

CRIAÇÃO DA VERSÃO PORTUGUESA DO MOS SF-36

Parte II- Testes de validação

PEDRO LOPES FERREIRA

Faculdade de Economia da Universidade de Coimbra. Coimbra.

RESUMO

Neste artigo é descrito o estudo realizado com o objectivo da validação da versão portuguesa do instrumento de medição denominado MOS SF-36. Começaremos por apresentar os resultados da aplicação da versão portuguesa do MOS SF-36 a uma amostra de 930 mulheres grávidas e os resultados de testes das escalas, incluindo a determinação dos valores de coerência interna e de fiabilidade. No entanto, como um instrumento fiável não é necessariamente válido são também apresentados resultados referentes a vários testes de validade. Por fim, concluiremos este artigo com a recomendação de utilização da versão portuguesa do instrumento de medição de estado de saúde SF-36.

SUMMARY

Creation of the Portuguese Version of MOS SF-36 Part II - Validation Tests

This paper describes the study aimed at validating the Portuguese version of the MOS SF-36 instrument of assessment. It starts by presenting the results of the implementation of this instrument in a sample of 930 pregnant women and the results of scaling tests, including the values of internal consistency and reliability. However, since a reliable instrument is not necessarily a valid one, the results of several validity tests are also presented. Finally, this paper ends by recommending the use of the Portuguese version of the SF-36 instrument of assessment.

INTRODUÇÃO

O objectivo deste estudo foi proceder à validação da versão portuguesa do instrumento de medição MOS SF-36. Está inserido num projecto mais amplo que visa detectar alguns amplificadores da história natural da gravidez. Para além dos aspectos físicos e biológicos da gravidez deixados à consideração e ao cuidado dos prestadores, os aspectos psicológicos e de funcionalidade da mulher carecem de maior atenção. Assim, pretende-se com este estudo: (1) obter maior conhecimento acerca da história natural da gravidez; e (2) provar a utili-

dade das medidas de estado funcional da grávida para apoiar as tomadas de decisão clínica.

AMOSTRA

A amostra usada para validar o MOS SF-36 consistiu em 930 mulheres grávidas ou no período pós-parto. Foram seleccionadas todas as grávidas que acorreram à Maternidade e aos Centros de Saúde. Assim, no prazo de um mês, os médicos distribuíram os questionários contendo as versões portuguesas do MOS SF-36 e COOP Grávidas, tendo sido possível fazer-se uma estratificação

referente aos trimestres da gravidez e ao período pós-parto. As grávidas não sentiram qualquer dificuldade em preencher os questionários, tendo-os considerados fáceis de entender.

Quadro 1 - Distribuição dos locais de recolha de informação

Local	Nº de Grávidas
Maternidade Dr. Daniel de Matos - Consulta Externa	670
Maternidade Dr. Daniel de Matos - Internamento	158
Centros de Saúde de Arganil, Cantanhede e Lousã	102
TOTAL	930

A distribuição etária foi normal com média 26.9 anos e com um desvio padrão de 5.3 anos (ver *Figura 1*).

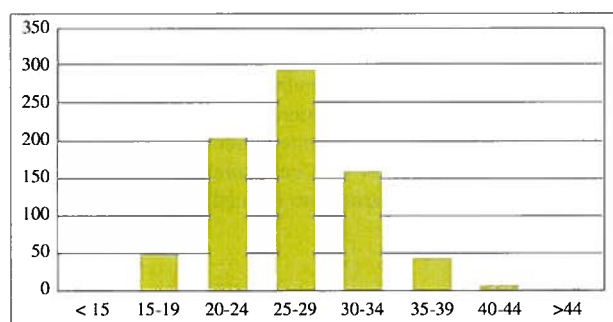


Figura 1 - Distribuição etária das grávidas

Destas 930 mulheres, 704 (76%) não tinham tido qualquer complicação durante a gravidez, 126 (14%) estavam no primeiro trimestre da gravidez, 241 (27%) no segundo trimestre da gravidez, 302 (35%) no fim da gravidez e 208 (24%) estavam no período pós-parto. A *Figura 2* apresenta graficamente a distribuição das grávidas por período de gravidez.

