

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS: PSIQUIATRIA**

**DOUTORADO**

**AVALIAÇÃO DO DESEMPENHO DO WHOQOL-BREF  
EM PACIENTES DEPRIMIDOS PROVENIENTES DE SERVIÇOS  
DE CUIDADOS PRIMÁRIOS DE DIFERENTES PAÍSES  
USANDO A ANÁLISE DE RASCH**

**NEUSA SICA DA ROCHA**

**Porto Alegre, novembro 2008**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS: PSIQUIATRIA**

**DOUTORADO**

**AVALIAÇÃO DO DESEMPENHO DO WHOQOL-BREF  
EM PACIENTES DEPRIMIDOS PROVENIENTES DE SERVIÇOS  
DE CUIDADOS PRIMÁRIOS DE DIFERENTES PAÍSES  
USANDO A ANÁLISE DE RASCH**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Médicas: Psiquiatria da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Doutor em Psiquiatria.

**NEUSA SICA DA ROCHA**

**ORIENTADOR  
PROF. DR. MARCELO PIO DE ALMEIDA FLECK**

**Porto Alegre, novembro 2008**

*Este trabalho é dedicado à educação pública, gratuita, livre e democrática como instrumento de construção da autonomia e geração de conhecimento.*

*Aos meus pais, que não tiveram a oportunidade de freqüentar a universidade, pela luta incansável pela minha educação, de meus 9 irmãos e tantas outras crianças que se beneficiaram ao longo de décadas, estudando e aprendendo nas mais de 90 escolas públicas que eles foram os responsáveis pela construção.*

*Ao Marcelo, pelo exemplo de professor dedicado e altruísta.*

## **AGRADECIMENTOS**

Ao Professor Marcelo Fleck, por ser um modelo de professor e pesquisador, ético e dedicado, por todos os ensinamentos, pelo incentivo e suporte.

Ao meu marido Alexandre, pelo amor, carinho, companheirismo e compreensão pelos momentos de afastamento do seu convívio.

A minha mãe Laís, pelo exemplo de amor e dedicação e constante apoio e cuidados.

Ao meu pai Tapir (in memoriam), pelo exemplo amor e de luta, e constante desafio e incentivo para que eu seguisse estudando e pesquisando.

Aos meus 9 irmãos, Ceci Iara, Nadia, Lígia, Marta, Guacira, Tapir Filho, Soraya, Leonel e Taís e aos meus 13 sobrinhos, Viviane, Marcus, Natália, Tapir Neto, Leonel, Bruno, Ana Caroline, Gabriel, Gabriela, Laís, Luis, Giulia e Lohan pelo amor e por serem sempre uma fonte de aconchego, segurança, alegria e espontaneidade.

A minha irmã gêmea Taís, pelo amor, pelo incentivo e por ser sempre uma excelente e divertida companhia.

A minha irmã Lígia, pelo exemplo de solidariedade e por ser a minha segunda mãe.

Aos meus pacientes, pela gratificação e constantes desafios para novas descobertas.

Aos colegas, Alayde Barcellos, Ana Paula Reolon, Gustavo Schestatsky, Ricardo Silveira, Márcia Dieterisch, pelo coleguismo e por cuidarem tão bem dos meus pacientes nos momentos que por compromissos acadêmicos tive de me ausentar de Porto Alegre.

Ao Professor Mick Power, da Universidade de Edimburgo, coordenador do grupo WHOQOL no mundo, por todos os ensinamentos e por me proporcionar a oportunidade de aprender a análise de Rasch num dos melhores centros de pesquisa do mundo e conhecer melhor o grupo WHOQOL.

Ao grupo WHOQOL no mundo, pelo exemplo de colaboração entre pesquisadores.

Ao centro coordenador do LIDO, Health Research Associates, pela concessão dos dados internacionais do estudo para que eu pudesse analisar e realizar esta tese.

Ao pesquisador Donald Bushnell, pela colaboração na revisão dos meus artigos.

Ao Professor Alan Tennant, do Laboratório de Psicometria para Ciências da Saúde da Universidade de Leeds-Reino Unido, pela colaboração na revisão dos meus artigos.

Aos professores Luis Augusto Rohde e Cláudio Eizirik, pelo apoio para que eu fizesse o doutorado sanduíche.

Ao professor Luis Augusto Rohde, pela oportunidade e convite para revisar os artigos da Revista Brasileira de Psiquiatria.

Ao colega Flavio Shansis, pelo convite para integrar o conselho editorial da Revista de Psiquiatria do Rio Grande do Sul.

A professora Isa Brenner, da escola de línguas Britania, pelo carinho e incentivo para que eu conseguisse aprender a língua inglesa e me tornar proficiente.

Aos colegas do grupo WHOQOL no Brasil, Ana Flavia Lima, Rogério Zimpel e Luciane Cruz pelo agradável convívio e companheirismo.

Aos colegas da Secretaria da Saúde de Viamão, Bernardo Kern, Rosane Viegas, Oscar Basso e Gislaine Thompson, pela amizade e sintonia na angústia na busca da solução dos graves problemas de saúde pública daquela cidade.

Aos meus amigos do colégio, Tais, Adriana(s), Lucia (in memoriam), Monique, Rosa, Ana Paula, Andréia, Luciana e Márcio; e os da faculdade Andry, Claudia, César, Ana Paula, Letícia, Vivien, Gustavo e Sandra por estarem sempre ao meu lado nos mais diferentes momentos da minha vida, sendo parte da minha família.

*“...Para além da enfermaria, não tinha com que desocupar o tempo. De tal maneira que deixei de sonhar. Só os pesadelos me visitavam. Eu estava aleijada desse órgão que segrega as matérias do sonhar. Eu estava doente sem doenças. Sofria destas maleitas que só Deus padece. Aconteceu assim: primeiro me acabou o riso; depois os sonhos; por fim, as palavras. É essa a ordem da tristeza, o modo como o desespero nos encerra num poço húmido. ...Foi nesse afundamento que me apaixonei por Vasto. O amor não é o irremediável remédio?”*

*A confissão de Marta em  
“A varanda do frangipani”  
Mio Couto  
(escritor moçambicano)*

## RESUMO

**Introdução:** O WHOQOL-Bref, a versão abreviada da medida genérica de qualidade de vida (QV) desenvolvida pela Organização Mundial da Saúde, foi desenvolvida simultaneamente em várias culturas e línguas sob a base da Teoria Clássica em psicometria. Este é composto por 26 itens que formam quatro domínios: físico, psicológico, social e ambiental. Uma vez que este pode ser considerado uma medida de QV subjetiva, alguns autores sugerem que a influência da sintomatologia depressiva no escore de QV possa ser verificada. Ademais, está bem adequado para examinar a influência da cultura na QV, mas sua validade transcultural em pacientes deprimidos continua não estudada. QV e depressão têm uma complexa relação conceitual. Até onde nós sabemos, não existe nenhum estudo avaliando a potencial sobreposição entre as medidas de QV e depressão utilizando dados transculturais. O “Longitudinal Investigation of Depression Outcomes” (LIDO) foi um estudo multi-cêntrico e transnacional, observacional que acompanhou pacientes com transtorno depressivo em serviços de cuidados primários por 12 meses em 6 países. Potencialmente, os dados do LIDO nos permitem testar empiricamente a relação entre as medidas de depressão e de QV. O modelo de Rasch, uma técnica estatística moderna, tem sido visto como uma diretriz que coloca em operação os axiomas das medidas aditivas. Este modelo apresenta uma série de análises para obter uma variável com uma estrutura aditiva, e, portanto, adequada para ser medida em nível intervalar. Também, é capaz de acessar a invariância transcultural de uma medida. **Objetivos:** O objetivo principal de nosso estudo é avaliar as



propriedades de medida do WHOQOL-Bref usando a análise de Rasch de seus domínios e itens em pacientes deprimidos de serviços de cuidados primários de diferentes países. Os objetivos secundários são: verificar se os itens do WHOQOL-Bref são invariantes entre os pacientes que têm um episódio depressivo atual, testar a invariância das medidas dos itens num contexto transcultural, e verificar a validade da versão brasileira do WHOQOL-Bref em adultos com depressão maior usando a análise de Rasch. Esta tese é composta por quatro estudos. O primeiro estudo (estudo 1) é um artigo de revisão que tem como objetivo apresentar as bases do modelo de Rasch, usando o exemplo prático de uma escala de depressão. O segundo (estudo 2) apresenta a avaliação da invariância dos itens do WHOQOL-Bref entre pacientes deprimidos e não deprimidos. O terceiro (estudo 3) enfoca nas propriedades de medida e na avaliação transcultural dos itens do WHOQOL-Bref e o quarto (estudo 4) é sobre a versão brasileira desta medida. **Métodos:** *Sujeitos:* Para o estudo 2, a amostra consistiu de 2359 sujeitos, dos quais 1193 tinham o diagnóstico confirmado de episódio depressivo atual e eram provenientes de seis países (Austrália, Brasil, Israel, Rússia, Espanha, e EUA) incluídos na avaliação basal do LIDO. Estes mesmos 1193 pacientes deprimidos fizeram parte da amostra do estudo 3. O estudo 4 incluiu os 208 pacientes deprimidos da amostra brasileira. *Medidas:* Nós utilizamos os seguintes instrumentos de medida: a Center for Epidemiological Studies Depression Scale (CES-D) para avaliar a severidade da depressão; o World Health Organization Quality of Life Instrument – versão abreviada (WHOQOL-Bref) como medida genérica de QV; e o Composite International Diagnostic Interview (CIDI), versão 2.1 para o diagnóstico de depressão. **Análise Estatística:** *Estudo 2:* A análise de Rasch foi utilizada para ver se os itens que exibiriam funcionamento diferencial (DIF), como uma maneira de

acessar a invariância em relação ao fator depressão definido como o diagnóstico de depressão usando o CIDI. *Estudos 3 e 4:* Uma análise de Rasch completa foi realizada para acessar as propriedades de medida dos domínios e dos itens do WHOQOL-Bref. As análises dos DIF foram feitas para verificar os vieses de idade, gênero e a invariância transcultural. **Resultados:** *Estudo 2:* Onze dos 26 itens do WHOQOL-Bref apresentaram DIF devido ao fator depressão e o domínio físico foi o que apresentou mais itens com DIF. *Estudo 3:* Os domínios físico, psicológico e ambiental necessitaram de ajustes para se adequar ao modelo de Rasch, principalmente devido à inadequação de itens individualmente ou DIF devido à idade. O domínio social manteve as suas fracas propriedades psicométricas demonstradas em estudos prévios. Apenas cinco itens foram totalmente invariantes entre os países, mas os vieses entre os outros itens desapareceram quando analisado juntos, indicando um cancelamento deste efeito. *Estudo 4:* Após usar a análise de Rasch na amostra brasileira, os 4 domínios do WHOQOL-Bref demonstraram medidas apropriadas de adequação ao modelo. Alguns itens necessitaram ajustes: 4 itens mudaram de escores ('dor'; 'finanças', 'serviços' e 'transporte'), 2 itens ('trabalho' e 'atividade') apresentaram dependência local nas respostas, e 1 item foi excluído ('sono'), mostrando multidimensionalidade. **Conclusões:** Nossos achados indicam que a maioria dos itens do WHOQOL-Bref não apresentam DIF para a ocorrência de um episódio atual de depressão maior e a variância associada com a depressão com esta medida genérica de QV é restrita a algumas facetas. Portanto, nós recomendamos este ajuste restrito em futuras análises desta medida para pacientes deprimidos. Também, nossos resultados demonstram que os domínios do WHOQOL-Bref com alguns ajustes apresentam propriedades de Rasch no contexto transcultural de seis países. Finalmente, após

sofrer ajustes pela análise de Rasch, o WHOQOL-Bref parece ser um instrumento psicometricamente válido para avaliar a QV de pacientes brasileiros de um serviço de cuidados primários.

**Palavras-chave:** Qualidade de vida, Depressão, Atenção primária à saúde, Questionários, Organização Mundial da Saúde, Estudos de Validação, Modelos Estatísticos.

## ABSTRACT

**Background:** The WHOQOL-Bref, the abbreviated generic measure of quality of life (QOL) developed by the World Health Organization, was developed simultaneously in several cultures and languages under the framework of Classical Test Theory. It has 26 items covering four domains: Physical, Psychological, Social Relationships and Environment. Since it can be considered a measure of subjective QOL, some authors have suggested that its assessment should be checked for the influence of depressive symptomatology on the QOL score. Moreover, it is well placed to examine the influence of culture on QOL, but its cross-cultural validity in depressed individuals remains unstudied. Quality of life and depression have a complex conceptual relationship. To our knowledge, there is no study addressing the potential overlap of QOL and depression measuring using a cross-cultural database. The Longitudinal Investigation of Depression Outcomes (LIDO) study was a multicenter, cross-national observational study that followed patients with depressive disorders in primary care settings for 12 months in six countries. Potentially, the LIDO database allows us to test empirically the relationship between depression and QOL measurement. The Rasch model, a modern statistical technique, is seen as a template which puts into operation the axioms of additive conjoint measurement. This model presents a set of analyses to obtain a variable with additive structure and, hence is suitable to be measured on an interval scale. Also, this able to assess the cultural invariance of a measure. **Objectives:** The main purpose of our study is to look at the measurement

properties of the WHOQOL-Bref using Rasch analysis of the domains and items of WHOQOL-Bref) in depressed patients from primary care services from different countries. Secondary purposes are: to verify whether the items of the WHOQOL-Bref are invariant among patients having a current major depressive episode, to test the invariance of item measures in cross-cultural settings, and to assess the validity of the Brazilian version of WHOQOL-Bref in adults with major depression using the Rasch analysis. This thesis is composed by four studies. The first study is a review article aiming at presenting the basic features of the Rasch model using a practical example of a depression scale. The second study (study 2) presents an evaluation of the invariance of the WHOQOL-Bref items between depressed and non depressed patients. The third study (study 3) focus on the measurement properties and the cross-cultural evaluation of the items of the WHOQOL-Bref and the fourth study (study 4) is on the Brazilian version of this measure. **Methods:** *Subjects:* In the study 2, the sample consisted of 2,359 subjects, of which 1,193 had a confirmed diagnosis of a current major depressive episode from six countries (Australia, Brazil, Israel, Russia, Spain, and the USA) involved in the baseline assessment of the LIDO study. These same 1,193 major depressed patients consisted the sample of the study 3. Study 4 had 208 major depressed patients from Brazilian sample. **Measures:** We used the following measurement instruments: the Center for Epidemiological Studies Depression Scale (CES-D) to assess severity of depression; the World Health Organization Quality of Life Instrument - Abbreviated version (WHOQOL-Bref) was used as a generic quality of life (QOL) instrument; and the Composite International Diagnostic Interview (CIDI), version 2.1 for the diagnosis of depression. **Analysis:** *Study 2:* The Rasch analysis was used to look at items exhibiting differential item functioning (DIF) as a way of assessing invariance in relation to a depression factor

defined by the diagnosis of depression using the Composite International Diagnostic Interview (CIDI). *Study 3 and 4:* A complete Rasch analysis was run to assess measurement properties of the WHOQOL-Bref domains and items. DIF analysis was made to verify bias by age, gender and cross-cultural invariance. **Results:** *Study 2:* Eleven out of the 26 items of the WHOQOL-Bref showed DIF due to the depression factor and the Physical domain presented more items displaying DIF. *Study 3:* The Physical, Psychological and Environment domains of the WHOQOL-Bref required adjustments to conform to the Rasch model expectations, mainly because of individual item misfit, or Differential Item Functioning (DIF) due to age. The Social domain maintained its poor psychometric properties evidenced by previous studies. Only five items were totally invariant across countries, but the bias of other items disappeared when pooled, indicating cancellation effects. *Study 4:* After using Rasch analysis, the 4 domains of WHOQOL-Bref showed appropriate fit to this model. Some items needed adjustments: 4 items were rescored ('Pain'; 'Finances', 'Services' and 'Transport'), 2 items ('Work' and 'Activity') identified they have dependency of responses, and 1 item was deleted ('Sleep'), showing multidimensionality. **Conclusions:** Our findings indicate that the majority of WHOQOL-Bref items do not exhibit DIF for a current major depressive episode and the variance associated with depression in this generic QOL measure is restricted to some facets of this construct. Thus, we recommend this restricted adjustment for depression in future analyses of this measure. Also, our results support the Rasch properties of the domains of the WHOQOL-Bref, with some modification, as a measure of generic subjective QOL in the context of primary care depressed patients in six countries worldwide. Finally, the WHOQOL-Bref, after Rasch adjustments, seems to be a psychometrically valid

instrument that it is suitable for evaluating the QOL of Brazilian depressed outpatients from primary care setting.

**Key words:** Quality of Life, Depression, Primary Health Care, Questionnaires, World Health Organization, Validation Studies, Statistical Models.

## LISTA DE QUADROS E TABELAS

Quadro 1: Pressupostos do Modelo de Rasch e os testes estatísticos para sua identificação (Andrich, Sheridan et al., 2004) .....	27
Tabela 1: Critérios de doses de antidepressivos potencialmete adequadas para tratamento de depressão no contexto primário do LIDO .....	44
Quadro 2: Instrumentos e padrão de utilização das medidas no LIDO .....	46
Quadro 3: Resumo dos achados principais dos estudos publicados internacionalmente utilizando os dados do LIDO .....	55
 <b>Estudo 2 - Is there a measurement overlap between depressive symptomatology and quality of life?</b>	
Table 1: General descriptions of baseline sample of LIDO study .....	104
Table 2: Differential Item Functioning (DIF) analysis in LIDO baseline sample N=2359 for depression factor WHOQOL-Bref domains .....	105
Table 3: Theoretical QOL and depression domains stemmed from WHOQOL-Bref original items .....	106
 <b>Estudo 3 - Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis</b>	
Table 1: General descriptions of baseline sample of depressed patients enrolled in the LIDO study .....	135
Table 2: Fit of the WHOQOL-Bref items to Rasch Model for baseline sample of depressed patients enrolled in the LIDO study .....	136



Table 3: The WHOQOL-Bref items displaying DIF for more than one person factor for the baseline sample of depressed patients enrolled in the LIDO study.....	137
Table 4: Fit of the WHOQOL items adjusted by Rasch model for baseline sample of depressed patients enrolled in the LIDO study .....	138
Table 5: Post hoc analysis of DIF by country for the WHOQOL-Bref items in depressed patients in the LIDO study.....	139

**Estudo 4 - Validity of the Brazilian version of WHOQOL-Bref in primary care depressed patients using Rasch modelling**

Table 1: General descriptions of baseline sample of depressed patients enrolled in Longitudinal Investigation of Depression Outcomes study in Brazil .....	158
Table 2: Fit of WHOQOL-Bref items to Rasch Model of baseline sample of depressed patients enrolled in Longitudinal Investigation of Depression Outcomes study in Brazil.....	159
Table 3: WHOQOL-Bref response scale adapted for 4 items after Rasch analysis.....	160
Table 4: Fit of WHOQOL items adjusted by Rasch model of baseline sample of depressed patients enrolled in LIDO study in Brazil .....	161

## LISTA DE FIGURAS

Figura 1: Artigos pesquisados no Medline na primeira busca por Quality of life, Item Response Theory, Rasch analysis, validity, depression, WHOQOL .....	37
Figura 2: Fluxograma descritivo dos pacientes recrutados para participar do Longitudinal Investigation of Depression Outcomes.....	47
<b>Estudo 1 - The perspectives of the use of Rasch analysis in psychometric research</b>	
Figure 1: Item Characteristic Curve of Item 14 (Appearance) of the Beck Depression Inventory (BDI) .....	78
Figure 2: Probability Categories Curves of the Item 8 of the BDI- Example of Centralized Thresholds.....	79
Figure 3: Probability Categories Curves of the Item 13 (make decisions) of the BDI- Example of disordered thresholds .....	80
<b>Estudo 3 - Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis</b>	
Figure 1: Person-Item location distribution map for the WHOQOL-Bref domains for baseline sample of depressed patients enrolled in the LIDO study.....	140

## SUMÁRIO

1 INTRODUÇÃO .....	21
2 REVISÃO DA LITERATURA .....	30
2.1 Avaliação de qualidade de vida em pacientes deprimidos .....	30
2.2 O Longitudinal Investigation of Depression Outcomes.....	38
2.2.1 Definição das variáveis de tratamento.....	43
2.2.2 Definição das categorias ou tipos de serviços.....	45
2.2.3 Cálculo de tamanho das amostras .....	45
2.2.4 Características gerais da amostra basal .....	48
2.2.4.1 Uso de serviços de saúde entre os centros.....	49
2.2.4.2 Barreiras ao acesso aos serviços.....	49
2.2.5 Estudos publicados internacionalmente utilizando as amostras do LIDO... 50	
3 JUSTIFICATIVA .....	57
4 OBJETIVOS .....	58
4.1 Objetivo geral .....	58
4.2 Objetivos específicos .....	58
5 HIPÓTESES.....	60
6 ARTIGOS .....	61
6.1 Estudo 1 - The perspectives of the use of Rasch analysis in psychometric research.....	63
6.2 Estudo 2 - Is there a measurement overlap between depressive symptomatology and quality of life? .....	83
6.3 Estudo 3 - Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis .....	107
6.4 Estudo 4 - Validity of the Brazilian version of WHOQOL-Bref in primary care depressed patients using Rasch modelling .....	142
7 CONSIDERAÇÕES FINAIS .....	162

8 CONCLUSÃO.....	168
9 REFERÊNCIAS BIBLIOGRÁFICAS .....	169
APÊNDICES.....	180

# 1 INTRODUÇÃO

A avaliação de qualidade de vida tem recebido atenção crescente tanto no campo da prática clínica quanto da pesquisa (Leplege e Hunt, 1997). Esta surgiu diante da necessidade de se avaliar outros aspectos da vida dos pacientes que não a mera identificação de sintomas. Além disto, o avanço tecnológico conquistado pela Medicina fez com que as medidas tradicionais de morbi-mortalidade fossem se tornando insuficientes para descrever as necessidades e expectativas das pessoas quando acometidas por enfermidades, bem como para a definição de tratamentos e políticas de saúde (Fleck, 2008).

Neste ambiente científico, vários instrumentos para medir qualidade de vida foram sendo construídos e estudados em diferentes populações (Bergner, Bobbitt *et al.*, 1976; Kaplan, Atkins *et al.*, 1984; Read, Quinn *et al.*, 1987; Jenkinson, Fitzpatrick *et al.*, 1988; Kaplan, Anderson *et al.*, 1989; Aaronson, Ahmedzai *et al.*, 1993; Jenkinson, Coulter *et al.*, 1993; Hurst, Kind *et al.*, 1997; Coons, Rao *et al.*, 2000). Um exemplo importante foi o trabalho iniciado na década de 90 pelo World Health Organization Quality of Life Group (WHOQOL Group) que passou a desenvolver um instrumento que valorizasse a percepção subjetiva do indivíduo, que fosse multidimensional e que contivesse aspectos positivos e negativos a serem considerados (WHOQOL Group, 1995; WHOQOL Group, 1996).

**O World Health Organization Quality of Life Instrument (WHOQOL)** é um instrumento para avaliar qualidade de vida, desenvolvido a partir do conceito de qualidade de vida criado pela Organização Mundial da Saúde (OMS), sendo

qualidade de vida definida como “a percepção de um indivíduo de sua posição na vida, no contexto de cultura e sistemas de valores sobre os quais vive e em relação a suas metas e expectativas. Isto é obviamente influenciado pela saúde física, status psicológico, nível de independência, relações sociais, fatores ambientais e crenças pessoais” (WHOQOL Group, 1995).

Inicialmente foi desenvolvido o instrumento para avaliação de qualidade de vida (QV) genérica da Organização Mundial da Saúde, o WHOQOL-100, que é composto por 100 itens dispostos em 6 domínios e 24 facetas.

Utilizando as questões do WHOQOL-100, foi construída a versão breve do WHOQOL, o WHOQOL-Bref. O modelo inicial do WHOQOL-Bref foi desenhado para ter seis domínios: saúde física, psicológico, relacionamentos sociais, ambiental, nível de independência e espiritualidade. Em análises posteriores, o domínio nível de independência foi agregado com o domínio saúde física e o espiritual foi agregado ao psicológico, gerando o modelo com quatro domínios. Dos 26 itens, os que questionam QV em geral e satisfação com a saúde não são incluídos em outros domínios, mas formam uma faceta da QV geral. O WHOQOL-Bref foi testado em 23 países (n=11830) e possui boas a excelentes propriedades psicométricas, confirmando a alta qualidade do instrumento em quase todas as línguas onde foi estudado (WHOQOL Group, 1998a; Skevington, Lotfy *et al.*, 2004). As versões nacionais precisam ser estudadas para estabelecer se estão ligadas à cultura majoritária, negligenciando os idiomas ou tendo problemas específicos para minorias (Saxena, Carlson *et al.*, 2001).

Além da preocupação com a melhor metodologia a ser utilizada para assegurar um processo adequado de validação, também há a necessidade de se buscar o melhor plano de análise estatística para estes estudos. Os instrumentos

WHOQOL-100 (WHOQOL Group, 1998b), WHOQOL-Bref (WHOQOL Group, 1998a; Skevington, Lotfy *et al.*, 2004), WHOQOL-HIV (WHOQOL Group, 2003) e WHOQOL-SRPB (Srpb, 2006) foram construídos dentro do referencial da Teoria Clássica da psicometria. Dos instrumentos do WHOQOL, somente o WHOQOL-Old (Power, Quinn *et al.*, 2005) incluiu também análise de Teoria da Resposta ao Item (análise de Rasch) no seu processo de desenvolvimento.

Existem apenas três estudos que testaram a estrutura dos instrumentos genéricos do WHOQOL utilizando modelos de Teoria da Resposta ao Item. O primeiro testou a formação de um índice de QV a partir dos itens do WHOQOL-100, o que resultou em apenas um item com equivalência cultural para todos países envolvidos (Leplege, 2000). Recentemente, alguns pesquisadores testaram a confiabilidade, estrutura de quatro domínios e a validade de constructo do WHOQOL-Bref utilizando a análise de Rasch multidimensional e constataram que a exclusão de sete itens tornou esta medida um constructo unidimensional (Wang, Yao *et al.*, 2006). Já em uma amostra dinamarquesa utilizando a análise de Rasch e um modelo de Teoria da Resposta ao Item de dois parâmetros, determinou que a estrutura de quatro domínios é uma expressão mais adequada da QV do que um escore total dos 26 itens (Noerholm, Groenvold *et al.*, 2004). A invariância da estrutura deste instrumento também já foi estudada em diferentes grupos de patologias (Yao e Wu, 2005).

Até recentemente, o modo de validação de instrumentos de avaliação mais comumente utilizado é a chamada Teoria Clássica (TC). A TC avalia fidedignidade através da consistência interna pelo alfa de Cronbach e as validades de construto e de critério e conteúdo (Pasquali, 1999; 2003; Streiner e Norman, 2003). Seu enfoque está na escala como um todo, sendo que todos os itens têm igual importância no

escore total do instrumento. Além disso, permite que se some os escores de cada um dos itens, uma vez avaliada a dimensionalidade da escala, para se obter o escore total, mesmo que as questões ou itens versem sobre conteúdos de diferente repercussão para a pontuação de um determinado fenômeno. O escore total é considerado o escore real onde o erro associado a este e a cada pessoa são considerados como não correlacionados e são normalmente distribuídos. (Pasquali, 1999; 2001; 2003; Streiner e Norman, 2003).

A partir da constatação das limitações da TC, é possível utilizar um modelo de análise diferente para verificar o desempenho dos instrumentos levando em consideração parâmetros e características com o enfoque nos seus itens. Esta tem sido chamada de Teoria da Resposta ao Item (TRI) (Wright e Stone, 1979; Andrade, Tavares *et al.*, 2000; Embretson e Reise, 2000; Baker, 2001).

Atualmente existem uma série de modelos matemáticos ou equações de TRI que, além de avaliarem mais de um parâmetro, também estudam diferentes tipos de itens e escalas de respostas. Os principais são os logísticos de um, dois e três parâmetros. O modelo de um parâmetro leva em consideração apenas a *dificuldade* do item, já o de dois parâmetros leva também em consideração a *discriminação* e por fim o de três parâmetros avalia, além da *dificuldade* e da *discriminação*, a possibilidade de *acerto ao acaso* ou pela *adivinhação* (“chute”) (Wright e Stone, 1979; Andrade, Tavares *et al.*, 2000; Baker, 2001).

Dentre estes modelos o que tem sido mais amplamente utilizado é o desenvolvido por Georg Rasch (Rasch, 1960). Neste modelo o parâmetro da discriminação do modelo logístico de dois parâmetros é fixado no valor de 1 para todos os itens, somente o parâmetro da *dificuldade* pode ter valores diferentes para cada item. Por isso, o modelo de Rasch é freqüentemente considerado um modelo



de um parâmetro (*dificuldade*) (Andrich, 1978). Os autores que utilizam o modelo de Rasch defendem que, embora matematicamente possam ser semelhantes, existe uma diferença de paradigma entre o modelo de Rasch e o modelo de um ou dois parâmetros da TRI. Segundo eles, ao se testar os dados usando o modelo de Rasch estamos partindo de uma hipótese definida a priori, isto é, já temos o modelo definido e o testaremos empiricamente com nossos dados. Já na perspectiva da TRI tentaríamos identificar qual o melhor modelo que se ajusta aos dados podendo ser este de um, dois ou três parâmetros (Andrich, 2004). Além disso, a análise de Rasch em relação requer amostras menores para estimação dos parâmetros o que a torna mais facilmente aplicável em pesquisa clínica (Conrad e Smith, 2004).

A análise de Rasch pode contribuir de várias formas para melhorar o desempenho de itens nas escalas de medida (Conrad e Smith, 2004): (a) equalizar as respostas de diferentes grupos de itens que pretendem medir um mesmo constructo; (b) desenvolver uma medida com unidade com intervalos iguais; (c) incorporar os dados perdidos (“missing”) por usar um método de estimação que recai sobre a suficiência estatística e métodos de estimação que simplesmente sumarizam as observações não perdidas (“nonmissing”) que são relevantes para cada parâmetro e compara com suas expectativas; (d) conduzir a avaliação da validade e da confiabilidade em uma mesma análise para a calibração dos itens e para a medida das pessoas; (e) estimar a habilidade das pessoas independente da distribuição da amostra dos itens; (f) estimar a dificuldade do item independente da distribuição da amostra das pessoas; (g) conduzir testes adaptativos computadorizados para aumentar a eficiência da medida; (h) expressar a calibração das medidas das pessoas e dos itens em uma escala linear comum; e (i) focar na performance do itens e pessoas individualmente e não nas estatísticas do grupo.

Um instrumento que funcione como uma escala Rasch deve obedecer a alguns pressupostos avaliados por este modelo (Wright e Mok, 2000). Os principais pressupostos são a unidimensionalidade, a invariância, a independência local e a aditividade.

A unidimensionalidade diz respeito à característica que o instrumento deve ter de acessar apenas um constructo ou traço latente (Embretson e Reise, 2000; Tesio, 2003; Smith, 2004; Bond e Fox, 2007). A invariância identifica se a escala funciona da mesma maneira independentemente de outros fatores que possam estar influenciando a amostra. A independência local é a propriedade que cada item tem de medir o que se propõe sem precisar de outros itens (Stevens, 1946; Smith, 2000). A aditividade refere-se a característica da unidade de medida de ser do mesmo tamanho (intervalo) ao longo do contínuo inteiro a ser medido quando os dados se adequam ao modelo. Essas unidades são chamadas de logits (“logarithm of odds units”) e são uma função linear da probabilidade de obter um certo escore para pessoa com uma determinada habilidade. Cada um destes pressupostos é avaliado por um teste estatístico (Tennant, 2004b; a; 2005), conforme a tabela abaixo:

**Quadro 1: Pressupostos do Modelo de Rasch e os testes estatísticos para sua identificação (Andrich, Sheridan *et al.*, 2004)**

Pressuposto	Teste estatístico
<b>Unidimensionalidade</b>	<p><b>Medidas gerais de adequação do modelo</b></p> <p><math>\chi^2</math> geral; P  Ideal <math>\chi^2</math> baixo e Probabilidade &gt;0.05  Primeiro componente da Matriz de Componente dos resíduos. Comparação entre os resíduos positivos e negativos dentro do Primeiro componente pelo Teste de t; P  Ideal Probabilidade &gt;0.05</p> <p><b>Medidas de Adequação dos Itens individuais</b></p> <p>Resíduos; ideal até <math>\pm 2.5</math>  <math>\chi^2</math>; P  Ideal <math>\chi^2</math> baixo e Probabilidade &gt;0.05</p>
<b>Invariância</b>	<p>Análise de Item com funcionamento diferencial (DIF)  ANOVA</p>
<b>Independência Local</b>	<p>Matriz de correlação dos resíduos. Itens com correlação &gt; +0,3, são indicativos de dependência local. Caso ocorram, estes itens devem ser colocados em uma análise de subteste, se isto melhorar as medidas gerais do modelo, estamos diante de itens com dependência local.</p>
<b>Aditividade</b>	<p>As escalas de resposta devem ser organizadas e se que considera a possibilidade de existência de um limiar de resposta (Threshold). Caso haja, limiares desordenados, é possível colapsar as categorias de respostas (Masters, 1982).</p>

Com o uso do modelo de Rasch é possível estabelecer um nível de medida fundamental (Bond e Fox, 2007). Para se ter um nível de medida fundamental é necessário que se construa um instrumento de medida, bem como a sua unidade de medida. Já a unidade de medida deve ser a mesma ao longo do instrumento de medida. Uma vez que se chegue à unidade de medida é possível operacionalizar a aditividade destas unidades. Então, pode-se usar o sistema de números naturais e determinar a correspondência entre unidades e números (Stone, 1996; Pasquali, 2003).

