

(図表15)に、病棟種別の(入院料、検査処置料、リハビリ料)の出来高換算医療費の比較を示す。

入院費に相当する部分は、亜急性病床が16343円、回復期リハビリテーション病棟が16303円と大差無く、標準偏差も小さい。検査処置料の平均は、亜急性病床が3056円、回復期リハビリテーション病棟が1736円であり、亜急性が1300円ほど高い。標準偏差が大きく、検査処置料は、特に亜急性病床において症例により大きなばらつきが見られる。

一番下の表とグラフは、出来高で支払われるリハビリテーションの一日あたりの医療費の分布を示す。亜急性病床は、2千円から4千円にピークがあるが、回復期リハビリテーション病棟は、2千円から1.4万までの範囲に広く分布しており、平均が亜急性病床が3095円、回復期リハビリテーション病棟が7134円である。

亜急性・回復期入院時の入院料出来高換算(1日あたり)

	1-1.2万	1.2-1.4万	1.4-1.6万	1.6-1.8万	1.8-2万	2-2.2万	2.2-2.4万	2.4-3万	3万以上	合計	平均入院料	SD
合計	1	14	173	118	36	2	1	2	1	348	16325	1954
亜急性病床		3	109	58	17	2	1		1	191	16343	2052
回復期リハ	1	11	64	60	19			2		157	16303	1833

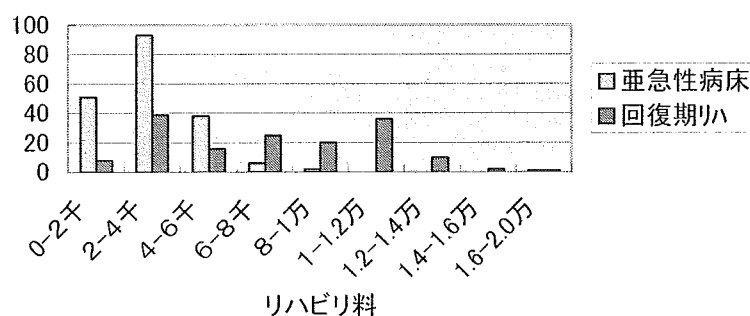
亜急性・回復期入院時の検査処置料出来高換算(1日あたり)

	0-5千	5千-1万	1万-1.5万	1.5-2万	2-2.5万	2.5-3万	3-3.5万	3.5-4万	4-4.5万	4.5-5万	5-7万	7万以上	合計	平均	SD
合計	327	8	2	8		1			1			1	348	2460	6487
亜急性病床	177	3	2	6		1			1			1	191	3056	8453
回復期リハ	150	5		2									157	1736	2371

亜急性・回復期入院時のリハビリ料出来高換算(1日あたり)

	0-2千	2-4千	4-6千	6-8千	8-1万	1-1.2万	1.2-1.4万	1.4-1.6万	1.6-2.0万	合計	平均	SD
合計	59	132	54	31	22	36	10	2	2	348	4912	3518
亜急性病床	51	93	38	6	2				1	191	3095	1892
回復期リハ	8	39	16	25	20	36	10	2	1	157	7134	3762

亜急性回復期リハ入院時リハビリ料(1日当たり)分布



(図表15: 病棟種別出来高換算医療費(入院料、検査処置料、リハビリ料)の比較)

C7 ケースミックス

(i) 病院・病棟種別ケースミックス別症例数

(図表16)に、病院・病棟種別ケースミックス別症例数を示す。今回の対象病院であるA病院は、股関節症、膝関節症の手術で有名な病院であり、B病院は熊本県内の頸髄・胸髄・腰髄損傷患者の多くが集まる脊髄損傷治療センター的役割を担う施設である。

その結果、A病院では股関節症、膝関節症の術後患者が亜急性病病床に入院(82例)し、B病院の回復期リハビリ病棟には、頸髄損傷の発症後46例、胸腰髄損傷の発症後21例が入院していることが、今回の調査対象病院のケースミックス上の最大の特徴といえる。