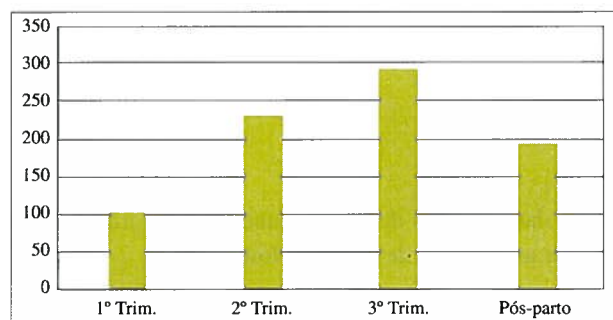


Figura 2 - Distribuição das grávidas por período de gravidez

Foram também recolhidas outras variáveis socio-demográficas e as relacionadas com a condição de gravidez (variáveis de contexto). De entre estas seleccionámos as variáveis referentes ao número de gravidez e partos anteriores. No caso da mulher ter sido internada foi também recolhida informação referente às datas de admissão e de

alta, à razão da admissão, aos diagnósticos, intervenções efectuadas e ao grupo de diagnóstico homogéneo (GDH) correspondente.

Para além dos dados correspondentes às características sócio-demográficas da amostra, procedemos também a uma descrição das respostas ao instrumento SF-36. Assim, quase 50% das respondentes consideraram a sua saúde boa, embora esta percentagem seja ligeiramente inferior um anos atrás. Em relação à saúde e actividades diárias, as percentagens de percepções de grandes limitações variam de 12,7% a 34,5%. Em termos de sentimentos (de vitalidade e de saúde mental) o perfil das doentes é francamente positivo. Enquanto que, em termos de desempenho físico, as respondentes se distribuíram igualmente, em relação ao desempenho emocional houve menos pessoas debilitadas. Só menos de 10% das respondentes afirmaram que tiveram, durante as quatro semanas anteriores, dores fortes ou muito fortes e que as limitaram bastante ou imenso. Em relação à função social, para cerca de 50% das respondentes a saúde não teve qualquer interferência nem as limitou no seu relacionamento normal com a família, amigos, vizinhos ou outras pessoas.

TESTES DAS ESCALAS E COERÊNCIA INTERNA

O primeiro grupo de testes que efectuámos nos dados começou por um teste às escalas que englobou não só o estudo das distribuições dos vários itens dentro de uma mesma escala e a correspondente dispersão, mas também a análise da relação linear entre os valores de cada item e os da escala a que pertencem (coerência interna), assim como o facto de cada item ser mais bem utilizado para medir uma determinada escala do que qualquer outra (validade discriminante). Estes últimos testes foram executados após termos recorrido às técnicas multitrajet de escalas.

A coerência interna é, portanto, determinada analisando as correlações entre cada item e a escala hipotética, considerando-se, em princípio, uma consistência interna grande a que apresenta valores de correlação para além de 0,401. Para além disso, a taxa geral de êxito para uma determinada escala é igual ao número de êxitos de escalonamento dividido pelo número total de testes de escalonamento. Por exemplo, para a escala FF foram executados dez testes, um por cada item^{2,3}.

A validação do agrupamento feito passa pelo teste da validade discriminante, isto é, passa pela análise das diferenças entre as correlações de cada item com a sua escala e as correlações do mesmo item com outras escalas; as primeiras deverão ser, em princípio, maiores

do que as segundas. Também aqui a taxa da validade discriminante foi calculada dividindo o número total de êxitos pelo número de testes realizados. Por exemplo, para a escala FF, foram realizados 80 testes, oito por cada um dos dez itens.

O Quadro II apresenta as médias e os desvios padrão dos itens, assim como as correlações entre cada um deles e as respectivas escalas hipotéticas. Daí se pode ver que as variâncias dos itens podem ser comparadas.