Podemos, portanto, identificar se os dados provenientes de escalas ordinais podem ser considerados intervalares, e, com isso ter segurança da validade das operações matemáticas realizadas com estes. A racionalização da quantificação é uma pré-condição necessária para a psicometria como uma ciência racional quantitativa. Os proponentes da análise de Rasch consideram que os seus modelos além de permitir esta racionalização, podem contribuir para as Ciências Sociais com aquilo que contribuiu a medida fundamental para as Ciências Físicas (Andrich, 2004).

Dentre as aplicações da análise de Rasch podemos destacar: os estudos de validade e confiabilidade; o desenvolvimento de escalas, a equalização; a avaliação da dimensionalidade; a geração de banco de itens; o desenvolvimento de testes adaptativos computadorizados; apresentação e interpretação de escores; a identificação de itens com funcionamento diferencial e equivalência transcultural, a otimização de escalas, a avaliação de mudança; entre outras (Conrad e Smith, 2004).

Finalmente, concluímos que o uso da análise de Rasch pode tanto aprimorar os processos de validação que utilizaram apenas TC quanto servir como modelo

único neste processo (Tesio, 2003). A maioria dos instrumentos de QV que utiliza este tipo de metodologia, o faz unicamente (Tennant, Mckenna *et al.*, 2004), ou em conjunto com a TC (Prieto, Alonso *et al.*, 2003; Sloan e Mandrekar, 2005). Também tem sido o modelo padrão atual para avaliar a validade transcultural de uma medida (Tennant, Penta *et al.*, 2004). Além disso, por construir medidas que são independentes da amostra e do instrumento, sua aplicação é especialmente valorizada na validação de instrumentos de QV em pacientes deprimidos. Embora existam todas estas técnicas estatísticas para aprimorar e avaliar os instrumentos, nenhuma delas deve sobrepujar o modelo teórico no qual se baseia a construção de um instrumento (Van Alphen, Halfens *et al.*, 1994).

## **2 REVISÃO DA LITERATURA**

### **2.1 Avaliação de qualidade de vida em pacientes deprimidos**

Embora o termo “qualidade de vida”, venha-se popularizando e pareça intuitivamente de fácil compreensão, não há um consenso sobre como defini-lo e tampouco como mensurá-lo. Também, não há um instrumento que seja considerado o padrão-ouro para estudos com pacientes com transtorno mental. Katschnig (Katschnig, 2006) considera que muitos dos instrumentos utilizados são de validade questionável, o que dificulta a comparabilidade dos achados obtidos. Também argumenta que, especialmente quando usados em pacientes psiquiátricos, é necessário que o instrumento contemple a avaliação de aspectos subjetivos e objetivos, inclua sintomas psicopatológicos para serem controlados, contenha a avaliação do bem-estar e satisfação tanto do funcionamento quanto do ambiente, e avalie cada área da vida do paciente separadamente e em tempos diferentes. Com isso, seria facilitada a transferência dos achados de pesquisa para a prática clínica.

Interessantemente, um estudo com pacientes psiquiátricos severamente doentes demonstrou que os médicos e seus pacientes tendiam a coincidir mais nos aspectos clínicos (sintomas e funções) da QV dos que nos sociais - ocupacional e relações sociais (Sainfort, Becker *et al.*, 1996). Neste mesmo sentido, outra pesquisa apontou que apesar da percepção do indivíduo ser um ponto central da QV, as medidas objetivas de QV acabam sendo mais sensíveis às intervenções pelos

tratamentos, porque muitos destes não têm o objetivo de melhorar os componentes subjetivos da QV (Ruggeri, Bisoffi *et al.*, 2001).

A falta de consenso em relação a qual melhor medida e qual melhor conceito que deve ser utilizado na pesquisa de QV em pacientes psiquiátricos também é realidade quando se estuda pacientes deprimidos. Dentre os instrumentos genéricos de QV o SF-12 (12-Item Short Form Health Survey), e o Q-LES-Q (Quality of Life Enjoyment and Satisfaction Questionnaire Measures of quality of life) foram comparados e concluiu-se que o primeiro media a percepção do paciente com sua funcionalidade e o segundo media a satisfação dos pacientes com os aspectos mencionados em seus itens (Wisniewski, Rush *et al.*, 2007).

Existe uma carência de modelos teóricos específicos para a aferição da QV em pacientes deprimidos. Leval (De Leval, 1999) sugere que a teoria das três dimensões baseada no horizonte individual do paciente deprimido, em que o presente representa a doença, o passado a sua vida saudável e o futuro a expectativa em relação ao tratamento, pode servir como um ponto de partida para a construção de um modelo teórico de QV neste grupo de pacientes.

Todos os estudos que avaliaram a QV de pacientes deprimidos concluíram que a QV deles está severamente comprometida (Blumenthal e Dielman, 1975; Bruce, Seeman *et al.*, 1994; Teasdale, Taylor *et al.*, 1995; Atkinson, Zibin *et al.*, 1997; Skevington, 1998; Demyttenaere, De Fruyt *et al.*, 2000; Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Naumann e Byrne, 2004; Papakostas, Petersen *et al.*, 2004; Kuehner e Buerger, 2005; Tylee e Gandhi, 2005; Ay-Woan, Sarah *et al.*, 2006; Diehr, Derleth *et al.*, 2006; Trompenaars, Masthoff *et al.*, 2006), mesmo quando comparados com outras doenças clínicas (Wells, Stewart *et al.*, 1989; Bonicatto, Dew *et al.*, 2001).

Quando avaliada em pacientes deprimidos, a medida QV é questionada pela possibilidade de haver uma sobreposição dos construtos de QV e depressão (Aigner, Forster-Streffleur *et al.*, 2006), podendo estar influenciada pelo viés afetivo (Heinonen, Aro *et al.*, 2004), pelo pobre insight e por eventos de vida recentes (Atkinson, Zibin *et al.*, 1997). Kuehner e Buerger (Kuehner e Buerger, 2005) sugerem que os prejuízos nos domínios psicológico e social da QV em pacientes deprimidos podem ter uma contribuição da baixa auto-estima, estilo de resposta ao humor deprimido e fraco apoio social e não à contribuição da severidade de sintomas. Com isso, a terapia deveria ser focada não apenas em diminuir sintomas, mas em ajudar o paciente a manter e estabelecer relacionamentos apoiadores, melhorar a sua auto-apreciação e desenvolver habilidades de lidar com o humor negativo.

Mesmo diante das limitações apontadas, a avaliação da QV tem sido utilizada como medida de desfecho de uma série de estudos que avaliam a resposta aos tratamentos para depressão maior (Papakostas, Petersen *et al.*, 2004). Existem estudos com o uso dos mais diversos medicamentos (Skevington e Wright, 2001) e outras terapias. Como exemplos, podemos enumerar aqueles que avaliaram: o uso da sertralina em comparação com imipramina para depressões não-melancólicas (Baca, Gonzalez-De-Chavez *et al.*, 2003); a segurança e efetividade da venlafaxina de liberação lenta em pacientes deprimidos no contexto de atenção primária (Baldomero e Enguix, 2003; Cervera-Enguix, Soutullo *et al.*, 2003; Kroenke, Messina *et al.*, 2006); o uso de escitalopram em pacientes deprimidos ambulatoriais (Burke, Gergel *et al.*, 2002); a estratégia de potencialização da mirtazapina (Carpenter, Yasmin *et al.*, 2002); a efetividade da venlafaxina de liberação lenta em comparação com o escitalopram (Fernandez, Montgomery *et al.*, 2005; Montgomery e Andersen, 2006) e com sertralina (Sir, D'souza *et al.*, 2005; Shelton, Haman *et al.*, 2006); a



eficácia da fluvoxamina em uma amostra de pacientes chineses (Gu, Cai *et al.*, 2004); a eficácia do uso de duloxetina na prevenção de recaída (Perahia, Gilaberte *et al.*, 2006), a efetividade da paroxetina em pequenas doses de liberação controlada (Trivedi, Pigott *et al.*, 2004); e a efetividade da terapia cognitiva tradicional comparada com a terapia feita de maneira compromissada (“Acceptance and commitment therapy”) e a privação parcial de sono combinada com a sertralina (Caliyurt e Guducu, 2005).

Um exemplo da importância da avaliação de QV como resultado do tratamento antidepressivo foi o estudo norte-americano STAR\*D (Sequenced Treatment Alternatives to Relieve Depression) (Rush, Fava *et al.*, 2004) que recrutou mais de 4000 pacientes ambulatoriais oriundos de cuidados primários e psiquiátricos de serviços públicos e privados com depressão maior não psicótica que utilizou entre a suas medidas de desfecho a avaliação de QV pelo Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) e pelo 12-item Short Form Health Survey (SF-12). Este estudo teve como objetivo comparar a efetividade do citalopram (Trivedi, M.-H., Rush, A. J. *et al.*, 2006), mais sete diferentes tratamentos incluindo quatro opções de mudança (venlafaxina, bupropiona, sertralina e terapia cognitivo-comportamental) e três opções de aumento de resposta (terapia cognitivo comportamental, bupropiona ou buspirona adicionada a citalopram) (Fava, Rush *et al.*, 2003; Rush, Fava *et al.*, 2004). Seus resultados sugeriram que a medida de QV juntamente com a avaliação dos sintomas é a maneira mais adequada de acessar a verdadeira severidade da depressão maior, demonstrada pela ampliação do entendimento da situação do paciente a partir da avaliação multidimensional oferecida por estas medidas (Trivedi, M. H., Rush, A. J. *et al.*, 2006).

Não encontramos nenhum estudo de validação do WHOQOL-Bref em pacientes deprimidos provenientes do contexto primário de atenção, seja usando a psicometria clássica, seja usando a análise de Rasch e/ou Teoria da resposta ao Item, seja integrando as duas metodologias.

A revisão da literatura foi realizada no período no Medline/Psych Info de 1997-março 2008 usando os termos: “quality of life” (QOL), “health status”, “well being”, “Item response theory” (IRT), “Rasch analysis”, “validity”, “depression”, “major depression”, “primary care”. O processo de revisão da literatura está sumarizado na figura 1.

Dos 343 artigos encontrados, 32 foram excluídos por não tratarem de QV, 88 por versarem sobre instrumentos específicos de QV, 20 por utilizarem crianças/adolescentes/idosos como sujeitos de pesquisa, 108 por tratarem de outras condições clínicas que não a depressão, 42 usaram instrumento diferente do WHOQOL-Bref ou WHOQOL-100, 5 foram escritas em línguas que não o inglês e o português (italiano, japonês, russo, alemão e húngaro), 11 foram realizados em países outros que não os estudados por nós e, finalmente, 10 tinham outros transtornos psiquiátricos. Portanto, chegamos a um total de 26 estudos neste momento. Na presente revisão, não encontramos nenhum estudo de validação desta medida em pacientes deprimidos no contexto de atenção primária, seja usando a psicometria clássica, seja usando a análise de Rasch e/ou Teoria da resposta ao Item, seja integrando as duas metodologias.

Uma grande parte destes se referia à avaliação de QV e depressão em outras condições clínicas. Uma grande parte dos estudos que avaliaram a QV em pacientes deprimidos foi realizada em amostras brasileiras. Um estudo recente demonstrou que o tratamento com antidepressivos por 12 semanas melhorou todos os domínios

da QV de pacientes severamente deprimidos (Berlim, Pargendler *et al.*, 2007). Outro sugeriu que os sintomas depressivos mesmo que em nível subsindrômico podem afetar negativamente a QV e que a severidade da depressão é fator com maior impacto na QV (Da-Silva-Lima e De-Almeida-Fleck, 2007). O estudo de validação e de confiabilidade do WHOQOL-Bref em pacientes deprimidos ambulatoriais demonstrou boas propriedades psicométricas desta medida neste contexto (Berlim, Pavanello *et al.*, 2005).

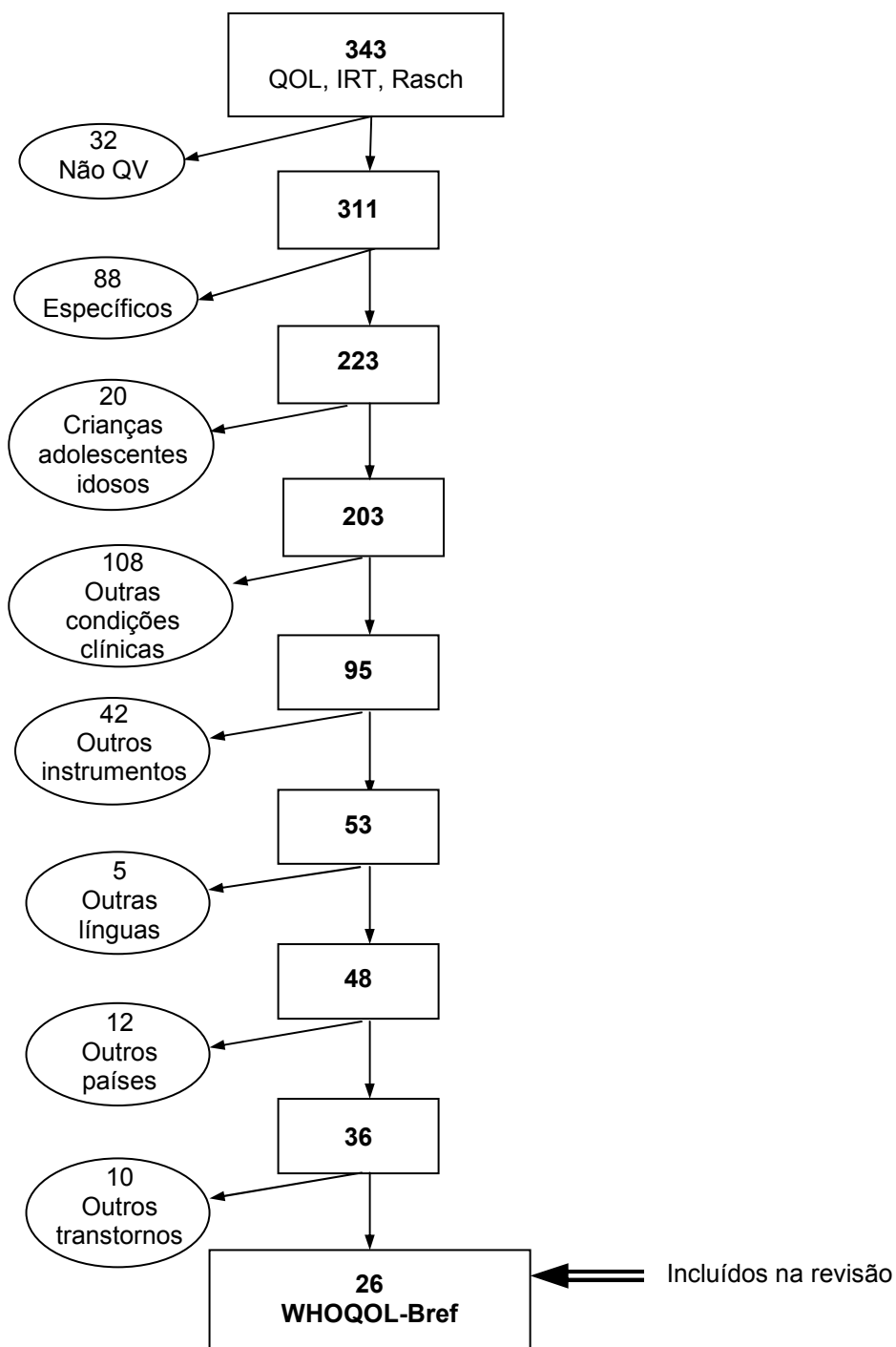
Utilizando outras amostras diferentes das brasileiras, os escores de QV também podem estar associados negativamente com a presença de psicopatologia e sua severidade e a presença de transtorno de personalidade (Masthoff, Trompenaars *et al.*, 2006).

Os efeitos de uma política de desinstitucionalização na Itália foram avaliadas pelo uso do WHOQOL-Bref (Picardi, Rucci *et al.*, 2006). Os pacientes que residiam em moradias asilares tiveram escores semelhantes aos pacientes esquizofrênicos ambulatoriais e escores menores que indivíduos normais. Os escores mais baixos de QV estavam associados com esquizofrenia, depressão unipolar, psicoses não afetivas, ansiedade ou transtornos somatoformes, bem como com pior saúde física e menor duração da doença e nenhuma participação em atividades internas. Neste mesmo estudo, pode-se identificar que o domínio da QV em que a diferença era mais marcada entre os sujeitos hígidos e os moradores foi o domínio social. Este achado será importante para direcionar as ações de ressocialização deste grupo de pacientes.

Um estudo realizado em Taiwan testou modelos preditivos de QV em pacientes deprimidos (Ay-Woan, Sarah *et al.*, 2006), levando em consideração as variáveis clínicas e demográficas e a competência ocupacional percebida. Dentre as

variáveis preditoras de QV, destacaram-se a satisfação com a performance ocupacional, a severidade da depressão, as atividades de vida diária e a idade. Os dados apontam para a importância de se realizar algum tipo de terapia ocupacional para melhoria da QV de pacientes deprimidos.

**Figura 1: Artigos pesquisados no Medline na primeira busca por Quality of life, Item Response Theory, Rasch analysis, validity, depression, WHOQOL**



## 2.2 O Longitudinal Investigation of Depression Outcomes

O Longitudinal Investigation of Depression Outcomes (LIDO) foi um estudo no qual participaram seis países – Brasil, Austrália, Israel, Espanha, Estados Unidos e Rússia- e teve por objetivo geral verificar a associação entre QV e desfechos econômicos em pacientes deprimidos oriundos de serviços de cuidados primários acompanhados por um período de um ano (Chisholm, Amir *et al.*, 2001).

Foi desenhado como um estudo transcultural para examinar a prevalência e os correlatos socioeconômicos de Transtornos Depressivos no contexto de cuidados primários. Avaliou também o impacto econômico na QV associado à depressão, os padrões atuais para tratamento de depressão, a associação destes tratamentos e resultados clínicos, bem como a validade transcultural e aplicabilidade das medidas de QV. A metodologia utilizada para obter esta amostra de pesquisa, bem como a comparação entre os centros envolvidos, características de seus serviços e populações incluídas, já foram extensamente descritos por pelos seus investigadores principais (Chisholm, Amir *et al.*, 2001; Herrman, Patrick *et al.*, 2002).

Algumas características da metodologia utilizada no LIDO serão resumidas a seguir, bem como na seção de método dos artigos.

### 1ª) Características das amostras:

Pacientes que estavam sendo atendidos nos serviços de cuidados primários dos centros envolvidos foram convidados a participar do estudo e avaliados para rastreamento de Transtorno Depressivo Maior através do uso da escala CES-D. O critério de elegibilidade para o segundo estágio de avaliação diagnóstica incluiu ter

um escore na escala CES-D  $\geq 16$ , não receber tratamento para depressão nos últimos três meses e não ter planos de se mudar nos próximos 12 meses. Utilizou-se, além de outros instrumentos (Tabela 2), o Composite International Diagnostic Interview (CIDI), que é uma entrevista psiquiátrica estruturada que possibilita o diagnóstico de depressão pelos critérios do DSM-IV (Weiller, Lecrubier *et al.*, 1994). Todos os pacientes que preencheram critérios para depressão tanto pelo CIDI quanto pelo CES-D na avaliação inicial foram convidados a participar das seguintes em três, nove e 12 meses. A avaliação de depressão aos nove meses foi novamente composta pelo diagnóstico pelo CIDI e os escores no CES-D. Os pacientes que mantiveram o diagnóstico de depressão pelo CIDI foram considerados como tendo depressão clínica persistente. Aqueles que não apresentavam nem o diagnóstico pelo CIDI e CES-D  $< 16$  foram considerados em remissão. Os demais pacientes que não tinham o diagnóstico pelo CIDI, mas tinham CES-D  $\geq 16$ , foram considerados em remissão parcial.

O primeiro centro iniciou o rastreamento em fevereiro de 1999 e o último centro que completou a inclusão de pacientes o fez em agosto de 1999. Os centros precisaram entre doze e dezesseis meses para completar a etapa de rastreamento para atingir o número suficiente de pacientes a serem incluídos.

2ª) Descrição dos centros e dos serviços de cuidados primários incluídos no estudo:

Os serviços onde foram coletados os dados foram escolhidos pelos pesquisadores principais de cada centro envolvido. Todos os centros participaram do teste de campo para o desenvolvimento do WHOQOL. A seleção dos centros se

baseou na distribuição global, limitando-se a dois centros com língua inglesa e evitando-se a concentração de centros na Europa ocidental.

A seguir apresentamos a descrição dos centros e serviços com as características da época do estudo.

#### 1) Barcelona, Espanha

Barcelona é a cidade principal da região autônoma da Catalunha e tinha em 1998 uma população de 1,7 milhões de habitantes. O tamanho da população tem se mantido estável por 25 anos, mas proporção de idosos tem crescido. Quarenta por cento da população trabalha em turno integral e a taxa de desemprego é de 10%. Três serviços urbanos de cuidados primários fizeram parte deste estudo, cada um cobrindo a população de 25 mil habitantes que atende em média 7800 pacientes por mês. A equipe da atenção primária era composta por enfermeiros, assistentes sociais, funcionários administrativos e médicos que atendem em média 520 pacientes por mês. Médicos e enfermeiros realizam visitas domiciliares uma vez por semana. Dentre as causas das visitas: 25% são devido a doenças e traumas agudos, 60% devido a doenças crônicas e 10-15% devido a problemas psicológicos.

#### 2) Be'er Sheva, Israel

Be'er Sheva tem uma população de 440 mil habitantes que consiste predominantemente de jovens adultos (16-45 anos) e crianças, o que reflete o baixo nível sócio-econômico e alto número de unidades familiares desta região de Israel. A taxa de desemprego é de 9,5%. Centros com diferentes tamanhos participaram do



estudo. Dimona foi o centro onde foi recrutada a maioria dos pacientes, neste são atendidos aproximadamente 5600 pacientes por mês, numa razão de 1: 800 médicos/pacientes. Destes pacientes, 60% apresentam doenças crônicas e 10% sofrem de doenças psicológicas.

### 3) Melbourne, Austrália

O local onde os pacientes foram recrutados corresponde a Melbourne metropolitana que apresenta aproximadamente uma população de 3 milhões de habitantes. É caracterizada pela diversidade cultural, um quarto de seus habitantes nasceu em um país que não fala inglês. A economia local é orientada para realização de serviços, a taxa de desemprego é de 10%. North Yarra Community centre foi o principal centro envolvido no estudo e atende 6500 pacientes por mês. Tem 5 médicos, que atendem em média 800 pacientes por mês com a colaboração de 3 enfermeiros. Os problemas psicológicos correspondem a 20% dos atendimentos.

### 4) Porto Alegre, Brasil

O local onde foram recrutados os pacientes corresponde a zona norte da capital do Estado do Rio Grande do Sul. Este local tem uma população de 100 mil habitantes, sendo 68% adultos jovens (16-45 anos) e crianças, 35% da população ganha menos de 200 dólares por mês, 10% está desempregado. O serviço de cuidados primários que participou do estudo era composto por 13 unidades que incluiu uma unidade de 30 leitos que servem 120 mil pessoas. Este serviço conta

com 44 médicos e 11 enfermeiros que atendem em média 17400 pacientes por mês, numa razão 1:395 médico/pacientes. Os problemas psicológicos correspondem a 12% da demanda.

#### 5) Seattle, Estados Unidos

O recrutamento de pacientes foi realizado na área metropolitana da cidade que apresenta uma população de aproximadamente 1,7 milhões de habitantes. É uma cidade com uma renda per capita relativamente alta e tem uma baixa taxa de desemprego (4%). Os atendimentos mensais variaram entre 500 a 1200 pacientes numa razão médico/paciente de 1:350. A maioria (70%) dos atendimentos foi atribuída às doenças crônicas e os restantes, a problemas agudos e psicológicos respectivamente (15% cada).

#### 6) St. Petersburg, Rússia

É uma cidade com 4,7 milhões de residentes com 70% da população entre 16-65 anos. A sua taxa de mortalidade é alta (65/100 000 habitantes com idades abaixo de 65 anos). O desemprego aumentou muito nos últimos anos, apesar das estatísticas oficiais apresentarem a taxa de 2%. As clínicas de onde foram recrutados os pacientes servem a aproximadamente 40 mil residentes e atendem uma média de 28 mil pacientes por mês. A razão médico/paciente é de 1:1270. Os atendimentos atribuíveis a causas psicológicas representam apenas 4% do total.

### 2.2.1 Definição das variáveis de tratamento

O número de medicamentos prescritos e/ou o número de consultas ou visitas médicas ou a outros profissionais foram os dois principais indicadores utilizados para descrever a natureza dos cuidados recebidos para a depressão. Seguindo a sugestão de Bellantuono, Arreghini et al (Bellantuono, Arreghini *et al.*, 1989) e Wells e Sherbourne (Wells e Sherbourne, 1999) o tratamento antidepressivo apropriado (derivado das informações registradas nos prontuários médicos) foi definido como sendo uma dose diária terapêutica: qualquer dia nos 30 dias prévios, ou uso diário por pelo menos um mês no período de seis meses (Tabela 2). Aconselhamento foi definido como o tempo de contato com o provedor durante o período de doze meses de observação. Por fim, um descritor adicional de cuidado foi designado para quem recebeu benzodiazepínicos.

**Tabela 1: Critérios de doses de antidepressivos potencialmete adequadas para tratamento de depressão no contexto primário do LIDO\*#**

Antidepressivo	Dose mínima (mg/dia)	Antidepressivo	Dose mínima (mg/dia)
Amitripitilina	75	Maprotilina	100
Amoxapina	100	Mianserina	60
Bupropiona	200	Moclobemida	300
Citalopram	20	Nefazodone	150
Clomipramina	75	Nortriptilina	40
Desipramina	75	Paroxetina	20
Dotiepina	75	Protiptilina	20
Doxepina	75	Sertralina	50
Fenelzina	45	Tianeptina	400
Fluoxetina	10	Tranilcipromina	20
Fluvoxamina	100	Trazodone	150
Imipramina	75	Trimipramina	75
Isocarboxazida	30	Venlafaxina	100
Lítio	600	Viloxazine	300
Lofepramida	140		

\*Doses foram baseadas na menor dose recomendada pelo protocolo da Agência Norte-americana para políticas de cuidados de saúde e pesquisa para tratamento de depressão na atenção primária e nas estimativas dos investigadores.

#Extraído com permissão de Simon, G. E., M. Fleck, *et al.* Prevalence and predictors of depression treatment in an international primary care study. *Am.J.Psychiatry*, v.161, n.9, p.1626-1634. 2004.

### 2.2.2 *Definição das categorias ou tipos de serviços*

Foram definidas três principais categorias de serviços: serviços de cuidados primários ambulatoriais que cobriam o maior contato com os profissionais de cuidados primários e de saúde mental, serviços de cuidados diários, oferecidos para os pacientes severamente enfermos e geralmente combinando o tratamento para os problemas relacionados com o transtorno mental, e, finalmente, serviços de internação hospitalar, que incorporava as admissões médicas e psiquiátricas. Esta definição serviu de base para posteriormente determinar a utilização dos serviços.

### 2.2.3 *Cálculo de tamanho das amostras*

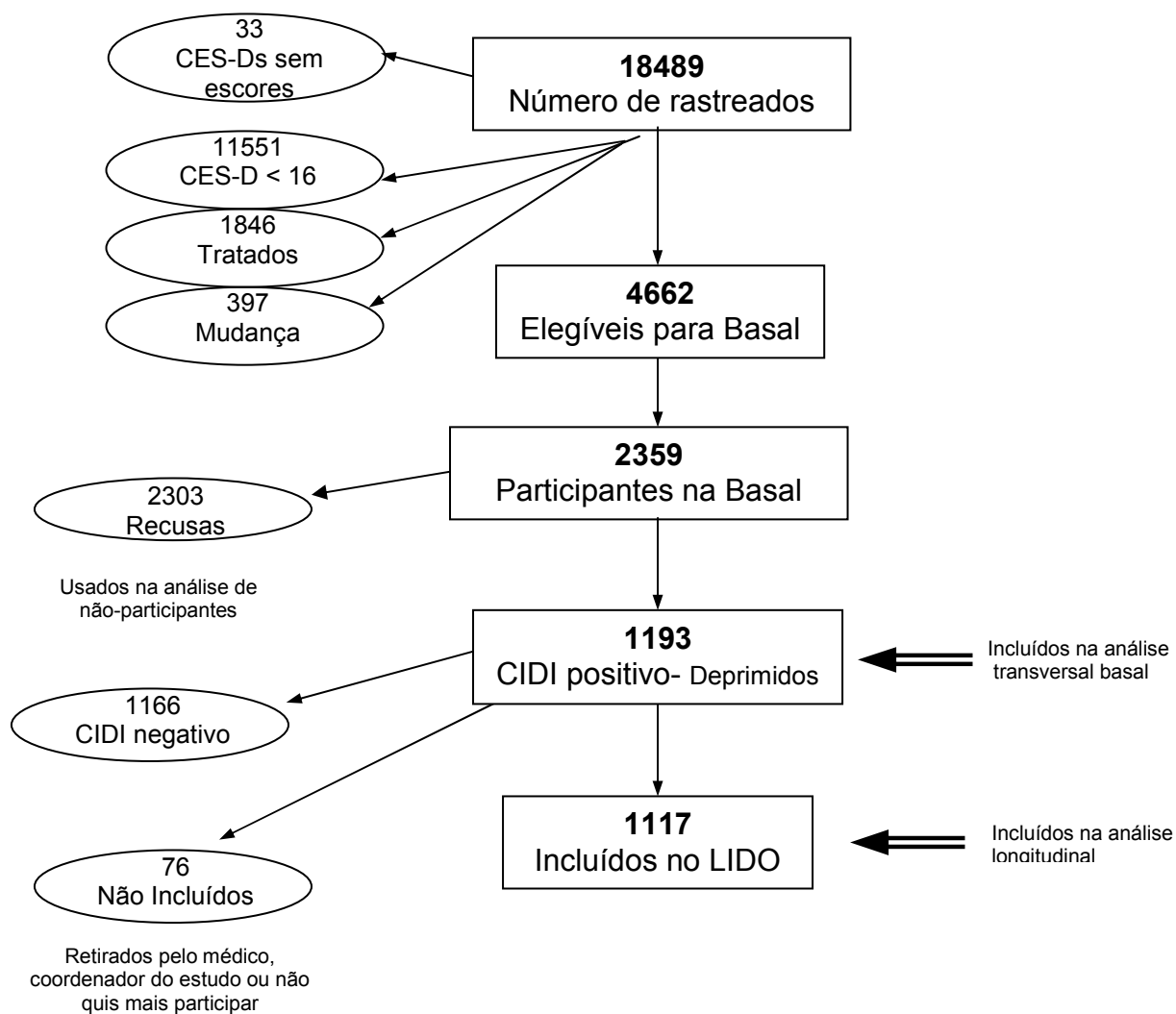
Cada centro teve como objetivo recrutar 150 pacientes (incluindo um mínimo de 50 homens), para determinar diferenças mínimas que poderiam ser detectadas com 80% de poder para as principais hipóteses do estudo. O cálculo foi baseado nos dados do Cardiovascular Health Study (Diehr, Patrick *et al.*, 1995), que coletou dados anuais de CES-D em idosos. Foi estimado, então, que aproximadamente 4288 indivíduos teriam que ser rastreados em cada serviço de cuidado primário, para poder identificar e incluir um número suficiente de pacientes deprimidos e assegurar que 150 participantes completariam o protocolo do estudo.

