, 2011). Bahkan untuk mengetahui risiko terhadap dampak terjadinya hipotiroid dan hipertiroid, perlu dilanjutkan dengan pemeriksaan TSH dan *free thyroxine* atau *free T4* (FT4). Namun, karena terkendala pada biaya yang mahal dan sulitnya teknis pelaksana-

naan, hal tersebut belum memungkinkan dilakukan secara massal.

Penelitian ini sebagai alternatif untuk mengembangkan *thyroid dysfunction questionnaire* (TDQ) berbasis gejala biopsikososial yang merupakan dampak dari kondisi hipotiroid dan hipertiroid. Model biopsikososial menyatakan bahwa kesehatan, sakit dan penyakit adalah hasil dari interaksi antara faktor biologis, psikologis dan sosial, melalui mekanisme perbedaan antara proses patofisiologi, yang menyebabkan penyakit dan persepsi seseorang tentang kesehatan dan akibat-akibatnya, yang disebut penyakit (Carrio, Suchman, & Epstein, 2004; Brown, Bonello, & Pollard, 2005; Novack, Cameron, Epe, Ader, Waldstein, Levenstein, Antoni, & Wainer, 2007; Bruns & Warren, 2011). Pendekatan gejala-gejala biopsikososial pada pengembangan TDQ ini didasarkan pada beberapa hasil penelitian, yang menyatakan adanya keterkaitan antara hasil pemeriksaan klinis, psikologis dan hasil pemeriksaan biokimia, baik pada anak-anak maupun pada orang dewasa yang menderita defisiensi yodium dan hipotiroid (Indra, Patil, Joshi, Pai, & Kalantri, 2004; Ocal, UnceoUlu, Aycan, BerberoUlu, AdYyaman, Ergun, EvliyaoUlu, & Akhurin, 2004; Agha, Walker, Perry, Drake, Chew, Jenkins, Grossman, & Monson, 2007). Sementara itu, berkaitan dengan pengukuran klinis, Mutalazimah (2010) menemukan bahwa sebesar 28% anak sekolah, mempunyai skor klinis yang tergolong defisiensi yodium dan terdapat hubungan yang signifikan, antara status klinis dan status yodium dalam urin. Hal ini berarti bahwa anak-anak yang mengalami defisiensi yodium, menunjukkan gejala klinis seperti *goiter*, kulit bersisik, muka sembab dan lamban, sedangkan gejala psikologis meliputi fungsi kognitif yang rendah, yang

didasarkan pada hasil pengukuran *colour progressive matrices* (CPM) dan hasil rata-rata nilai rapor dibawah rata-rata, serta gejala sosial seperti apatis dan sulit berinteraksi dengan orang lain.

Brown, dkk. (2005), menemukan beberapa domain dari psikososial yang berkaitan dengan hipotiroid, yakni *mood disorder* dan *stress* dengan berbagai gejala depresi, seperti gangguan tidur, penurunan aktivitas, kurang berenergi, penurunan konsentrasi dan gangguan nafsu makan. Gonen, Kisakol, Cilli, Dikbas, Gungor, Inal, dan Kaya (2004) menambahkan, bahwa dua pertiga dari penderita gangguan fungsi tiroid mengalami gangguan psikologis, seperti mudah panik, cemas, depresi, fobia, iritabilitas dan mudah marah. Sementara itu, Hermann, Hower, dan Lederbogen (2004) menyatakan 16 gejala dalam domain depresi, yang berkaitan dengan gangguan tiroid di daerah endemis defisiensi yodium, bahkan menimbulkan rendahnya rasa percaya diri, mudah terlibat konflik dengan orang lain, kurang bisa memahami orang lain, serta membatasi interaksi dengan orang lain (Watt, 2009).

TDQ yang akan digunakan untuk mendeteksi gangguan tiroid pada penelitian ini, dikembangkan melalui berbagai tahapan konstruksi instrumen. Tahap awal mengembangkan instrumen yang bermanfaat untuk mencapai tujuan pengukuran adalah dengan mengidentifikasi domain dari konstruk yang akan dicapai (Fung, 2010). Menurut Azwar (2009) langkah dasar dalam konstruksi instrumen diawali dengan mengidentifikasi tujuan ukur, mengoperasionalkan konsep dan melakukan penskalaan. Identifikasi domain dan aitem pada TDQ dilakukan menggunakan metode *expert judgment*, termasuk proses penskalaan untuk menghasilkan skor setiap aitem yang dikembangkan.

Penggunaan *expert judgment* dalam penelitian sudah diterapkan pada berbagai bidang, termasuk bidang kesehatan, dan diperlukan untuk menghindari adanya data yang bersifat subjektif dan berpeluang menimbulkan bias, terhadap keaslian dan akurasi (Muxika, Borja, & Bald, 2007; Bruce, Langley, & Tjale, 2008). Dengan demikian, *expert judgment* direkomendasikan untuk meminimalkan bias dan meningkatkan akurasi hasil penelitian. Lebih khusus, *expert judgment* ditujukan untuk menguji validitas isi yang berkaitan dengan substansi penelitian (Watt, Hegedüs, Groenvold, Bjorner, Rasmussen, Bonnema, & Rasmussen, 2007; Parsa, Kandiah, Nasir Hejar, & Afah, 2008; Fung, 2010; Flynn, O'Neill, McInerney, & Kenned, 2011). Sementara itu, penskalaan merupakan bagian dari pengukuran dalam proses konstruksi instrumen yang mengaitkan konstruk kualitatif menjadi satuan metrik kuantitatif, yang berguna bagi pemaknaan skor aitem dari sebuah instrumen (Salzberger, 2013).