Quadro II - Médias e desvios padrão dos itens e correlações com as escalas

ITEM	MÉDIA	DesvPd	FF	DF	DC	SG	VT	FS	DE	SM
3a	1,87	,74	,69	,21	,15	,03	,23	,03	,01	,13
3b	2,22	,67	,67	,24	,14	,17	,20	,25	,05	,11
3c	2,18	,69	,66	,22	,17	,17	,20	,23	,09	,15
3d	2,08	,71	,45	,23	,15	,09	,20	,10	,12	,13
3e	2,43	,68	,68	,18	,09	,19	,14	,28	,22	,13
3f	2,11	,69	,61	,18	,19	,17	,24	,21	,17	,16
3g	2,30	,74	,69	,25	,15	,19	,16	,26	,19	,09
3h	2,45	,74	,70	,13	,03	,24	,08	,29	,22	,09
3i	2,47	,80	,60	,09	,01	,21	,03	,28	,23	,07
3j	2,47	,80	,60	,09	,01	,21	,03	,28	,23	,07
4a	1,52	,50	,21	,48	,23	,11	,21	,10	,40	,10
4b	1,43	,50	,18	,58	,31	,15	,35	,16	,43	,22
4c	1,39	,49	,25	,59	,31	,17	,35	,16	,28	,19
4d	1,38	,48	,18	,49	,32	,15	,39	,06	,27	,22
7	4,33	1,22	,11	,34	,76	,29	,48	,15	,19	,36
8	4,95	,96	,18	,41	,76	,30	,48	,23	,25	,33
1	3,17	,84	,14	,17	,32	,54	,37	,12	,21	,37
11a	3,89	,91	,28	,15	,21	,56	,25	,26	,21	,28
11b	3,75	,90	,14	,12	,17	,59	,25	,18	,18	,25
11c	3,88	,82	,24	,11	,22	,51	,23	,16	,21	,29
11d	3,46	,85	,12	,17	,24	,62	,37	,14	,21	,33
9a	3,67	1,32	,15	,34	,40	,32	,57	,20	,23	,45
9e	3,38	1,32	,11	,37	,40	,34	,59	,11	,25	,49
9g	3,92	1,16	,21	,31	,39	,31	,62	,22	,19	,40
9i	3,61	1,18	,24	,32	,41	,26	,57	,26	,19	,38
6	4,29	,86	,15	,14	,25	,25	,27	,19	,20	,41
10	4,17	1,29	,32	,11	,09	,14	,14	,19	,20	,09
5a	1,63	,48	,27	,43	,18	,20	,21	,18	,51	,24
5b	1,55	,50	,22	,46	,24	,22	,26	,22	,62	,32
5c	1,56	,50	,15	,19	,13	,23	,17	,19	,40	,25
9b	4,20	1,23	,17	,16	,29	,31	,41	,23	,29	,63
9c	5,15	1,07	,17	,19	,31	,35	,41	,26	,32	,69
9d	3,90	1,32	,12	,23	,35	,36	,52	,21	,28	,67
9f	4,87	1,11	,15	,16	,27	,30	,42	,25	,26	,64
9h	4,63	1,24	,06	,18	,22	,27	,38	,12	,17	,51
2	3,23	,86	,03	,11	,21	,20	,23	,06	,09	,24

Quadro III - Testes de coerência interna e validade discriminante

Escala	Nº de Itens	Amplitude de Correlações		Testes de Consistência Interna ^c		Testes de Validade Discriminante ^d	
		Consistência Interna ^a	Validade Discriminante ^b	Nº Êxitos / Total	Taxa de Êxito (%)	Nº Êxitos/ Total	Taxa de Êxito (%)
FF	10	,45 - ,70	,01 - ,29	9/10	90	75/80	94
DF	4	,48 - ,59	,06 - ,43	4/4	100	32/32	100
DC	2	,76	,11 - ,48	2/2	100	16/16	100
SG	5	,51 - ,62	,12 - ,37	5/5	100	40/40	100
VT	4	,57 - ,62	,11 - ,49	4/4	100	32/32	100
FS	2	,19	,09 - ,41	0/2	0	9/16	56
DE	3	,40 - ,62	,13 - ,46	3/3	100	24/24	100
SM	5	,51 - ,69	,06 - ,52	5/5	100	40/40	100

a - Correlações entre os itens e a escala hipotética; b - Correlação entre os itens e as outras escalas; c - Número = 0,40; d - Número de correlações significativamente altas / número total de correlações

Como também se pode ver neste quadro, quase todas as correlações entre cada item e a sua escala igualam ou excedem o ponto de corte 0,40, o que nos conduz a umas taxas gerais de êxito de coerência interna quase perfeitas (Quadro III). Também os testes de validade discriminante resultaram em grande êxito por todos os itens da versão portuguesa do SF-36.

TESTES DE FIABILIDADE

Fiabilidade é a medida segundo a qual um instrumento de medição fornece resultados de uma forma consistente e precisa. De uma outra maneira, consiste em determinar quanto da variação em pontos é verdadeira ou apenas devida a erros de natureza aleatória.

Para testarmos a fiabilidade da versão portuguesa do MOS SF-36 usámos o procedimento do teste-reteste, isto é, administrámos o mesmo questionário duas vezes aos mesmos indivíduos. Uma outra forma de testar a fiabilidade foi calcular a correlação da divisão em metade, partindo-se do pressuposto de que, se se está a medir uma determinada característica, ambas as metades do teste deverão fornecer resultados sensivelmente equivalentes. Por fim, uma outra técnica análoga para a determinação da fiabilidade foi o coeficiente alfa que consiste na média de todas as possíveis fiabilidades de divisão em metade ajustadas para o número dos itens⁴.