## Quadro 2: Instrumentos e padrão de utilização das medidas no LIDO

Medida	Objetivo	Quando foi Administrada
CES-D (Lyness, Noel <i>et al.</i> , 1997)	Medida de Rastreamento de Depressão; Indicador de severidade baseado em sintomas	Todas as avaliações
CIDI (Weiller, Lecrubier <i>et al.</i> , 1994)	Medida de diagnóstico de depressão	Basal; Mês 9
WHOQOL-Bref (WHOQOL Group, 1998a; Skevington, Lofy <i>et al.</i> , 2004)	Medida Genérica de QV	Basal; Semana 6; Mês 3, 9
QLDS (Hunt e Mckenna, 1992; Mckenna e Hunt, 1992)	Medida de QV específica para depressão	Basal; Semana 6; Meses 3, 9
Lista de condições comórbidas	Indicadores de condições comórbidas	Basal; Mês 12
Questões demográficas	Descrição, Controle de confundidores e variáveis econômicas	Rastreamento; Basal; Meses 3, 9, 12
Questões sobre utilização de recursos	Descrição, Controle de confundidores e variáveis econômicas	Rastreamento; Basal; Meses 3, 9, 12

LIDO: Longitudinal Investigation of Depression Outcomes

**Figura 2: Fluxograma descritivo dos pacientes recrutados para participar do Longitudinal Investigation of Depression Outcomes**



#### 2.2.4 Características gerais da amostra basal

A amostra de rastreamento foi composta de 18.489 indivíduos, sendo que 37% destes tiveram escores na CES-D acima de 16, perfazendo um total de 4.662 e 28% acima de 20. Do universo de pacientes elegíveis, 2.359 (51%) completaram avaliação diagnóstica, dos quais 1.193 pacientes (também 51%) apresentaram-se com depressão clínica atual, preenchendo o critério A para depressão pelo DSM (Figura 2). Comparados com os que seguiram no estudo, aqueles que não participaram eram levemente mais jovens (média de idade= 40,5 vs. 41,6 anos,  $P=0,022$ ), menos freqüentemente eram do sexo feminino -65% vs. 68,6%,  $P=0.01$ ), tinham escores menores na CES-D (24,7 vs. 26,4,  $P=0,01$ ) e tinham um escore maior no componente físico do SF-12 (43,7 vs. 42,6,  $P<0,01$ ). Como se esperava, a recuperação da depressão esteve associada com escores menores de depressão na avaliação basal (Simon, Chisholm *et al.*, 2002).

A idade média dos sujeitos foi de 40,3 (DP 14,5) anos e 11,6 (DP 3,5) anos de estudo. Houve diferenças entre os centros tanto na idade, quanto nos anos de estudo, variando de 37,9 a 46,8 na idade e 9,0 a 13,2 nos anos de estudo, sendo os maiores valores reportados em St. Petersburgo. Houve uma grande semelhança no percentual de mulheres, variando de 65% a 76% em cada centro e 70,3% do total. A proporção de sujeitos casados variou de 25,7% em Melbourne a 69,4% em Be'er Sheva, com média total de 47.2%. O percentual de desemprego variou de 6,3% em Barcelona a 25,7% em Melbourne.



#### 2.2.4.1 Uso de serviços de saúde entre os centros

As taxas de utilização de serviços de saúde na avaliação basal descritas a seguir referem-se aos 3 meses anteriores ao início do estudo. O número médio de visitas a atendimento de atenção primária e ambulatorial variou de 5,6 (DP 6,3) em Barcelona a 11,8 (DP 14,4) em Seattle, com uma média total de 8,5 (DP 10,0) visitas. Já o número médio de atendimento em hospital-dia ou grupo de apoio foi menor que 1 em todos os centros. O único centro onde o número de dias de internação hospitalar foi maior que 1 por sujeito foi em St. Petersburg (2,6 dias, DP 7.8). Estes achados demonstram que a identificação dos pacientes com diagnóstico de depressão depende dos serviços de cuidados primários, apesar da não especialização do atendimento e fraco encaminhamento para atendimento secundário.

#### 2.2.4.2 Barreiras ao acesso aos serviços

Os principais fatores identificados pelos sujeitos foram o custo dos cuidados (acima de 20% em todos os centros, 76% em St. Petersburg), sentir-se constrangido/envergonhado (20% ou mais em Be'er Sheva, Melbourne, Seattle, St. Petersburg) e preocupações com os efeitos adversos da medicação (variou entre 18-44%). Com um nível menor de preocupação, apareceram o risco para seu trabalho, a inconveniência de chegar até os centros de tratamento e o desencorajamento da família para procurar tratamento.

### 2.2.5 Estudos publicados internacionalmente utilizando as amostras do LIDO

Até o presente momento foram publicados 10 estudos sendo que 8 deles utilizaram amostras internacionais e 2 utilizaram somente a amostra brasileira. Os pesquisadores do LIDO foram divididos em grupos para análise dos dados: grupo coordenador, grupo de análise econômica, grupo de depressão, grupo de QV e centros. Segundo esta divisão, o grupo de análise econômica e de depressão publicaram 3 artigos cada um, e os grupos coordenador e do Brasil também publicaram 2 artigos cada.

O primeiro estudo publicado foi o conduzido por (Chisholm, Amir *et al.*, 2001) utilizou a amostra basal (n=2359) e centrou-se na descrição sócio-demográfica dos centros, das amostras, dos serviços e na utilização dos mesmos. Comparação dos centros demonstrou que há uma variabilidade em termos de gastos e provimento dos serviços, bem como dos modelos de financiamento. Houve diferenças esperadas nas variáveis demográficas e no padrão de utilização dos serviços. Concluiu-se que a pesquisa multinacional de serviços é complicada pela heterogeneidade dos sistemas de serviços que refletem a realidade dos cuidados.

Já o segundo estudo publicado, utilizou a amostra de rastreamento (n=18489) e teve como achado principal que pacientes com escores de depressão mais altos têm pior saúde, estado funcional, QV e maior utilização de serviços. Dentre aqueles com CES-D $\geq$ 16, os que tiveram tratamento para depressão relataram mais frequentemente insatisfação com a saúde e tiveram escore mais altos de depressão (Herrman, Patrick *et al.*, 2002). Estes achados são consistentes com outros estudos que identificaram uma associação significativa entre depressão e incapacidade

social (Ormel, Von Korff *et al.*, 1993). O estudo que utilizou a amostra brasileira de rastreamento encontrou resultados muito semelhantes aos do estudo internacional (Fleck, Lima *et al.*, 2002). A aplicabilidade transcultural das associações encontradas já tinha também sido evidenciada no World Health Organization International Study on Psychological Problems in General Health Care (Ormel, Vonkorff *et al.*, 1994; Ormel, Vonkorff *et al.*, 1999).

O terceiro estudo publicado teve como objetivo a avaliação dos desfechos em depressão e seus correlatos econômicos (Simon, Chisholm *et al.*, 2002). Utilizando a amostra que foi acompanhada por 12 meses, identificou que depois de 9 meses de acompanhamento aproximadamente um terço dos pacientes ainda mantinham o diagnóstico de depressão e outro um terço ainda tinha sintomas residuais. Entre aqueles que persistiram deprimidos, o custo dos serviços de saúde se manteve o mesmo que na avaliação basal. Já entre aqueles que tiveram remissão completa, os custos diminuíram pela metade, o que mostra a relevância de se investir no tratamento destes pacientes.

O quarto estudo internacional (Bech, Lucas *et al.*, 2003) que teve como objetivo avaliar associação entre o subgrupo de depressão, tipo de tratamento e a retenção no estudo ao longo de 12 meses. Não houve diferença estatisticamente significativa na associação entre os três diferentes subgrupos e a taxa de saída do estudo, mas os pacientes mais velhos eram mais propensos a abandonarem, mas se estivessem usando antidepressivos eram menos propensos ao abandono. Encontrou-se uma relação inversa entre estar em uso de antidepressivo e sua propensão a abandonar o estudo. Dentre os três diferentes subgrupos, os pacientes com depressão moderada e comorbidade com condições médicas severas apresentaram a pior QV, a menor taxa de tratamento com antidepressivo e a maior

insatisfação com sua saúde. A taxa de abandono variou entre os centros de 6% em São Petersburgo a 28,1% em Melbourne. São Petersburgo e Be'er Sheva tiveram o menor percentual de tratamento de qualquer tipo, e também o menor percentual de tratamento com antidepressivos. Estes dois centros também tiveram os menores níveis de depressão na avaliação basal medida pela CES-D. Melbourne apresentou a maior quantidade de visitas de aconselhamento, e São Petersburgo a menor. Concluiu-se que o subgrupo de pacientes com depressão moderada e comorbidades médicas é o que deve merecer maior atenção visto que apresenta a menor taxa de tratamento com antidepressivo e pior QV e insatisfação com sua saúde quando comparados com pacientes com mesmo nível de depressão.

O quinto estudo internacional (Chisholm, Diehr *et al.*, 2003) utilizou a amostra basal (n=2359) para explorar a associação entre estado depressivo, perda de dias de trabalho e custo da utilização dos cuidados de saúde. As taxas de contato com os serviços entre os centros foram de 5-14% para visitas dos profissionais de saúde mental, 94-100% para visitas médicas ou de cuidados primários e de 5-18% para admissões hospitalares. A incidência de dias ausentes ao trabalho variou de 20% em Porto Alegre a 55% em Seattle, onde a média de dias perdidos nos 3 meses prévios foi de 1,4 dias em Porto Alegre a 7,6 em São Petersburgo. Na maioria dos centros o custo dos dias de trabalho perdidos foi menor que o custo total dos cuidados de saúde, mas representou 15-40% do custo total combinando dias de trabalho perdidos e custos com a saúde. Controlando para o efeito das variáveis sócio-demográficas e características clinicamente relevantes, a comorbidade médica esteve associada com 17-46% do aumento dos custos com os cuidados da saúde na maioria dos centros. Utilizando a equivalência poder de compra para converter em dólares americanos o consumo dos serviços de saúde por cada sujeito revelou

que o peso econômico de uma depressão não tratada no contexto primário ou se aproxima ou excede os gastos médios per capita com cuidados de saúde.

O sexto estudo internacional (Simon, Fleck *et al.*, 2004) que utilizou a amostra de seguimento (n=1117) teve como objetivo avaliar os preditores e a prevalência de tratamento para depressão. A proporção de tratamento não excedeu 40% em nenhum dos centros. Dentre os pacientes com depressão persistente ao longo dos 9 meses de acompanhamento, o percentual de qualquer tratamento antidepressivo variou de 37% em Seattle a 0% em São Petersburgo. A preocupação quanto aos custos do tratamento foi a maior barreira ao tratamento, seguido das preocupações com os efeitos adversos da medicação. Já a severidade da depressão esteve associada com tratamento adequado em apenas um dos quatro centros pesquisados. Concluiu-se que as baixas taxas de tratamento relacionadas às barreiras financeiras podem ser controladas com a adoção de políticas de saúde específicas.

O sétimo estudo publicado que utilizou a amostra de seguimento (n=968) (De Almeida Fleck, Simon *et al.*, 2005) teve como objetivo definir que fatores da avaliação basal estavam associados com os desfechos clínicos nos nove meses nos seis países envolvidos no estudo. Para todas as variáveis com diferenças estatisticamente significativas, o grupo que atingiu remissão completa teve melhor desempenho na avaliação basal. Este grupo também apresentou menores níveis de depressão medidos pela CES-D, e melhor QV e funcionamento, medido pela QLDS, pelo WHOQOL e pelo SF-12 respectivamente, bem como teve menos episódios prévios de depressão. Concluiu-se que a depressão e QV podem ser diferentes facetas de um mesmo constructo, sendo este foi o primeiro estudo a identificar a avaliação da QV como preditor de remissão para depressão.

Por último, foi publicado um artigo utilizando a amostra de seguimento (n=968) (Diehr, Derleth *et al.*, 2006) que teve como objetivo determinar se as medidas de estado físico, mental e QV mudavam ao longo do tempo nas pessoas que estavam deprimidas na avaliação basal, e se esta mudança era sincrônica. Todas as medidas melhoraram significativamente ao longo do tempo, sendo que o estado mental foi o que mais melhorou e o físico o que menos, a QV teve uma melhora intermediária. Este achado está em acordo com outros estudos (Ormel, Von Korff *et al.*, 1993; Skevington e Wright, 2001) que também demonstraram uma sincronia entre as mudança nos sintomas depressivos e as medidas de QV e estado de saúde.

Mais recentemente, o centro brasileiro publicou um artigo (Da Silva Lima e De Almeida Fleck, 2007) que teve como objetivo avaliar a associação entre QV e depressão subsindrômica. Trinta e cinco por cento dos pacientes apresentaram depressão subsindrômica. Os achados sugerem que a depressão subsindrômica causa um impacto na QV de pacientes deprimidos provenientes de serviço de cuidados primários no Brasil.

Embora tenha sido planejada como objetivo secundário do LIDO, a avaliação das propriedades psicométricas dos instrumentos utilizados neste estudo, não tinha até então sido realizada.

Os estudos que compõem esta tese utilizaram para análise a amostra total de participantes da avaliação basal, incluindo os pacientes com CIDI negativo (n=2359) e a amostra utilizada para análise transversal, excluindo os pacientes com CIDI negativo (n=1193).

O quadro a seguir resume os principais achados das publicações realizadas a partir do LIDO.

### Quadro 3: Resumo dos achados principais dos estudos publicados internacionalmente utilizando os dados do LIDO

AUTORES	TÍTULO	OBJETIVOS	GRUPO DE ANÁLISE/AMOSTRA	RESULTADOS/CONCLUSÕES
da Silva Lima AF, de Almeida Fleck MP. (2007)	Subsyndromal depression: an impact on quality of life?	Demonstrar a associação entre QOL e depressão subsindrômica em uma amostra de pacientes de cuidados primários do Brasil.	Centro Brasileiro Amostra Basal Brasileira	35.2% apresentam depressão subsindrômica que provoca prejuízos na QV.
Diehr PH, Derleth AM, McKenna SP, Martin ML, Bushnell DM, Simon G, Patrick DL. (2006)	Synchrony of change in depressive symptoms, health status, and quality of life in persons with clinical depression.	Determinar se as medidas de estado físico e mental e qualidade de vida mudam ao longo do tempo nos pacientes que estavam clinicamente deprimidos na avaliação basal e determinar se esta mudança tem sincronia com as outras medidas.	Grupo coordenador Amostra de seguimento	Status mental foi medida com maior mudança, seguido da QV e status físico. Nos deprimidos, as medidas de saúde mental e física, assim como QV mostraram associações longitudinais sincrônicas com as medidas de sintomas depressivos.
de Almeida Fleck MP, Simon G, Herman H, Bushnell D, Martin M, Patrick D; Longitudinal Investigation of Depression Outcomes Group. (2005)	Major depression and its correlates in primary care settings in six countries. 9-month follow-up study.	Definir que fatores da avaliação basal estavam associados com os desfechos clínicos em nove meses nos seis países.	Grupo de depressão Amostra de seguimento	25% a 48% apresentaram remissão completa. Nível educacional, eventos de vida e escore na QLDS foram preditores moderados de remissão completa.
Simon GE, Fleck M, Lucas R, Bushnell DM (2004)	Prevalence and predictors of depression treatment in an international primary care study.	Avaliar a prevalência e preditores de tratamento para depressão nos diferentes centros.	Grupo de depressão Amostra de seguimento	Entre 38 a 0 % receberam tratamento antidepressivo Problemas financeiros foram a maior barreira ao tratamento.  O rastreamento e a notificação ao medico não é suficiente para promover o tratamento. A probabilidade de receber tratamento está mais influenciada por características dos serviços do que sintomas clínicos.
Bech P, Lucas R, Amir M, Bushnell D, Martin M, Buesching D; LIDO Group. (2003)	Association between clinically depressed subgroups, type of treatment and patient retention in the LIDO study.	Avaliar se existe uma associação entre severidade da depressão e retenção no estudo e redução de sintomas.  Avaliar a repercussão da comorbidade de doenças físicas nesta associação.	Grupo de depressão Amostra de seguimento	Não houve diferença de abandono.  Pacientes com depressão moderada e com comorbidades receberam menos antidepressivo e apresentaram pior QV.  Os pacientes com comorbidades receberam menos antidepressivo e pior QV, apesar da mesma gravidade de sintomas depressivos.

Chisholm D, Diehr P, Knapp M, Patrick D, Treglia M, Simon G (2003)	Depression status, medical comorbidity and resource costs. Evidence from an international study of major depression in primary care (LIDO).	Explorar a relação entre estado depressivo, perda de dias de trabalho e custos dos cuidados de saúde.	Grupo de análise econômica Amostra basal	Comorbidade médica esteve associada com 17-47% do aumento dos custos com a saúde em 5 dos 6 centros.  As conseqüências econômicas da depressão são mais fortemente influenciadas pela presença das comorbidades médicas do que pela severidade dos sintomas isoladamente.
Simon GE, Chisholm D, Treglia M, Bushnell D; LIDO Group. (2002)	Course of depression, health services costs, and work productivity in an international primary care study.	Examinar os diferentes desfechos de depressão e seus correlatos econômicos.	Grupo de análise econômica Amostra de seguimento	36 % apresentaram depressão persistente, 29% remissão parcial, e apenas 35% remissão completa.  Menos dias perdidos de trabalho e menor utilização dos recursos de saúde estavam associados quadros depressivos mais favoráveis.
Fleck MP, Lima AF, Louzada S, Schestasky G, Henriques A, Borges VR, Camey S. (2002)	Association of depressive symptoms and social functioning in primary care service, Brazil].	Avaliar a associação entre sintomas depressivos e funcionamento social em uma amostra brasileira.	Grupo do Centro Brasileiro Amostra brasileira de rastreamento	Sintomas depressivos apresentam uma forte associação com pobre funcionamento social, prejuízos na QV e alta utilização dos recursos de saúde em cuidados primários.
Herrman, H., Patrick, D. L., Diehr, P., Martin, M. L., Fleck, M., Simon, G. E., & Buesching, D. P. (2002)	Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms.	Verificar a associação entre sintomas depressivos e estado funcional, QOL global e uso dos serviços de saúde entre os diferentes centros.  Entre os pacientes deprimidos, verificar que fatores estão associados com tratamento recente para depressão.	Grupo de análise econômica Amostra de rastreamento	Maiores escores de depressão estão associados com pior saúde, estado funcional e QV e aumento do uso dos serviços, mas não com variáveis demográficas.
Chisholm, D., Amir, M., Fleck, M., Herrman, H., Lomachenkov, A., Lucas, R., Patrick, D.(2001)	Longitudinal Investigation of Depression Outcomes (The LIDO Study) in primary care in six countries: comparative assessment of local health systems and resource utilization.	Descrever os métodos de utilizados na pesquisa de utilização de serviços.  Comparar os centros, sistemas de serviços e as populações incluídas no LIDO.	Grupo Coordenador Amostra basal	Comparação dos centros demonstrou que há uma variabilidade em termos de gastos e provimento dos serviços, bem como de seus modelos de financiamento. Houve diferenças nas variáveis demográficas e no padrão de utilização dos serviços.  A pesquisa multinacional de serviços é complicada pela heterogeneidade dos sistemas de serviços que refletem a realidade dos cuidados. Para serem comparáveis, os estudos devem ser cuidadosamente construídos.

---



### **3 JUSTIFICATIVA**

O WHOQOL-Bref foi construído e desenvolvido utilizando o referencial da Teoria Clássica em psicometria, há poucos estudos utilizando psicometria moderna para análise deste instrumento. Há uma carência de estudos de avaliação do desempenho do WHOQOL-Bref em pacientes deprimidos provenientes do contexto de atenção primária do ponto de vista transcultural. Por fim, não há estudos para avaliar a extensão da sobreposição entre as medidas de depressão e QV.

## **4 OBJETIVOS**

### **4.1 Objetivo geral**

Avaliar o desempenho do WHOQOL-Bref como instrumento de avaliação de qualidade genérica para pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.

### **4.2 Objetivos específicos**

1. Apresentar o modelo de Rasch e sua aplicação nos estudos de avaliação de instrumentos de pesquisa.
2. Avaliar a extensão da sobreposição das medidas de depressão e QV em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.
3. Avaliar o desempenho transcultural do WHOQOL-Bref em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.

4. Avaliar o desempenho do WHOQOL-Bref em pacientes deprimidos provenientes de um serviço de cuidados primários no Brasil usando a análise de Rasch.

## 5 HIPÓTESES

- O WHOQOL-Bref se mostrará uma medida com desempenho adequado para avaliação de QV geral em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.
- O WHOQOL-Bref se mostrará com uma medida transculturalmente adequada para avaliação de QV geral em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.
- O WHOQOL-Bref se mostrará uma medida com itens de sobreposição entre as medidas de depressão e QV em geral em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando a análise de Rasch.
- O WHOQOL-Bref se mostrará uma medida adequada para avaliação de QV geral em pacientes deprimidos provenientes de um serviço de cuidados primários do Brasil usando a análise de Rasch.

## 6 ARTIGOS

A presente tese será composta por 4 artigos que pretendem desenvolver o objetivo geral e cada um dos objetivos específicos:

### 6.1 Estudo 1

*The perspectives of the use of Rasch analysis in psychometric research*

Apresentar a análise de Rasch e sua aplicação nos estudos de avaliação do desempenho de escalas.

### 6.2 Estudo 2

*Is there a measurement overlap between depressive symptomatology and quality of life?*

Avaliar a extensão da sobreposição dos constructos de depressão e QV em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando o modelo de Rasch.

Aceito para publicação: **Comprehensive Psychiatry**, 2008.

### 6.3 Estudo 3

*Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis*

Avaliar o desempenho transcultural do WHOQOL-Bref em pacientes deprimidos provenientes de serviços de cuidados primários de 6 países no mundo usando o modelo de Rasch.

Pré-aceito, em revisão: **Medical Decision Making**, 2008.

### 6.4 Estudo 4

*Validity of the Brazilian version of WHOQOL-Bref in primary care depressed patients using Rasch modelling*

Avaliar o desempenho do WHOQOL-Bref em pacientes deprimidos provenientes de um serviço de cuidados primários no Brasil usando o modelo de Rasch.

Aceito para publicação: **Revista de Saúde Pública**, 2008.

## 6.1 Estudo 1

### The perspectives of the use of Rasch analysis in psychometric research

Neusa Sica da Rocha<sup>1,2</sup> and Marcelo Pio de Almeida Fleck<sup>1,3</sup>

1 Hospital de Clinicas de Porto Alegre, Programa de Pós Graduação em Ciências Médicas: Psiquiatria. Universidade Federal do Rio Grande do Sul (UFRGS)

2 Psychiatrist, Programa de Pós Graduação em Ciências Médicas: Psiquiatria. Universidade Federal do Rio Grande do Sul (UFRGS).

3 Assistant Professor, Departamento de Psiquiatria e Medicina Legal da Universidade Federal do Rio Grande do Sul (UFRGS).

Correspondent author:

Neusa Sica da Rocha

Hospital de Clinicas de Porto Alegre

Serviço de Psiquiatria

Avenida Ramiro Barcelos, 2350 4º andar

Phone: 51 21018413

e-mail: [neusa-rocha@via-rs.net](mailto:neusa-rocha@via-rs.net)

**Abstract**

The present article aims at reviewing the main characteristics of the Rasch analysis. We present a synthetic review of the main features of the Rasch analysis, using as an example the latent variable of depressive symptoms and the illustrative analysis of some items of the Beck Depression Inventory. With the use of the Rasch, we can obtain with a higher validity for the measurement of scale adequacy, the meaning of the numbers attributed to scales, the amount of latent traits that the numbers represent and the adequacy of statistical operations used to analyze them. Rasch analysis has a great potential of application in clinical research, e.g.: (a) the possibility of quality improvement of research instruments; (b) the building of computerized adaptive tests and finally (c) for the study of the validity of classification systems diagnostic criteria.

**Key words:** logistic models, Rasch analysis, Scales, Psychometrics; Validity of Tests.



## Introduction

There has been a growing search to develop methods of quantifying mental states in the last decades. Thus, the majority of the studies, either descriptive, testing hypothesis or assessing new treatments, have been using assessment scales. However, often the large use of health assessment scales has been performed uncritically, starting from the principle that many of the controversies regarding the validity of the scales are resolved in literature. Marshall et al (Marshall, Lockwood, Bradley, *et al*, 2000), examining a number of controlled trials in schizophrenia found that nearly 40 per cent reported that the treatment were effective when they use unpublished scales rather than validated ones. At the same time, it is as if we had built a big building with a concrete that we have used without paying attention to its quality. Lastly, the solidity of our studies depends on the quality (validity) of the tools that we are using.

Differently from when one wants to measure the size of a concrete object (e.g. a table), there are more complex challenges when the nature of what we are going to measure is a subjective variable such as health, depression, etc. Actually, we are searching to measure a construct, something that we intend to operate through the descriptive characteristics that we technically call a “latent variable”. The usual procedure has been to build a scale with a certain number of items that intend to assess some behaviors related to the construct (Tesio, 2003). Therefore, if we would like to measure these constructs by a psychometric test, we would have to look for indicators (to construct items) which are related to the construct in a way to be specified by a theory.

A second challenge is that when we propose to verify the quantity of this latent trait that the individual has, we face a series of technical difficulties among which the definition of the measurement scale, the meaning of the numbers attributed to the scales, the quantity of latent trait that the numbers represent and the adequacy of the statistical operations used to analyze them..

When someone responds to a certain question or item he/she has a certain quantity of the latent trait or ability(Baker, 2001). For example, as in a test that assesses the level of knowledge presented by the student, the questionnaire (e.g. depression) assesses the level or quantity of “ability” (in the case, the latent trait “depressive symptoms”), of the individuals that respond to it. Thus, if in a test of proficiency in English such as TOEFL(Chyn, 1995) (Test of English as a Foreign Language) depending on the responses that we give in the beginning, the level of difficulty increases successfully it is possible through its total score to know the level of proficiency of the students, in a Beck Inventory for depression we can through the score, identify the levels of depression. “Ability” here is not to be understood literally in the Rasch analysis, but as a technical term representing any hypothetical construct measured by Rasch models. Therefore, persons do not have a “certain quantity of the latent trait”, rather numbers, i.e. the person parameter values, are allocated to them. These numbers are just values of latent variables, i.e. mathematical constructs to explain differences between persons and relationships *between* persons’ item responses.

Recently there has been a growing use of more advanced psychometric strategies to better study and describe the performance of scales and of the

items that are part of a scale. The present article aims (a) to describe in a simplified manner the use of the Rasch analysis and (b) to present its potential utility in the improvement of subjective state measurements using as an example the items of a scale.

### **The Classical Test Theory x Item Response Theory**

The way to validate assessment tools most commonly used until recently is the so-called Classical Test Theory (CTT). The CTT assesses reliability as part of validity studies through the internal consistency by the Cronbach alpha and the construct validity, criterion and content validity. Its main approach is on the scale as a whole, being that all items have equal importance in the total score of the instrument. The focus of CTT lies on the measured value, its decomposition into true score and measurement error, and the estimation of reliability. CTT provides no information about the measurement of hypothetical constructs by manifest indicators (it does not provide us with a measurement model), so that it is impossible to check on the base of CTT assumptions alone whether the assumptions that we make if we use the total score as ability estimator are correct.

Besides, it allows that, once assessed the dimensionality of the scale, we can add up the scores of each one of the items to obtain the total score even if the questions or items describe contents of different repercussion to the punctuation of a certain phenomenon. For example, in a depression scale the item that assesses suicidal ideation and that assesses inattention have the same absolute value. We know that clinically a depressive syndrome with

suicidal ideation is more severe and that this symptom should have a larger discrimination power in our assessment.

From the evidence of the limitations of CTT it is possible to utilize a different model of analysis to verify the validity of the instruments taking into consideration parameters and characteristics with the focus on their items. It has been called the Item Response Theory.

### **The Item Response Theory (IRT) and the Rasch Model**

The first IRT models appeared in the 50's in the education area based on the need to built tests that would be at the same time simple, valid and with high discrimination power. They were at first models to assess a sole ability from only one group where the test had dichotomic responses.

The IRT tries to identify in the items three proprieties or parameters that are difficulty, discrimination and possibility of success by chance (guessing). The *difficulty* is also denominated "index of location" or location(Baker, 2001). The *discrimination* describes how much an item can differentiate people that have abilities either below or above the item location. The parameters of difficulty and discrimination of the items are independent of each other and vary according to the level of ability of the respondent. This means that we can have items of different difficulties and discriminations. On the other hand, the parameter of guessing does not depend on the level of ability since both the examined of high and low ability have the same probability of succeeding the item by chance.

The IRT models can be graphically represented by the Item Characteristic Curve (ICC). It is based on the fact that individuals with more

ability (latent trait) have more chance of succeeding the item. As we can observe the slope reminds us of an “S”(Baker, 2001). The ICC presented below was built from a Rasch model (Andrich, 1978; Rasch, 1960).

**Please insert Figure 1 here**

In the graph above, extracted from the statistical software RUMM 2020(Andrich, Sheridan & Luo, 2004), the curve corresponds to the ICC of the item “Appearance” of the Beck Inventory for depression. On axis X we have the person location and on axis Y we have the expected value in case the data is adjusted to the model. The slope that is formed represents the model and the dots on the line, the sample data. Since the dots are on the line we can conclude that the item represented above is adjusted to the Rasch model. Besides, we can identify in the same graphic that the value of its location is  $-0,463$  (Locn= $-0,463$ ) given on the right side after the character “Descriptor for item 14”. The verification of the adequacy of the item to the model is performed through the assessment of the residual and the qui-square. For such, it is defined that the residual should be greater or smaller that  $\pm 2.5$  and the chi-square greater than 0.05, showing that there is no significant difference between the data regarding the item and the proposed model. In the case of the previous example, the Fit residual is  $-1.129$  and the probability of the qui-square is 0.189 (ChiSq[Pr]=0.189). All these numbers point to an adequacy of this item to the model, for the residual is below  $-2.5$  and the probability of the qui-square is lower than 0.05.

Today, there is a series of mathematic models or IRT equations that besides assessing more than a parameter, also study different types of items and response scales. The main ones are the logistics 1, 2 and 3 parameters. The model of 1 parameter takes into account only the *difficulty* of the item, the one with 2 parameters also takes into account the *discrimination* and finally the one with 3 parameters assesses besides the *difficulty* and the *discrimination*, the possibility of guessing.

Among those models the one that has been widely used is the one developed by Georg Rasch. In this model the parameter of discrimination of the logistic model of two parameters is fixed in the value of 1 for all the items, only the parameter of *difficulty* can have different values for each item. That's why the Rasch model is frequently considered a model of 1 parameter (*difficulty*)(Baker, 2001; Rasch, 1960).

One of the forms commonly utilized to search for an improvement in the tool is to verify the best response scale possible for the item. We are used to considering the data coming from response scales as if they were at intervals, but in truth they are in the majority, ordinal data where it is not adequate the use of mean and standard deviation but the use of percentages. When we work with response scales (polytomic items), we know that they produce ordinal data that should be transformed in interval data.

There are some statistic softwares that can make calculations. We can highlight BILOG, Winsteps (Linacre & Wright, 1999) and the RUMM 2020 (Andrich, Sheridan & Luo, 2004). The specificities and particularities of each one are not the objective of this article. In the statistic software RUMM 2020 the measure of the variability is the Fit Residual and we consider the variation of  $\pm$

2.5 as a value above or below which, the items are not adequate to the model. A positive residual indicates that the observation was greater than what was expected and a negative residual indicates that the observation was smaller than what was expected.

### **The example of the construct DEPRESSIVE SYMPTOMS measured by the Beck Depression Inventory.**

We will use the Beck Inventory for the assessment of Depressive Symptoms (BDI) and some of its items to discuss methodological implications of data analysis coming from form assessment scales using Rasch analysis. Thus, we will try to show that the variables are forms of representing the relation between the probability of an individual to give a response to one of the items and his/her latent traits, proficiencies or abilities in the assessed area.

In the BDI, as the majority of the scales in Psychiatry we make a simple sum of the values attributed by the interviewer to each question (item), and we reach the measurement of the total intensity of the “depressive symptoms”. Similarly, when we measure a length of an object for example, we use a ruler and we measure directly its length from the sum of our metric unit: the centimeter.

Making a parallel between a ruler and the depression scale (BDI) we are going to raise some differences that are pertinent in this comparison. If we use the metric system for the length, we admit that there are uniform measurement intervals and that a 20 cm length represents twice the 10 cm length. However, can we, with the same precision, assure that a person that has a total score of 20 points in the BDI twice more “depression” than who had obtained a score

10? Obviously, our clinical experience impedes us from making this statement and also the CTT. Using the concepts of CTT, simply added to the values attributed to the items without questioning if in fact our measurement unit is calibrated. This way the CTT doesn't take into consideration a central metric question: in what way can we turn the depression scale closer to a ruler so that mathematical and statistical operations made with their scores can be as valid and representative as a ruler?

Through the Rasch model we have a way to build interval measurements from raw data such as the ones coming from response scales like the Likert or the ones used in the symptoms assessment scales. For example, in the case of the BDI, of "I am not sad/I am said/I am always sad and I can't escape from that/ I am so said or unhappy that I can't stand" respectively 0,1,2,3 might not have the same distance among them and thus any conventional statistics such as mean and standard deviation does not give us a legitimate idea of what happens with those data that may not have regular intervals (Tesio, 2003; Wright & Mok, 2000).

In Figure 2 below we present a graphic representation of one item where the thresholds were centralized, i.e., each category of response had the same chance of being responded, thus, is adequate to the Rasch model and we can consider an interval measure. As we can observe, each slope in the graphic represents a category of response, the slope with the number 0 represents a category of response "I don't feel un any way worse than the others", 1 represents "I am critic in relation to myself because of my weaknesses and mistakes", 2 represents "I blame myself for my mistakes" and 3 represents "I blame myself for all bad things that happen" represents respectively 0,1,2,3. We



can see that each slope has a part of its area independent from the others and there is no overlap of curves, indicating that there is no need to group the categories.