Berkaitan dengan analisis aitem hasil *expert judgment*, Azwar (2009) menjelaskan bahwa untuk menghitung nilai skala bagi aitem yang mewakili *rating* dari keseluruhan *expert*, digunakan rumus dengan dasar estimasi harga median (Q_2) dari proporsi kumulatif yang berisi nilai 0,5. Dengan demikian, setiap aitem akan mempunyai nilai dalam skala interval. Selain itu, bisa diinterpretasikan adanya homogenitas antar *expert* dalam menyepakati aitem, melalui pendekatan perhitungan kuartil (selisih kuartil satu dan kuartil tiga, relatif kecil).

Penggunaan *expert judgment* seringkali menjadi pilihan untuk melakukan penskalaan. Ketika merancang sebuah skala *rating*, peneliti harus menentukan jumlah titik pada skala. *Thurstone* menggunakan metode *equal appearing interval* dengan 11

kategori (Krosnick & Presser 2009; Azwar, 2009; Anderson, Neafsey, & Peabody, 2011). Penskalaan *Thurstone*, diawali dengan melakukan *issue generation* berkaitan dengan substansi (identifikasi aitem), memilih para ahli sesuai dengan kualifikasi untuk melakukan *judgment*, meminta para ahli untuk melakukan *rating* dengan 11 kategori, menghitung hasil *rating* menggunakan pendekatan nilai median, membuang aitem yang tidak memenuhi kesepakatan para ahli, memilih aitem-aitem dengan tingkat kesepakatan terbaik, membangun skala berdasarkan aitem yang terpilih (Kumar, 1999; Azwar, 2009).

Penelitian ini bertujuan untuk menganalisis aitem-aitem berbasis gejala biopsikososial dalam TDQ, yang meliputi gejala biologis seperti domain *goiter*, gejala klinis khas hipotiroid dan gejala klinis khas hipertiroid. Sementara itu, gejala psikososial pada domain hipotiroid dengan indikator seperti depresi, penurunan daya ingat, apatis dan malas berinteraksi dengan orang lain, sedangkan gejala psikososial pada domain hipertiroid dengan indikator gangguan emosional, kecemasan, kebingungan dalam berfikir dan mudah berkonflik dengan orang lain. Aitem-aitem dalam TDQ yang telah dikembangkan berdasarkan indikator pada setiap domain diperoleh dari proses *expert judgment* melalui penskalaan *Thurstone*, sehingga didapatkan skor setiap aitem dan tingkat kesepakatan *expert* terhadap aitem-aitem dalam TDQ tersebut. Proses penskalaan aitem TDQ ini akan bermanfaat untuk memudahkan pengguna dalam menginterpretasikan skor TDQ dalam menentukan kondisi subjek yang terindikasi hipotiroid atau hipertiroid.

Metode

Jenis Penelitian

Penelitian ini merupakan bagian dari penelitian pengembangan, karena melakukan langkah awal konstruksi instrumen TDQ yakni menganalisis skor aitem menggunakan penskalaan *Thurstone* berbasis *expert judgment*. Subjek penelitian, penelitian ini menggunakan dua kelompok ahli, kelompok pertama adalah tiga orang ahli bidang tiroid dan endokrinologi untuk pengujian validitas isi dan kelompok kedua adalah 88 orang praktisi bidang kedokteran, yakni 44 dokter residen penyakit dalam di RSUP dr. Sardjito Yogyakarta dan 44 adalah dokter Puskesmas yang tersebar di Puskesmas Kabupaten Sleman yang dipilih secara *sequential random sampling*, untuk menjadi ahli pada penskalaan aitem. Kriteria subjek ahli untuk validasi isi aitem TDQ adalah guru besar, mempunyai keahlian, pengalaman dan penelitian bidang endokrinologi sekurang-kurangnya 10 tahun. Kriteria subjek ahli untuk penskalaan aitem yakni dokter yang sedang mengambil pendidikan spesialis penyakit dalam tingkat madya dan mandiri, sedang duduk di semester IV-IX dan telah menempuh stase endokrinologi. Sementara itu, dokter-dokter Puskesmas dipilih yang mempunyai pengalaman bekerja sebagai dokter fungsional, sekurang-kurangnya mempunyai pengalaman bekerja selama dua tahun.

Prosedur Penelitian

Penelitian diawali menguji validitas isi dari 33 aitem TDQ oleh tiga ahli bidang tiroid dan endokrinologi. Selanjutnya, analisis aitem dilakukan dengan meminta

justifikasi 88 dokter terhadap 33 aitem TDQ tersebut, yang masing-masing aitem disediakan respons dengan 11 kategori (1 – 11). Analisis data, pengujian validitas isi dianalisis menggunakan perhitungan *content validity ratio* (CVR) dan *content validity index* (CVI), sedangkan analisis penskalaan aitem metode *Thurstone* menggunakan pendekatan rumus median (Q_2) untuk menentukan skor setiap aitem. Skor aitem yang dipilih adalah skor aitem yang memiliki nilai lebih tinggi pada kelompok dokter rumah sakit dan Puskesmas, sedangkan skor akhir aitem yang benar-benar digunakan untuk implementasi TDQ merupakan hasil pembulatan angka desimal di belakang koma dari setiap skor aitem ($< 0,5$ dibulatkan ke bawah dan $\geq 0,5$ dibulatkan ke atas). Selanjutnya ditentukan juga tingkat homogenitas ahli dalam memberikan pembobotan aitem, yang dianalisis menggunakan nilai selisih antara kuartil tiga (Q_3) dan kuartil satu (Q_1), bila nilainya relatif kecil maka dinyatakan homogen.