Os valores de fiabilidade encontrados para as escalas do SF-36 estão apresentados no Quadro IV. Todos os estimadores excedem os padrões aceitáveis usados. Para cada escala, a mediana dos coeficientes de fiabilidade

igual ou excede 0,80, com excepção da função social (a média para esta escala com dois itens é 0,76).

Quadro IV - Estimadores de fiabilidade para as escalas do SF-36

Escala	Cronbach α	Teste-Reteste	Divisão em metade
FF	0,8731	0,676	0,6881
DF	0,7511	0,786	0,6944
DC	0,8441	0,452	0,8441
SG	0,8745	0,781	0,4501
VT	0,8264	0,722	0,7664
FS	0,6031	0,557	0,6031
DE	0,7104	0,481	0,7543
SM	0,6446	0,779	0,4468

Os elementos da amostra usada (n=930) diferem segundo características socio-demográficas e segundo o período de gravidez em que se encontram.

TESTES DE VALIDADE

Validade é a medida segundo a qual um instrumento de medição mede aquilo que se pretende que meça. No entanto, embora esta frase seja entendível pela grande parte das pessoas, existem diversas estratégias para se testar a validade. Estes estudos sobre a validade permitem aumentar a compreensão do que constituem as diferenças ou as alterações nas pontuações médias.

No campo da saúde, um dos primeiros testes de validade que normalmente é feito é o da validade de conteúdo, isto é, a determinação se o instrumento de medição cobre extensivamente o universo dos objectos ou

domínio de medição. A distribuição das principais escalas é analisada com um nível de um dos itens previamente fixado.

A Figura 3 representa a distribuição dos pontos da escala de Função Física (FF) correspondente à respostas das mulheres que afirmaram serem capazes de percorrer um ou mais quarteirões sem qualquer limitação devida à sua saúde.

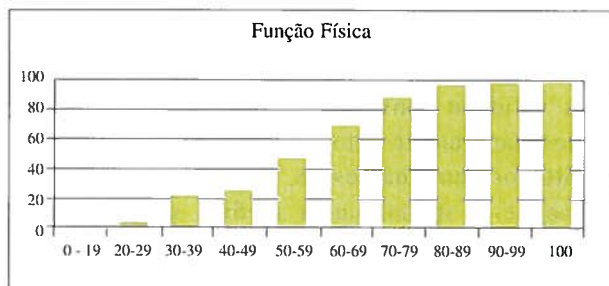


Figura 3 - % das mulheres que conseguem andar um ou mais quarteirões

Como se pode ver, nos últimos três níveis da escala FF, praticamente todas as mulheres (98% a 100%) não sentiam qualquer limitação ao andar, enquanto que nos primeiros dois níveis só cerca de 0% a 3% andavam sem limitação. Nos cinco níveis intermédios 22%, 26%, 48%, 71% e 88% indicaram serem capazes de percorrer um quarteirão. Assim, um salto de 45 para 55 na escala FF representa uma diferença de cerca 22 pontos percentuais (48%-26%) na capacidade de percorrer um quarteirão sem limitação.

Todos os restantes nove itens que fazem parte desta escala foram analisados. Realçaremos aqui apenas duas situações. A primeira é a correspondente ao item 3a, relativo às actividades extenuantes, onde todas as mulheres da amostra (100%) no mais alto nível de Função Física não sentiam qualquer limitação (ver Figura 4). As percentagens de respostas para os restantes nove níveis da escala variaram de 10,6% a 26,6%. Isto demonstra que este item 3a nos informa se o respondente pontua no mais alto nível; em relação aos outros níveis não há grandes variações.

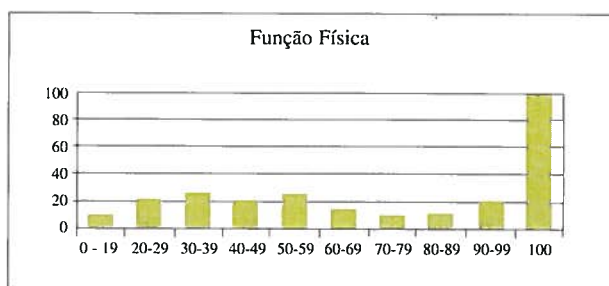


Figura 4 - % das mulheres que conseguem realizar actividades extenuantes

Tal como também se esperava, as limitações em tomar banho ou vestir-se (item 3j) só se fazem sentir em sujeitos com fracos scores de Função Física (ver Figura 5). De facto, os últimos seis níveis da escala FF têm percentagens de limitações que vão de 0% a 23,5%, o que nos permite concluir que este item nos define a parte mais baixa da escala.