**Please insert Figure 2 here**

However, the item 13 (take decisions), with a scale of “I take decisions as well as I used to do/ I postpone decision taking more than I used to/I have more difficulty in taking decisions than before/I absolutely can’t take any decision anymore” represented respectively by 0,1,2,3, presented a disordered threshold. Thus, was not adequate to the Rasch model, forcing us to make changes in its response scale. We can observe that category 1 (identified in the slope with numeral 1), had a distribution that can be included both on category 0 and 2. This indicates that this category of response (category 1) has to be suppressed and included in one of the categories mentioned before, i.e., “I postpone decision taking more than I used to” can be included both in category “I take decisions as well as I used to do” and in category “I have more difficulty in taking decisions than before”.

**Please insert Figure 3 here**

We can identify that the area of the slope of category 1 was included in the areas of categories 0 and 2; thus can be included both in one and in other category of response, tending more to category 2.

Other presuppositions of the Rasch model are the unidimensionality, additivity, local independence and invariance. This means that if the BDI scale fits those presuppositions we would have to ensure that we are measuring a sole construct (unidimensionality) and thus the value attributed to each question (item) of the scale can be adequately added to the value of the other, each item is measuring a relevant aspect and does not depend on another item to have this information (local independence) and even if this item is administered to other respondent it will continue measuring the same ability (invariance) (Tesio, 2003).

Following the example of our ruler, we would have to test if BDI is effectively measuring one sole construct (depression), i.e., if all the items of the scale are equally important and not redundant (humor, insomnia, lack of appetite, suicidal ideation). Besides, we verify that if we administer the samples with different levels of depression we can identify how much the fact that they are depressed can have influenced the response to the item. According to this parallel, it is as if each item of the scale (symptom of depression) represented a range of this ruler. The unit measure of the Rasch model is the logit (log-odd units), that in the case of the ruler is the centimeter. When the data fit the model, the items are located along the same logit according to the difficulty that they present to the individuals. On the other hand, the individuals are positioned according to the ability demonstrated on these items. This technique results in a scale-free measure of the person and a sample-free calibration of the items.

One of the advantages of the Rasch analysis in relation to CTT is the possibility of making most easily an *equating*. The *equating* allows the

comparison both from different items coming from different tests and abilities coming from different groups of the same metrics. Thus, we could compare different depression scales and compare their items for example.

It is worth to mention that the use of the Rasch analysis intends to turn the values attributed to the items, mathematically valid and thus assure the validity both of the mathematic operations done with those values and its validation in relation to the content that is being measured. In the CTT, the score 20 in the BDI does not represent twice the depression than the ones that have the score 10, but with the use of the Rasch analysis we are at least sure that the number 20 really represents 20 and the 10, 10. The validity of any mathematical operation that uses those data is assured. Besides, when we calibrate the measure unit, we can assure that each item of the spectrum will bring us new information, each space of this ruler will be filled out by a new content. Coming back to the analogy of the ruler, if we have two items that ask differently about motor slowness, for example, it is as if we had duplicated the space of the centimeter 2 of the ruler, not influencing the total measurement of the object. This means that each information (logit) should be as relevant for the scale as the space for each centimeter is for the ruler.

### **Conclusion: Rasch analysis applications in clinical research**

The Rasch analysis is potentially an important methodological tool in the perfecting and assessment of the already existing quantitative measurements and for the development of new measure. In some contemporary research areas such as quality of life assessment, according to the information broadcasted in a symposium of the 12<sup>th</sup> International Society for Quality of Life

Research Annual Meeting, IRT is already present in the majority of the new instruments that are being developed (Sloan & Mandrekar, 2005).

An important application of the Rasch analysis in clinical research is the possibility of building measures composed of a shorter number of items that have a good internal consistency and with items that are not redundant. Since the Cronbach alpha value (main measure for the internal consistency of reliability on the CTT) is highly influenced by the number of items that are part of a scale, we could have instruments with high internal consistency, but, at the cost of redundant items that do not have the same discriminating capacity (Mallinson, Stelmack & Velozo, 2004). When we use the Rasch analysis, we can build measurement instruments that can be made of a smaller number of items, being assured that each item will bring another piece of information in the spectrum that is being measured (Prieto, Alonso & Lamarca, 2003). With this, for example, the measures of the outcomes of clinical trials can be assessed by less extensive scales, needing a smaller number of participants to identify significant differences (Fava & Balaise, 2005).

Within psychiatric assessment, the Rasch analysis can be applied in the assessment of the diagnostic criteria of the main Psychiatric Disorders since it is possible through its use, to identify which criteria is more or less discriminating, as well as if they add or not information to the diagnosis.

Another application would be in the construction of computerized adapted tests (CATs) where we apply to the respondents questions with different difficulty levels (or discrimination) aiming at shortening the assessment time. For example, as we mention before in the TOEFL test, depending on the answers we give to questions in the beginning the level of difficulty gets higher

successively. Thus, using a scale for depression diagnosis, for example, depending on how the person answers the question about depressive humor then the computer puts the following questions (Gardner, Shear, Kelleher, *et al*, 2004).

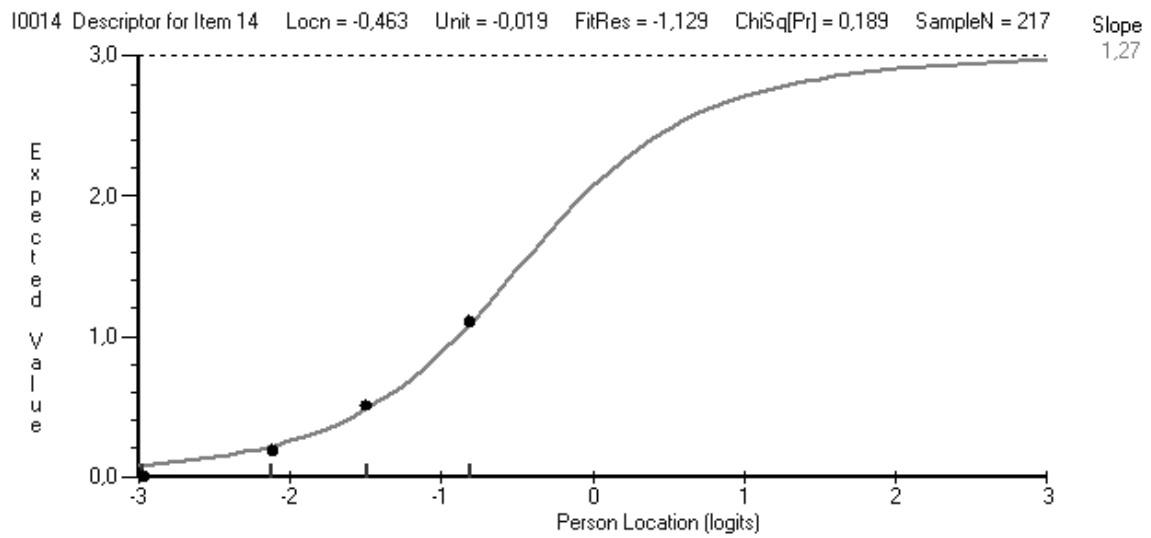
The use of the Rasch analysis has also been used as a construction method for measures that need a transcultural validation allowing the identification of items that would not be differently answered in different cultures to assure the universality of the construct that is being measured (Tennant, Penta, Tesio, *et al*, 2004). The example that is mostly studied in literature is the validation and construction of tools that would assess the quality of life (McHorney & Monahan, 2004; Tennant, McKenna & Hagell, 2004).

The present article was proposed to briefly revise the Rasch analysis. Even though it is complex in a first approach, it is necessary that progressively professionals that use scales, either on clinical activity or in research be introduced to this important statistical technique that would fatally become part of the revision of traditional scales and of the new scales that are being created.

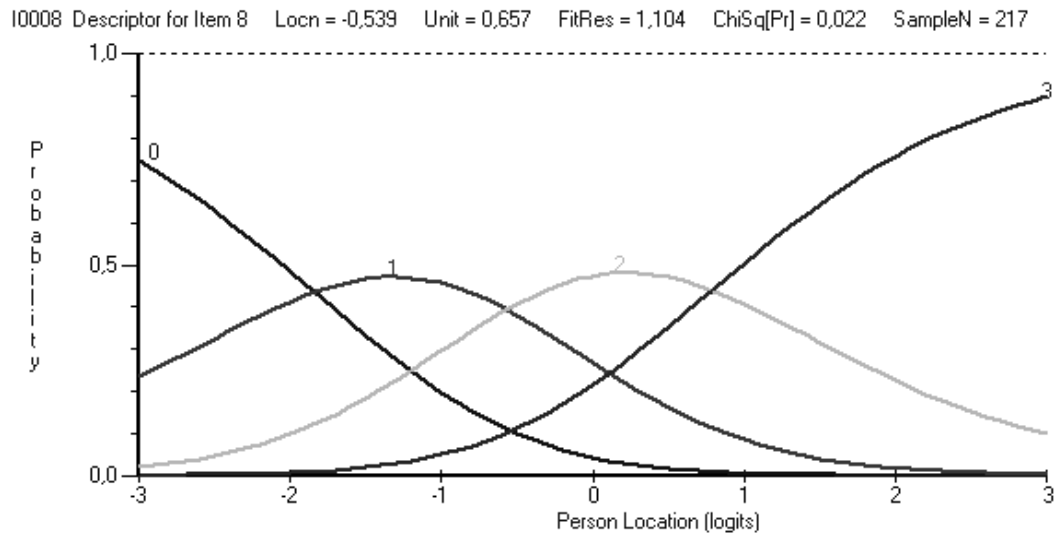
### **Acknowledgements**

This study was partially funded by FIPE-HCPA, CAPES and the University of Edinburgh.

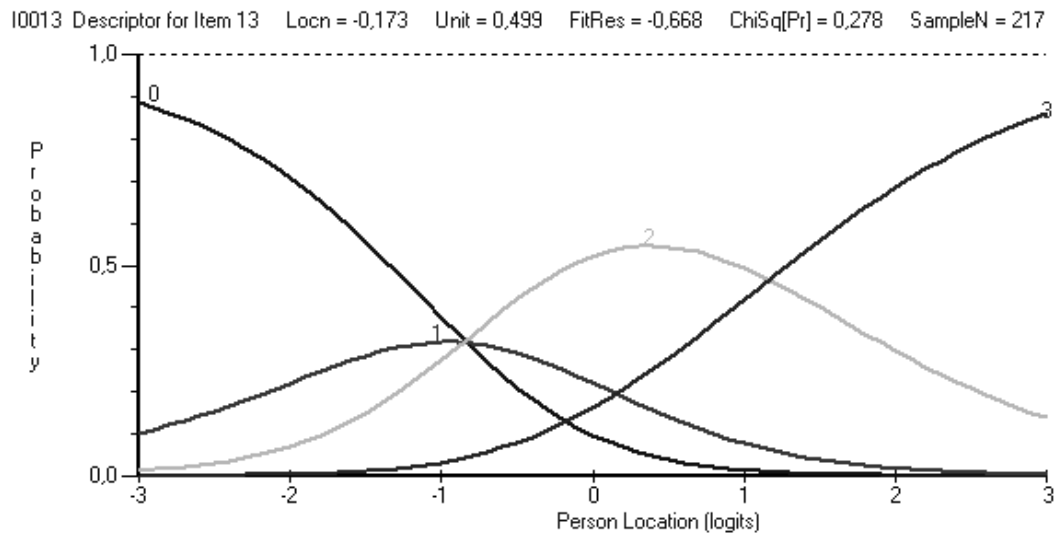
**Figure 1: Item Characteristic Curve of Item 14 (Appearance) of the Beck Depression Inventory (BDI)**



**Figure 2: Probability Categories Curves of the Item 8 of the BDI- Example of Centralized Thresholds**



**Figure 3: Probability Categories Curves of the Item 13 (make decisions) of the BDI- Example of disordered thresholds**





## REFERENCES

- Andrich, D. (1978)** A rating formulation for ordered response categories. *Psichometrika*, **43**, 561-573.
- Andrich, D., Sheridan, B. & Luo, G. (2004)** RUMM: a Windows program for analysing item response data according to Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM Laboratory.
- Baker, F. B. (2001)** *The Basics of Item Response Theory*. Madison: ERIC Clearinghouse on Assessment and Evaluation.
- Chyn, S. (1995)** An investigation of IRT-based assembly of TOEFL Test. TOEFL Technical Report Educational Testing Service.
- Fava, G. A. & Balaise, C. (2005)** A discussion on the role of clinimetrics and the misleading effects of psychometric theory. *Journal of Clinical Epidemiology*, **58**, 753-756.
- Gardner, W., Shear, K., Kelleher, K. J., et al (2004)** Computerized adaptive measurement of depression: a simulation study. *BMC Psychiatry*, **4**, 1-11.
- Glass, G. V., Staley, J. C., Glass, G. V., et al (1970)** Measurement, Scales, and Statistics. In *Statistical Methods in Education and Psychology*, pp. 7-23. New Jersey: Prentice-hall Inc.
- Linacre, J. M. & Wright, B. D. (1999)** *Winsteps computer program*. Chicago: MESA Press.
- Mallinson, T., Stelmack, J. & Velozo, C. (2004)** A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care*, **42**, 117-124.
- Marshall, M., Lockwood, A., Bradley, C., et al (2000)** Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br.J.Psychiatry*, **176**, 249-252.
- McHorney, C. A. & Monahan, P. O. (2004)** Postscript: Applications of Rasch analysis in health care. *Med Care*, **42**, 173-178.

- Michell, J. (2003)** Measurement: a Beginner's Guide. *Journal of Applied Measurement*, **4**, 298-308.
- Norquist, J. M., Fitzpatrick, R., Dawson, J., et al (2004)** Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care*, **42**, 125-136.
- Prieto, L., Alonso, J. & Lamarca, R. (2003)** Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual.Life Outcomes.*, **1**, 27.
- Rasch, G. (1960)** *Probabilistic models for some intelligence and attainment tests*. Copenhagen:: Danish Institute for Educational Research.
- Sloan, J. & Mandrekar, S. (2005)** Item Response Theory: when is it useful and where does classical test theory fit in? *Quality of Life Research*, **14**, 1982.
- Smith, J. E. (2000)** Metric Development and Score Reporting in Rasch Measurement. *Journal of Applied Measurement*, **1**, 303-326.
- Stevens, S. S. (1946)** On theory of scales of measurement. *Science*, **103**, 677-680.
- Tennant, A., McKenna, S. P. & Hagell, P. (2004)** Application of Rasch analysis in the development and application of quality of life instruments. *Value.Health*, **7 Suppl 1**, S22-S26.
- Tennant, A., Penta, M., Tesio, L., et al (2004)** Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care*, **42**, 137-148.
- Tesio, L. (2003)** Measuring Behaviors and Perceptions: Rasch Analysis as a Tool for rehabilitation Research. *J Rehabil Med*, **35**, 105-115.
- Van Alphen, A., Halfens, R., Hasman, A., et al (1994)** Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing*, **20**, 196-201.
- Wright, B. D. & Mok, M. (2000)** Rasch Models Overview. *Journal of Applied Measurement*, **1**, 83-106.

## 6.2 Estudo 2

### **Is there a measurement overlap between depressive symptomatology and quality of life?**

**Neusa Sica da Rocha<sup>1,2</sup>, Marcelo Pio de Almeida Fleck<sup>1,3</sup>,  
Mick J. Power<sup>4</sup>, Donald M. Bushnell<sup>5</sup>**

1 Hospital de Clinicas de Porto Alegre, Graduate Program in Medical Sciences:  
Psychiatry-UFRGS-Brazil

2 Psychiatrist, Graduate Program in Medical Sciences: Psychiatry-UFRGS-Brazil

3 Adjunct Professor, Department of Psychiatry and Legal Medicine-UFRGS-Brazil

4 Professor, Section of Health and Clinical Psychology, University of Edinburgh-UK

5 Associate Director, Health Research Associates, USA

Corresponding author:

Neusa Sica da Rocha  
Avenida Ramiro Barcelos, 2350 4º andar  
Psychiatric Service  
Porto Alegre-RS-Brazil  
CEP 90003-035  
Phone: 51 21018413  
e-mail: [neusa-rocha@via-rs.net](mailto:neusa-rocha@via-rs.net)

#### **ACKNOWLEDGMENTS**

This study was partially funded by FINEP-HCPA (Brazil), CAPES (Brazil) and the University of Edinburgh (United Kingdom). This paper was presented as a poster at the 160<sup>th</sup> Annual Meeting of the American Psychiatric Association, 2007.

## **ABSTRACT**

**Background:** Previous studies found that depression is associated with a broad impairment in quality of life (QOL), this finding might be associated to a measurement overlap.

**Methods:** The objective of this study was to verify whether the items of the WHOQOL-Bref, a measure of generic QOL, are invariant among patients having a current major depressive episode who come from primary care services. We investigated data from primary care services from six countries (Australia, Brazil, Israel, Russia, Spain, and the USA) involved in the baseline sample of the Longitudinal Investigation of Depression Outcomes (LIDO). The Rasch Model was used to analyze items exhibiting differential item functioning (DIF) as a way of assessing invariance in relation to a depression factor defined by the diagnosis of depression using the Composite International Diagnostic Interview (CIDI).

**Results:** The sample consisted of 2,359 subjects, of which 1,193 had a confirmed diagnosis of a current major depressive episode. Eleven out of the 26 items of the WHOQOL-Bref showed DIF due to the depression factor and the Physical domain presented more items displaying DIF.

**Conclusions:** Our findings indicate that the majority of WHOQOL-Bref items do not exhibit DIF for a current major depressive episode and the variance associated with depression in this generic QOL measure is restricted to some facets of this construct. Thus, we recommend this restricted adjustment for depression in future analyses of this measure.

## INTRODUCTION

Major depression is one of the most frequent mental disorders (Weissman, Bland *et al.*, 1996; Kessler, Berglund *et al.*, 2003; Ustun, Yuso-Mateos *et al.*, 2004) and is associated with a variety of limitations. (Fleck, Lima *et al.*, 2002) Currently, it is the fourth leading cause of disability worldwide and projections estimate it will be the second leading cause by year 2020. (Ustun, Yuso-Mateos *et al.*, 2004) The effectiveness of the treatment for this disorder has been assessed using the evaluation of quality of life (QOL) as an outcome measure. (Sturm e Wells, 1995; Wells, Sherbourne *et al.*, 2000; Skevington e Wright, 2001; Papakostas, Petersen *et al.*, 2004; Asarnow, Jaycox *et al.*, 2005)

Research with depressed patients in different contexts has mostly concluded (Ay-Woan, Sarah *et al.*, 2006) that QOL is impaired even more than in patients with other chronic clinical conditions (Stewart, Greenfield *et al.*, 1989; Wells, Stewart *et al.*, 1989; Hays, Wells *et al.*, 1995; Bonicatto, Dew *et al.*, 2001). These findings force us to face a methodological issue associated with the validity of measures of the construct of depression, (Sainfort, Becker *et al.*, 1996; Atkinson, Zibin *et al.*, 1997; Leidy, Palmer *et al.*, 1998; Berlim, Pavanello *et al.*, 2005; Scocco, Fantoni *et al.*, 2006) given that these same studies point to the possibility of an overlap of the constructs of depression and QOL. (Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Heinonen, Aro *et al.*, 2004; Naumann e Byrne, 2004; Kuehner e Buerger, 2005; Masthoff, Trompenaars *et al.*, 2006)

Currently, there are many generic QOL scales in use. One example is the WHOQOL-Bref, (Whoqol, 1998; Skevington, Lotfy *et al.*, 2004) the abbreviated

version of the QOL instrument developed by the World Health Organization. The WHOQOL-Bref was constructed in a cross-cultural simultaneous methodology to ensure its validity for a number of cultures and languages. This instrument stemmed from the WHOQOL-100 (Whoqol, 1995) and has therefore been studied in several groups of diseases (Yao e Wu, 2005) including major depression. (Skevington e Wright, 2001; Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Naumann e Byrne, 2004; Kuehner e Buerger, 2005; Rocha, Fleck *et al.*, 2005; Aigner, Forster-Streffleur *et al.*, 2006; Ay-Woan, Sarah *et al.*, 2006; Scocco, Fantoni *et al.*, 2006) It is particularly interesting that it has been evaluated in samples of depressed subjects mainly because it is a self-report scale that measures different aspects of QOL. Although it is based on the perception of the respondent, and some of its items can be characterized essentially as subjective (psychological, social), some other items are closer to objective evaluations (physical, environment). Commonly, the findings of poor QOL in depressed patients are interpreted as a result of Beck's cognitive triad (Beck, 1979) of feelings of pessimistic helplessness about themselves, the world and their future. (Aigner, Forster-Streffleur *et al.*, 2006) Some authors used Beck's conceptual model of depression to demonstrate that depression and QOL are tautological measures. (Katsching, 1997; Angermeyer, Holzinger *et al.*, 2002) Despite the fact that QOL measures have been questioned when used in depressed patients due to this 'depressive view', there is a lack of studies seeking to assess the extension of this bias in samples of depressed patients.

The analytic method we chose to address this question was that of Rasch modelling. More well-known in education assessment, Rasch modelling can be

used to develop a measure that could be both non-influenced by the sample and by the scale. (Rasch, 1960; Andrich, 1978; Smith, 2000) It means that we can identify whether items that are part of a scale have different meanings across groups, for instance, patients presenting with a depressive episode and those who are not.

We hypothesize that the majority of the WHOQOL-Bref items will present different functioning comparing patients having a current major depressive episode with those who despite presenting depressive symptoms, are not suffering from a depressive episode.

The present study was designed to verify whether the WHOQOL-Bref items have different functioning in primary care depressed patients who have been diagnosed as having a current major depressive episode when compared to those who have not. Rasch analyses were applied on data from a cross-sectional sample of the Longitudinal Investigation of Depression Outcomes (LIDO), a multicenter, cross-national observational study which followed patients with depressive disorders in primary care settings in six countries (Chisholm, Amir *et al.*, 2001) for 12 months .

## **METHOD**

The design, instruments and methodology as well as the sample size estimation of the LIDO study are described in detail elsewhere. (Chisholm, Amir *et al.*, 2001; Herrman, Patrick *et al.*, 2002) In summary, patients presenting at primary care centres in six sites (Barcelona, Spain; Be'er Sheva, Israel; Melbourne, Australia; Porto Alegre, Brazil; Seattle, USA, and St Petersburg, Russia) were screened for symptoms of depression. Those who met inclusion criteria for longitudinal assessment- new and/or untreated episode and a score over 16 on the Centre for Epidemiological Studies- Depression scale (CES-D) (Radloff, 1977)- were interviewed and assessed with the Composite International Diagnostic Interview (CIDI)(Robins, Wing *et al.*, 1988), a standardized diagnostic instrument for major depression.

### **1. Sample selection**

Participants had to be 18-75 years old, be a patient in a participating primary care setting and meet CIDI criteria for current major depression. They also had to be able and willing to participate in planned visits and/or study-required contacts; provide adequate contact details to ensure follow-up; give written informed consent; read, understand and complete the self-administered surveys in the primary language at the site; and plan to be available for the 12 months of the study. They were not included if they: (a) were currently receiving treatment for depression or had been treated for depression over the past 3 months; (b) had a known major psychiatric disorder or psychosis; (c) had a diagnosis of dementia, Alzheimer's disease or organic brain syndrome.



Selection of primary care settings by investigators was pragmatic, and based primarily on good working relationships with the primary care physicians and clinic managers. (Herrman, Patrick *et al.*, 2002)

## **2. Measures**

### **a) Demographic characteristics**

Continuous variables were age and years of education; binary variables were gender, marital status, employment, and self-report of health.

### **b) Severity of depression**

The CES-D is a self-rated 20-item scale designed to measure symptoms of depression in community populations. Respondents rate items by recalling the past week and using a four-point response scale, with higher scores indicating the presence and persistence of symptoms. (Radloff, 1977)

### **c) Generic Quality of Life**

The World Health Organization Quality of Life Instrument- Abbreviated version (WHOQOL-Bref) (Skevington, Lotfy *et al.*, 2004) is a self-rated 26-item measure taken from the larger WHOQOL-100, a multilingual assessment for generic quality of life, which was developed concurrently across 15 international field centres. The 26 items of the WHOQOL-Bref distribute into four domains (physical, psychological, social relationships and environment) and are answered using individualised five-point scales. Each subscale is scored positively. The WHOQOL group define QOL as “the individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns”. (WHOQOL Group, 1995)

#### **d) Depression diagnostic instrument**

The Composite International Diagnostic Interview, version 2.1 (**CIDI**)(Weiller, Lecrubier *et al.*, 1994) is a completely structured psychiatric diagnostic assessment developed for use in cross-national epidemiologic studies. Data from CIDI were used to assess diagnostic criteria for depression from the American Psychiatric Association DSM-IV.(American Psychiatric, 1994)

To be considered as having a current major depressive episode, patients should have a CES-D score of  $\geq 16$ , and a positive CIDI interview for the DSM-IV criteria for major depression. This diagnosis was utilized to define the “depression factor” for Rasch analysis.

### **3. Statistical Analysis**

#### **Rasch analysis**

The Rasch model is a uni-dimensional model, first used in educational assessment, which asserts that the easier the item the more likely it will be passed, and the more able the person, the more likely they will pass an item compared to a less able person. The Rasch model assumes that items and persons can be located on the same latent trait continuum, that the person's ability parameter and the items' threshold parameters are estimated, and that the probability of responding to an item-answer category is modeled.

There are some principles in Rasch modelling that need to be evaluated to ensure that an instrument has Rasch properties. The most commonly assessed properties are unidimensionality, local independence and invariance. (Tesio, 2003) For the purpose of the present study, only the property of invariance was analysed using the “depression factor” (i.e. diagnosis of a current major depressive episode) as the main factor.

To ensure the invariance of a measure, the items of a particular scale should work in the same way, irrespective of which group is assessed. This type of analysis has been called Differential Item Functioning (DIF). That is, individuals with the same score on the test are expected to have the same probability of response to an item, irrespective of any group membership. The statistical test used for detecting DIF is an ANOVA of the person-item deviation residuals with person factors (e.g., age, gender, country, diagnosis of major depression) and class intervals (e.g., group along the trait) as factors. Two kinds of DIF can be identified: uniform and non-uniform. In the former, there is a constant difference between groups in the probability of affirming an item across the trait (main effect) and in the latter the difference varies across the trait (interaction effect). DIF analysis is most commonly used for assessing cross-cultural validity. (Tennant, Penta *et al.*, 2004) When an item displays DIF for the 'depression factor', it will be defined as not invariant, that is, it is a sign of a 'depression biased' item.

Analysing only patients having the confirmed diagnosis of major depression, all items were checked for DIF by gender, educational level and age. Adjustment for cross-cultural (site) validity is the focus of a separate publication. (Rocha, Fleck *et al.*, 2008)

Since QOL is a multi-dimensional construct, Rasch analysis was performed preserving the original sub-scale structure (4 domains plus overall) as if each domain were a separate scale. (Orley, Saxena *et al.*, 1998; Heinonen, Aro *et al.*, 2004)

Rasch analysis was undertaken using the RUMM 2020 package. (Andrich, Sheridan *et al.*, 2004)

#### **4. Ethics**

All the individuals invited to participate in the study completed a written consent which presented the objectives of the study. This study was approved by the sites' Ethics Research Committees.

## RESULTS

Across the six sites, 18489 screening surveys were administered, out of which 4382 (23.7%) met initial eligibility criteria for baseline assessment. A total sample of 2359 patients participated in the baseline assessment. Of the 2359 completing the baseline assessment, 1193 (50.6%) patients with current clinical depression (i.e. met DSM-IV criteria for current major depressive episode) were invited to enrol in the longitudinal study. Compared to those not meeting criteria for major depressive episode (N=1166; 49%), patients having the confirmed diagnosis of major depression had poorer QOL in all WHOQOL-Bref domains as evidenced by mean values of the total scores for each domain (physical: 49.9 vs. 62.7,  $P \leq 0.01$ ; psychological: 45.6 vs. 60.7,  $P \leq 0.01$ ; environment 51.1 vs. 59.4,  $P \leq 0.01$ ; social 48.5 vs. 61.1,  $P \leq 0.01$ ). Patients who had CES-D  $\geq 16$  but did not have the major depressive episode confirmed by the CIDI (N=1166) were used as the comparison group in the present analysis. Data presented in Table 1 provide a summary description of the baseline sample in terms of age, gender, occurrence of major depression, self-report of health along with sample sizes.

### **Insert Table 1 here**

The DIF analysis for “depression factor” showed that 11 out of 26 items of the WHOQOL-Bref displayed DIF for this factor. All domains had at least one item displaying DIF. The most affected domain was the Physical one (5 out of 7 items) and the least was Environment (2 out of 8 items), as can be seen in Table 2.

### **Insert Table 2 here**

We divided the WHOQOL-Bref items into two theoretical domains (Table 3): “quality of life items not related to depression” (12 items), including overall quality of life, general health, medication, safety, environment, body, finances, information, mobility, home, services and transport) and “quality of life depression items” (14 items). The following were items presented as DSM-IV criteria for depression and/or as Hamilton Rating Scale for Depression items (Hamilton, 1960): pain, positive feelings, meaning in life, thinking, energy, leisure, sleep, daily life activities, relationships, support, work, self-esteem, sex and negative feelings).

### **Insert Table 3 here**

As shown in table 3, from the 11 items that display DIF for the depression factor, 7 items could be seen as diagnostic criteria for a major depression episode.

In this depressed sample (n=1193), no items presented DIF for gender and age concomitantly, which would be a violation of Rasch assumptions of unidimensionality and invariance. Four items displayed DIF for age ( $\leq 45$  years old and  $\geq 45$  years old: the median of the sample): Medication, Daily Life Activities, Positive feelings and Thinking. Only the Body item displayed DIF for gender.

## DISCUSSION

Our findings indicate that the WHOQOL-Bref items are mostly invariant in this sample of primary care major depressed patients from 6 countries worldwide. This is evidenced by the fact that 11 out of 26 items (42%) of the WHOQOL-Bref had different functioning for major depressed patients. Also, the single domain that has more items affected than non-affected by depression was the physical domain. This is the first research study designed to assess this possibility in a population with these characteristics. Other studies that tested the validity of QOL instruments used traditional test theory, had reduced sample sizes, and involved patients from single-culture clinical settings (Atkinson, Zibin *et al.*, 1997; Kuehner, 2002; Naumann e Byrne, 2004; Berlim, Pavanello *et al.*, 2005) thereby limiting the generalization of their results.

Several studies (Katsching, 1997; Demyttenaere, De Fruyt *et al.*, 2000; Bonicatto, Dew *et al.*, 2001; Skevington e Wright, 2001; Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Heinonen, Aro *et al.*, 2004; Naumann e Byrne, 2004; Aigner, Forster-Streffleur *et al.*, 2006) have considered a possible overlap of depression and QOL constructs, when analysing QOL of depressed patients. Nevertheless, it is important to differentiate a conceptual from a metrical overlap. The fact that there are some items that could measure both QOL and depression does not mean that QOL and depression are the same construct. For example, the fact that body temperature is both a measure of fever and the intensity of an infection does not mean that fever and infection are the same concept.

QOL is neither the opposite of depression, nor is euthymia a synonym of QOL. We agree with the assertion of Orley (Orley, Saxena *et al.*, 1998) about the

WHOQOL concept of QOL that underlies the WHOQOL-Bref measurement: “a QOL scale is a much broader assessment and although affect-laden, it represents a subjective evaluation of oneself and one’s social and material world.”

The discussion about which conceptual models would be more suitable for describing the study of QOL in depressive disorder is of great interest. Particularly, because neither the QOL (Lepelge e Hunt, 1997) nor the depressive disorder (Parker, 2000; Fink, Bolwig *et al.*, 2007) conceptual models show a consensus in the literature. Commonly, the findings of poor QOL of depressed patients are interpreted as a result of Beck’s cognitive triad. (Beck, 1979) of feelings of pessimistic helplessness about themselves, the world and their future. (Aigner, Forster-Streffleur *et al.*, 2006) Some authors used Beck’s conceptual model of depression to demonstrate that depression and QOL are tautological measures. (Katsching, 1997; Angermeyer, Holzinger *et al.*, 2002) Nevertheless, if so, it would be difficult to explain why the physical domain is so affected by depression while the environment domain is not. Cognitive dysfunction presented by depressive patients pointed by Beck’s theory should have equally influenced the responses (view) to the physical and the environment domains. Also, it would be difficult to explain why only 11 out of 26 items (42%) had DIF for depression and not 100% or at least 70%, as should be expected if QOL and depression were the same construct.

The fact that the Physical domain was the one mainly affected by depression may be associated with the Primary Care settings, where patients were recruited in an appointment with a general practitioner for physical complaints which may or may not have been related to their depressive episode.