H a s i l

TDQ dikembangkan berbasis pada gejala-gejala gangguan tiroid yang meliputi gejala gangguan biologis, psikologis dan sosial, yang bertujuan sebagai alat skrining hipotiroid dan hipertiroid pada wanita usia produktif. Dalam proses identifikasi domain, didapatkan tiga domain utama yakni domain *goiter* (3 aitem), domain hipotiroid (14 aitem) dan domain hipertiroid (16 aitem). Berdasarkan justifikasi ahli, disepakati bahwa *goiter* merupakan *common aitem* yang bisa timbul pada kondisi hipotiroid maupun hipertiroid, sedangkan domain hipotiroid dan hiper-

tiroid merupakan domain yang aitem-aitemnya bersifat dikotomi. Perbedaan jumlah aitem pada setiap domain dikarenakan aitem-aitem yang mewakili indikator dari setiap domain merupakan gejala-gejala yang secara empiris diperoleh para ahli ketika melakukan pemeriksaan terhadap pasien. Dengan demikian jumlah aitem setiap domain tidak dapat dibuat sama karena gejala-gejala yang berkaitan dengan hipotiroid atau hipertiroid memang berbeda.

Hasil pengujian validitas isi dengan menggunakan metode *Thurstone* yang dilakukan sebelum proses penskalaan tersaji pada Tabel 1.

Pengujian validitas isi pada penyusunan TDQ ini, dilakukan dengan meminta justifikasi kepada tiga ahli, mengenai relevansi setiap aitem sebagai gejala hipotiroid dan hipertiroid, dengan lima respons jawaban, yakni: sangat tidak relevan, tidak relevan, cukup relevan, relevan dan sangat relevan. Nilai CVR, digunakan sebagai analisis validasi isi terhadap aitem-aitem gejala gangguan tiroid. Perhitungan CVR menggunakan rumus $[n_e - (N/2)]/(N/2)$. Pada rumus tersebut n_e adalah jumlah ahli atau *subject matter expert raters* (SMEs), yang menyatakan bahwa sebuah aitem sangat penting, N adalah jumlah seluruh ahli.

Setelah pengujian validitas isi, maka selanjutnya dilakukan penskalaan *Thurstone* pada TDQ, melalui justifikasi 44 dokter rumah sakit dan 44 dokter Puskesmas. Adapun hasil perhitungan skala *Thurstone* pada penskalaan aitem-aitem TDQ tersaji pada Tabel 2.

Tabel 1
 Nilai CVR aitem-aitem dalam TDQ

Domain	Gejala	CVR
A. Goiter	A. 1. Pembesaran kelenjar tiroid	1,00
	A. 2. Gangguan pada kerongkongan (sakit, terasa penuh, tertekan, terasa ada gumpalan, gangguan menelan)	1,00
	A. 3. Perubahan suara menjadi serak/parau	0,33
B. Hipotiroid	B. 1. Sensitif terhadap suhu dingin	1,00
	B. 2. Edema (sembab muka/tangan/kaki)	1,00
	B. 3. Kulit kering/bersisik	0,33
	B. 4. Bradikardi	0,33
	B. 5. Konstipasi	1,00
	B. 6. Kram otot/nyeri sendi	1,00
	B. 7. Rambut kering/rontok/kuku mudah patah	1,00
	B. 8. Menstruasi berlebihan	0,33
	B. 9. Berat badan naik	1,00
	B.10. Ataksia (gangguan koordinasi gerak)	1,00
	B.11. Sering merasa sangat lelah/letih/lamban	1,00
	B.12. Merasa kesulitan mengingat/kehilangan konsentrasi	1,00
	B.13. Sering sedih, tertekan, mudah menangis, mudah putus asa, tidak percaya diri, membebani orang lain	0,33
	B.14. Malas, apatis, inisiatif dan aktivitas sosial berkurang	1,00
C. Hipertiroid	C. 1. Tangan gemetar/tremor	1,00
	C. 2. Keringat berlebihan	1,00
	C. 3. Hiperaktivitas/energetik/terengah-engah	1,00
	C. 4. Takikardi/palpitasi	1,00
	C. 5. Sensitif terhadap suhu panas	1,00
	C. 6. Sering BAB	0,33
	C. 7. Mata membelalak	1,00
	C. 8. Kulit sangat lembek	1,00
	C. 9. Lemah otot	1,00
	C.10. Menstruasi tidak rutin	1,00
	C.11. Berat badan turun (nafsu makan normal)	1,00
	C.12. Sering merasa bingung ketika berpikir	0,33
	C.13. Sering merasa cemas (gugup, gelisah, tegang)	0,33
	C.14. Sulit beristirahat (tidur tidak nyenyak/sulit tidur)	0,33
	C.15. Sulit mengendalikan emosi, mudah tersinggung, mudah marah dan mudah bermasalah dengan orang lain	0,33
	C.16. Mudah berkonflik, kurang bisa memahami orang lain	0,33