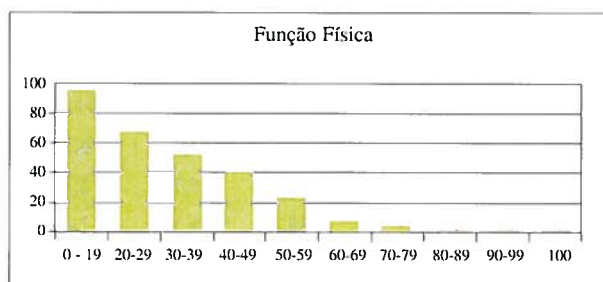


Figura 5 - % das mulheres que têm limitações a tomar banho ou vestir-se

Mas também analisámos a escala Saúde em Geral. O Quadro V apresenta a distribuição dos níveis da escala SG para os indivíduos que responderam, respectivamente, *ótima* (resposta 1), *boa* (resposta 3) e *razoável ou fraca* (respostas 4 ou 5) à pergunta 1.

Quadro V - Valores da escala SG segundo a opinião dos respondentes

Níveis	Escala SG		Percentagem		
	Média	N	Ótima	Boa	Razoável/Fraca
91 - 100	94,8	68	26,5	10,3	0,0
81 - 90	85,3	122	13,2	34,7	0,0
71 - 80	74,1	170	6,5	61,5	0,0
61 - 70	64,9	209	1,9	78,7	6,8
51 - 60	55,5	131	3,1	60,8	23,0
41 - 50	47,2	95	0,0	22,0	78,0
31 - 40	38,5	40	5,1	2,6	92,3
21 - 30	25,7	7	0,0	0,0	100,0
11 - 20	20,0	1	0,0	0,0	100,0
0 - 10	7,5	2	0,0	0,0	100,0

Por fim, ao estudar a escala Vitalidade, comparámos a sua distribuição em relação às grávidas que se sentiram sempre ou a maior parte do tempo (respostas 1 ou 2) cansadas (item 9i) ou com muita energia (item 9e). O Quadro VI apresenta estas distribuições.

TESTES DE SENSIBILIDADE

As medidas em saúde, para além de terem de ser

Quadro VI - Valores da escala VT segundo a opinião dos respondentes

Níveis	Escala VT		Porcentagem	
	Média	N	Cansada	Muita energia
91 - 100	96,5	10	0,0	100,0
81 - 90	86,3	42	0,0	97,6
71 - 80	77,7	90	1,1	82,0
61 - 70	67,3	128	6,3	38,3
51 - 60	57,4	148	4,7	11,0
41 - 50	47,8	182	11,0	6,7
31 - 40	37,9	113	25,7	2,7
21 - 30	28,3	83	65,1	2,4
11 - 20	18,8	37	94,6	0,0
0 - 10	6,5	18	100,0	0,0

fiáveis e válidas, necessitam também de ser sensíveis. E para testarmos isso, fomos ver como é que as várias dimensões de saúde medidas pelo SF-36 se comportavam quando passávamos pelas sub-amostras dos vários períodos da gravidez e pós-parto.

A validade da construção é testada através da especificação do domínio das variáveis, estabelecendo a estrutura interna das variáveis observadas e verificando relações teóricas entre os valores das escalas e

critérios externos. Isto consegue-se através de comparações entre grupos de indivíduos para os quais é de esperar que existam diferenças. Por exemplo, restringindo-nos somente ao cohort das mulheres grávidas, é de esperar que mulheres no último trimestre da sua gravidez apresentem valores mais baixos de funcionalidade física devido, essencialmente, ao peso.

Esta mesma validade de construção, muitas vezes é testada através dos testes da validade convergente e da validade discriminante. Obtém-se uma validade convergente quando métodos diferentes para medir a mesma construção fornecem resultados semelhantes. A validade discriminante consegue-se quando uma medida de uma construção fornece valores diferentes da medida de outra construção.

A validade de critério demonstra se os valores do teste estão sistematicamente relacionados com um ou mais critérios de resultados. Por exemplo, é de esperar que as mulheres no 3º trimestre de gravidez tenham alguns indicadores de saúde mais baixos dos outros períodos.

O Quadro VII a seguir apresenta, para cada um dos períodos da gravidez, os valores da média e do desvio padrão para cada uma das dimensões do SF-36.