(Aragones, Labad *et al.*, 2005; Tylee e Gandhi, 2005) An international study reported percentages of depression with somatic presentation from 45% to 95% in the countries studied. (Simon, Vonkorff *et al.*, 1999) Several studies have shown that almost one in four patients treated in general practice presents psychological complaints, predominantly anxiety and/or depression. If we agree that anxiety symptoms may be somatically focused, our findings for the physical domain may be explained either by the occurrence of somatic depression, or the comorbidity with anxiety disorders and complaints. (Runkewitz, Kirchmann *et al.*, 2006)

Our study was not aimed at assessing the construct validity of the WHOQOL-Bref for this sample. Since energy, sleep and activity were facets affected by depression and can also be clues to a possible presence of a melancholic or psychotic depression (American Psychiatric, 1994; Parker, 2005; Shorter, 2007), future investigations will be able to verify the hypothesis that the closer the patients' symptoms are to a melancholic depression, the lower is the possibility of an overlap between the constructs of depression and QOL.

When assessing QOL of primary care depressed patients using the WHOQOL-Bref, researchers must be aware of this extension of the influence of major depression on this measure. Some authors could recommend that we should broadly control depressive symptoms when using instruments for assessing subjective QOL. Our study suggests that the extension of the variance attributed to depression is restricted to some items that are repeatedly measured in depression studies. We suggest the specific control of these items (otherwise, assessing QOL in depressed patients could result in conclusions that QOL is strongly influenced by depression symptoms).

Finally, we conclude that the WHOQOL-Bref, as a generic QOL instrument, is based on a broad concept of QOL with most of its items not being influenced by the presence of a depressive episode. In addition, this instrument is considered by some authors as a measure of internal life satisfaction or subjective enjoyment in life and a suitable measure of global QOL. Further studies would be necessary and interesting to test our hypothesis that other generic QOL instruments would result in similar findings to our current research.

## References

- Aigner, M., S. Forster-Streffleur, *et al.* What does the WHOQOL-Bref measure? Measurement overlap between quality of life and depressive symptomatology in chronic somatoform pain disorder. Soc.Psychiatry Psychiatr Epidemiol., v.41, n.1, p.81-86. 2006.
- American Psychiatric, A. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association. 1994.
- Andrich, D. A rating formulation for ordered response categories. Psychometrika, v.43, p.561-573. 1978.
- Andrich, D., B. Sheridan, *et al.* RUMM: a Windows program for analysing item response data according to Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM Laboratory 2004.
- Angermeyer, M. C., A. Holzinger, *et al.* Depression and quality of life: results of a follow-up study. Int J Soc.Psychiatry, v.48, n.3, p.189-199. 2002.
- Aragones, E., A. Labad, *et al.* Somatized depression in primary care attenders. J.Psychosom.Res., v.58, n.2, p.145-151. 2005.
- Asarnow, J. R., L. H. Jaycox, *et al.* Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: a randomized controlled trial. JAMA, v.293, n.3, p.311-319. 2005.
- Atkinson, M., S. Zibin, *et al.* Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. Am.J Psychiatry, v.154, n.1, p.99-105. 1997.
- Ay-Woan, P., C. P. Sarah, *et al.* Quality of life in depression: predictive models. Qual.Life Res, v.15, n.1, p.39-48. 2006.
- Beck, A. T. R., A.J; Shaw, B.F.; Emery, G. Cognitive therapy of depression: a treatment manual. New York: Guilford Press. 1979.
- Berlim, M. T., D. P. Pavanello, *et al.* Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual.Life Res, v.14, n.2, p.561-564. 2005.
- Bonicatto, S. C., M. A. Dew, *et al.* Adult outpatients with depression: worse quality of life than in other chronic medical diseases in Argentina. Soc.Sci.Med, v.52, n.6, p.911-919. 2001.
- Chisholm, D., M. Amir, *et al.* Longitudinal Investigation of Depression Outcomes (The LIDO Study) in primary care in six countries: comparative assessment of local health systems and resource utilization. Int J Methods Psychiatr Res, v.10, n.2, p.59-71. 2001.

Demyttenaere, K., J. De Fruyt, *et al.* Measuring quality of life in depression. Current Opinion in Psychiatry, v.15, p.89-92. 2000.

Fink, M., T. G. Bolwig, *et al.* Melancholia: restoration in psychiatric classification recommended. Acta Psychiatr.Scand., v.115, n.2, p.89-92. 2007.

Fleck, M. P., A. F. Lima, *et al.* [Association of depressive symptoms and social functioning in primary care service, Brazil]. Rev.Saude Publica, v.36, n.4, p.431-438. 2002.

Hamilton, M. A rating scale for depression. J Neurol Neurosurg Psychiatry, v.23, Feb, p.56-62. 1960.

Hays, R. D., K. B. Wells, *et al.* Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. Arch.Gen.Psychiatry, v.52, n.1, p.11-19. 1995.

Heinonen, H., A. R. Aro, *et al.* Is the evaluation of the global quality of life determined by emotional status? Qual.Life Res, v.13, n.8, p.1347-1356. 2004.

Herrman, H., D. L. Patrick, *et al.* Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. Psychol.Med., v.32, n.5, p.889-902. 2002.

Katsching, H. Quality of life in depression. In: (Ed.). Quality of life in mental disorders. London: John Wiley and Sons, 1997. Quality of life in depression, p.137-147.

Kessler, R. C., P. Berglund, *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA, v.289, n.23, p.3095-3105. 2003.

Kuehner, C. Subjective quality of life: validity issues with depressed patients. Acta Psychiatr Scand., v.106, n.1, p.62-70. 2002.

Kuehner, C. e C. Buerger. Determinants of subjective quality of life in depressed patients: the role of self-esteem, response styles, and social support. J Affect.Disord., v.86, n.2-3, p.205-213. 2005.

Leidy, N. K., C. Palmer, *et al.* Health-related quality of life assessment in euthymic and depressed patients with bipolar disorder. Psychometric performance of four self-report measures. J Affect.Disord., v.48, n.2-3, p.207-214. 1998.

Leplege, A. e S. Hunt. The problem of quality of life in medicine. JAMA, v.278, n.1, p.47-50. 1997.

Masthoff, E. D., F. J. Trompenaars, *et al.* Quality of life and psychopathology: investigations into their relationship. Aust.N.Z.J Psychiatry, v.40, n.4, p.333-340. 2006.

Naumann, V. J. e G. J. Byrne. WHOQOL-Bref as a measure of quality of life in older patients with depression. Int Psychogeriatr., v.16, n.2, p.159-173. 2004.

Orley, J., S. Saxena, *et al.* Quality of life and mental illness. Reflections from the perspective of the WHOQOL. Br.J Psychiatry, v.172, p.291-293. 1998.

Papakostas, G. I., T. Petersen, *et al.* Quality of life assessments in major depressive disorder: a review of the literature. Gen.Hosp.Psychiatry, v.26, n.1, p.13-17. 2004.

Parker, G. Classifying depression: should paradigms lost be regained? Am.J.Psychiatry, v.157, n.8, p.1195-1203. 2000.

\_\_\_\_\_. Beyond major depression. Psychol.Med., v.35, n.4, p.467-474. 2005.

Radloff, L. S. The CES-D scale: a self report depression scale for research in the general population. Applied Psychological Measurement, v.1, p.385-401. 1977.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen:: Danish Institute for Educational Research. 1960.

Robins, L. N., J. Wing, *et al.* The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch.Gen.Psychiatry, v.45, n.12, p.1069-1077. 1988.

Rocha, N. S., M. Fleck, *et al.* Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis. 2008.

Rocha, N. S., M. P. Fleck, *et al.* What are we measuring using WHOQOL-Bref in depressed patients? The LIDO experience. . Qual.Life Res, v.14, p.2016. 2005.

Runkewitz, K., H. Kirchmann, *et al.* Anxiety and depression in primary care patients: predictors of symptom severity and developmental correlates. J.Psychosom.Res., v.60, n.5, p.445-453. 2006.

Sainfort, F., M. Becker, *et al.* Judgments of quality of life of individuals with severe mental disorders: Patient self-report versus provider perspectives. Am.J Psychiatry, v.153, n.4, p.497-502. 1996.

Scocco, P., G. Fantoni, *et al.* Role of depressive and cognitive status in self-reported evaluation of quality of life in older people: comparing proxy and physician perspectives. Age Ageing, v.35, n.2, p.166-171. 2006.

Shorter, E. The doctrine of the two depressions in historical perspective. Acta Psychiatr.Scand.Suppl, n.433, p.5-13. 2007.

Simon, G. E., M. Vonkorff, *et al.* An international study of the relation between somatic symptoms and depression. N.Engl.J.Med., v.341, n.18, p.1329-1335. 1999.

Skevington, S. M., M. Lotfy, *et al.* The World Health Organization's WHOQOL-Bref quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual.Life Res, v.13, n.2, p.299-310. 2004.

Skevington, S. M. e A. Wright. Changes in the quality of life of patients receiving antidepressant medication in primary care: validation of the WHOQOL-100. Br.J Psychiatry, v.178, p.261-267. 2001.

Smith, J. E. Metric Development and Score Reporting in Rasch Measurement. Journal of Applied Measurement, v.1, n.3, p.303-326. 2000.

Stewart, A. L., S. Greenfield, *et al.* Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA, v.262, n.7, p.907-913. 1989.

Sturm, R. e K. B. Wells. How can care for depression become more cost-effective? JAMA, v.273, n.1, p.51-58. 1995.

Tennant, A., M. Penta, *et al.* Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care, v.42, n.1 Suppl, p.137-148. 2004.

Tesio, L. Measuring Behaviors and Perceptions: Rasch Analysis as a Tool for rehabilitation Research. J Rehabil Med, v.35, p.105-115. 2003.

The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc.Sci.Med, v.41, n.10, p.1403-1409. 1995.

Tylee, A. e P. Gandhi. The importance of somatic symptoms in depression in primary care. Prim.Care Companion.J.Clin.Psychiatry, v.7, n.4, p.167-176. 2005.

Ustun, T. B., J. L. Yuso-Mateos, *et al.* Global burden of depressive disorders in the year 2000. Br.J Psychiatry, v.184, p.386-392. 2004.

Weiller, E., Y. Lecrubier, *et al.* The relevance of recurrent brief depression in primary care. A report from the WHO project on Psychological Problems in General Health Care conducted in 14 countries. Eur.Arch.Psychiatry Clin.Neurosci., v.244, n.4, p.182-189. 1994.

Weissman, M. M., R. C. Bland, *et al.* Cross-national epidemiology of major depression and bipolar disorder. JAMA, v.276, n.4, p.293-299. 1996.

Wells, K. B., C. Sherbourne, *et al.* Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. JAMA, v.283, n.2, p.212-220. 2000.

Wells, K. B., A. Stewart, *et al.* The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA, v.262, n.7, p.914-919. 1989.

\_\_\_\_\_. Development of the World Health Organization WHOQOL- Bref quality of life assessment. The WHOQOL Group. Psychol.Med, v.28, n.3, p.551-558. 1998.

Yao, G. e C. H. Wu. Factorial invariance of the WHOQOL- Bref among disease groups. Qual.Life Res, v.14, n.8, p.1881-1888. 2005.

**Table 1: General descriptions of baseline sample of LIDO<sup>a</sup> study**

	N	Age± S.D.	Female (%)	Major depression <sup>b</sup> (%)	Good Health (%)
Total	2359	41.6±14.9	1619(68.6)	1193(50.6)	1407(59.8)
Be'er Sheva, Israel	383	41.4±14.3	237(61.9)	184(48)	244(64.2)
Barcelona, Spain	472	41.5±15.2	335(71.0)	214(45.3)	275(58.5)
Melbourne, Australia	437	39.4±14.3	284(65.0)	245(56.1)	253(57.9)
Porto Alegre, Brazil	391	39.9±13.6	294(75.2)	208(53.2)	340(87.0)
Seattle, USA	366	41.8±15.0	245(66.9)	175(47.8)	245(67.1)
St Petersburg, Russia	310	47.0±16.2	224(72.3)	167(53.9)	50(16.1)

<sup>a</sup> Longitudinal Investigation of Depression Outcomes; <sup>b</sup> CES-D score  $\geq 16$  and positive CIDI for major depression (DSM-IV criteria)



**Table 2: Differential Item Functioning (DIF) analysis in LIDO <sup>a</sup> baseline sample N=2359 for depression factor <sup>b</sup> WHOQOL-Bref domains**

Overall	F	P	Physical Domain	F	p	Psychological domain	F	p	Environment domain	F	p	Social domain	F	p
Overall QOL	5.05	0.02	Pain*	24.5	0.00001	Positive Feelings	4.2	0.04	Safety*	41.0	0.00005	Relationships*	47.3	0.00001
General Health	0.42	0.52	Medication*	55.8	0.00002	Meaning in life	1.9	0.17	Environment	0.4	0.51	Sex	0.7	0.42
			Energy*	57.4	0.00003	Think	0.2	0.65	Finances	0.3	0.65	Support*	12.9	0.0003
			Mobility	8.7	0.003	Body*	55.0	0.00005	Information	0.1	0.77			
			Sleep*#	17.8	0.00002	Esteem	4.0	0.05	Leisure	1.98	0.16			
			Daily Life Activities*	97.6	0.00005	Negative Feelings*	98.0	0.00001	Home	0.2	0.66			
			Work	2.1	0.15				Services	10.6	0.001			
									Transport*	27.7	0.00001			

<sup>a</sup> Longitudinal Investigation of Depression Outcomes; <sup>b</sup> CES-D score  $\geq 16$  and positive CIDI for major depression (DSM-IV criteria)

\* Uniform DIF; 2 way ANOVA; P <0.0001; # Non-uniform DIF; 2 way ANOVA; P <0.0001

**Table 3: Theoretical QOL and depression domains stemmed from WHOQOL-Bref original items**

<b>Content</b>	<b>QOL items (4/12)</b>	<b>Content</b>	<b>Depression items§ (7/14)</b>
<b>Overall QOL</b>	How would you rate your quality of life?	<b>Pain*</b>	To what extent do you feel pain prevents you from doing what you need to do?
<b>General Health</b>	How satisfied are you with your health?	<b>Positive Feelings</b>	How much do you enjoy life?
<b>Medication*</b>	How much do you need any medical treatment to function in your daily life?	<b>Meaning in life</b>	To what extent do you feel your life to be meaningful?
<b>Safety*</b>	How safe do you feel in your daily life?	<b>Think</b>	How well are you able to concentrate?
<b>Environment</b>	How healthy is your physical environment?	<b>Energy*</b>	Do you have enough energy for everyday life?
<b>Body*</b>	Are you able to accept your bodily appearance?	<b>Leisure</b>	To what extent do you have the opportunity for leisure activities?
<b>Finances</b>	Have you enough money to meet your needs?	<b>Sleep*#</b>	How satisfied are you with your sleep?
<b>Information</b>	How available to you is the information that you need in your day-to-day life?	<b>Daily Life Activities*</b>	How satisfied are you with your ability to perform your daily living activities?
<b>Mobility</b>	How well are you able to get around?	<b>Work</b>	How satisfied are you with your capacity for work?
<b>Home</b>	How satisfied are you with the conditions of your living place?	<b>Esteem</b>	How satisfied are you with yourself?
<b>Services</b>	How satisfied are you with your access to health services?	<b>Sex</b>	How satisfied are you with your sex life?
<b>Transport*</b>	How satisfied are you with your transport?	<b>Relationships*</b>	How satisfied are you with your personal relationships?
		<b>Support*</b>	How satisfied are you with the support you get from your friends?
		<b>Negative Feelings*</b>	How often do you have negative feelings, i.e. blue mood, despair, depression?

§ DSM-IV criterion for depression and/or Hamilton Rating Scale for Depression items

\* Uniform DIF; 2 way ANOVA; P <0.0001; # Non-uniform DIF; 2 way ANOVA; P <0.001

### 6.3 ESTUDO 3

#### **Cross-cultural evaluation of the WHOQOL-Bref domains in primary care depressed patients using Rasch Analysis**

**Neusa Sica da Rocha,<sup>1,2</sup> Marcelo Pio de Almeida Fleck,<sup>1,3</sup>  
Mick J. Power,<sup>4</sup> Donald M. Bushnell<sup>5</sup>**

<sup>1</sup>Hospital de Clínicas de Porto Alegre, Graduate Program in Medical Sciences: Psychiatry-UFRGS-Brazil; <sup>2</sup>Psychiatrist, Graduate Program in Medical Sciences: Psychiatry-UFRGS-Brazil; <sup>3</sup>Adjunct Professor, Department of Psychiatry and Legal Medicine-UFRGS-Brazil; <sup>4</sup>Professor, Section of Health and Clinical Psychology, University of Edinburgh-UK & Professor II, Dept of Psychology, University of Tromso-Norway; <sup>5</sup>Associate Director, Health Research Associates, USA

Corresponding author:

Neusa Sica da Rocha  
Avenida Ramiro Barcelos, 2350 4º andar  
Psychiatric Service  
Porto Alegre-RS-Brazil  
CEP 90003-035  
Phone: +55 51 21018413  
e-mail: [neusa-rocha@via-rs.net](mailto:neusa-rocha@via-rs.net)

This study was partially funded by FIPE-HCPA (Brazil), CAPES (Brazil) and the University of Edinburgh (United Kingdom). The authors acknowledge Professor Alan Tennant from University of Leeds for his commitment on teaching and revising methodological issues while this manuscript was being written.

## **ABSTRACT**

**Objective:** Our objective was to test the Rasch properties of the World Health Organization Quality of Life Instrument - Abbreviated version (WHOQOL-Bref) domains in depressed patients from primary care settings.

**Subjects:** The sample consisted of 1,193 subjects having a confirmed diagnosis of depression from six countries (Australia, Brazil, Israel, Russia, Spain, and the USA) involved in the Longitudinal Investigation of Depression Outcomes (LIDO) study.

**Measures:** We used the following measurement instruments: the Center for Epidemiological Studies Depression Scale (CES-D) to assess severity of depression; the World Health Organization Quality of Life Instrument - Abbreviated version (WHOQOL-Bref) was used as a generic quality of life (QOL) instrument; and the Composite International Diagnostic Interview (CIDI), version 2.1 for the diagnosis of depression.

**Results:** The Physical, Psychological and Environment domains of the WHOQOL-Bref required adjustments to conform to the Rasch model expectations, mainly because of individual item misfit, or Differential Item Functioning (DIF) due to age. The Social domain maintained its poor psychometric properties evidenced by previous studies. Only five items were totally invariant across countries, but the bias of other items disappeared when pooled, indicating cancellation effects.

**Conclusions:** Our findings support the Rasch properties of the domains of the WHOQOL-Bref, with some modification, as a measure of generic subjective QOL in the context of primary care depressed patients in six countries worldwide.

## INTRODUCTION

The WHOQOL-Bref, the abbreviated generic measure of quality of life (QOL) developed by the World Health Organization, was developed simultaneously in several cultures and languages under the framework of Classical Test Theory. It has 26 items covering four domains: Physical, Psychological, Social Relationships and Environment. Since it can be considered a measure of subjective QOL, some authors have suggested that its assessment should be checked for the influence of depressive symptomatology on the QOL score (Pyne, Patterson *et al.*, 1997; Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Naumann e Byrne, 2004; Papakostas, Petersen *et al.*, 2004; Aigner, Forster-Streffleur *et al.*, 2006; Ay-Woan, Sarah *et al.*, 2006). Moreover, it is well placed to examine the influence of culture on QOL, but its cross-cultural validity in depressed individuals remains unstudied.

Quality of life and depression have a complex conceptual relationship (Katsching, 1997; Berlim e Fleck, 2007). To our knowledge, there is no study addressing the potential overlap of QOL and depression measuring using a cross-cultural database. The Longitudinal Investigation of Depression Outcomes (LIDO) study was a multicenter, cross-national observational study that followed patients with depressive disorders in primary care settings for 12 months in six countries. All LIDO publications that aimed at assessing QOL (measured by the WHOQOL-Bref and the Quality of Life Depression Scale (QLDS)), concluded that depressed participants present poor QOL, even when other variables are taken into account such as demographic variables, health status, and severity of depression (Fleck, Lima *et al.*, 2002; Herrman, Patrick *et al.*, 2002; Bech, Lucas *et al.*, 2003; Diehr, Derleth *et al.*, 2006; Da Silva Lima e

De Almeida Fleck, 2007). The study of Fleck and colleagues (De Almeida Fleck, Simon *et al.*, 2005) identified that only QOL assessed by QLDS, life events and educational level were significant predictors of complete remission. Neither depressive symptoms, nor general QOL assessed by the WHOQOL-Bref domains appeared as significant remission factors.

Potentially, the LIDO database allows us to test empirically the relationship between depression and QOL measurement. The purpose of this paper is to test the measurement properties of the WHOQOL-Bref domains and the invariance of item measures in cross-cultural settings.

## **METHOD**

The design, instruments and methodology, including sample size estimation of the LIDO study, are described in detail elsewhere (Chisholm, Amir *et al.*, 2001; Herrman, Patrick *et al.*, 2002). In summary, patients attending primary care centers in six sites (Barcelona, Spain; Be'er Sheva, Israel; Melbourne, Australia; Porto Alegre, Brazil; Seattle, USA; and St. Petersburg, Russia) were screened for symptoms of depression. Those meeting inclusion criteria – a new and/or untreated episode and a score of over 16 on the Centre for Epidemiological Studies - Depression scale (CES-D) (Radloff, 1977) – were interviewed and assessed using a standardized diagnostic instrument for major depression, the Composite International Diagnostic Interview (CIDI) (Robins, Wing *et al.*, 1988).

### **1. Sample Selection**

For inclusion in the LIDO study, participants had to be 18-75 years old, be a patient in a participating primary care setting, and meet the CIDI criteria for current major depression. They also had to be able and willing to participate in planned visits and/or study-required contacts; provide adequate contact details to ensure follow-up; give written informed consent; read, understand and complete the self-administered surveys in the primary language at the site; and plan to be available for the 12 months of the study. Individuals were not included if they: (a) were currently receiving treatment for depression or had been treated for depression during the past 3 months; (b) had been diagnosed with a major psychiatric disorder or psychosis; (c) had a diagnosis of dementia, Alzheimer's disease or organic brain syndrome.

Selection of primary care settings by investigators was pragmatic, and based primarily on good working relationships with the primary care physicians and clinic managers (Herrman, Patrick *et al.*, 2002).

## 2. Measures

**a) Demographic characteristics.** Continuous variables were age and years of education; categorical variables were gender, marital status, employment, and self-reported health.

**b) Severity of depression.** The CES-D is a 20-item scale designed to measure symptoms of depression in community populations (Radloff, 1977).

**c) Generic Quality of Life.** The World Health Organization Quality of Life Instrument - Abbreviated version (WHOQOL-Bref) (Whoqol, 1998; Skevington, Lotfy *et al.*, 2004) is a 26-item instrument based on the larger WHOQOL-100, a multilingual assessment for generic quality of life, which was developed concurrently across 15 international field centers. The items of the WHOQOL-Bref form four domains (physical, psychological, social and environment) and are answered using individualized five-point scales. Each subscale is scored positively. This instrument was developed according to the WHO's definition of QOL as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" (WHOQOL Group, 1995).



**d) Depression diagnostic instrument.** The Composite International Diagnostic Interview, version 2.1 (CIDI) (Weiller, Lecrubier *et al.*, 1994) is a completely structured psychiatric diagnostic assessment developed for use in cross-national epidemiologic studies. Data from the CIDI were used to assess diagnostic criteria for depression from the American Psychiatric Association DSM-IV (American Psychiatric, 1994).

All the measures were obtained at the same time. A complete description of LIDO baseline measures can be found at Chisholm *et al.*, 2001 (Chisholm, Amir *et al.*, 2001). We opted to describe here specifically those instruments whose results were analyzed for the purposes of this paper.

### **3. Statistical Analysis**

#### **Rasch Analysis**

The Rasch methodology used in this paper is detailed elsewhere (Eyres, Carey *et al.*, 2005) and has been increasingly used. (Tennant, Mckenna *et al.*, 2004; Lawton, Lundgren-Nilsson *et al.*, 2006; Pallant, Miller *et al.*, 2006; Slade, Fear *et al.*, 2006; Tennant, Kucukdeveci *et al.*, 2006; Conaghan, Emerton *et al.*, 2007; Covic, Pallant *et al.*, 2007; Keenan, Redmond *et al.*, 2007; Kutlay, Kucukdeveci *et al.*, 2007; Mckenna, Doward *et al.*, 2007; Pallant e Tennant, 2007; Tennant e Conaghan, 2007; Rocha e Fleck, 2008). Briefly, the Rasch model is seen as a template which puts into operation the axioms of additive conjoint measurement (Pallant, Miller *et al.*, 2006). This model presents a set of analyses to obtain a variable with additive structure and, hence is suitable to be measured on an interval scale (Rasch, 1960). Dichotomous models can be extended to analyze items with more than two categories (polytomous), and this

involves a “threshold” parameter, where the threshold represents the equal probability point between any two adjacent categories within an item. The model used in the present analysis is a further derivation, the Partial Credit Model (Masters, 1982).

Three overall fit statistics are considered to determine model fit. Two are item-person interaction statistics distributed as a Z statistic with mean of zero and standard deviation (SD) of 1 (indicates perfect fit to the model). A third is an item-trait interaction statistic reported as  $\chi^2$ , reflecting the invariance across the trait (indicated by a non-significant  $\chi^2$ ). , In addition, individual item-fit statistics are presented as residuals (acceptable within the range  $\pm 2.5$ ) and as a  $\chi^2$  statistic (required also a non-significant  $\chi^2$ ).

An estimate of the internal consistency reliability of the scale was also provided, based on the Person Separation Index (PSI), where the estimates on the logit scale for each person were used to calculate reliability (Tennant, McKenna *et al.*, 2004). A value of 0.7 would indicate the ability to differentiate two groups and 0.8 three groups (Fisher, 1992). Traditionally, values above 0.7 would be adequate for group comparison, and above 0.85 for individual use (Streiner e Norman, 2003).

Some sources of deviation of model expectations can be identified, thus the scale can be modified and improved to better model fit. In this Rasch analysis, first, because we are working on polytomous items, they were analyzed in order to check if their response categories were ordered. The boundaries between categories are called “thresholds” (Andrich, 1978). “Disordered thresholds” may indicate that it is necessary to collapse adjacent categories. Usually, collapsing of categories where disordered thresholds occur

improves overall fit for the model, it may eliminate the inversion or disorder, and also improve psychometric properties of the scale (Bond & Fox, 2007).

Secondly, we assessed the presence of differential item functioning (DIF). The basis of the DIF approach lies in the item logistic function, e.g. the proportion of individuals at the same ability level who answer a given item correctly. To ensure the invariance of a measure, a scale should work in the same way, irrespective of which group is assessed. That is, individuals with the same score on the test are expected to have the same probability of response to an item, irrespective of any group membership. The statistical test used for detecting DIF is an ANOVA of the person-item deviation residuals with person (e.g., age, gender, country) and class intervals (e.g., group along the trait) as factors. Two kinds of DIF can be identified: uniform and nonuniform. In the first, there is a constant difference between groups in the probability of affirming an item across the trait, and in the second the difference varies across the trait. DIF analysis is most commonly used for assessing cross-cultural validity (Tennant, Penta *et al.*, 2004). Once there are 6 countries involved in the study, we use the Tukey test as a post hoc test to perform a pair-wise comparison of the means in order to identify where the difference(s) lie(s) among countries. Finally, creating a subset of “pure” items (free from DIF) and a subset of “impure” items (showing DIF), we ran a subtest analysis (anchoring) comparing both subsets and checked if the data fitted the model. We also checked DIF again to verify if they were cancelled out. If so, we considered it an invariant measure, despite the fact of having DIF initially. When items displayed DIF for more than one factor and had a poor fit to the model, adjustments were made to allow these items to vary by group, using the split procedure and deletion of

biased items. This procedure aims to remove items having the poorest fit to the Rasch model and the approach is repeated until an unbiased set of item is identified. For example, if the 'Pain' item has DIF for age, as a consequence, we could have 2 forms of counting the 'Pain' item, one for people >40 years and another for people <40 years old. These two forms of counting the item are re-analysed together with unsplit items. Fit is again reassessed, and if the item still displays misfit to the model, then they are removed from the scale and do not contribute to the person estimate (Tennant, Penta *et al.*, 2004).

Thirdly, looking at the residual correlation matrix, we considered whether or not any values exceeded + 0.3. This indicates the presence of local dependency. Thus, when the data fit the model, and the requirements of local independence are met, a unidimensional linear scale is derived from the ordinal raw score, and the scale is appropriate for valid parametric approaches (Wright e Stone, 1979).

Finally, after issues of threshold disordering, DIF, local independence and fit have been solved, a Principal Components Analysis (PCA) of the residuals detects any signs of multidimensionality (Smith, 2002). The absence of any meaningful pattern in the residuals will support the assumption of the unidimensionality of the scale. This is tested by allowing the correlation between items and the first factor to determine subsets of items and then testing, using a series of independent t-tests, to see if the person estimates derived from each subset significantly differ. For a unidimensional scale it would be expected that, given the difference in estimates are normally distributed, no more than 5% of such tests would be outside the range  $\pm 1.96$  (Tennant e Pallant, 2006).

### **Sample Size Estimation**

A sample size of 150 individuals is able to estimate the item difficulty, with  $\alpha = 0.01$  to within  $\pm 0.5$  logits. This size is also suitable to test for DIF, where a difference of 0.05 SD within the residuals for any two groups with  $\beta$  of 0.20 at an  $\alpha$  of 0.01 to be detected, and with the same  $\alpha$  and  $\beta$ , a difference in SD within the residuals can be identified where this difference varies by 0.1 logits across each of the six groups (Linacre, 1994).

Rasch analysis was undertaken using the RUMM 2020 package (Andrich, Sheridan *et al.*, 2004).

## RESULTS

A total sample of 2,359 patients (51 % of eligible cases) participated in the baseline assessment. Compared to those completing baseline assessment, those declining to participate were slightly younger (mean age 40.5 yrs vs. 41.6,  $P= 0.02$ ), less often female (65% vs. 68.6%,  $P\leq 0.01$ ), and had lower CES-D scores (24.7 vs. 26.4,  $P< 0.01$ ). Of the 2,359 completing the baseline assessment, 1,193 (51%) patients with current clinical depression (i.e. met DSM-IV criteria for major depressive episode) were invited to enrol in the longitudinal study. Compared to those not meeting criteria for major depressive episode ( $N=1166$ ; 49%), depressed patients had poorer QOL in all WHOQOL-Bref domains as evidenced by mean values of total scores of each domain (physical: 49.9 vs. 62.7,  $P\leq 0.01$ ; psychological: 45.6 vs. 60.7,  $P\leq 0.01$ ; environment 51.1 vs. 59.4,  $P\leq 0.01$ ; social 48.5 vs. 61.1,  $P\leq 0.01$ ). Data presented in table 1 provide summarized descriptions of the baseline sample in terms of age, gender, self-reported health and general QOL, severity of depression, and sample sizes. The sample comprised predominantly female, middle aged patients in good physical health, with moderate levels of depression, and presenting a broad range of self-rated general QOL that varied among countries. Patients having poor and very poor QOL varied from 8% in Seattle to 37.7 % in St. Petersburg.

**Insert Table 1 here**

## 1. Rasch analysis of the WHOQOL-Bref Domains

Preliminary analysis showed that the WHOQOL-Bref domains did not meet the expectations of the Rasch model, indicating that they required adjustments. As shown in table 2, overall measures of model fit for each domain pointed to misfit: physical (total item  $\chi^2 = 161.5$ ;  $P = 0.0001$ ; item fit residual = 0.49; SD = 2.59; person fit residual = -0.37; SD = 1.23); psychological (total item  $\chi^2 = 123.5$ ;  $P = 0.0001$ ; item fit residual = 0.57; SD 2.86; person fit residual = -0.41; SD = 1.25); environment (total item  $\chi^2 = 101.8$ ;  $P = 0.01$ ; item fit residual = 0.70; SD = 1.23; person fit residual = -0.37; SD = 1.31), and social (item fit residual = 0.93; SD = 2.11; person fit residual = -0.49; SD = 1.07; total item  $\chi^2 = 65.9$ ;  $P = 0.0001$ ). In addition, physical, psychological and environment domains had at least one item not fitting the model: sleep (residual = 3.5;  $\chi^2 = 21.98$ ;  $P = 0.01$ ); activity (residual = -3.46;  $\chi^2 = 51.39$ ;  $P = 0.0001$ ); think (residual = 2.84;  $\chi^2 = 25.57$ ;  $P = 0.0001$ ); body (residual = 4.54;  $\chi^2 = 31.33$ ;  $P = 0.0001$ ) esteem (residual = -2.5;  $\chi^2 = 22.53$ ;  $P = 0.0001$ ) and transport (residual = 2.87;  $\chi^2 = 16.73$ ;  $P = 0.05$ ). The property of unidimensionality was violated in physical (t-test  $P = 0.08$ ) and environment domains (t-test  $P = 0.06$ ). Internal consistency reliability estimates varied among domains: physical (PSI = 0.78), psychological (PSI = 0.74), environment (PSI = 0.73), and social (PSI = 0.58). The majority of WHOQOL-Bref items were found to have an additive structure and an interval level scale, only two items needed to be rescored: Sleep and Home, because of disordered thresholds of category responses.