Tabel 2
Hasil penskalaan aitem gejala gangguan fungsi tiroid

Domain/aitem	Q ₂ S	Q ₂ P	(Q ₃ -Q ₁) S	(Q ₃ -Q ₁) P	Nilai aitem terpilih	Nilai akhir aitem
Goiter						
Pembesaran kelenjar tiroid	9,07	9,34	1,83	1,82	9,34	9
Gangguan rasa pada kerongkongan	7,50	6,30	2,69	3,10	7,50	8
Perubahan suara/parau	3,64	4,10	3,09	3,71	4,10	4
Hipotiroid						
Sensitif suhu dingin	8,68	7,64	2,50	4,77	8,68	9
Edema (muka/tangan/kaki)	6,21	6,25	3,00	4,07	6,25	6
Kulit kering/bersisik	7,25	7,25	4,42	3,00	7,25	7
Rambut rontok/kuku mudah patah	7,25	7,25	4,42	3,00	7,25	7
Bradikardi	7,78	7,00	5,25	4,58	7,78	8
Konstipasi	7,32	6,49	3,70	4,83	7,32	7
Kram otot/nyeri sendi	5,36	6,14	4,33	3,42	6,14	6
Menstruasi berat	4,37	5,33	3,92	3,83	5,33	5
Berat badan naik	8,00	6,00	4,36	4,50	8,00	8
Ataksia (gangguan koordinasi gerak)	5,50	5,72	4,62	3,37	5,72	6
Sangat lelah/letih	9,30	8,03	3,50	2,15	9,30	9
Mudah lupa/kurang konsentrasi	6,50	8,33	4,52	3,08	8,33	8
Mudah sedih, tidak percaya diri	8,50	9,00	4,90	4,33	9,00	9
Malas/apatis/kurang partisipasi	7,50	8,17	5,28	3,39	8,17	8
Hipertiroid						
Gemetar/tremor	8,87	8,17	2,33	2,60	8,87	9
Keringat berlebihan	8,80	8,50	2,39	2,93	8,80	9
Hiperaktivitas/energetik/terengah-engah	8,81	6,50	2,67	3,57	8,81	9
Takikardi/palpitasi	9,50	9,04	2,73	2,58	9,50	10
Sensitif suhu panas	9,50	8,50	2,45	3,96	9,50	10
Sering BAB	8,28	6,80	2,82	2,94	8,28	8
Mata membelalak	9,73	9,00	1,88	3,17	9,73	10
Kulit sangat lembek	7,50	6,37	4,50	3,33	7,50	8
Lemah otot	7,50	6,50	3,08	3,25	7,50	8
Menstruasi tidak rutin	6,36	6,21	3,48	3,59	6,36	6
Berat badan turun	9,19	8,25	3,25	2,23	9,19	9
Sering merasa bingung ketika berfikir	7,00	7,62	4,20	2,40	7,62	8
Sering merasa cemas (gugup, gelisah, tegang)	9,17	8,83	3,50	2,54	9,17	9
Sulit beristirahat (tidur tidak nyenyak)	8,75	8,75	2,98	2,50	8,75	9
Sulit mengendalikan emosi, mudah tersinggung, mudah marah	8,70	8,25	3,43	3,80	8,70	9
Mudah berkonflik, kurang bisa memahami pada orang lain	6,90	7,64	4,64	4,62	7,64	8

Keterangan:

Q₂S : Nilai median pada kelompok dokter rumah sakit

Q₂P : Nilai median pada kelompok dokter puskesmas

(Q₃-Q₁)S : Selisih kuartil tiga dan kuartil satu pada kelompok dokter rumah sakit

(Q₃-Q₁)P : Selisih kuartil tiga dan kuartil satu pada kelompok dokter Puskesmas

Penskalaan aitem diawali dengan menyusun 33 aitem dalam sebuah format justifikasi yang terangkum dalam tiga domain, yakni domain *goiter* tiga aitem, domain hipotiroid 14 aitem dan domain hipertiroid 16 aitem. Masing-masing aitem disediakan 11 kategori respons jawaban (1 - 11) yang harus diisi oleh para dokter yang terdiri dari dua kelompok, yakni kelompok dokter residen spesialis penyakit dalam RSUP dr. Sardjito (S) dan kelompok dokter Puskesmas se-Kabupaten Sleman (P). Pemilihan dokter residen spesialis penyakit dalam, karena kelompok ini mempunyai relevansi mendalam endokrinologi, termasuk masalah gangguan tiroid yang banyak dijumpai kasusnya secara klinis di rumah sakit, sebagai pelayanan kesehatan tersier. Sementara itu, pemilihan dokter Puskesmas, karena mempunyai pengalaman menangani berbagai permasalahan gangguan tiroid, yang terjadi pada tingkat pelayanan kesehatan primer berbasis komunitas.

Diskusi

Nilai CVR yang dihitung berdasarkan justifikasi ahli berkisar antara -1 sampai +1, bila nilai negatif maka kurang dari satu perdua ahli menyatakan bahwa aitem tersebut penting dan bila hasilnya positif maka lebih dari satu perdua ahli menyatakan bahwa aitem tersebut penting (Tabel 1). Sementara itu, berdasarkan nilai rerata CVR, diperoleh nilai CVI sebesar 78% yang berarti "*adequate agreement*", yakni para ahli cukup sepakat terhadap relevansi keseluruhan aitem, sebagai gejala gangguan fungsi tiroid.

Validitas isi merupakan pertimbangan utama dalam metodologi pengembangan kuesioner (Rudge, Calderon, Dias, Lopes, Barbosa, Maestá, Odland, & Rudge, 2009), yang dibangun dari tiga sumber, yakni:

kajian pustaka, keterwakilan dari populasi yang relevan dan pendapat ahli (*experts*), sedangkan tahapannya adalah mengembankan dan melakukan *judgment* (Mortel, 2009). Tujuan *expert judgment* adalah mencari pertimbangan mengenai relevansi dari isi sebuah instrumen, yang terdiri dari beberapa domain dan mengenai representativitas, bahwa isi didalam aitem-aitem mewakili keseluruhan domain tersebut (Fung, 2010).