Quadro VII - Valores das escalas do SF-36 nos diversos períodos da gravidez

Período		FF	DF	DC	SG	VT	FS	DE	SM
1º Trimestre	n	118	116	117	115	112	117	111	112
	média	67.4	59.5	76.3	67.5	57.5	81.3	61.9	71.7
	desvdpd	56.3	37.8	20.2	15.1	19.5	19.7	38.1	18.8
2º Trimestre	n	218	218	227	220	219	224	210	216
	média	64.2	52.0	73.2	68.2	56.1	80.2	65.4	71.6
	desvdpd	22.5	37.1	22.8	16.9	18.4	21.0	37.2	19.4
3º Trimestre	n	292	289	295	284	291	294	279	289
	média	57.7	35.7	61.5	65.2	49.1	77.8	53.0	69.7
	desvdpd	22.5	34.1	21.1	15.7	19.1	21.1	39.5	17.0
Pós-parto	n	202	203	205	202	206	206	198	205
	média	61.4	30.3	61.4	67.4	50.7	76.8	51.0	70.8
	desvdpd	28.5	31.7	24.7	16.2	19.1	23.4	38.5	19.1

Entre o primeiro e segundo trimestre nada foi possível ser detectado apenas pelo questionário SF-36. Todas as diferenças encontradas eram estatisticamente não significativas. No entanto, entre o segundo e o último trimestre de gravidez, encontrou-se, à excepção das dimensões de saúde mental e função social, uma diferença significativa entre as médias dos vários valores. Isto é apresentado na Figura 6 a seguir. De novo, entre o último período da gravidez e o período pós-parto, nada foi encontrado.

Em resumo, excepto na função social e na saúde mental, onde os valores das médias nos vários períodos são quase constantes, assiste-se a uma descida no valor médio de saúde em relação a todos os restantes, com especial incidência no desempenho físico e na dor corporal. Isto está graficamente representado na Figura 7.