**Insert Table 2 here**

All items were checked for DIF for age, gender and country, as shown in table 3. Some items showed DIF for two factors: medication (age/country), positive feelings (age/country), think (age/country), body (gender/country), and activity (age/country). Items mostly displayed DIF for country (twenty of 26 items), age (four of 26 items) and gender (one of 26 items). No item showed DIF for more than two factors.

**Insert Table 3 here**

## **2. WHOQOL-Bref domains after Rasch Adjustments**

Since all domains did not meet model requirements, data presented below correspond to Rasch adjustments using each stage of the analysis protocol described before.

**Insert Table 4 here**

### **a) Physical Domain Items**

Deleting the items Sleep and Activity for people < 40 years old resulted in the best overall measures of fit for the physical domain, where the item fit residual modified from 0.49 (SD 2.59) to 0.58 (SD 1.05) ; person fit residual from -0.37 (SD 1.23) to -0.35 (SD 1.14); total  $\chi^2$  from 161.5 to 62.5; P from 0.0001 to 0.19. The Sleep item was rescored, collapsing category responses by suppressing the middle category. This item was removed because after rescoring, it persisted in showing misfit (residual = 3.56;  $\chi^2 = 21.8$ ; P = 0.01). No item from the physical domain had correlations > 0.3 on the person item



residual correlation matrix, indicating local independency of responses. Because the Activity item displayed DIF for age, it was split and Activity for people < 40 years old was removed (residual = -3.59;  $\chi^2 = 35.8$ ;  $P = 0.0001$ ). However, even splitting the Medication item by age group (overall  $\chi^2 = 104.9$ ;  $P = 0.0007$ ), or removing this item for people < 40 years (residual = 0.21;  $\chi^2 = 25.9$ ;  $P = 0.002$ ), did not improve overall measures of fit (total item  $\chi^2 = 71.5$ ;  $P = 0.06$ ; item fit residual = 0.72; SD = 1.18; person fit residual = -0.39; SD = 1.18), as compared to the best solution. For the best Rasch adjustment for the physical domain, the probability of the test of unidimensionality improved from 0.08 to 0.05. The purification procedure resulted in canceling out DIF for age. Internal consistency reliability estimate (PSI) modified from 0.78 to 0.75. As can be seen from the item-person location map (Figure 1a), physical domain items located from -0.65 to 0.50 logit, while persons from -4.18 to 4.08 logits, but the percentage outside the range  $\pm 1$  logit was 17.7%, then this domain access 83.3% of depressed patients.

### **b) Psychological Domain Items**

Because the original psychological domain did not fit the Rasch model as shown in table 3, and the Body item showed the worst fit (residual = 4.54;  $\chi^2 = 31.3$ ;  $P = 0.001$ ) and displayed DIF for gender (uniform and non-uniform), it was removed from the following analysis. No item displayed correlations  $> +0.3$  at Residual Correlation Matrix pointing to local independence of responses. The Think item displayed DIF for age, it was split, and when the Think item for people > 40 years old was removed (residual = 3.43;  $\chi^2 = 21.07$ ;  $P = 0.01$ ), overall measures of fit improved. The item fit residual modified from 0.57 (SD

2.86) to 0.51 (SD 1.5); person fit residual from -0.41 (SD 1.25) to -0.48 (SD 1.22);  $\chi^2 = 123.5$  to 54.1; P from 0.0001 to 0.16. Besides the previous procedure, splitting the Positive Feelings item for age proved to be the best Rasch solution for the psychological domain (item fit residual modified from 0.51 (SD 1.5) to 0.38 (SD 1.27); person fit residual from -0.48 (SD 1.22) to -0.47 (SD 1.14);  $\chi^2 = 54.1$  to 50.4; P from 0.16 to 0.57). Purification procedure cancelled out DIF for age. The final solution compared to the initial one maintained equal internal consistency reliability estimate of 0.75. When we split items in the final solution, the test of unidimensionality could not be run. As can be seen from the item-person location map (Figure 1b), psychological domain items located from -0.52 to 0.32 logit, while persons from -4.83 to 3.73 logit, but the percentage outside the range  $\pm 1$  logit was 17.7%, then this domain assesses 83.3% of depressed patients.

### **c) Environment Domain Items**

Removing the Transport item, rescored Home item (collapsing category responses by suppressing the middle category) improved overall measures of fit, because the item fit residual modified from 0.70 (SD 1.23) to 0.52 (SD 1.12); person fit residual from -0.37 (SD 1.31) to -0.38 (1.27); total  $\chi^2$  from 101.8 to 80.57; P from 0.01 to 0.07. The residual correlation matrix did not display any correlation at  $\pm 0.3$ . The probability of the test of unidimensionality improved from 0.06 to 0.05. Internal consistency reliability estimate (PSI) modified from 0.73 to 0.71. As can be seen from the item-person location map (Figure 1c), items located from -0.59 to 0.57 logit, while persons from -3.30 to 3.36 logits, but the

percentage outside the range  $\pm 1$  logit was 19.1%, then this domain access 80.9% of depressed patients.

#### **d) Social Domain Items**

Although the Individual item fit of the Social domain items did not show any item displaying a fit residual  $> 2.5$ , overall fit statistics (item fit residual = 0.93 (SD 2.11); person fit residual = -0.49 (SD 1.07); total  $\chi^2 = 65.9$ ;  $P = 0.00001$ ) pointed to a misfit to the model. The probability of the test of unidimensionality was 0.01. The internal consistency reliability estimate (PSI) was 0.58. As can be seen from the item-person location map (Figure 1d), items located from -0.16 to 0.30 logits, while persons from -1.99 to 2.17 logits, and the percentage outside the range  $\pm 1$  logit was 25.4%, then this domain access 74.6% of depressed patients.

**Insert Figure 1 here**

### **3. Cross-cultural invariance of the WHOQOL-Bref items-results from DIF Analysis**

#### **a) Results from each country**

The WHOQOL-Bref domains showed a good fit to the Rasch model, when each country sample was analyzed separately. The most significant adjustment done was rescoring, which needed to be done for all samples for different items. The Sleep item had to be removed from all samples, except for the USA and Israel. Local dependency of responses was identified between the Activity and Work items for the Brazilian (Rocha e Fleck, 2008) and Spanish

samples. The same was true for the Pain and Medication items for the Russian data.

### **b) Results of pooling data**

Using only ANOVA, we found that 15 items showed DIF among different countries. If we look at post hoc analysis, we can observe that most items have DIF only between one or two countries versus the remainder. Only one item displayed DIF for more than two countries. Five items were free from any DIF due to country. The Tukey test was not carried out for split or deleted items, because the analysis program does not provide required estimates for this calculus.

### **Insert Table 5 here**

Looking at individual countries, three countries scored differently for two items: Russia (Pain, Information), Spain (Safety, Finances) and Israel (Environment, Esteem). Brazil (Negative Feelings), the USA (Energy), and Australia (Services) had different scores in relation to others for one item.

The USA and Australia differed between each other for three items (Work, Home, and Relationships). No other pair of countries differed between each other more than once.

After purification procedure using subtest analysis, comparing pure and impure items, all DIF for country was cancelled out for all domains, as shown by overall measures of model fit for each domain: physical (total item  $\chi^2 = 17.2$ ;  $P = 0.51$ ; item fit residual = 0.26; SD = 3.46; person fit residual = -0.64; SD = 0.67); psychological (total item  $\chi^2 = 14.4$ ;  $P = 0.64$ ; item fit residual = -0.45; SD = 3.47;

person fit residual = -0.60; SD = 0.62); environment (total item  $\chi^2 = 5.8$ ; P = 0.99; item fit residual = -3.54; SD = 14.78; person fit residual = -0.55; SD = 0.97), and social (item fit residual = -0.001; SD = 6.52; person fit residual = -0.59; SD = 1.12; total item  $\chi^2 = 29.9$ ; P = 0.02).

## DISCUSSION

Our findings point to the measurement properties of the WHOQOL-Bref domains as a measure of generic QOL and invariance of items using Rasch analysis in the context of primary care depressed patients in six countries worldwide. This is the first study to assess this issue in a population with such characteristics. Other studies that had a similar purpose used traditional test theory, reduced sample sizes, or patients from a clinical setting from only one culture (Atkinson, Zibin *et al.*, 1997; Kuehner, 2002; Naumann e Byrne, 2004; Berlim, Pavanello *et al.*, 2005), thereby limiting generalization of their results. Although several studies have affirmed a possible overlap between depression and QOL constructs, this may be not so evident in our results, when we analyze the functioning of each item, with the focus on the instrument as a whole.

Another relevant finding is the influence of age as an intervening factor on item functioning, in agreement with some current research. The importance of diagnosis of depression in old age is well-recognized. Therefore, any QOL measure should maintain this correspondence by using specific measures for these populations. This has been a concern of some authors who are working on the design of specific instruments for older people (Power, Quinn *et al.*, 2005) and for children (Schmidt, Debensason *et al.*, 2006).

The item Body Image has been excluded, despite the fact that it fit the Rasch model when the WHOQOL-Bref was analyzed using individual country samples. A possible explanation is that the concept of body image cannot be compared between cultures and gender for depressed primary care patients. The exclusion of the Sleep item may be related to its poor fitting that is not related to cross-cultural issues, once it functioned poorly at country level as well

as when the entire sample was analyzed. The exclusion of the Transport item must be seen with caution, because this item had a good fit at country level and its exclusion did not much improve overall measures of model fit.

The dilemma of deciding on whether or not to delete items can, in some way, compromise construct validity if relevant content from the conceptual framework are likely to be excluded. Our challenge as researchers is to reconcile the conceptual model and the empirical evidence. Although inclusion of these excluded items has a conceptual motivation, this was not revealed in this empirical study, in agreement with other similar recent data (Wang, Yao *et al.*, 2006). Obviously, exclusion\inclusion of items reduces comparability of measures between different populations. And again we have to face the need to make concessions depending on the research purpose. Other studies have found relevant issues related to mental health and social adjustment that could be considered in QOL measures for depression, such as issues associated with stigma, barriers to treatment (Saxena, Jane-Llopis *et al.*, 2006), existential well-being (autonomy, freedom) and religiosity, which are not part of the short form of the WHOQOL. The inclusion of these aspects could broaden the spectrum of measurement of the instrument (i.e., items presenting locations outside  $\pm 1$ ); moreover, it could improve the reliability and the validity (Smith, 2001), mainly by improving the social domain's inadequate psychometric performance.

The WHOQOL-Bref has proven to be a good cross-cultural measure of QOL in depressed patients. This conclusion can be seen by the number of items that were free from DIF (five items) and others that displayed DIF for only one country or a pair of countries (sixteen items). Only one item related to relationships was affected by more than two countries. These findings point to

the cultural equivalence of this measure. Another study that had a similar purpose, found that only one item of the WHOQOL-100 had cross-cultural equivalence among the four cultures studied (France, Hong Kong, Argentina and the United Kingdom) using the Rasch model (Leplege, 2000). There are a large number of statistical procedures to examine item cultural equivalence of tests. Item response theory is one of the three most known and useful to identify item bias due to cultural differences (Hambleton e Patsula, 1998).

The five items that were equivalent for all countries (general health, meaning in life, leisure, sex and support) are almost all relational concepts, the exception being general health. This reinforces the importance, at least for the study of QOL in depressed patients, of the humanistic perspective of this concept. Our findings are in agreement with those of authors who have argued that a truly humanistic measure of QOL should take into account people's necessity for being free, moving its concept more to an existential than utilitarian approach of human needs (Leplege e Hunt, 1997).

The present study did not retest the original four-domain structure of the WHOQOL-Bref, which may be a limitation, regardless of the fact that there is no consensus about the most suitable method to evaluate instrument dimensionality (Coste, Bouee *et al.*, 2005). As a consequence, we decided to maintain the four-domain structure, because besides being the most conservative approach, it is the most studied one (Yao e Wu, 2005). To analyze all WHOQOL-Bref items together to see if they form a unidimensional and meaningful measure would not be a reasonable strategy; future investigations may be necessary to test the multidimensional structure of WHOQOL-Bref under a multidimensional Rasch model.



A valid measure of QOL in depressed patients has important clinical implications. It allows the evaluation of relevant facets that are not part of the patient's routine assessment by providing some additional information that could contribute to better understanding and diagnosis(Trivedi, Rush *et al.*, 2006). We consider that clinical improvement depends on other factors that are not accessible by most commonly used therapeutic approaches, as well as exclusive symptomatic assessment does not reflect a reliable and valid evaluation of patients' improvement or impairment during their treatment.

Finally, a generic QOL measure in addition to being valid and easily applicable for outcome, can contribute to advances, as well as to the understanding of new dimensions that can be the target for care of the depressed patient.

## REFERENCES

- Aigner, M., S. Forster-Streffleur, *et al.* What does the WHOQOL-Bref measure? Measurement overlap between quality of life and depressive symptomatology in chronic somatoform pain disorder. Soc.Psychiatry Psychiatr Epidemiol., v.41, n.1, p.81-86. 2006.
- American Psychiatric, A. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association. 1994.
- Andrich, D. A rating formulation for ordered response categories. Psychometrika, v.43, p.561-573. 1978.
- Andrich, D., B. Sheridan, *et al.* RUMM: a Windows program for analysing item response data according to Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM Laboratory 2004.
- Angermeyer, M. C., A. Holzinger, *et al.* Depression and quality of life: results of a follow-up study. Int J Soc.Psychiatry, v.48, n.3, p.189-199. 2002.
- Atkinson, M., S. Zibin, *et al.* Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. Am.J Psychiatry, v.154, n.1, p.99-105. 1997.
- Ay-Woan, P., C. P. Sarah, *et al.* Quality of life in depression: predictive models. Qual.Life Res, v.15, n.1, p.39-48. 2006.
- Bech, P., R. Lucas, *et al.* Association between clinically depressed subgroups, type of treatment and patient retention in the LIDO study. Psychol.Med., v.33, n.6, p.1051-1059. 2003.
- Berlim, M. T. e M. Fleck. Quality of life and major depression. In: M. S. Ritsner e A. G. Awad (Ed.). Quality of life in Schizophrenia, Mood and Anxiety Disorders. Dorcrecht: Springer, 2007. Quality of life and major depression, p.241-252.
- Berlim, M. T., D. P. Pavanello, *et al.* Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual.Life Res, v.14, n.2, p.561-564. 2005.
- Bond, T. G. e C. M. Fox. Applying the Rasch model- Fundamental measurement in the human sciences. Mahwah: Lawrence Erlbaum Associates. 2007.
- Chisholm, D., M. Amir, *et al.* Longitudinal Investigation of Depression Outcomes (The LIDO Study) in primary care in six countries: comparative assessment of local health systems and resource utilization. Int J Methods Psychiatr Res, v.10, n.2, p.59-71. 2001.

Conaghan, P. G., M. Emerton, *et al.* Internal construct validity of the Oxford Knee Scale: evidence from Rasch measurement. Arthritis Rheum, v.57, n.8, Dec 15, p.1363-7. 2007.

Coste, J., S. Bouee, *et al.* Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. Qual.Life Res, v.14, n.3, p.641-654. 2005.

Covic, T., J. F. Pallant, *et al.* A longitudinal evaluation of the Center for Epidemiologic Studies-Depression scale (CES-D) in a rheumatoid arthritis population using Rasch analysis. Health Qual Life Outcomes, v.5, p.41. 2007.

Da Silva Lima, A. F. e M. P. De Almeida Fleck. Subsyndromal depression: an impact on quality of life? J Affect Disord, v.100, n.1-3, Jun, p.163-9. 2007.

De Almeida Fleck, M. P., G. Simon, *et al.* Major depression and its correlates in primary care settings in six countries. 9-month follow-up study. Br.J.Psychiatry, v.186, p.41-47. 2005.

Diehr, P. H., A. M. Derleth, *et al.* Synchrony of change in depressive symptoms, health status, and quality of life in persons with clinical depression. Health Qual.Life Outcomes, v.4, p.27. 2006.

Eyres, S., A. Carey, *et al.* Construct validity and reliability of the Rivermead Post-Concussion Symptoms Questionnaire. Clin Rehabil, v.19, n.8, Dec, p.878-87. 2005.

Fisher, W. P. Reliability statistics. Rasch Measurement Transations, v.6, p.238. 1992.

Fleck, M. P., A. F. Lima, *et al.* [Association of depressive symptoms and social functioning in primary care service, Brazil]. Rev.Saude Publica, v.36, n.4, p.431-438. 2002.

Hambleton, R. K. e L. Patsula. Adapting tests for use in multiple languages and cultures. Social Indicators Research, v.45, p.153-171. 1998.

Herrman, H., D. L. Patrick, *et al.* Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. Psychol.Med, v.32, n.5, p.889-902. 2002.

Katsching, H. Quality of life in depression. In: (Ed.). Quality of life in mental disorders. London: John Wiley and Sons, 1997. Quality of life in depression, p.137-147.

Keenan, A. M., A. C. Redmond, *et al.* The Foot Posture Index: Rasch analysis of a novel, foot-specific outcome measure. Arch Phys Med Rehabil, v.88, n.1, Jan, p.88-93. 2007.

Kuehner, C. Subjective quality of life: validity issues with depressed patients. Acta Psychiatr Scand., v.106, n.1, p.62-70. 2002.

Kutlay, S., A. A. Kucukdeveci, *et al.* Validation of the Middlesex Elderly Assessment of Mental State (MEAMS) as a cognitive screening test in patients with acquired brain injury in Turkey. Disabil Rehabil, v.29, n.4, Feb 28, p.315-21. 2007.

Lawton, G., A. Lundgren-Nilsson, *et al.* Cross-cultural validity of FIM in spinal cord injury. Spinal Cord, v.44, n.12, Dec, p.746-52. 2006.

Leplege, A. e S. Hunt. The problem of quality of life in medicine. JAMA, v.278, n.1, p.47-50. 1997.

Leplege, A. E., E. Methodological issues in using the Rasch model to select cross culturally equivalent items in order to develop a Quality of Life index: the analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, United Kingdom). Journal of Applied Measurement v.1, p.372-392. 2000.

Linacre, J. M. Sample size and item calibration stability. Rasch Measurement Transactions, v.7, p.28. 1994.

Masters, G. N. Rasch model for partial credit scoring. Psychometrika, v.47, p.149-174. 1982.

Mckenna, S. P., L. C. Doward, *et al.* Quality of life in infants and children with atopic dermatitis: addressing issues of differential item functioning across countries in multinational clinical trials. Health Qual Life Outcomes, v.5, p.45. 2007.

Naumann, V. J. e G. J. Byrne. WHOQOL-Bref as a measure of quality of life in older patients with depression. Int Psychogeriatr., v.16, n.2, p.159-173. 2004.

Pallant, J. F., R. L. Miller, *et al.* Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. BMC Psychiatry, v.6, p.28. 2006.

Pallant, J. F. e A. Tennant. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol, v.46, n.Pt 1, Mar, p.1-18. 2007.

Papakostas, G. I., T. Petersen, *et al.* Quality of life assessments in major depressive disorder: a review of the literature. Gen.Hosp.Psychiatry, v.26, n.1, p.13-17. 2004.

Power, M., K. Quinn, *et al.* Development of the WHOQOL-old module. Qual.Life Res, v.14, n.10, p.2197-2214. 2005.

Pyne, J. M., T. L. Patterson, *et al.* Assessment of the quality of life of patients with major depression. Psychiatr Serv., v.48, n.2, p.224-230. 1997.

Radloff, L. S. The CES-D scale: a self report depression scale for research in the general population. Applied Psychological Measurement, v.1, p.385-401. 1977.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen:: Danish Institute for Educational Research. 1960.

Robins, L. N., J. Wing, *et al.* The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch.Gen.Psychiatry, v.45, n.12, p.1069-1077. 1988.

Rocha, N. S. e M. Fleck. Validity of the Brazilian version of WHOQOL-Bref in primary care depressed patients using Rasch modelling. Rev Saude Publica. 2008.

Saxena, S., E. Jane-Llopis, *et al.* Prevention of mental and behavioural disorders: implications for policy and practice. World Psychiatry, v.5, n.1, p.5-14. 2006.

Schmidt, S., D. Debensason, *et al.* The DISABKIDS generic quality of life instrument showed cross-cultural validity. J Clin.Epidemiol., v.59, n.6, p.587-598. 2006.

Skevington, S. M., M. Lotfy, *et al.* The World Health Organization's WHOQOL-Bref quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual.Life Res, v.13, n.2, p.299-310. 2004.

Slade, A., J. Fear, *et al.* Identifying patients at risk of nursing home admission: The Leeds Elderly Assessment Dependency Screening tool (LEADS). BMC Health Serv Res, v.6, p.31. 2006.

Smith, E. V. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of Applied Measurement, v.3, n.205-231. 2002.

Smith, E. V., Jr. Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. Journal of Applied Measurement, v.2, n.3, p.281-311. 2001.

Streiner, D. L. e G. R. Norman. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press. 2003.

Tennant, A. e P. G. Conaghan. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum, v.57, n.8, Dec 15, p.1358-62. 2007.

Tennant, A., A. A. Kucukdeveci, *et al.* Assessing normative cut points through differential item functioning analysis: an example from the adaptation of the Middlesex Elderly Assessment of Mental State (MEAMS) for use as a cognitive screening test in Turkey. Health Qual Life Outcomes, v.4, p.18. 2006.

Tennant, A., S. P. Mckenna, *et al.* Application of Rasch analysis in the development and application of quality of life instruments. Value.Health, v.7 Suppl 1, p.S22-S26. 2004.

Tennant, A. e J. F. Pallant. Unidimensionality matters!(a tale of two Smiths?). Rasch Measurement Transactions, v.20, n.1, Summer, p.1048-1051. 2006.

Tennant, A., M. Penta, *et al.* Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care, v.42, n.1 Suppl, p.I37-I48. 2004.

The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc.Sci.Med, v.41, n.10, p.1403-1409. 1995.

Trivedi, M. H., A. J. Rush, *et al.* Factors associated with health-related quality of life among outpatients with major depressive disorder: a STAR\*D report. J Clin Psychiatry, v.67, n.2, Feb, p.185-95. 2006.

Wang, W. C., G. Yao, *et al.* Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-Bref using multidimensional Rasch analysis. Qual.Life Res, v.15, n.4, p.607-620. 2006.

Weiller, E., Y. Lecrubier, *et al.* The relevance of recurrent brief depression in primary care. A report from the WHO project on Psychological Problems in General Health Care conducted in 14 countries. Eur.Arch.Psychiatry Clin.Neurosci., v.244, n.4, p.182-189. 1994.

The WHOQOL Group. Development of the World Health Organization WHOQOL-Bref quality of life assessment. The WHOQOL Group. Psychol.Med, v.28, n.3, p.551-558. 1998.

Wright, B. D. e M. H. Stone. Best test design. Chicago: MESA press. 1979.

Yao, G. e C. H. Wu. Factorial invariance of the WHOQOL-Bref among disease groups. Qual.Life Res, v.14, n.8, p.1881-1888. 2005.

**Table 1: General descriptions of baseline sample of depressed patients enrolled in the LIDO study**

	Total	Be'er Sheva, Israel	Barcelona, Spain	Melbourne, Australia	Porto Alegre, Brazil	Seattle, USA	St Petersburg, Russia
N	1193	184	214	245	208	175	167
Age ± SD	40.3±14.5	39.7±13.7	37.9±14.3	39.2±13.4	39.4±13.6	46.4±13.9	40.5±16.7
CES-D mean ± SD	30.1± 9.6	27.5± 8.2	29.9±9.8	30.9±9.8	31.7±9.8	30.7±10.2	29.4±9.1
Female (%)	843 (70.7)	121 (65.8)	156 (72.9)	159 (64.9)	156 (75)	123 (70.3)	128 (76.6)
Good Health (%)*	629 (52.9)	629 (52.9)	106 (49.8)	120 (49.0)	172 (82.7)	108 (62.1)	27 (16.2)
General QOL (%)*							
Very Poor	50 (4.2)	6 (3.3)	3 (1.4)	11 (4.5)	2 (1.0)	7 (4.0)	21 (12.6)
Poor	135 (11.3)	16 (8.7)	23 (10.7)	37 (15.1)	10 (4.8)	7 (4.0)	42 (25.1)
Neither poor nor good	481 (40.3)	52 (28.3)	137 (64.0)	71 (29.0)	76 (36.5)	62 (35.4)	83 (49.7)
Good	455 (38.1)	85 (46.2)	47 (22.0)	107 (43.7)	105 (50.5)	90 (51.4)	21 (12.6)
Very Good	66 (5.5)	23 (12.5)	4 (1.9)	16 (6.5)	15 (7.2)	8 (4.6)	

\*self-rated

**Table 2: Fit of the WHOQOL-Bref items to Rasch Model for baseline sample of depressed patients enrolled in the LIDO study**

Item	Location	Residual	$\chi^2$	P	Summary of overall measures of model fit statistics	
<b>Overall</b>						
Overall QOL	-0.37	1.25	4.42	0.62		
General Health	0.37	-0.66	12.63	0.05		
<b>Physical</b>						
Pain	-0.43	2.16	8.68	0.47	Item Fit Residual; SD	0.49; 2.59
Medication	-0.61	2.16	16.64	0.05	Person Fit Residual; SD	-0.37; 1.23
Energy	0.28	-0.50	14.97	0.09	Total Item $\chi^2$	161.5
Mobility	-0.85	1.82	13.01	0.16	Chi-square P	0.0001
<b>Sleep*</b>	<b>0.75</b>	<b>3.50</b>	<b>21.98</b>	<b>0.01</b>	t-test P <sup>a</sup>	0.08
<b>Activity</b>	<b>0.50</b>	<b>-3.46</b>	<b>51.39</b>	<b>0.0001</b>		
Work	0.35	-2.22	34.86	0.0001		
<b>Psychological</b>						
Positive Feelings	0.30	-1.70	19.11	0.02	Item Fit Residual, SD	0.57; 2.86
Meaning in life	-0.30	-1.48	17.44	0.04	Person Fit Residual, SD	-0.41; 1.25
<b>Think</b>	<b>-0.21</b>	<b>2.84</b>	<b>25.57</b>	<b>0.0001</b>	Total Item $\chi^2$	123.48
<b>Body</b>	<b>-0.39</b>	<b>4.54</b>	<b>31.33</b>	<b>0.0001</b>	Chi-square P	0.0001
<b>Esteem</b>	<b>0.16</b>	<b>-2.50</b>	<b>22.53</b>	<b>0.0001</b>	t-test P	0.05
Negative Feelings	0.44	1.70	7.51	0.58		
<b>Environment</b>						
Safety	0.17	0.53	9.38	0.40		
Environment	-0.01	0.73	6.74	0.66	Item Fit Residual; SD	0.70; 1.23
Finances	0.59	-0.98	9.65	0.38	Person Fit Residual; SD	-0.37; 1.31
Information	-0.54	0.17	13.52	0.14	Total Item $\chi^2$	101.8
Leisure	0.51	1.70	11.03	0.27	Chi-square P	0.01
Home*	-0.05	-0.57	27.39	0.0001	t-test P <sup>a</sup>	0.06
Services	-0.44	1.16	7.38	0.60		
<b>Transport</b>	<b>-0.23</b>	<b>2.87</b>	<b>16.73</b>	<b>0.05</b>		
<b>Social</b>						
Relationships	0.06	-1.45	51.21	0.0001	Item Fit Residual	0.93; 2.11
Sex	0.27	2.30	6.60	0.58	Person Fit Residual	-0.49; 1.07
Support	-0.33	2.13	8.04	0.43	Total Item $\chi^2$	65.9
					Chi-square P	0.0001
					t-test P	0.01

\*items that needed to be rescored because of disordered thresholds for category responses

**Bold: items that do not fit in the model.**

<sup>a</sup>P of test for unidimensionality, independent t-test probability must be 0.05 or less.



**Table 3: The WHOQOL-Bref items displaying DIF for more than one person factor for the baseline sample of depressed patients enrolled in the LIDO study**

Items	Factors		
	Age 4/26	Gender 1/26	Country 5/26
Medication	✓		Brazil vs. Russia
Positive Feelings	✓		Australia vs. USA
Think	✓		Australia vs. Russia
Body		✓*	Spain vs. all others
Activity	✓		USA vs. all others

✓uniform DIF; P <0.001

\*Nonuniform DIF; P <0.001

**Table 4: Fit of the WHOQOL items adjusted by Rasch model for baseline sample of depressed patients enrolled in the LIDO study**

Item	Location	SD	Residual	$\chi^2$	P	Summary of overall measures of model fit statistics	
Overall							
Overall QOL	-0.37		1.25	4.42	0.62		
General Health	0.37		-0.66	12.63	0.05		
Physical							
Mobility	-0.71	0.04	1.92	11.03	0.27	Item Fit Residual; SD	0.58; 1.05
Medication	-0.46	0.03	1.14	11.51	0.24	Person Fit Residual; SD	-0.35; 1.14
Pain	-0.29	0.03	0.71	5.33	0.80	Total Item $\chi^2$	62.5
Energy	0.42	0.04	1.06	8.29	0.51	Chi-square P	0.19
Work	0.49	0.03	-0.56	13.22	0.15	t-test P <sup>a</sup>	0.05
Activity >40 years	0.55	0.05	-0.8	13.18	0.15	PSI	0.75
Psychological							
Meaning in life	-0.52	0.04	-0.57	10.34	0.32	Item Fit Residual; SD	0.38; 1.33
Think <40 years	-0.16	0.06	1.82	8.11	0.52	Person Fit Residual; SD	-0.47; 1.14
Esteem	0.02	0.04	-0.02	6.36	0.70	Total Item $\chi^2$	50.5
Positive Feelings		0.06				Chi-square P	0.57
<40 years	0.03		-0.52	10.67	0.30		
Negative Feelings	0.31	0.04	2.31	4.24	0.89	PSI	0.75
Positive Feelings		0.06					
>40 years	0.32		-0.70	10.81	0.21		
Environment							
Information	-0.59	0.04	-0.16	11.62	0.24	Item Fit Residual	0.52; 1.12
Services	-0.47	0.03	2.45	18.70	0.03	Person Fit Residual	-0.38; 1.27
Home	-0.08	0.04	-0.20	21.28	0.01	Total Item $\chi^2$	80.57
Environment	-0.05	0.04	0.61	5.32	0.81	Chi-square P	0.07
Safety	0.14	0.04	0.41	4.11	0.90	t-test P <sup>a</sup>	0.05
Leisure	0.49	0.03	1.45	10.38	0.32	PSI	0.71
Finances	0.57	0.03	-0.90	9.17	0.42		
Social							
Support	-0.33	0.03	2.13	8.04	0.43	Item Fit Residual	0.93; 2.11
Relationships		0.03			0.001	Person Fit Residual	-0.49; 1.07
Sex	0.06		-1.45	51.21			
	0.27	0.03	2.30	6.60	0.58	Total Item $\chi^2$	65.9
						Chi-square P	0.0001
						t-test P <sup>a</sup>	0.02
						PSI	0.58

<sup>a</sup> P of test for unidimensionality (independent t-test), probability must be 0.05 or less.