Analisis CVR dan CVI telah banyak digunakan dalam pengembangan kuesioner bidang kesehatan. Hasil CVR dan CVI pada setiap pengujian validitas isi aitem dari sebuah kuesioner akan menghasilkan nilai rentang yang bervariasi, bergantung pada muatan isi, jumlah dan justifikasi para ahli yang memberikan penilaian relevansi setiap aitem. Misalnya aitem-aitem TDQ mempunyai nilai CVR antara 0,33 - 1,00, dibandingkan dengan kuesioner *Rating of Cognitive Failures at Work* yang dikembangkan oleh Allahyari, Rangi, Khosravi, dan Zayeri (2011) dengan CVR - 0,75 - 1,00 dan kuesioner *Health Promotion Survey* Stephens, Lowman, Graham, Morris, Kohler, dan Waugh (2013) mempunyai CVR -0,64 - 1,00. Sementara itu nilai CVI tidak jauh berbeda dengan TDQ sebesar 78%, yakni: *Rating of Cognitive Failures at Work* 70% dan *Health Promotion Survey* 72%. Nilai CVR dan CVI juga berpeluang mempunyai hasil yang mendekati sempurna, seperti pada kuesioner *Modified Polycystic Ovary Syndrome* yakni rata-rata CVR 0,92 dan CVI mencapai 96% (Bazarganipour, Ziaei, Montazeri, Faghihzadeh, & Frozanfard, 2012).

Tabel 2 merupakan hasil perhitungan akhir dari metode penskalaan *Thurstone* oleh para dokter, yang telah dihitung nilai median dengan rumus $Q_2 = Bb + \{(0,5 - pkb)/pi\}$, dengan penjelasan *Bb* merupakan batas bawah nilai yang berisi median, *pkb*

adalah nilai p kumulatif pada batas bawah nilai yang berisi median, p merupakan nilai proporsi (f/n) pada angka yang berisi median dan i adalah luas interval (bernilai 1). Sementara itu, untuk mengetahui homogenitas antar ahli dalam memberi justifikasi nilai pada setiap aitem, digunakan $Q_3 - Q_1$ dengan rumus masing-masing $Q_3 = Bb + \{(0,75 - pkb)/p\}i$ dan $Q_1 = Bb + \{(0,25 - pkb)/p\}i$. Selanjutnya hasil perhitungan Q_2 sebagai skor aitem, sedangkan $Q_3 - Q_1$ sebagai dasar pengguguran aitem.

Melalui perhitungan tersebut maka didapatkan hasil seperti tersaji pada Tabel 2, yang berisi perbandingan hasil Q_2 antara kelompok ahli rumah sakit (Q_2S) dan kelompok ahli Puskesmas (Q_2P). Nilai Q_2 setiap aitem yang dipilih adalah yang lebih tinggi antar kedua kelompok tersebut, karena semakin tinggi nilai Q_2 berarti semakin tinggi relevansi aitem tersebut dalam mewakili gejala hipotiroid atau hipertiroid. Berdasarkan hasil pemilihan terhadap nilai Q_2 tertinggi dalam setiap aitem, didapatkan nilai aitem berkisar antara 4,10 – 9,73. Nilai akhir aitem merupakan pembulatan dari nilai aitem terpilih, sehingga nilai akhir aitem yang akan digunakan untuk dasar penyusunan skor pada TDQ adalah berkisar 4 – 10. Pembulatan nilai aitem tersebut didasarkan pada pertimbangan bahwa tujuan penyusunan TDQ adalah sebagai alat skrining hipotiroid dan hipertiroid. Dengan demikian pengguna TDQ dalam hal ini petugas kesehatan memerlukan cara skoring dan interpretasi yang lebih mudah dan cepat, karena TDQ merupakan bagian dari *short questionnaire* yang implementasinya dibatasi waktu, sehingga nilai pembulatan akan lebih memudahkan dalam penjumlahan skor dan menentukan kondisi subjek yang diperiksa.

Bila dikaji lebih detail pada setiap domain, terlihat bahwa pada domain

goiter yang terdiri atas tiga aitem (pembesaran kelenjar tiroid, gangguan pada kerongkongan dan perubahan suara) yang mempunyai skor aitem tertinggi adalah aitem pembesaran kelenjar tiroid, berasal dari nilai median hasil justifikasi kelompok dokter Puskesmas dengan nilai Q_2 sebesar 9,34 dengan nilai akhir hasil pembulatan sembilan. Skor-skor aitem pada domain hipotiroid juga bervariasi antara 6 – 10, untuk aitem-aitem yang mewakili gangguan psikososial dijustifikasi cukup baik oleh para ahli dengan skor 8 – 9 seperti mudah sedih, malas berinteraksi atau berpartisipasi dengan orang lain. Demikian juga dengan domain hipertiroid, skor setiap aitem berkisar 6 – 10, dengan aitem-aitem psikososial masing-masing mempunyai skor yang cukup tinggi sebesar 8 – 9 seperti mudah merasa bingung, sering merasa cemas, mudah tersinggung dan kurang bisa memahami orang lain. Kesemuanya ini merupakan bukti empiris berdasarkan pengalaman para ahli, bahwa gejala biopsikososial memang melekat pada gangguan tiroid.

Homogenitas antar ahli terhadap seluruh aitem dapat diinterpretasikan berdasarkan selisih Q_3 dan Q_1 , yang kesemuanya relatif kecil berkisar antara 1,82 – 5,28. Kisaran nilai tersebut masih berada dalam rentang yang cukup homogen diantara nilai respons 1 – 11. Dengan demikian, dapat disimpulkan bahwa tidak ada pengguguran aitem, sehingga semua aitem, yakni 33 aitem yang dijustifikasi oleh ahli dapat digunakan sebagai dasar perancangan TDQ pada tahap desain instrumen berikutnya. Tidak adanya pengguguran aitem pada TDQ ini juga dikarenakan seluruh aitem yang dijustifikasi oleh para ahli merupakan gejala-gejala pada domain-domain biologis dan psikososial, sehingga proses penskalaan hanya bertujuan untuk mendapatkan urutan *severity of*

symptom, sehingga diperoleh bobot skor pada masing-masing aitem untuk menentukan interpretasi secara keseluruhan pada nilai total TDQ untuk mengindikasikan kondisi hipotiroid atau hipertiroid.