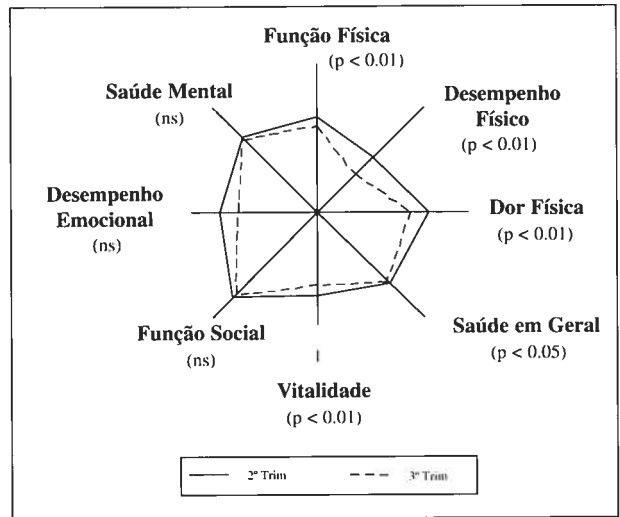


Figura 6 - Sensibilidade das escalas SF-36 nos diversos períodos da gravidez

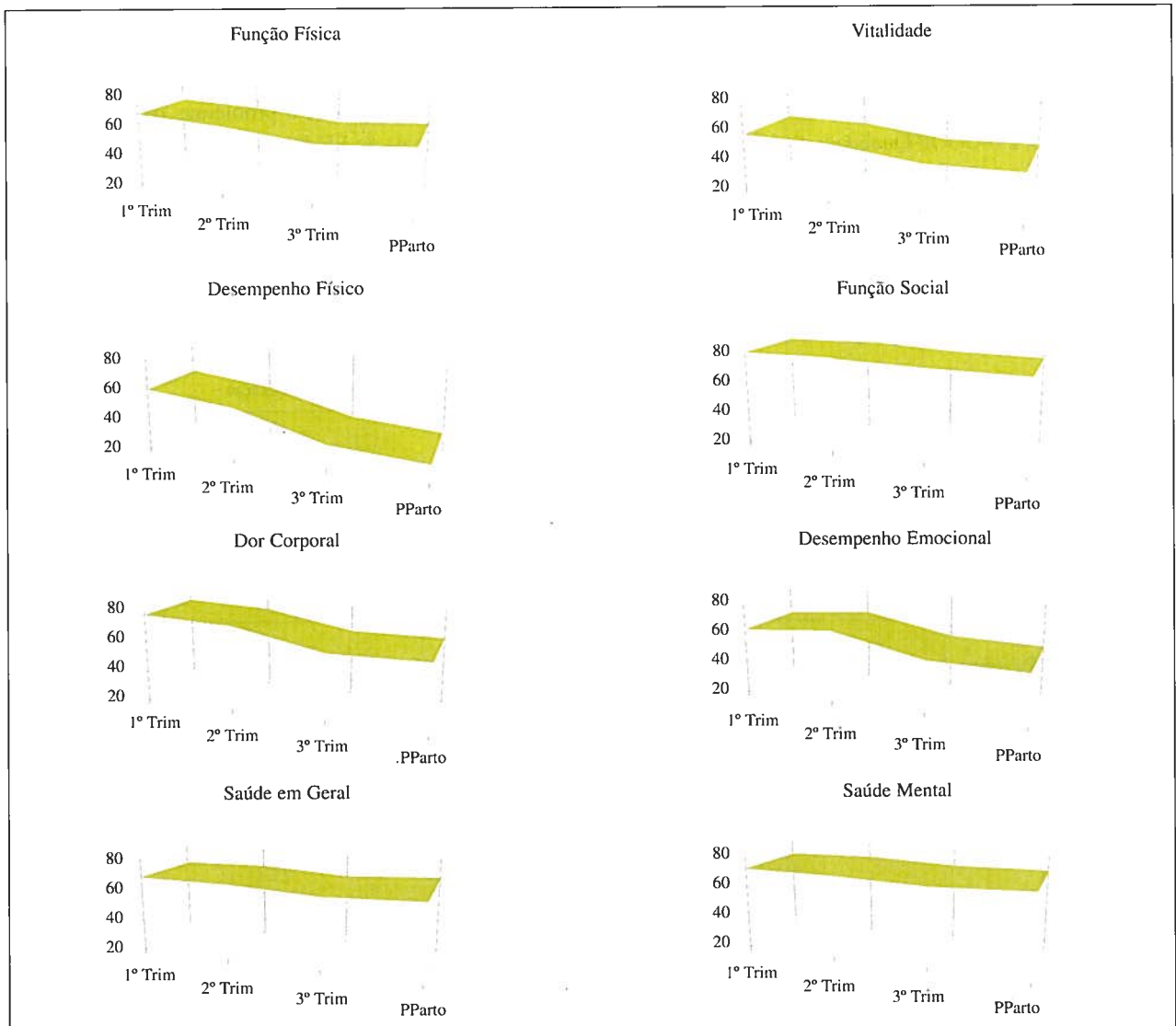


Figura 7 - Escalas do SF-36 nos diversos períodos da gravidez

COMPARAÇÃO COM COOP GRÁVIDAS

Por fim comparámos os resultados obtidos neste instrumento com os obtidos, para as mesmas pessoas, com o COOP Grávidas, um instrumento especialmente desenvolvido pela Faculdade de Medicina de Dartmouth, EUA especialmente para este grupo de doentes.

O objectivo deste instrumento de medição é detectar problemas de saúde e problemas sociais durante a gravidez e o período pós-parto, um pré-requisito para a gestão de qualquer problema psicossocial. Ao analisar os factores hipotéticos subjacentes aos dez COOP Grávidas podemos ver que os COOPs Dor/Mal-estar e Actividades Diárias medem o impacto da gravidez na percepção do bem estar físico. Os COOPs Sentimentos, Stress, Estado Emocional e Amor Próprio medem dimensões da saúde mental. O COOP de Apoio Social revela se a mulher se apercebe se alguém está disponível para a ajudar. Os COOP Hábitos de Saúde medem o grau de envolvimento em comportamentos que podem ser perigosos ou benéficos para a sua saúde ou para a do feto. Finalmente, o COOP Preparação examina se será capaz de lidar com as exigências de um recém-nascido, e o COOP Qualidade de Vida permite-lhe integrar todas as dimensões de bem-estar numa única escala. Cada COOP tem cinco opções de resposta: uma pontuação de 5 corresponde a uma má pontuação.

O Quadro VIII apresenta os resultados das correlações da comparação do SF-36 com o COOP Grávidas. A análise destes resultados fornece-nos alguma prova relativamente à capacidade do instrumento em distinguir entre grupos diferentes de doentes.