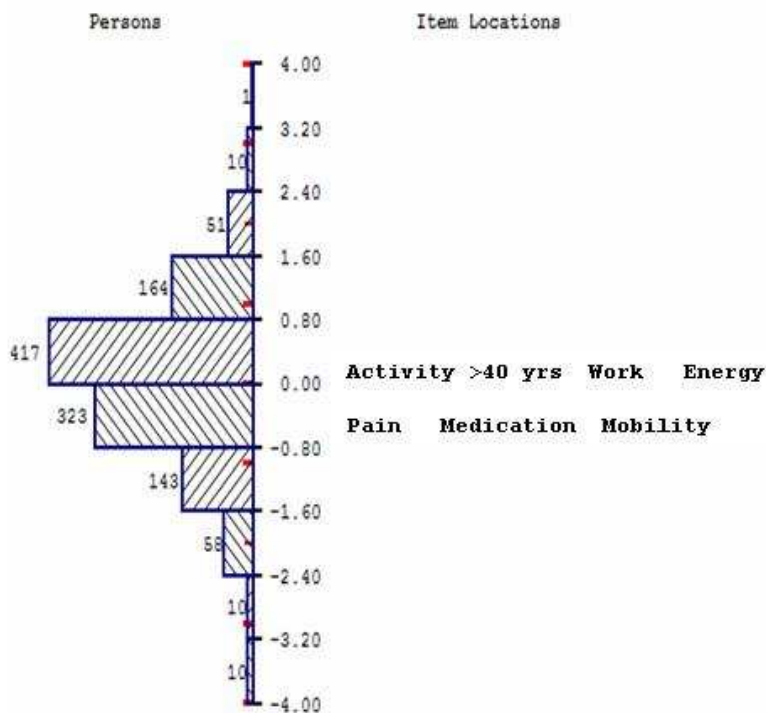
**Table 5: Post hoc analysis of DIF by country for the WHOQOL-Bref items in depressed patients in the LIDO study**

<b>Items*</b>			
Free from DIF (5/20)	DIF for one country (9/20)	DIF between two countries (5/20)	DIF for more than two countries (1/20)
General Health	Pain <sup>1</sup>	Overall QOL <sup>2,3</sup>	Relationships <sup>3,4,5,6</sup>
Meaning in life	Safety <sup>2</sup>	Medication <sup>1,6</sup>	
Leisure	Environment <sup>3</sup>	Mobility <sup>1,4</sup>	
Sex	Energy <sup>4</sup>	Work <sup>4,5</sup>	
Support	Finances <sup>2</sup>	Home <sup>4,5</sup>	
	Information <sup>1</sup>		
	Esteem <sup>3</sup>		
	Services <sup>5</sup>		
	Negative Feelings <sup>6</sup>		

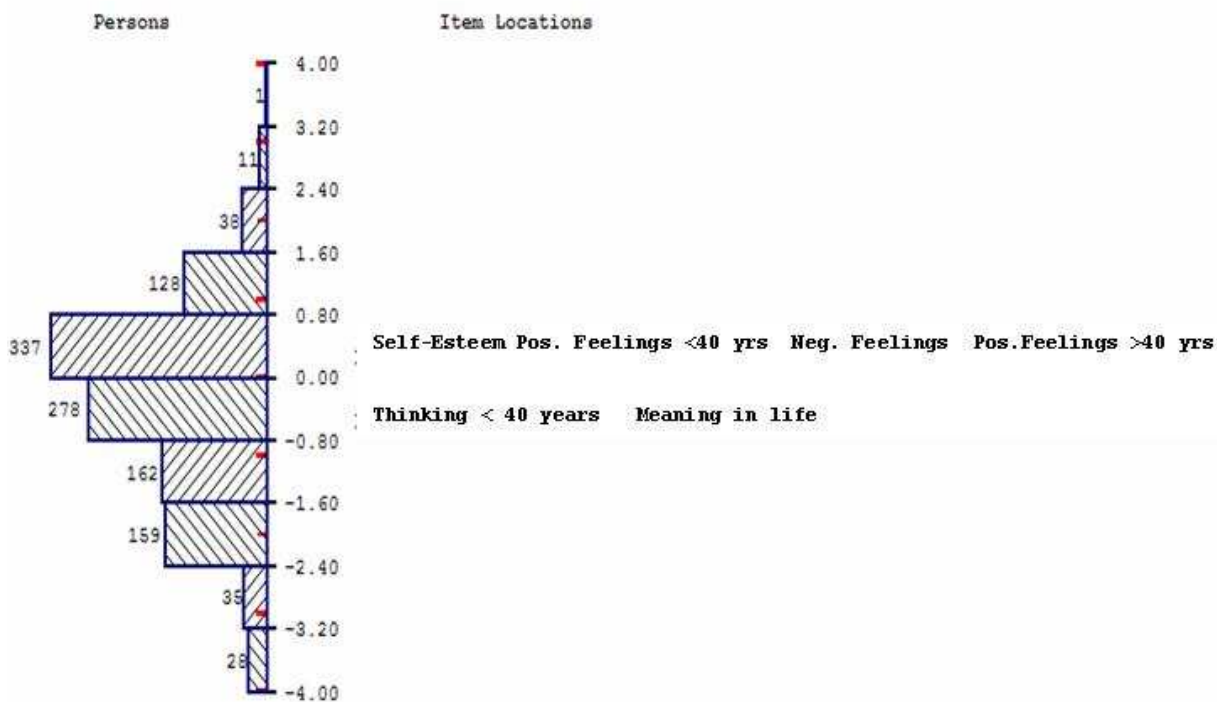
\*Excluding deleted or split items: think, body, activity, positive feelings, sleep, transport  
 1 Russia, 2 Spain, 3 Israel, 4 USA, 5 Australia, 6 Brazil

**Figure 1: Person-item location distribution map for the WHOQOL-Bref domains for baseline sample of depressed patients enrolled in the LIDO study**

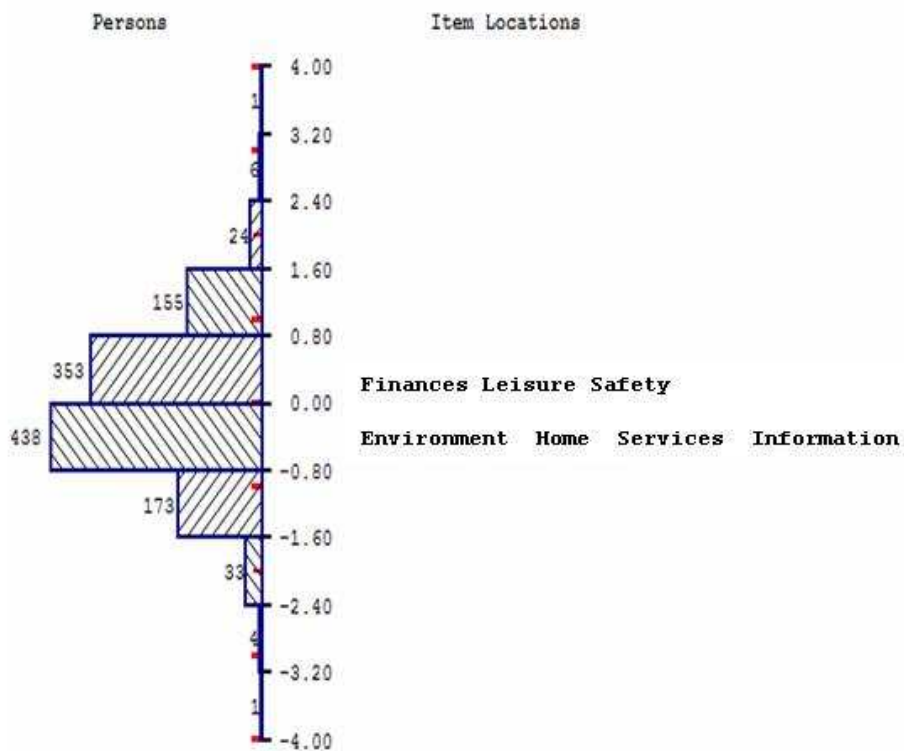
**a) Physical Domain**



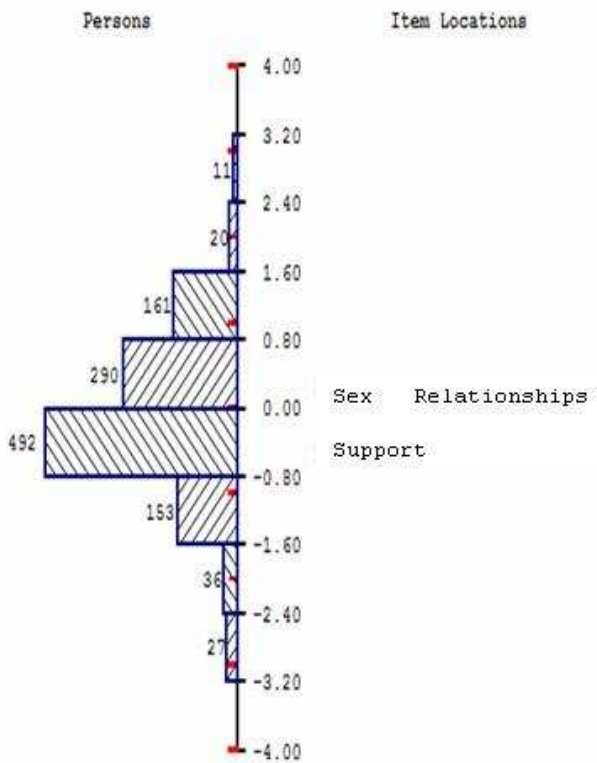
**b) Psychological Domain**



c) Environment Domain



d) Social Domain



## 6.4 Estudo 4

### **Validity of the Brazilian version of WHOQOL-Bref in primary care depressed patients using Rasch modelling**

### **Validade da versão brasileira do WHOQOL-Bref em pacientes deprimidos da atenção primária usando o modelo de Rasch**

**Neusa Sica da Rocha<sup>1,2</sup>, Marcelo Pio de Almeida Fleck<sup>1,3</sup>**

1 Hospital de Clínicas de Porto Alegre, Post Graduation Program in Medical Sciences: Psychiatry-UFRGS-Brazil

2 Psychiatrist, Post Graduation Program at Medical Sciences: Psychiatry-UFRGS-Brazil

3 Assistant Professor at the Department of Psychiatry and Legal Medicine-UFRGS-Brazil

Correspondent author:

Neusa Sica da Rocha  
Avenida Ramiro Barcelos, 2350 4º andar  
Psychiatric Service  
Porto Alegre-RS-Brazil - 90003-035  
Phone: 55 51 21018413  
e-mail: [neusa-rocha@via-rs.net](mailto:neusa-rocha@via-rs.net)

This study was partially funded by FIPE-HCPA, CAPES and the University of Edinburgh.

## **Abstract**

**Objective:** To assess the validity of the Brazilian version of WHOQOL-Bref in adults with major depression from Brazilian baseline sample of the Longitudinal Investigation of Depression Outcomes (LIDO) using Rasch modelling.

**Methods:** 208 major depressed patients were recruited in a primary care service. We used as measures the Center for Epidemiological Studies Depression Scale (CES-D) to assess intensity of depression; the World Health Organization Quality of Life Instrument- Abbreviated version (WHOQOL-Bref) as generic quality of life instrument; and the Composite International Diagnostic Interview (CIDI), version 2.1 for diagnostic of depression. The statistical method was the Rasch analysis.

**Results:** After using Rasch analysis, the 4 domains of WHOQOL-Bref showed appropriate fit to this model. Some items needed adjustments: 4 items were rescored ('Pain'; 'Finances', 'Services' and 'Transport'), 2 items ('Work' and 'Activity') identified they have dependency of responses, and 1 item was deleted ('Sleep'), showing multidimensionality.

**Conclusion:** The WHOQOL-Bref, after Rasch adjustments, seems to be a psychometrically valid instrument that it is suitable for evaluating the quality of life of Brazilian depressed outpatients from primary care setting.

**Keywords:** Statistical models, primary health care, depression, evaluation of research programs and tools

## Resumo

**Objetivo:** testar a validade da versão brasileira do WHOQOL-Bref em adultos com depressão maior da amostra brasileira basal do “Longitudinal Investigation of Depression Outcomes (LIDO)” usando o modelo de Rasch.

**Métodos:** 208 pacientes com depressão maior foram recrutados em um serviço de atenção primária. Os instrumentos utilizados foram: a “Center for Epidemiological Studies Depression Scale” (CES-D) para avaliar a intensidade da depressão; o “World Health Organization Quality of Life Instrument- Abbreviated version (WHOQOL-Bref)” versão brasileira, como medida de qualidade de vida genérica; e a “Composite International Diagnostic Interview (CIDI), version 2.1” para o diagnóstico de depressão. O método estatístico foi o da análise de Rasch.

**Resultados:** Após usar a análise de Rasch, os 4 domínios do WHOQOL-Bref demonstraram estar adequados ao modelo de Rasch. Alguns itens necessitaram de ajustes: 4 itens foram recodificados (‘Dor’; ‘Finanças’, ‘Serviços’ e ‘Transporte’), 2 itens (‘Trabalho’ e ‘Atividade’) demonstraram dependência de respostas, e 1 item foi retirado (‘Sono’), por apresentar sinal de multidimensionalidade.

**Conclusão:** O WHOQOL-Bref, após ajustes do modelo de Rasch, parece ser um instrumento psicometricamente válido e que está adequado para avaliar a qualidade de vida de pacientes deprimidos brasileiros, provenientes de um serviço de atenção básica.

**Descritores** Modelos Estatísticos, cuidados primários, depressão, avaliação de programas e instrumentos de pesquisa.



## INTRODUCTION

The WHOQOL-Bref is a generic quality of life (QOL) measure, which has been developed simultaneously in many cultures and languages by the World Health Organization. Although WHOQOL-Bref had been developed using this methodology to ensure its cross-cultural validity, this was established in previous research using only Classical Test Theory (CTT) (Whoqol, 1995; 1998; Skevington, Lotfy *et al.*, 2004). In consonance with international WHOQOL-Bref studies of validation, the Brazilian Portuguese version was also validated using CTT (Fleck, Louzada *et al.*, 1999; Berlim, Pavanello *et al.*, 2005).

Modern statistical analysis such as Rasch analysis has been pointed as a useful statistical method to yield measures that could be at the same time not influenced by sample nor by the scale (Rasch, 1960; Andrich, 1978; Smith, 2000; Pallant e Tennant, 2007). Also this technique may be a complementary tool for validation studies that used CCT (Prieto, Alonso *et al.*, 2003; Mallinson, Stelmack *et al.*, 2004; Norquist, Fitzpatrick *et al.*, 2004). Based on the assumption of invariability posed by Rasch analysis, we can identify whether or not items that are part of a scale are affected by external factors such as presence of a depressive episode, age, gender and culture. Some authors have questioned the validity of QOL measure in depressed patients (Sainfort, Becker *et al.*, 1996; Atkinson, Zibin *et al.*, 1997; Leidy, Palmer *et al.*, 1998; Scocco, Fantoni *et al.*, 2006), with the possibility of existing an overlap of depression and QOL constructs (Angermeyer, Holzinger *et al.*, 2002; Kuehner, 2002; Heinonen, Aro *et al.*, 2004; Naumann e Byrne, 2004; Kuehner e Buerger, 2005; Masthoff, Trompenaars *et al.*, 2006).

Since Major Depression is an important public health problem in Brazil, where the estimates of point prevalence of this disorder are between 3.5 to 9.7% and lifetime rate may be as high as 15% in Brazilian population (Mari, 2007), a valid QOL instrument is of great interest for these people. Independently of the research context, any health policy or test of new treatment for these patients would benefit from information given by this measure that is not only focused on symptoms nor functionality (Fleck, 2008).

This study was planned to establish the validity of WHOQOL-Bref using Rasch modelling by analysing a cross-sectional Brazilian baseline sample of the Longitudinal Investigation of Depression Outcomes (LIDO), a multicenter, cross-national observational study which followed patients with depressive disorders in primary care settings for 12 months in six countries (Chisholm, Amir *et al.*, 2001).

## METHOD

Patients presenting at a primary care service in Porto Alegre, Brazil, were screened for symptoms of depression. Those meeting inclusion criteria- new and/or untreated episode and a score of over 16 on the Centre for Epidemiological Studies-Depression scale (CES-D); (Radloff, 1977) - were interviewed and assessed with a standardized diagnostic instrument for major depression, the Composite International Diagnostic Interview(CIDI); (Robins, Wing *et al.*, 1988).

### Measures

**Demographic characteristics.** Continuous variables were: age and years of education; categorial variables were: gender, marital status and self-report of health.

**Intensity of depression.** The CES-D is a 20-item scale designed to measure symptoms of depression in community populations (Radloff, 1977).

**Generic Quality of Life.** The World Health Organization Quality of Life Instrument- Abbreviated version (WHOQOL Group, 1998; Fleck, Louzada *et al.*, 2000; Skevington, Lotfy *et al.*, 2004). The 26 items of WHOQOL-Bref distribute into four domains (physical, psychological, social relationships and environment) and are answered using individualised five-point scales. Each subscale is scored positively.

**Depression diagnostic instrument. Composite International Diagnostic Interview, version 2.1** is a completely structured psychiatric diagnostic assessment developed for use in cross-national epidemiologic studies. Data from CIDI were used to assess diagnostic criteria for depression from the American Psychiatric Association DSM-IV (American Psychiatric, 1994).

## **Statistical Analysis**

### **Rasch analysis**

#### **Rasch Model**

The Rasch model is a one-dimensional model, first used in educational assessment (Rasch, 1960), which asserts that the easier the item the more likely it will be passed, and the more able the person, the more likely they will pass an item compared with a less able person. When testing QOL of patients, they have a given QOL level, and they are presented with a range of items corresponding to differing facets of QOL. Thus, in the context of QOL measurement, a person with better QOL will have a greater probability of answering positively (where positive reflects better QOL) than someone with a lower QOL.

This model can be extended to analyse items with more than 2 categories, and this involves a “threshold” parameter, where the threshold represents the equal probability point between any two adjacent categories within an item. The model used in the present analysis is a further derivation, the Partial Credit Model (Masters, 1982).

#### **Overall measures of fit**

Three overall fit statistics are considered to determine the model fit. Two are item-person interaction statistics distributed as a Z statistic with mean of zero and SD of 1 (indicates perfect fit to the model). The third one is an item-trait interaction statistic reported as a  $\chi^2$ , reflecting the invariance across the trait (indicated by a non significant  $\chi^2$ ). Besides, individual item-fit statistics are presented as residuals (acceptable within the range  $\pm 2.5$ ) and as a  $\chi^2$  statistic (required also a non significant  $\chi^2$ ).

The boundaries between categories of responses are called “thresholds” and “disorder thresholds” may indicate that it will be necessary to collapse adjacent categories. Following this, data are fitted to the model to determine overall fit, and how well each item fits the model.

### **Rasch model properties:**

Rasch model has some assumptions that need to be evaluated to ensure that an instrument has Rasch properties. The most commonly Rasch assumptions assessed are: a) unidimensionality; b) local independence; and c) invariance.

#### **a) Unidimensionality**

Unidimensionality is used to ensure if a single latent trait explain all the variance in the data. The residuals are what remain when the ‘Rasch Factor’ has been removed from the data, and therefore, the first factor of the Principal Component Analysis (PCA) is the primary contributor to the variance of the data, with the ‘Rasch Factor’ discounted. We take the items presenting the highest positive correlation with the first component of PCA of the residuals, and the items with the highest negative loading items, and derive estimates for these two sets. These are compared to test if the assumption of unidimensionality holds by applying an independent t-test to each person pair of estimates. If less than 5% of these estimates are outside the range of  $\pm 1.96$ , the scale is deemed to be unidimensional.

#### **b) Local Independence**

Local independence means that when the ability is influencing the performance is constant, responses to any pair of items are statistically independent. Looking at the residual correlation matrix, we see if any values exceed + 0.3. This will indicate the presence of local dependency. If items are correlated in the residuals, we

merge them into a 'super' item through the subtest procedure, and see if improved fit is obtained. If so, it is a sign of local dependency and a violation of one of Rasch assumptions.

### **c) Invariance**

Invariability implies that the parameters that characterize an item do not depend on ability distribution of the persons and the parameters that characterize the persons do not depend on the set of test items. To ensure the invariability of this measure, we conduct a type of analysis which has been called Differential Item Functioning (DIF). The statistical test used for detecting DIF is an ANOVA of the person-item deviation residuals with person factors (e.g., age, gender, country) and class intervals (e.g., group along the trait) as factors. All items were checked for DIF by gender, age and educational level as person factors. Items that do not yield the same item response function for 2 or more groups are violating the requirement for unidimensionality and invariability.

### **Reliability**

An estimate of the internal consistency reliability of the scale is also provided, based on the Person Separation Index (PSI), where the estimates on the logit scale for each person are used to calculate reliability.

Rasch analysis was undertaken using the RUMM 2020 package (Andrich, Sheridan *et al.*, 2004).

All the individuals invited to participate in the study completed a written consent presenting the objectives of the study. The local Ethics Research Committee approved the project.

## RESULTS

Data presented in table 1 provide summary descriptions of baseline sample. Our sample is predominantly female, married, middle aged, primary school educated, with good health and moderate levels of depression.

### Table 1

Looking at all domains of WHOQOL-Bref, only Physical did not meet the expectations of the Rasch model, assessed by the summary of overall measures of fit statistics, where: total item  $\chi^2$  is 86.9; Chi-square P is 0.02; PSI is 0.79; P (independent t-test) is between 0.06-and 0.12.

In the same way, when analysing individual item fit of all items, only 'Sleep' item (residual=3.09;  $\chi^2$ = 23.66; P=0.0001) did not fit the Rasch model.

### Table 2

Category of response scale was checked. From the 26 items of Bref, 'Pain', 'Finances', 'Services' and 'Transport' displayed disordered thresholds of category responses, and to accord with Rasch properties, they needed to be rescored. Suppressing middle category of responses for all these items resulted in ordering their thresholds. Observe that for rescored items, this solution shortened response scale to 1-4, differently from all remaining items that maintained original scoring 1-5.

### Table 3

All items were checked for DIF for age (45 years old less vs. 45 years old plus), gender and educational level (at least primary level vs. primary level plus). Items 'Positive feelings' and 'Support' displayed DIF for age and 'Energy' for gender. No item displayed DIF for educational level. No item has shown DIF for more than one factor.

As Physical domain's items did not fit the Rasch model, all items were re-analysed to conform to Rasch model assumptions. Analysing Physical Domain, 'Activity' and 'Work' items have shown correlations of 0.34 on Person item residual correlation matrix, indicating local dependency of responses (correlation >0.3), then the subtest analysis was performed. After the subtest analysis, therefore, Physical domain overall measures of fit have improved, total item  $\chi^2$  modified from 86.9 to 38.9; P from 0.02 to 0.57; PSI from 0.73 to 0.76; P (independent t-test) from 0.06-0.12 to -0.02-0.04. Despite the improvement of all overall measures, 'Sleep' item remained showing signals of misfit to the model (residual= 2.58;  $\chi^2$ = 10.4; P=0.32). Deleting 'Sleep' item resulted in the best overall measures for Physical domain: total item  $\chi^2$  modified from 51.6 to 38.9; P from 0.57 to 0.72; PSI maintained 0.76; P (independent t-test) from -0.02-0.04 to 0.04-0.10. 'Sleep' item misfit was caused by multidimensionality.

Psychological Domain, Environment Domain items and Social Domain maintained its fit to Rasch model, as illustrated in Table 4.

**Table 4**



## DISCUSSION

Our findings point to the validity of WHOQOL-Bref as a measure of generic subjective QOL in the context of primary care depressed patients in Brazil. Other studies that had similar purposes, used CCT, small sample sizes (from 41 to 81 participants) and patients from clinical setting (Atkinson, Zibin *et al.*, 1997; Kuehner, 2002; Naumann e Byrne, 2004; Berlim, Pavanello *et al.*, 2005).

Examining all WHOQOL-Bref items in this Brazilian sample, only 7 items (Sleep, Activity, Work, Pain, Finances, Services and Transport) did not meet Rasch requirements. An interesting aspect is that four of them are Physical domain items. This finding may be associated with Primary Care setting, where patients were recruited in an appointment with a general practitioner for physical complains which may or may not have been related to their depressive episode (Aragones, Labad *et al.*, 2005; Tylee e Gandhi, 2005). An international study reported percentages of depression with somatic presentation from 45% to 95% in the countries studied (Simon, Vonkorff *et al.*, 1999). 'Sleep' item deletion and inclusion as a separate item may be justifiable by its multidimensionality, since 'Sleep' may be a content from physical or/and psychological domains.

The conciliation of conceptual model and empiric evidence is our defiance as researchers. Although, inclusion of these excluded items had a conceptual reason, this was not revealed in this empirical study. Obviously, the exclusion and inclusion of items diminish the comparability of measures between different populations; here again we need to make concessions, depending on research purpose.

This research did not re-test the original 4 domains structure of WHOQOL-Bref, which can be a limitation, regardless of the fact that there is not a literature consensus on the most suitable method to evaluate instrument dimensionality

(Coste, Bouee *et al.*, 2005). As a consequence, we opted to maintain the 4-domain structure, because besides being the most conservative approach, it is the most studied one (Yao e Wu, 2005).

Finally, the validation of WHOQOL-Bref Brazilian version using Rasch analysis serves as a complement to its previous studies of validation, evidencing the robustness of this instrument as a generic cross-cultural QOL measure.

## References

- American Psychiatric, A. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association. 1994.
- Andrich, D. A rating formulation for ordered response categories. Psichometrika, v.43, p.561-573. 1978.
- Andrich, D., B. Sheridan, *et al.* RUMM: a Windows program for analysing item response data according to Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM Laboratory 2004.
- Angermeyer, M. C., A. Holzinger, *et al.* Depression and quality of life: results of a follow-up study. Int J Soc.Psychiatry, v.48, n.3, p.189-199. 2002.
- Aragones, E., A. Labad, *et al.* Somatized depression in primary care attenders. J.Psychosom.Res., v.58, n.2, p.145-151. 2005.
- Atkinson, M., S. Zibin, *et al.* Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. Am.J Psychiatry, v.154, n.1, p.99-105. 1997.
- Berlim, M. T., D. P. Pavanello, *et al.* Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual.Life Res, v.14, n.2, p.561-564. 2005.
- Chisholm, D., M. Amir, *et al.* Longitudinal Investigation of Depression Outcomes (The LIDO Study) in primary care in six countries: comparative assessment of local health systems and resource utilization. Int J Methods Psychiatr Res, v.10, n.2, p.59-71. 2001.
- Coste, J., S. Bouee, *et al.* Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. Qual.Life Res, v.14, n.3, p.641-654. 2005.
- Fleck, M. Problemas conceituais em qualidade de vida. In: M. P. D. A. Fleck (Ed.). A avaliação de qualidade de vida: guia prático para profissionais de saúde. Porto Alegre: Artmed, 2008. Problemas conceituais em qualidade de vida, p.19-28.
- Fleck, M. P., S. Louzada, *et al.* [Application of the Portuguese version of the instrument for the assessment of quality of life of the World Health Organization (WHOQOL-100)]. Rev.Saude Publica, v.33, n.2, p.198-205. 1999.
- \_\_\_\_\_. [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-Bref]. Rev Saude Publica, v.34, n.2, Apr, p.178-83. 2000.
- Heinonen, H., A. R. Aro, *et al.* Is the evaluation of the global quality of life determined by emotional status? Qual.Life Res, v.13, n.8, p.1347-1356. 2004.

Kuehner, C. Subjective quality of life: validity issues with depressed patients. Acta Psychiatr Scand, v.106, n.1, p.62-70. 2002.

Kuehner, C. e C. Bueger. Determinants of subjective quality of life in depressed patients: the role of self-esteem, response styles, and social support. J Affect.Disord., v.86, n.2-3, p.205-213. 2005.

Leidy, N. K., C. Palmer, *et al.* Health-related quality of life assessment in euthymic and depressed patients with bipolar disorder. Psychometric performance of four self-report measures. J Affect.Disord., v.48, n.2-3, p.207-214. 1998.

Mallinson, T., J. Stelmack, *et al.* A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care, v.42, n.1 Suppl, p.117-124. 2004.

Mari, J. J. J., M.R.; Kohn, R. . Epidemiologia dos transtornos psiquiátricos em adultos. In: M. F. M. Mello, A. A. F.; Kohn, R. (Ed.). Epidemiologia da saúde mental no Brazil. Porto Alegre: Artmed, 2007. Epidemiologia dos transtornos psiquiátricos em adultos.

Masters, G. N. Rasch model for partial credit scoring. Psychometrika, v.47, p.149-174. 1982.

Masthoff, E. D., F. J. Trompenaars, *et al.* Quality of life and psychopathology: investigations into their relationship. Aust.N.Z.J Psychiatry, v.40, n.4, p.333-340. 2006.

Naumann, V. J. e G. J. Byrne. WHOQOL-Bref as a measure of quality of life in older patients with depression. Int Psychogeriatr., v.16, n.2, p.159-173. 2004.

Norquist, J. M., R. Fitzpatrick, *et al.* Comparing alternative Rasch-based methods vs raw scores in measuring change in health. Med Care, v.42, n.1 Suppl, p.125-136. 2004.

Pallant, J. F. e A. Tennant. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol, v.46, n.Pt 1, Mar, p.1-18. 2007.

Prieto, L., J. Alonso, *et al.* Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual.Life Outcomes., v.1, n.1, p.27. 2003.

Radloff, L. S. The CES-D scale: a self report depression scale for research in the general population. Applied Psychological Measurement, v.1, p.385-401. 1977.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen:: Danish Institute for Educational Research. 1960.

Robins, L. N., J. Wing, *et al.* The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch.Gen.Psychiatry, v.45, n.12, p.1069-1077. 1988.

Sainfort, F., M. Becker, *et al.* Judgments of quality of life of individuals with severe mental disorders: Patient self-report versus provider perspectives. Am.J Psychiatry, v.153, n.4, p.497-502. 1996.

Scocco, P., G. Fantoni, *et al.* Role of depressive and cognitive status in self-reported evaluation of quality of life in older people: comparing proxy and physician perspectives. Age Ageing, v.35, n.2, p.166-171. 2006.

Simon, G. E., M. Vonkorff, *et al.* An international study of the relation between somatic symptoms and depression. N.Engl.J.Med., v.341, n.18, p.1329-1335. 1999.

Skevington, S. M., M. Lotfy, *et al.* The World Health Organization's WHOQOL-Bref quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual.Life Res, v.13, n.2, p.299-310. 2004.

Smith, J. E. Metric Development and Score Reporting in Rasch Measurement. Journal of Applied Measurement, v.1, n.3, p.303-326. 2000.

Tylee, A. e P. Gandhi. The importance of somatic symptoms in depression in primary care. Prim.Care Companion.J.Clin.Psychiatry, v.7, n.4, p.167-176. 2005.

The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc.Sci.Med, v.41, n.10, p.1403-1409. 1995.

\_\_\_\_\_. Development of the World Health Organization WHOQOL-Bref quality of life assessment. The WHOQOL Group. Psychol.Med, v.28, n.3, p.551-558. 1998.

Yao, G. e C. H. Wu. Factorial invariance of the WHOQOL-Bref among disease groups. Qual.Life Res, v.14, n.8, p.1881-1888. 2005.

**Table 1: General descriptions of baseline sample of depressed patients enrolled in Longitudinal Investigation of Depression Outcomes study in Brazil**

	Depressed N=208
<b>Gender (%)</b>	
Male	52 (25)
Female	156(75)
<b>Marital Status (%)</b>	
Married	106 (51)
Widowed	17 (8.2)
Separated	22 (10.6)
Divorced	12 (5.8)
Never Married	51 (24.5)
<b>Health Status (%)</b>	
Excellent	6 (2.9)
Very good	34 (16.3)
Good	132 (63.5)
Fair	31 (14.9)
Poor	5 (2.4)
<b>Years of Education</b>	
Mean ± S.D	9.1±3.3
<b>Age</b>	
Mean ± S.D.	39.2±13.6
<b>CES-D score</b>	
Mean ± S.D.	31.7±9.9

**Table 2: Fit of WHOQOL-Bref items to Rasch Model of baseline sample of depressed patients enrolled in Longitudinal Investigation of Depression Outcomes study in Brazil**

Item	Location	Residual	X <sup>2</sup>	P	Summary of overall fit statistics	
<b>Overall</b>						
Overall QOL	-0.68	1.09	4.83	0.44		
General Health	0.68	-0.94	6.28	0.28		
<b>Physical</b>						
Pain*	-0.48	1.23	6.22	0.72	Total Item x <sup>2</sup>	86.9
Medication	-0.33	0.94	7.64	0.57	Chi-square P	0.02
Energy	0.32	0.16	12.73	0.18	PSI	0.79
Mobility	-0.67	0.01	13.70	0.13	t-test P <sup>b</sup>	0.06-0.12
<b>Sleep</b>	<b>0.41</b>	<b>3.09</b>	<b>23.66</b>	<b>0.0001</b>		
Activity	0.56	-1.28	12.33	0.20		
Work	-0.48	-1.51	10.61	0.30		
<b>Psychological</b>						
Positive					Total Item x <sup>2</sup>	47.35
Feelings	0.57	0.35	5.40	0.80	Chi-square P	0.72
Spirituality	-0.40	-0.94	6.63	0.68	PSI	0.73
Think	-0.11	1.33	9.32	0.41	t-test P <sup>b</sup>	0.01-0.07
Body	-0.45	0.71	6.99	0.64		
Esteem	0.10	-0.13	6.48	0.69		
Negative						
Feelings	0.29	0.69	12.53	0.19		
<b>Environment</b>						
Safety	0.36	0.46	1.80	0.99	Total Item x <sup>2</sup>	49.3
Environment	0.17	1.46	5.29	0.81	Chi-square P	0.98
Finances*	0.81	0.42	7.75	0.56	PSI	0.63
Information	-0.67	0.59	6.43	0.70	t-test P <sup>b</sup>	0.05-0.11
Leisure	0.50	0.02	7.34	0.60		
Home	-0.41	-0.99	6.33	0.71		
Services*	-0.51	-0.16	9.34	0.41		
Transport*	-0.24	1.74	5.02	0.83		
<b>Social</b>						
Relationships	-0.14	-0.58	7.83	0.45	Total Item x <sup>2</sup>	24.9
Sex	0.30	2.03	9.50	0.30	Chi-square P	0.35
Support	-0.16	0.05	7.60	0.37	PSI	0.58
					t-test P <sup>b</sup>	0.01-0.07

\*items that needed to be rescored because of disordered thresholds of categories response

**Bold: items that do not fit in the model.**

<sup>b</sup> P of test for unidimensionality, shown as CI95%, probability of 0.05 must be included in the interval.