Berbagai hasil pengujian validitas isi dan penskalaan aitem berbasis *expert judgment* yang dilakukan dalam penyusunan TDQ ini menggunakan *direct methods*, karena menurut Burgman, Fidler, McBride, Walshe, dan Wintle (2006) merupakan metode yang paling sederhana dan memungkinkan untuk memaksimalkan hasil justifikasi, karena peneliti dapat berinteraksi secara langsung dengan para ahli. Selanjutnya pemilihan metode *Thurstone* untuk penskalaan karena mempunyai kelebihan yakni selain skor aitem yang diperoleh berada dalam skala interval yang memudahkan pengguna untuk menginterpretasikan, juga karena mempunyai 11 respons yang lebih panjang dari skala lainnya. Dikemukakan juga bahwa panjang dari skala merupakan salah satu cara untuk memaksimalkan reliabilitas dan validitas (Krosnick & Presser, 2009).

Ada beberapa hal yang mendasari kelebihan respons dengan skala yang lebih panjang, bahwa secara teoritis responsden akan memilih respons yang paling representatif dari keadaan yang sebenarnya. Misalnya, ketika sebuah aitem mempunyai dua respons, yakni suka dan tidak suka, hal ini akan lebih menyulitkan responsden menjawab bila responsden mempunyai jawaban kadang-kadang suka atau kadang-kadang tidak suka. Penambahan kategori respons, akan lebih mewakili keadaan responsden yang sesungguhnya. Selain itu, panjang skala juga menentukan hal-hal lain yang lebih rinci, seperti lebih mudah untuk dimengerti, pemahaman arti setiap kategori respons, menghindari apriori terhadap respons dan adanya kemiripan antar kategori pada respons.

Namun, semua itu bergantung pada karakteristik individu, mencakup kemampuan kognitif dan motivasi dari subjek yang akan diukur (Krosnick & Presser, 2009). Penelitian lain yang menggunakan metode *Thurstone* dalam penskalaan aitem adalah Plotnick, Gomez, White, dan Turoff (2007), mengemukakan bahwa metode *Thurstone* memberikan nilai dengan urutan peringkat yang lebih jelas dan mempunyai keuntungan karena setiap aspek dilihat dalam pandangan yang lebih luas. Selain itu lebih memberikan kekuatan pendapat para ahli, yang diperoleh melalui perhitungan dari perbedaan setiap aspek dalam skala interval, dengan demikian akan diperoleh informasi yang lebih lengkap daripada penggunaan skala yang lebih sederhana.

Kesimpulan

Berdasarkan berbagai uraian hasil dan pembahasan, dapat diambil beberapa kesimpulan dari hasil penelitian ini yakni: pengujian validitas isi berbasis *expert judgment* melalui CVR memperoleh nilai dengan rentang 0,33 – 1,00 yang berarti terdapat aitem-aitem yang disepakati relevansinya oleh dua pertiga ahli dan terdapat pula aitem-aitem yang disepakati relevansinya oleh seluruh ahli. Diperoleh pula nilai CVI sebesar 78% dengan interpretasi para ahli cukup sepakat terhadap relevansi 33 aitem dalam TDQ. Hasil penskalaan *Thurstone* pada aitem-aitem TDQ mendapatkan homogenitas para ahli dalam memberikan justifikasi pembobotan dengan hasil akhir skor aitem berkisar antara 4 - 10. Skor-skor aitem TDQ yang berada pada skala interval ini akan memudahkan dalam implementasi TDQ dan memberikan interpretasi kepada subjek yang akan diperiksa dan ditentukan dalam kategori normal, terindikasi hipotiroid atau terindikasi hipertiroid. Tindak lanjut

yang bisa dilakukan dari hasil penelitian ini adalah tahap perancangan TDQ dan uji coba pada subjek di daerah endemis defisiensi yodium untuk menganalisis reliabilitasnya.