またB病院の亜急性病棟では、慢性疾患の状態悪化や脱水などの内科の救急的入院患者やその他の非リハビリ目的の患者が数多く入院しているが、A病院は、亜急性病棟も、全てリハビリ目的の患者により使用されている。

他の多くの亜急性、回復期病棟と同様に、脳血管疾患や大腿骨骨折後のリハビリ目的の入院比率が高い。

		合計	病院 A	病院 B	亜急 性	回復 リハ
リ ハ ビ リ 目 的	11:脳血管疾患の発症後	38	35	3	9	29
	12:頸髄損傷の発症後	46		46		46
	13:胸腰髄損傷の発症後	21		21		21
	14:大腿骨頸部、下肢または骨盤等の骨折	99	50	49	51	48
	15:上肢、脊椎など②以外の骨折	6	4	2	6	
	16:関節・脊椎等の手術後	82	81	1	82	
	17:慢性関節リウマチ、各種神経筋疾患等の憎悪	4	1	3	3	1
	18:心疾患・呼吸器疾患の急性期加療後	6		6	4	2
そ の 他 入 院	31:肺炎などの救急入院	1		1	1	
	53:慢性疾患の状態悪化や脱水	10		10	10	
	55:その他	23		23	23	
総合計		336	171	165	189	147
欠損			12	12	2	10

(図表16:病院・病棟種別ケースミックス別症例数)

(ii) ケースミックス別入院日数と医療費

(図表17)に、ケースミックス別入院日数と医療費を示す。表の「総日数」とは、“一般病棟の入院日数+亜急性/回復期の入院日数”を、「亜急性回復リハ日数」とは、亜急性病床または回復期リハビリテーション病棟だけの入院日数を意味する。また「入院費」、「診断治療費」は、現状包括払いであるが、出来高払いで支払われた場合の算定値を示す。「リハ費」は、現状の出来高で支払われている額を示している。「包括部分出来高」は、入院費、診断治療費の出来高払いで支払われた場合の合計値である。

「亜急性・回復リハ日数」は、非リハビリ目的の入院は短く、脊髄損傷のリハビリが90日以上と長い。ケースミックスごとの平均値の差が大きく、また各区分内の標準偏差も平均値の半分程度であり、区分内のばらつきも比較的小さいといえる。この区分を用いて「亜急性・回復リハ日数」の分散の29.6% ($R^2=0.296$) が説明できる。

一方、区分毎の「包括部分出来高」は、最小17525円、最大22788円と差が小さく、また区分内のばらつきが大きい。この区分を用いて「亜急性・回復リハ日数」の分散の1.7% ($R^2=0.0172$) しか説明できず、包括出来高の予測モデルとして使用できない。

	症例数	総日数	SD	亜急性回復リハ日数	SD	入院費	診断治療費	リハ費	包括部分出来高	
リハビリ目的	11: 脳血管疾患発症後	38	63.6	39.2	51.4	38.9	16591	1701	6243	18292
	12: 頸髄損傷の発症後	46	105.2	49.7	94.1	49.7	16283	1487	7356	17770
	13: 胸腰髄損傷の発症後	21	107.9	54.4	99.1	52.2	15699	1826	6560	17525
	14: 大腿骨頸部、下肢または骨盤等の骨折	99	62.6	33.6	48.7	32.2	16642	2051	4686	18693
	15: 上肢、脊椎など②以外の骨折	6	60.5	23.9	41.6	24.3	15855	1623	3603	17508
	16: 関節・脊椎等の手術後	82	58.3	21.7	33.4	17.5	15930	4013	3928	19943
	17: 慢性関節リウマチ、各種神経筋疾患等の憎悪	4	73.0	45.5	57.5	39.6	16758	6030	4545	22788
	18: 心疾患・呼吸器疾患の急性期加療後	6	82.2	52.5	47.8	33.9	16155	1212	3240	17367
入院その他	31: 肺炎などの救急入院	1	45.0		14		15350	2730	0	18080
	53: 慢性疾患の状態悪化や脱水	10	45.6	16.7	28.8	11.4	16568	2922	2207	19490
	55: その他	23	54.0	23.0	38.8	24.0	16727	1567	2650	18285
総合計	336									