**Table 3: WHOQOL-Bref response scale adapted for 4 items after Rasch analysis**

Items	Response Scale				
Pain	1	2	3	3	4
Finances	1	2	3	3	4
Services	1	2	2	3	4
Transport	1	2	2	3	4
All remaining items	1	2	3	4	5



**Table 4: Fit of WHOQOL items adjusted by Rasch model of baseline sample of depressed patients enrolled in LIDO study in Brazil**

Item	Location	Residual	$\chi^2$	P	Summary of overall fit statistics for domains	
<b>Overall</b>						
Overall QOL	-0.68	1.09	4.83	0.44		
General Health	0.68	-0.94	6.28	0.28		
<b>Physical</b>						
Activity+Work <sup>a</sup>	0.63	-0.82	3.65	0.93	Total Item $\chi^2$	38.9
Pain	-0.56	0.92	9.43	0.40	Chi-square P	0.72
Medication	-0.13	0.71	11.65	0.23	PSI	0.76
Energy	0.52	0.44	8.05	0.53	t-test P <sup>b</sup>	0.04-0.10 <sup>b</sup>
Mobility	-0.47	0.31	6.19	0.72		
<b>Psychological</b>						
Positive					Total Item $\chi^2$	47.35
Feelings	0.57	0.35	5.40	0.80	Chi-square P	0.72
Spirituality	-0.40	-0.94	6.63	0.68	PSI	0.73
Think	-0.11	1.33	9.32	0.41	t-test P <sup>b</sup>	0.01-0.07 <sup>b</sup>
Body	-0.45	0.71	6.99	0.64		
Esteem	0.10	-0.13	6.48	0.69		
Negative						
Feelings	0.29	0.69	12.53	0.19		
<b>Environment</b>						
Safety	0.38	0.27	1.55	1.00	Total Item $\chi^2$	43.1
Environment	0.16	1.26	7.36	0.60	Chi-square P	0.99
Finances	0.83	0.52	10.84	0.29	PSI	0.63
Information	-0.67	0.35	4.47	0.88	t-test P <sup>b</sup>	0.05-0.11 <sup>b</sup>
Leisure	0.51	0.02	5.94	0.75		
Home	-0.43	-1.09	4.78	0.85		
Services	-0.56	-0.09	3.77	0.93		
Transport	-0.22	1.11	4.44	0.88		
<b>Social</b>						
Relationships	-0.14	-0.58	7.83	0.45	Total Item $\chi^2$	24.9
Sex	0.30	2.03	9.50	0.30	Chi-square P	0.35
Support	-0.16	0.05	7.60	0.37	PSI	0.58
					t-test P <sup>b</sup>	0.01-0.07 <sup>b</sup>

<sup>a</sup> subtest analysis

<sup>b</sup> P of test for unidimensionality, shown as CI95%, probability of 0.05 must be included in the interval

## 7 CONSIDERAÇÕES FINAIS

O estudo do desempenho do WHOQOL-Bref em pacientes deprimidos provenientes de serviços de cuidados primários de diferentes países utilizando a análise de Rasch nos permite chegar a algumas conclusões.

Uma conclusão mais genérica é que, apesar do WHOQOL-Bref ter sido concebido, construído e desenvolvido utilizando os conceitos da psicometria clássica, destaca-se o fato de que apenas poucas modificações tenham tido que ser feitas para que este instrumento se adequasse aos pressupostos da análise de Rasch. Este achado indica que apesar dos paradigmas das duas teorias clássica e modelo de Rasch ou Teoria da Resposta ao Item serem diferentes e considerados como incompatíveis (Andrich, 2004), isto não se evidenciou em nosso estudo. Foi possível a integração dos achados até então mostrados pela Teoria Clássica com os achados da análise de Rasch. Diferente de defender a oposição entre métodos alguns autores também propõem a integração (Prieto, Alonso *et al.*, 2003), embora se tenha noção que esta nem sempre é possível e provavelmente não serão todos os instrumentos e estudos que permitirão tal integração (Norquist, Fitzpatrick *et al.*, 2004). Ainda, acreditamos que isto somente foi possível, porque quando estudamos o WHOQOL-Bref existe uma base conceitual sólida (Van Alphen, Halfens *et al.*, 1994) construída através do esforço conjunto de varias nações e grupos de pesquisa que sustenta a interpretação dos achados das análises (WHOQOL Group, 1995; WHOQOL Group, 1996; WHOQOL Group, 1998b; a; Power, Bullinger *et al.*, 1999; Skevington e Tucker, 1999; WHOQOL Group, 2003; Skevington, Lotfy *et al.*, 2004; Skevington, Sartorius *et al.*, 2004). O nosso estudo foi o primeiro a se preocupar com

a avaliação desta medida usando a análise de Rasch em pacientes deprimidos de diferentes países.

Um segundo ponto importante que os nossos resultados puderam demonstrar é que o estudo do WHOQOL-Bref em pacientes deprimidos deve levar em consideração a interferência dos sintomas depressivos na avaliação da medida da QV. Nossos dados apontam para a necessidade de se controlar o efeito da depressão na avaliação de QV em pacientes com depressão. Alguns autores recomendam um controle genérico destes sintomas (Aigner, Forster-Streffleur *et al.*, 2006). Os nossos achados permitem delimitar esta interferência a alguns aspectos da QV, como dor, energia, sono, atividades da vida diária, relações sociais, apoio dos outros e sentimentos negativos, que podem ser considerados também como critérios diagnósticos de depressão (American Psychiatric Association, 1994) ou itens de escalas de depressão (Hamilton, 1960). Concordamos com a afirmação de Orley (Orley, Saxena *et al.*, 1998) que ‘ uma escala de QV é uma avaliação muito mais ampla e apesar de influenciada pelo afeto, representa uma avaliação de si mesmo e de mundo material e social’. A interferência dos sintomas depressivos na QV pode ser provavelmente generalizável para a maioria dos pacientes deprimidos, independente do contexto em que fizeram parte, internados, ambulatoriais ou da atenção básica.

Ainda no que diz respeito à interferência da ocorrência de um episódio depressivo na avaliação da medida genérica de QOL, encontramos diferenças entre os domínios. O domínio físico foi o mais afetado pela ocorrência de depressão parece estar associado com o contexto de cuidados primários, onde os pacientes foram recrutados em consultas com o clínico geral por queixas físicas que poderiam estar ou não relacionadas com o episódio depressivo (Aragones, Labad *et al.*, 2005;

Tylee e Gandhi, 2005). Alguns estudos apontam que um em quatro pacientes tratados na prática clínica geral se apresentam com queixas psicológicas, predominantemente ansiedade e/ou depressão (Simon, Vonkorff *et al.*, 1999). Uma vez que os sintomas de ansiedade podem ser somáticos, nossos achados para o domínio físico também podem ser explicados tanto pela ocorrência de depressão com sintomas somáticos, quanto pela comorbidade com transtornos de ansiedade ou suas queixas (Runkewitz, Kirchmann *et al.*, 2006)

Outra conclusão importante são os achados da avaliação transcultural dos domínios do WHOQOL-Bref. A validade transcultural do WHOQOL-Bref já havia sido demonstrada em estudos utilizando a psicometria clássica (WHOQOL Group, 1998b; a; Skevington, Lotfy *et al.*, 2004). O nosso estudo é o primeiro a avaliar esta medida utilizando a análise de Rasch. Podemos mostrar, a partir de uma amostra de pacientes deprimidos, que o WHOQOL-Bref com poucos ajustes é uma medida transculturalmente adequada para as culturas estudadas porque apenas um dos itens não foi equivalente para mais do que dois países. Outro estudo que teve um objetivo semelhante usou o WHOQOL-100 demonstrou que apenas um item tinha equivalência transcultural para as quatro culturas estudadas (Franca, Hong Kong, Argentina e Reino Unido) usando a análise de Rasch (Leplege, 2000).

Analisando os cinco itens que foram completamente equivalentes entre os países (saúde geral, sentido da vida, lazer, vida sexual e apoio dos outros) podemos identificar que são todos conceitos relacionais, com exceção da saúde geral. Tal achado reforça a importância, pelo menos para o estudo da QV em pacientes deprimidos, da perspectiva humanista deste conceito. Este resultado está em concordância com outros autores que argumentam que uma medida de QV verdadeiramente humanista deve considerar as necessidades das pessoas de

serem livres, movendo o conceito para uma visão mais existencial do que utilitária (Leplege e Hunt, 1997).

Como identificamos um restrito espectro de medida do WHOQOL-Bref (locations no limite de  $\pm 1$ ) nos pacientes deprimidos estudados, os nossos resultados permitem sugerir para investigação de futuras pesquisas, a incorporação nas medidas de QV para pacientes deprimidos, questões relacionadas à saúde mental e ajustamento social como estigma, barreiras econômicas ao tratamento (Saxena, Jane-Llopis *et al.*, 2006), bem-estar existencial (autonomia, liberdade) e religiosidade (Koenig, George *et al.*, 1998), que não fazem parte desta medida de QOL. A inclusão destes aspectos além de aumentar o espectro de medida do instrumento (isto é, inclusão de itens que poderiam apresentar locations fora do limite de  $\pm 1$ ); além disto, poderia provavelmente melhorar a sua confiabilidade e validade (Smith, 2001), principalmente por melhorar a performance inadequada do domínio social.

Ainda, quando analisamos o desempenho do WHOQOL-Bref em cada um dos países envolvidos no estudo, optamos por descrever a análise dos dados brasileiros por ser do nosso país de origem. Assim como na análise transcultural, para o Brasil também não foi necessário fazer extensas modificações para que este instrumento de adequasse ao modelo de Rasch. Apenas sete itens sofreram modificações. Estes achados apontam para a confirmação da validade desta medida para esta população brasileira, embora outros estudos que tinham objetivos semelhantes usaram a teoria clássica com número pequeno de participantes (de 41 a 81 participantes) recrutados em clínicas especializadas (Atkinson, Zibin *et al.*, 1997; Kuehner, 2002; Naumann e Byrne, 2004; Berlim, Pavanello *et al.*, 2005).

O nosso estudo apresentou algumas dificuldades e limitações. Não testamos a estrutura geral de quatro domínios do WHOQOL-Bref, porque além de não haver um consenso na literatura de qual o método mais adequado para avaliar a dimensionalidade de uma escala (Coste, Bouee *et al.*, 2005), esta estrutura é a mais estudada na literatura e mostrou-se invariável entre diferentes grupos de patologias (Yao e Wu, 2005). Também sabemos que este instrumento tem uma estrutura multidimensional e que estaria mais adequado o uso de um modelo multidimensional. Até onde temos conhecimento, ainda não existem recursos computacionais disponíveis que realizem este tipo de análise. Um grupo de pesquisa norte-americano pretende disponibilizar para final de 2009, um programa estatístico que possa realizar análises de constructos multidimensionais usando modelos multidimensionais baseados na Teoria da resposta ao item.

Por fim, como o Transtorno Depressivo Maior é altamente prevalente no contexto de atenção primária (Kessler, Berglund *et al.*, 2003) e diante da realidade descrita pelos estudos prévios do LIDO no seguimento de doze meses, onde se identificou que menos da metade atingem remissão completa (De Almeida Fleck, Simon *et al.*, 2005), sendo que apenas um percentual pequeno (até 40 %) dos pacientes deprimidos (Simon, Fleck *et al.*, 2004) receberam algum tipo de tratamento, isto nos dá uma idéia da relevância de se investir em pesquisas e políticas de saúde direcionadas para esta população. Portanto, o nosso estudo tem o mérito de se preocupar em estudar esta população, enfocando no refinamento de uma medida de QV que além de poder ser uma ferramenta útil para avaliação clínica, pois outros estudos (Rush, Trivedi *et al.*, 2003; Rush, Fava *et al.*, 2004; Trivedi, Rush, Wisniewski, Nierenberg *et al.*, 2006; Trivedi, Rush, Wisniewski, Warden *et al.*, 2006) sugeriram que a avaliação de QV contribui para acessar a

severidade e comprometimento que tal transtorno pode resultar na vida destas pessoas, pode servir para monitorar a efetividade de políticas de saúde orientadas para estes pacientes. Ainda, por se tratar de uma medida para qual a avaliação transcultural foi estudada, permite que realizemos comparações entre nações provenientes de diferentes culturas com uma segurança maior da validade de nossos resultados.

## 8 CONCLUSÕES

1. A análise de Rasch é uma ferramenta que complementa as análises prévias da TC.
2. A sobreposição das medidas de QV e depressão se restringe a alguns aspectos que devem ser considerados em análises futuras, buscando medir QV independente da severidade da depressão.
3. Os domínios do WHOQOL-Bref, com alguns poucos ajustes, mostraram-se uma medida transculturalmente adequada para avaliar a QV de pacientes deprimidos da atenção básica para os países estudados.
4. Os domínios do WHOQOL-Bref, também com alguns poucos ajustes, mostraram-se uma medida adequada para avaliar a QV de pacientes deprimidos da atenção básica num serviço do Brasil.



## 9 REFERÊNCIAS BIBLIOGRÁFICAS

Aaronson, N. K., S. Ahmedzai, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst, v.85, n.5, Mar 3, p.365-76. 1993.

Aigner, M., S. Forster-Streffleur, *et al.* What does the WHOQOL-Bref measure? Measurement overlap between quality of life and depressive symptomatology in chronic somatoform pain disorder. Soc.Psychiatry Psychiatr Epidemiol., v.41, n.1, p.81-86. 2006.

Andrade, D. F., H. R. Tavares, *et al.* Teoria da Resposta ao Item: conceitos e aplicações. 2000 (14| SINAPE).

Andrich, D. A rating formulation for ordered response categories. Psichometrika, v.43, p.561-573. 1978.

\_\_\_\_\_. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care, v.42, n.1 Suppl, p.17-16. 2004.

Andrich, D., B. Sheridan, *et al.* RUMM: a Windows program for analysing item response data according to Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM Laboratory 2004.

Angermeyer, M. C., A. Holzinger, *et al.* Depression and quality of life: results of a follow-up study. Int J Soc.Psychiatry, v.48, n.3, p.189-199. 2002.

Atkinson, M., S. Zibin, *et al.* Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. Am.J Psychiatry, v.154, n.1, p.99-105. 1997.

Ay-Woan, P., C. P. Sarah, *et al.* Quality of life in depression: predictive models. Qual.Life Res, v.15, n.1, p.39-48. 2006.

Baca, E., M. Gonzalez-De-Chavez, *et al.* Sertraline is more effective than imipramine in the treatment of non-melancholic depression: Results from a multicentre, randomized study. Progress in Neuro Psychopharmacology and Biological Psychiatry, v.27, n.3, p.493-500. 2003.

Baker, F. B. The Basics of Item Response Theory. Madison: ERIC Clearinghouse on Assessment and Evaluation. 2001.

Baldomero, E.-B. e S. C. Enguix. Calidad de vida en pacientes con depresion tratados en Atencion Primaria. Efectividad y seguridad de la venlafaxina retard / Quality of life, in depressed patients in primary health care setting. Effectiveness and safety of venlafaxine extended release. Actas Espanolas de Psiquiatria, v.31, n.6, p.331-338. 2003.

Bech, P., R. Lucas, *et al.* Association between clinically depressed subgroups, type of treatment and patient retention in the LIDO study. Psychol.Med., v.33, n.6, p.1051-1059. 2003.

Bellantuono, C., E. Arreghini, *et al.* Psychotropic drug prescription in Italy. A survey in general practice. Soc Psychiatry Psychiatr Epidemiol, v.24, n.4, Jul, p.212-8. 1989.

Bergner, M., R. A. Bobbitt, *et al.* The sickness impact profile: validation of a health status measure. Med Care, v.14, n.1, Jan, p.57-67. 1976.

Berlim, M. T., J. Pargendler, *et al.* Significant improvement in the quality of life of Brazilian depressed outpatients 12 weeks following the start of antidepressants. Psychiatry Res, v.153, n.3, Dec 3, p.253-9. 2007.

Berlim, M. T., D. P. Pavanello, *et al.* Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual.Life Res, v.14, n.2, p.561-564. 2005.

Blumenthal, M. D. e T. E. Dielman. Depressive symptomatology and role function in a general population. Arch.Gen.Psychiatry, v.32, n.8, p.985-991. 1975.

Bond, T. G. e C. M. Fox. Applying the Rasch model- Fundamental measurement in the human sciences. Mahwah: Lawrence Erlbaum Associates. 2007.

Bonicatto, S. C., M. A. Dew, *et al.* Adult outpatients with depression: worse quality of life than in other chronic medical diseases in Argentina. Soc.Sci.Med, v.52, n.6, p.911-919. 2001.

Bruce, M. L., T. E. Seeman, *et al.* The impact of depressive symptomatology on physical disability: MacArthur Studies of Successful Aging. Am.J Public Health, v.84, n.11, p.1796-1799. 1994.

Burke, W.-J., I. Gergel, *et al.* Fixed-dose trial of the single isomer SSRI Escitalopram in depressed outpatients. Journal of Clinical Psychiatry, v.63, n.4, p.331-336. 2002.

Caliyurt, O. e F. Guducu. Partial sleep deprivation therapy combined with sertraline induces more rapid improvements in quality of life items in major depressive disorder. J Affect Disord, v.88, n.1, Sep, p.75-8. 2005.

Carpenter, L.-L., S. Yasmin, *et al.* A double-blind, placebo-controlled study of antidepressant augmentation with mirtazapine. Biological Psychiatry, v.51, n.2, p.183-188. 2002.

Cervera-Enguix, S., C.-A. Soutullo, *et al.* Quality of life in 833 outpatients with major depression treated with open-label venlafaxine extended release: An observational 24-week study. International Journal of Psychiatry in Clinical Practice, v.7, n.3, p.193-197. 2003.

Chisholm, D., M. Amir, *et al.* Longitudinal Investigation of Depression Outcomes (The LIDO Study) in primary care in six countries: comparative assessment of local health systems and resource utilization. Int J Methods Psychiatr Res, v.10, n.2, p.59-71. 2001.

Chisholm, D., P. Diehr, *et al.* Depression status, medical comorbidity and resource costs. Evidence from an international study of major depression in primary care (LIDO). Br.J.Psychiatry, v.183, p.121-131. 2003.

Conrad, K. J. e E. V. Smith, Jr. International conference on objective measurement: applications of Rasch analysis in health care. Med Care, v.42, n.1 Suppl, p.11-16. 2004.

Coons, S. J., S. Rao, *et al.* A comparative review of generic quality-of-life instruments. Pharmacoeconomics, v.17, n.1, Jan, p.13-35. 2000.

Da-Silva-Lima, A.-F.-B. e M.-P. De-Almeida-Fleck. Subsyndromal depression: An impact on quality of life? Journal of Affective Disorders, v.100, n.1-3, p.163-169. 2007.

Da Silva Lima, A. F. e M. P. De Almeida Fleck. Subsyndromal depression: an impact on quality of life? J Affect Disord, v.100, n.1-3, Jun, p.163-9. 2007.

De Almeida Fleck, M. P., G. Simon, *et al.* Major depression and its correlates in primary care settings in six countries. 9-month follow-up study. Br.J.Psychiatry, v.186, p.41-47. 2005.

De Leval, N. Quality of life and depression: symmetry concepts. Qual Life Res, v.8, n.4, Jun, p.283-91. 1999.

Demyttenaere, K., J. De Fruyt, *et al.* Measuring quality of life in depression. Current Opinion in Psychiatry, v.15, p.89-92. 2000.

Diehr, P., D. Patrick, *et al.* Including deaths when measuring health status over time. Med Care, v.33, n.4 Suppl, Apr, p.AS164-72. 1995.

Diehr, P. H., A. M. Derleth, *et al.* Synchrony of change in depressive symptoms, health status, and quality of life in persons with clinical depression. Health Qual.Life Outcomes., v.4, p.27. 2006.

Embretson, S. E. e S. P. Reise. Item Response Theory for Psychologists. Mahwan: Lawrence Erlbaum Associates, Inc. 2000.

Fava, M., A. J. Rush, *et al.* Background and rationale for the sequenced treatment alternatives to relieve depression (STAR\*D) study. Psychiatr Clin North Am, v.26, n.2, Jun, p.457-94, x. 2003.

Fernandez, J.-L., S. Montgomery, *et al.* Evaluation of the Cost Effectiveness of Escitalopram versus Venlafaxine XR in Major Depressive Disorder. Pharmacoeconomics, v.23, n.2, p.155-167. 2005.

Fleck, M. Problemas conceituais em qualidade de vida. In: M. P. D. A. Fleck (Ed.). A avaliação de qualidade de vida: guia prático para profissionais de saúde. Porto Alegre: Artmed, 2008. Problemas conceituais em qualidade de vida, p.19-28.

Fleck, M. P., A. F. Lima, *et al.* [Association of depressive symptoms and social functioning in primary care service, Brazil]. Rev.Saude Publica, v.36, n.4, p.431-438. 2002.

Gu, N., Z. Cai, *et al.* Fluvoxamine in the treatment of depression: Results from a Chinese population. Primary Care Psychiatry, v.9, n.3, p.89-94. 2004.

Heinonen, H., A. R. Aro, *et al.* Is the evaluation of the global quality of life determined by emotional status? Qual.Life Res, v.13, n.8, p.1347-1356. 2004.

Herrman, H., D. L. Patrick, *et al.* Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. Psychol.Med., v.32, n.5, p.889-902. 2002.

Hunt, S. M. e S. P. Mckenna. The QLDS: a scale for the measurement of quality of life in depression. Health Policy, v.22, n.3, Oct, p.307-19. 1992.

Hurst, N. P., P. Kind, *et al.* Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol, v.36, n.5, May, p.551-9. 1997.

Jenkinson, C., A. Coulter, *et al.* Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. Bmj, v.306, n.6890, May 29, p.1437-40. 1993.

Jenkinson, C., R. Fitzpatrick, *et al.* The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. Soc Sci Med, v.27, n.12, p.1411-4. 1988.

Kaplan, R. M., J. P. Anderson, *et al.* The Quality of Well-being Scale. Applications in AIDS, cystic fibrosis, and arthritis. Med Care, v.27, n.3 Suppl, Mar, p.S27-43. 1989.

Kaplan, R. M., C. J. Atkins, *et al.* Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. J Chronic Dis, v.37, n.2, p.85-95. 1984.

Katschnig, H. Quality of life in mental disorders: challenges for research and clinical practice. World Psychiatry, v.5, n.3, Oct, p.139-45. 2006.

Kroenke, K., N. Messina, Iii, *et al.* Venlafaxine Extended Release in the Short-Term Treatment of Depressed and Anxious Primary Care Patients With Multisomatoform Disorder. Journal of Clinical Psychiatry, v.67, n.1, p.72-80. 2006.

Kuehner, C. Subjective quality of life: validity issues with depressed patients. Acta Psychiatr Scand., v.106, n.1, p.62-70. 2002.

Kuehner, C. e C. Buerger. Determinants of subjective quality of life in depressed patients: the role of self-esteem, response styles, and social support. J Affect.Disord., v.86, n.2-3, p.205-213. 2005.

Leploge, A. e S. Hunt. The problem of quality of life in medicine. JAMA, v.278, n.1, p.47-50. 1997.

Lepke, A. E., E. Methodological issues in using the Rasch model to select cross culturally equivalent items in order to develop a Quality of Life index: the analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, United Kingdom). Journal of Applied Measurement v.1, p.372-392. 2000.

Lyness, J. M., T. K. Noel, *et al.* Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. Arch.Intern.Med, v.157, n.4, p.449-454. 1997.

Masters, G. N. Rasch model for partial credit scoring. Psychometrica, v.47, p.149-174. 1982.

Masthoff, E. D., F. J. Trompenaars, *et al.* Quality of life and psychopathology: investigations into their relationship. Aust.N.Z.J Psychiatry, v.40, n.4, p.333-340. 2006.

Mckenna, S. P. e S. M. Hunt. A new measure of quality of life in depression: testing the reliability and construct validity of the QLDS. Health Policy, v.22, n.3, Oct, p.321-30. 1992.

Montgomery, S.-A. e H.-F. Andersen. Escitalopram versus venlafaxine XR in the treatment of depression. International Clinical Psychopharmacology, v.21, n.5, p.297-309. 2006.

Naumann, V. J. e G. J. Byrne. WHOQOL-Bref as a measure of quality of life in older patients with depression. Int Psychogeriatr, v.16, n.2, p.159-173. 2004.

Noerholm, V., M. Groenvold, *et al.* Quality of life in the Danish general population--normative data and validity of WHOQOL-Bref using Rasch and item response theory models. Qual Life Res, v.13, n.2, Mar, p.531-40. 2004.

Ormel, J., M. Von Korff, *et al.* Depression, anxiety, and social disability show synchrony of change in primary care patients. Am J Public Health, v.83, n.3, Mar, p.385-90. 1993.

Ormel, J., M. Vonkorff, *et al.* Onset of disability in depressed and non-depressed primary care patients. Psychol Med, v.29, n.4, Jul, p.847-53. 1999.

\_\_\_\_\_. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. Jama, v.272, n.22, Dec 14, p.1741-8. 1994.

Papakostas, G. I., T. Petersen, *et al.* Quality of life assessments in major depressive disorder: a review of the literature. Gen.Hosp.Psychiatry, v.26, n.1, p.13-17. 2004.

Pasquali, L. Instrumentos psicológicos: manual prático de elaboração. Brasília: IBAPP. 1999.

\_\_\_\_\_. Técnicas de exame psicológico- TEP: manual. São Paulo: Casa do Psicólogo, v.1. 2001.

\_\_\_\_\_. Psicometria: teoria geral dos testes na psicologia e educação. Petrópolis: Vozes. 2003. 397 p.

Perahia, D.-G., I. Gilaberte, *et al.* Duloxetine in the prevention of relapse of major depressive disorder: Double-blind placebo-controlled study: Double-blind placebo-controlled study. British Journal of Psychiatry, v.188, n.4, p.346-353. 2006.

Picardi, A., P. Rucci, *et al.* The quality of life of the mentally ill living in residential facilities: findings from a national survey in Italy. Eur Arch Psychiatry Clin Neurosci, v.256, n.6, Sep, p.372-81. 2006.

Power, M., K. Quinn, *et al.* Development of the WHOQOL-old module. Qual.Life Res, v.14, n.10, p.2197-2214. 2005.

Prieto, L., J. Alonso, *et al.* Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual.Life Outcomes., v.1, n.1, p.27. 2003.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research. 1960.

Read, J. L., R. J. Quinn, *et al.* Measuring overall health: an evaluation of three important approaches. J Chronic Dis, v.40 Suppl 1, p.7S-26S. 1987.

Ruggeri, M., G. Bisoffi, *et al.* Subjective and objective dimensions of quality of life in psychiatric patients: a factor analytical approach: The South Verona Outcome Project 4. Br.J Psychiatry, v.178, p.268-275. 2001.

Rush, A. J., M. Fava, *et al.* Sequenced treatment alternatives to relieve depression (STAR\*D): rationale and design. Control Clin Trials, v.25, n.1, Feb, p.119-42. 2004.

Sainfort, F., M. Becker, *et al.* Judgments of quality of life of individuals with severe mental disorders: Patient self-report versus provider perspectives. Am.J Psychiatry, v.153, n.4, p.497-502. 1996.

Saxena, S., D. Carlson, *et al.* The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research. Qual Life Res, v.10, n.8, p.711-21. 2001.

Shelton, R.-C., K.-L. Haman, *et al.* A randomized, double-blind, active-control study of sertraline versus venlafaxine XR in major depressive disorder. Journal of Clinical Psychiatry, v.67, n.11, p.1674-1681. 2006.

Simon, G. E., D. Chisholm, *et al.* Course of depression, health services costs, and work productivity in an international primary care study. Gen.Hosp.Psychiatry, v.24, n.5, p.328-335. 2002.

Simon, G. E., M. Fleck, *et al.* Prevalence and predictors of depression treatment in an international primary care study. Am.J.Psychiatry, v.161, n.9, p.1626-1634. 2004.

Sir, A., R.-F. D'souza, *et al.* Randomized Trial of Sertraline Versus Venlafaxine XR in Major Depression: Efficacy and Discontinuation Symptoms. Journal of Clinical Psychiatry, v.66, n.10, p.1312-1320. 2005.

Skevington, S. M. Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. Pain, v.76, n.3, p.395-406. 1998.

Skevington, S. M., M. Lotfy, *et al.* The World Health Organization's WHOQOL-Bref quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL Group. Qual.Life Res, v.13, n.2, p.299-310. 2004.

Skevington, S. M. e A. Wright. Changes in the quality of life of patients receiving antidepressant medication in primary care: validation of the WHOQOL-100. Br.J Psychiatry, v.178, p.261-267. 2001.

Sloan, J. e S. Mandrekar. Item Response Theory: when is it useful and where does classical test theory fit in? Quality of Life Research, v.14, n.9, p.1982. 2005.

Smith, E. V., Jr. Introduction to Rasch Measurement. Maple Grove: JAM Press. 2004.

Smith, J. E. Metric Development and Score Reporting in Rasch Measurement. Journal of Applied Measurement, v.1, n.3, p.303-326. 2000.



Stevens, S. S. On theory of scales of measurement. Science, v.103, n.2684, p.677-680. 1946.

Stone, M. H. An essay on the ruler: what is measurement? Rasch Measurement Transactions, v.10, p.502-503. 1996.

Streiner, D. L. e G. R. Norman. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press. 2003.

Teasdale, J. D., M. J. Taylor, *et al.* Depressive thinking: shifts in construct accessibility or in schematic mental models? J Abnorm Psychol, v.104, n.3, Aug, p.500-7. 1995.

Tennant, A. Intermediate Rasch course using RUMM 2020. Leeds. 2004a.

\_\_\_\_\_. Introduction to Rasch analysis course using RUMM 2020. Leeds. 2004b.

\_\_\_\_\_. Advanced Rasch Course using RUMM 2020. Leeds. 2005.

Tennant, A., S. P. Mckenna, *et al.* Application of Rasch analysis in the development and application of quality of life instruments. Value.Health, v.7 Suppl 1, p.S22-S26. 2004.

Tennant, A., M. Penta, *et al.* Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care, v.42, n.1 Suppl, p.137-148. 2004.

Tesio, L. Measuring Behaviors and Perceptions: Rasch Analysis as a Tool for rehabilitation Research. J Rehabil Med, v.35, p.105-115. 2003.

THE WHOQOL SRPB GROUP. A cross-cultural study of spirituality, religion, and personal beliefs as components of quality of life. Soc Sci Med, v.62, n.6, Mar, p.1486-97. 2006.

The WHOQOL Group. Initial steps to developing the World Health Organization's Quality of Life Instrument (WHOQOL) module for international assessment in HIV/AIDS. AIDS Care, v.15, n.3, Jun, p.347-57. 2003.

The WHOQOL Group. What quality of life? World Health Forum. 17: 354-356 p. 1996.

The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc.Sci.Med, v.41, n.10, p.1403-1409. 1995.

Trivedi, M.-H., T.-A. Pigott, *et al.* Effectiveness of Low Doses of Paroxetine Controlled Release in the Treatment of Major Depressive Disorder. Journal of Clinical Psychiatry, v.65, n.10, p.1356-1364. 2004.

Trivedi, M.-H., A. J. Rush, *et al.* Evaluation of Outcomes With Citalopram for Depression Using Measurement-Based Care in STAR\*D: Implications for Clinical Practice. American Journal of Psychiatry, v.163, n.1, p.28-40. 2006.

Trivedi, M. H., A. J. Rush, *et al.* Factors associated with health-related quality of life among outpatients with major depressive disorder: a STAR\*D report. J Clin Psychiatry, v.67, n.2, Feb, p.185-95. 2006.

Trompenaars, F. J., E. D. Masthoff, *et al.* Relationship between mood related disorders and quality of life in a population of Dutch adult psychiatric outpatients. Depress Anxiety, v.23, n.6, p.353-63. 2006.

Tylee, A. e P. Gandhi. The importance of somatic symptoms in depression in primary care. Prim.Care Companion.J.Clin.Psychiatry, v.7, n.4, p.167-176. 2005.

Van Alphen, A., R. Halfens, *et al.* Likert or Rasch? Nothing is more applicable than good theory. Journal of Advanced Nursing, v.20, p.196-201. 1994.

Wang, W. C., G. Yao, *et al.* Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-Bref using multidimensional Rasch analysis. Qual.Life Res, v.15, n.4, p.607-620. 2006.

Weiller, E., Y. Lecrubier, *et al.* The relevance of recurrent brief depression in primary care. A report from the WHO project on Psychological Problems in General Health Care conducted in 14 countries. Eur.Arch.Psychiatry Clin.Neurosci., v.244, n.4, p.182-189. 1994.

Wells, K. B. e C. D. Sherbourne. Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. Arch Gen Psychiatry, v.56, n.10, Oct, p.897-904. 1999.

Wells, K. B., A. Stewart, *et al.* The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA, v.262, n.7, p.914-919. 1989.

\_\_\_\_\_. Development of the World Health Organization WHOQOL-Bref quality of life assessment. The WHOQOL Group. Psychol.Med, v.28, n.3, p.551-558. 1998a.

\_\_\_\_\_. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc.Sci.Med, v.46, n.12, p.1569-1585. 1998b.

Wisniewski, S.-R., A. J. Rush, *et al.* Comparison of quality of life measures in a depressed population. Journal of Nervous and Mental Disease, v.195, n.3, p.219-225. 2007.

Wright, B. D. e M. Mok. Rasch Models Overview. Journal of Applied Measurement, v.1, n.1, p.83-106. 2000.

Wright, B. D. e M. H. Stone. Best test design. Chicago: MESA press. 1979.

Yao, G. e C. H. Wu. Factorial invariance of the WHOQOL-Bref among disease groups. Qual.Life Res, v.14, n.8, p.1881-1888. 2005.

# APÊNDICES