Kepustakaan

- Agha, A., Walker, D., Perry, L., Drake, W. M., Chew, S. L., Jenkins, P. J., Grossman, A. B., & Monson J. P. (2007). Unmasking of central hypothyroidism following growth hormone replacement in adult hypopituitary patients. *Clinical Endocrinology*, 66(1), 72-77. Diunduh dari: www.ncbi.nlm.nih.gov/pubmed. tanggal 12 Maret 2012.
- Allahyari, T., Rangi, N. H., Khosravi, Y., & Zayeri, F. (2011). Development and evaluation of a new questionnaire for rating of cognitive failures at work. *International Journal of Occupational Hygiene*. IJOH 3: 6-11. Diunduh dari: www.journals.tums.ac.i. tanggal 12 Januari 2014.
- Anderson, E. H., Neafsey, P. J., & Peabody, S. (2011). Psychometrics of the Computer-Based Relationships With Health Care Provider Scale in Older Adults. *Journal of Nursing Measurement*, 19(1), 3-16. Diunduh dari: www.ncbi.nlm.nih.gov/pmc. tanggal 12 Januari 2014.
- Azwar, S. (2009). *Penyusunan skala psikologi*, Penerbit Pustaka Pelajar, Yogyakarta.
- Bayram, F., Beyazyildiz, A., & Gökçe, C., Budak, N., Erdoğan, N., Kurtoğlu, S., Kula, M., Ünlühızarıcı K., & Keleştimur S. (2009). The prevalence of iodine deficiency, serum thyroglobulin, anti-thyroglobulin and thyroid peroxidase antibody levels in the urban areas of Kayseri, Central Anatolia. *Experimental and Clinical Endocrinology & Diabetes*; 117(2), 64-68. Diunduh dari: <https://www.thieme-connect.de/>. tanggal 17 Nopember 2011.
- Bazarganipour, F., Ziaei, S., Montazeri, A., Faghihzadeh, S., Frozanfar, F. (2012). Psychometric properties of the Iranian version of modified polycystic ovary syndrome health-related quality-of-life questionnaire. *Human Reproduction*, 27(9), 2729-2736. Diunduh dari: <http://humrep.oxfordjournals.org>. tanggal 12 Januari 2014.
- Bogale, A., Abebe, Y., Stoecker, B. J., Abuye, C., Ketema, K., & Hambidge, K. M. (2009). Iodine status and cognitive function of women and their five year old children in rural Sidama, Southern Ethiopia. *East African Journal of Public Health*, 6(3), 299-302. Diunduh dari: <http://web.a.ebscohost.com/ehost/pdfviewer>. tanggal 12 September 2011.
- Bruce, J. C., Langley, G. C., & Tjale, A. A. (2008). The use of experts and their judgments in nursing research: an overview. *Curationis*. 31(4), 57-61. Diunduh dari: <http://www.ncbi.nlm.nih.gov/pubmed/19653538>. tanggal 16 Januari 2014.
- Bruns, D., & Warren, P. A. (2011). *Assessment of psychosocial contribution to disability*. Handbook Behavioral Health Disability: Innovations in Prevention and Management. Editor Pamela A. Warren. Springer New York.
- Brown, B. T., Bonello, R., & Pollard, H. (2005). The biopsychosocial model and hypothyroidism. *Chiropractic & Osteopathy*, 13(5). Diunduh dari: <http://www.biomedcentral.com>. tanggal 23 Agustus 2011.

- Burgman, M., Fidler, F., McBride, M., Walshe, T., & Wintle, B. (2006). *Eliciting Expert Judgments: Literature Review*. Australian Centre for Excellence in Risk Analysis (ACERA), University of Melbourne. Diunduh dari: <http://www.acer.unimelb.edu.au/>. tanggal 14 Januari 2014.
- Carrio, F. B., Suchman, A. L., & Epstein, R. M. (2004). The Biopsychosocial model 25 years later: principles, practice, and scientific inquiry. *Annals of Family Medicine*, 2, 576-582. Diunduh dari: <http://www.annfam.org/content>. tanggal 13 Nopember 2011.
- Djokomoeljanto, R. (2008). *Gangguan akibat kekurangan iodium*, Makalah Seminar Nasional Peningkatan Kualitas Sumber Daya Manusia Melalui Penanggulangan Gangguan Akibat Kekurangan Iodium (GAKI), Balai Litbang GAKI Depkes RI Magelang, 19 Januari 2008.
- Flynn, G., O'Neill, C., McInerney, C., & Kennedy, H. G. (2011). The DUNDRUM-1 structured professional judgment for triage to appropriate levels of therapeutic security: retrospective-cohort validation study. *BMC Psychiatry*, 11(43), 1-11. Diunduh dari: <http://pubmedcentralcanada.ca>. tanggal 21 Juni 2011.
- Fung, K. (2010). *The initial development and content validity of an asperger's syndrome self-screening instrument for adult*, Thesis, Department of Educational Psychology and Special Education, University of Saskatchewan, Saskatoon, Saskatchewan. Diunduh dari: <http://www.ecommons.usask.ca>. tanggal 28 Juni 2011.
- Gonen, M. S., Kisakol, G., Cilli, A. S., Dikbas, O., Gungor, K., Inal, A., & Kaya, A. (2004). Assessment of anxiety in subclinical thyroid disorder. *Endocrine Journal*, 51(3), 311-315. Diunduh dari: <https://www.jstage.jst.go.jp/article>. tanggal 26 Agustus 2011.
- Hermann, D., Hewer, W., Lederbogen, F. (2004). Testing the association between thyroid dysfunction and psychiatric diagnostic group in an iodine-deficient area. *Journal of Psychiatry and Neuroscience*, 29(6), 444-449. Diunduh dari: <http://www.ncbi.nlm.nih.gov>. tanggal 12 September 2011.
- Indra, R., Patil, S. S., Joshi, R., Pai, M., & Kalantri, S. P. (2004). Accuracy of physical examination in the diagnosis of hypothyroidism: a cross-sectional, double-blind study. *Journal of Postgraduate Medicine*, 50, 7-10. Diunduh dari: <http://www.jpjgmonline.com>. Tanggal 25 Mei 2011.
- Krosnick, J. A., & Presser, S. (2009). *Question and questionnaire design*. Handbook of survey research (2nd Edition) James D. Wright and Peter V. Marsden (Eds). San Diego, CA: Elsevier. Diunduh dari: <http://www.rds-eastmidlands.nihr.ac.uk>. tanggal 29 Juni 2011.
- Kumar, R. (1999). *Research methodology: a step by step guide for beginner*, Sage Publications, London-Thousand Oaks-New Delhi.
- Medani, A. M. M. H., Elnour, A. A., & Saeed, A. M. (2011). Endemic goitre in the Sudan despite long-standing programmes for the control of iodine deficiency disorders. *Bulletin of The World Health Organization*, 89, 121-126. Diunduh dari: <http://www.who.int/bulletin>. tanggal 20 Januari 2012.
- Mortel, T. (2009). Development of a questionnaire to assess health care students' hand hygiene knowledge,