欠損 12

$R^2= 0.296$ $R^2= 0.017$
 $Pr < 0.0001$ $Pr > 0.8392$

(図表17: ケースミックス別入院日数と医療費)

(図表 18) に、ケースミックス別入院日数のグラフを示す。先にも述べたが、31：肺炎などの救急入院、53：慢性疾患の状態悪化や脱水、55：その他などの非リハビリ目的の入院は短い。

一方、リハビリ目的の入院では、脊髄損傷とその他の入院に分けることができるようである。以下に示すように、

11：脳血管疾患発症後 51.4 日、

14：大腿骨頸部、下肢または骨盤等の骨折 48.7 日、

15：上肢、脊椎など②以外の骨折 41.6 日

18：心疾患・呼吸器疾患の急性期加療後 47.8 日

と主要なリハビリ目的の入院は、40-50 日に程度に集中している。

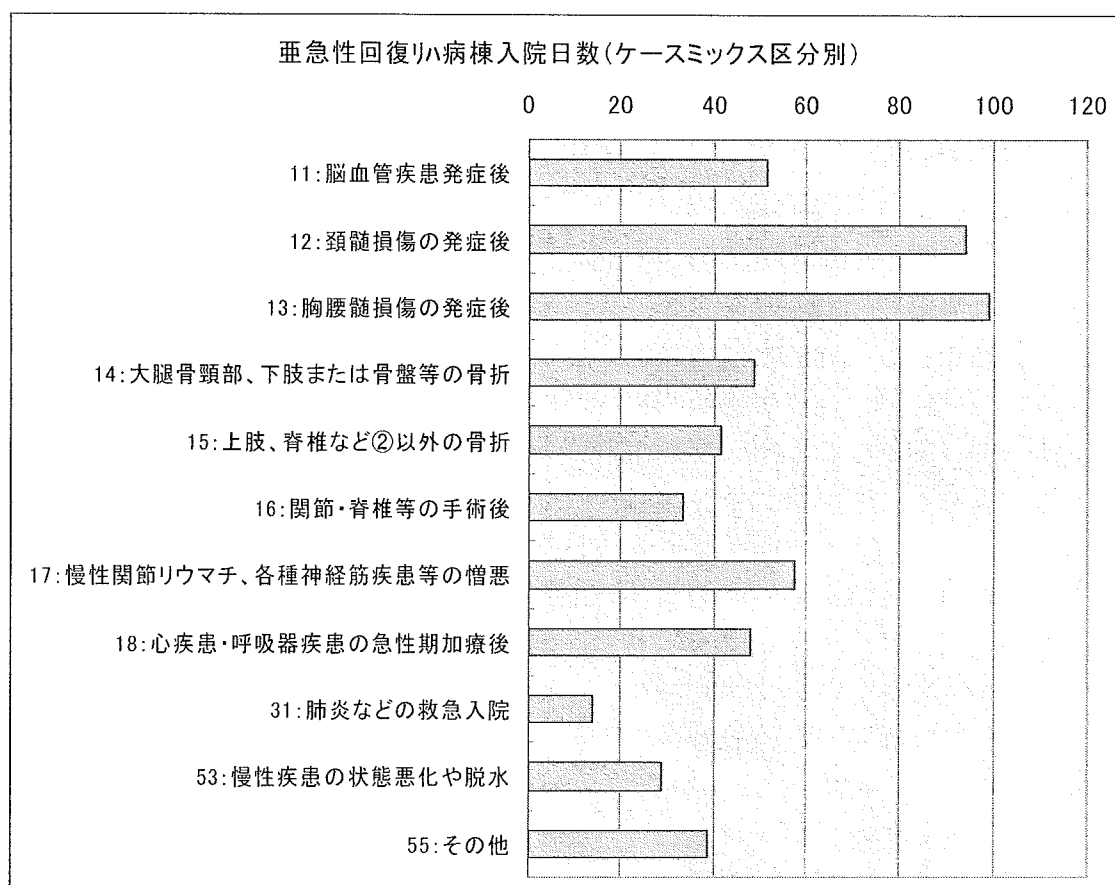
A 病院の亜急性病床に数多く入院している 16：関節・脊椎等の手術後のリハビリは 33.4 日と他のリハビリより入院期間が短い。