Reparar que ambos os instrumentos têm comportamentos semelhantes. No entanto,

- a dimensão COOP Social não corresponde a nenhuma dimensão no SF-36; contudo, está altamente correlacionada com a saúde mental e a vitalidade;
- O COOP Preparação é, de facto, uma dimensão nova; no entanto é interessante verificarmos que está negativamente correlacionada com a dimensão física;
- surpreendentemente, a dimensão física do SF-36 não está correlacionada com nenhuma dimensão COOP, o mesmo acontecendo com a dimensão Alteração de Saúde. Isto parece indicar que, pelo menos para as mulheres desta amostra, o bem-estar físico está mais relacionado com aquilo que são capazes de fazer no âmbito das suas actividades diárias do que em termos absolutos;
- As limitações devidas a problemas de saúde (física ou emocional) do SF-36 apareceram correlacionadas com os COOP físico e mental;
- A dimensão Saúde em geral do SF-36 correlaciona-se com o COOP Qualidade de Vida.

Quadro VIII - COOP Grávidas MOS SF-36

MOS SF-36	COOP Grávidas				
	Mental	Físico	Social	Preparação	Qualidade
Função física	.1959	.2420	.0778	-.0710	.1729
Saúde mental	.6645	.4202	.3858	.1658	.4814
Desempenho	.3426	.4439	.1713	.0195	.2624
Saúde geral	.3406	.3402	.2190	.0633	.4590
Vitalidade	.4286	.5019	.3292	.1888	.4174
Dor Física	.4156	.6937	.2292	.0981	.3566

Quadro IX - Resumo da informação sobre o SF-36

ESCALA	Nº DE ITEMS	Nº DE NÍVEIS	MÉDIA	DESVIO PADRÃO	FIABILIDADE
FF Função Física	10	21	61,4	24,9	,86
DF Desempenho físico	4	5	42,2	36,8	,74
DC Dor física	2	11	66,8	23,1	,85
SG Saúde geral	5	21	66,9	16,0	,79
VT Vitalidade	4	21	52,7	19,4	,78
FS Função social	2	9	78,7	21,5	,29
DE Desempenho emocional	3	4	57,0	38,8	,69
SM Saúde mental	5	26	70,9	18,5	,83

CONCLUSÃO

Este estudo permitiu obter a versão portuguesa do questionário de saúde SF-36. O quadro IX resume o que atrás foi dito relativamente à interpretação possível do MOS SF-36. As primeiras escalas medem a dimensão Função Física, as últimas, a dimensão Saúde Mental e as intermédias são as apresentaram uma validade moderada em relação às duas componentes do estado de saúde.

O número de itens e o número de níveis definidos por cada escala permitem distinguir as escalas mais precisas (FF, SG, VT e SM) com vinte ou mais níveis das menos precisas (DF e DE) com apenas quatro ou cinco níveis.

Embora três das escalas (FF, DF e DC) apresentem validade substancial como medidas de estado de saúde físico é possível distingui-las entre si. Assim, FF mede as limitações em desempenho comportamental nas actividades físicas quotidianas, DF mede a incapacidade em executar tarefas diárias devido a problemas físicos e DC mede essencialmente a gravidade da dor e as limitações resultantes.

Por outro lado, as escalas SM, DE e FS são as que melhor representam a componente mental do estado de saúde. Aqui, enquanto FS e DE medem as limitações ou as incapacidades sentidas pelo sujeito e atribuíveis a problemas pessoais ou emocionais, a escala SM é uma escala bipolar com um chão representando sempre nervosismo e depressão e um tecto sempre com alegria, paz e calma.

As duas escalas restantes apresentam-se concomitantemente sensíveis a resultados físicos e mentais. A escala

VT representa a energia e a fadiga sentidas pelo sujeito e a escala bipolar SG fornece uma pontuação média a quem não avalie desfavoravelmente a sua saúde em termos gerais.

Para tornar a validação da versão portuguesa do MOS SF-36 ainda mais completa, estamos neste momento a implementar um estudo tendente a estudar a sensibilidade a alterações no tempo e um outro tendente à obtenção de normas nacionais.

AGRADECIMENTOS

É nosso dever agradecer aos vários tradutores que nos apoiaram na fase da adaptação cultural do SF-36 para português e às doentes que tão pacientemente responderam aos vários questionários realizados. Por fim, quero exprimir o meu agradecimento a Elsa Vasco e Marta Rosete, e a Victor Raposo, que me apoiaram em diversas fases importantes deste estudo.

BIBLIOGRAFIA

1. HOWARD KI, FOREHAND G: A method for correcting item-total correlations for the effect of relevant item inclusion. *Educational and Psychological Measurement* 1962; 22: 731-735.
2. MCHORNEY CA, WARE JE, LU JFR, SHERBOURNE CD: The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* 1994; 32: 40-66.
3. WARE JE, SNOW KK, KOSINSKI M, GANDEK B: SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center, 1993.
4. CRONBACH LJ, WARRINGTON WG: Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika* 1951; 16: 167-188.