- beliefs and practices. *Australian Journal of Advanced Nursing*, 26(3), 9-16. Diunduh dari: <http://www.epubs.scu.edu.au>. tanggal 26 Juni 2011.
- Mutalazimah. (2010). Status klinis sebagai prediktor status yodium urin pada kasus *thyroid dysfunction* anak sekolah. *Jurnal Nutrisia*, 12(2), 79-85.
- Muxika, I., Borja, A., & Bald, J. (2007). Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European water framework directive. *Marine Pollution Bulletin*, 55, 16-29. Diunduh dari: <http://www.ncbi.nlm.nih.gov/pubmed>. tanggal 26 Agustus 2011.
- Novack, D. H., Cameron, O., Epe, E., Ader, R., Waldstein, S. R., Levenstein, S., Antoni, M. H., & Wainer, A. R. (2007). Psychosomatic medicine: the scientific foundation of the biopsychosocial model. *Academic Psychiatry*, 31, 388-401. Diunduh dari: <http://www.ncbi.nlm.nih.gov/pubmed>. tanggal 26 Agustus 2011.
- Ocal, G., UnceoUlu, H., Ayca, Z., BerberoUlu, M., AdYyaman, P., Ergun, EvliyaoUlu, O., & Akhurin, S. (2004). The effect of subclinical thyroid dysfunction and L-thyroxine treatment on spinal bone mineral density in clinically euthyroid goitrous children. *Turkish Journal of Endocrinology and Metabolism*, 4, 129-134. Diunduh dari: <http://www.turkjem.org>. tanggal 26 Agustus 2011.
- Parsa, P., Kandiah, M., Nasir M. T., Hejar, A.R., & Afah, N. M. Z. (2008). Reliability and validity of Champion's Health Belief Model Scale for breast cancer screening among Malaysian women. *Singapore Medical Journal*, 49(11), 897-903. Diunduh dari: <http://www.ncbi.nlm.nih.gov/pubmed>. tanggal 8 Juni 2011.
- Plotnick, L., Gomez, E. A., White, C., & Turoff, M. (2007). *Furthering development of a unified emergency scale using Thurstone's law of comparative judgment: a progress report*. Information Systems Department, New Jersey Institute of Technology Newark, NJ USA. Diunduh dari: http://www.dhs.gov/xlibrary/assets/hsas_unified_scale_feedback.pdf. tanggal 14 Januari 2014.
- Rudge, C. V. C., Calderon, I. M. P., Dias, A., Lopes, G. P., Barbosa, A. P., Maestá, I., Odland, J. O., & Rudge, M.V.C. (2009). Design and validity of a questionnaire to assess sexuality in pregnant women. *Reproductive Health*, 6(12). Diunduh dari: <http://www.ncbi.nlm.nih.gov/pubmed>. tanggal 29 Juni 2011.
- Salzberger, T. (2013). *Scaling and indexes, Quantitative Research Methods 1*, Vienna University of Economics and Business, October 30th, 2013. Diunduh dari: <http://www.statmath.wu.ca.at>. tanggal 5 Januari 2014.
- Sebotsa, M. L. D., Dannhauser, A., Mollentze, W.F., Oosthuizen, G.M., Mahomed F.A., & Jooste, P.L. (2009). Knowledge, attitudes and practices regarding iodine among patients with hyperthyroidism in the Free State, South Africa. *South African Journal of Clinical Nutrition*, 22(1), 18-21. Diunduh dari: <http://www.ajol.info>. tanggal 30 September 2011.
- Singh, M. B., Marwal, R., & Lakshminarayana, J. (2010). Assessment of iodine deficiency disorders in school age children in Jodhpur district of Rajasthan. *Journal of Human Ecology*; 32(2):79-83. Diunduh dari: <http://www.krepublishers.com>. tanggal 27 Agustus 2011.

- Stephens, J. L., Lowman, J. D., Graham, C. L., Morris, D. M., Kohler, C. L., & Waugh, J. B. (2013). Improving the Validity and Reliability of a Health Promotion Survey for Physical Therapists. *Cardiopulmonary Physical Therapy Journal*, 24(1), 14-23. Diunduh dari <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3677182/>. tanggal 12 Januari 2014.
- Tim GAKY Pusat, (2005). *Rencana aksi nasional kesinambungan program penanggulangan gangguan akibat kurang yodium*, Tim GAKY Pusat, Jakarta.
- Vejbjerg, P., Knudsen, N., Perrild, H., Laurberg, P., Andersen, S., Rasmussen, L.B, Ovesen, L., & Jorgensen, T. (2009). Estimation of iodine intake from various urinary iodine measurements in population studies. *Thyroid*. November, 19(11), 1281-1286. Diunduh dari www.ncbi.nlm.nih.gov/pubmed. tanggal 14 September 2011.
- Watt, T., Hegedüs, L., Groenvold, M., Bjorner, J. B., Rasmussen, A. K., Bonnema, S. J., & Rasmussen, U. F. (2007). Which domains of thyroid-related quality of life are most relevant? patients and clinicians provide complementary perspectives. *Thyroid*, 17(7), 647-654. Diunduh dari <http://www.ncbi.nlm.nih.gov/pubmed>. tanggal 31 Maret 2011.
- Watt, T. (2009). *Development of a Danish thyroid-specific quality of life questionnaire*, PhD Thesis, Department of Endocrinology, Copenhagen University Hospital Rigshospitalet and Health Service Research, Institute of Public Health. Diunduh dari: http://www.dadlnet.dk/dmb/dmb_phd/doc/PhDthesis_Torquil_Watt.pdf. tanggal 31 Maret 2011.