一方、脊髄損傷のリハビリ目的の入院の場合、

12：頸髄損傷の発症後 94.1 日

13：胸腰髄損傷の発症後 99.1 日

であり、ともに 90 日を越える。



(図表 18 : ケースミックス別入院日数)

D. 考察

(1) 亜急性病床と回復期リハビリテーション病棟との比較

前年度の研究では、亜急性病床の全国調査を行ない、亜急性病床の実態把握を行なった。今年度の調査の目標の一つとして、回復期リハビリテーション病棟の実態を把握し、亜急性病床との違いを明らかにすることをかかげた。当初の 5 病院以上の目標とは異なり、データ収集の技術的な障害により、対象病院が 2 病院になったが、以下に示すような、少なくとも調査対象 2 病院における亜急性病床と回復期リハビリテーション病棟の違いが明らかになった。

- (i) 両病院ともに回復期リハビリテーション病棟は、脳血管疾患や脊髄損傷後などのリハビリテーション目的の患者のみが入院している。亜急性病床に関しては、A病院は関節・脊椎等の手術後のリハビリ目的患者が多く、B病院は非リハビリ目的の患者が数多く入院している。
- (ii) 亜急性病床が設定している入院期間（90 日以内）が、回復期リハビリテーション病棟の設定している入院期間より短いため、亜急性病床の平均入院期間が 35.2 日、回復期リハビリテーション病棟の平均入院期間が 74.7 日と、入院期間に大きな差が認められた。特にB病院の回復期リハビリテーション病棟には数多くの脊髄損傷後リハビリ目的の患者が入院しているため、平均入院期間が 101.3 日となっている。
- (iii) 回復期リハビリテーション病棟に入院時の寝たきり度の低いリハビリ目的の患者がより多く入院し、しかも亜急性病床と比較してリハビリの密度が高い。その結果、入院時と退院時の寝たきり度の改善の割合が、亜急性病床が 34%、回復期リハビリテーション病棟が 66%であり、回復期リハビリテーション病棟の方が、寝たきり度の改善比率が有意に高い。
- (iv) 亜急性病床と回復期リハビリテーション病棟の入院費と検査処置費を出来高ベースで計算すると、入院費は大きな差が見られないが、検査処置費は亜急性病床のほうが高い。調査時の亜急性病床の包括部分の診療報酬が 20500 円、出来高換算の包括部分の（医療費（入院費＋検査処置費）は、19399 円であり、差し引き 1101 円のプラスとなる。亜急性病床は出来高ベースで運営するより包括払いのほうが有利ということになる。一方回復期リハビリテーション病棟は、包括部分の診療報酬が 16200 円、出来高換算の包括部分の医療費（入院費＋検査処置費）は、18039 円であり、差し引き 1839 円のマイナスとなる。回復期リハビリテーション病棟は、入院日数や長期入院の入院費の減算などの縛りがないならば、出来高ベースで運営するほうが有利ということになる。
- (v) 回復期リハビリテーション病棟でのリハビリの一日平均単価は 3095 円、亜急性病床のリハビリの一日平均単価は 7134 円であり、回復期リハビリテーション病棟でのリハビリ提供量が有意に多い。リハビリを含めた出来高換算の総医療費は、回復期リハビリテーション病棟の方が高い（亜急性 22494 円、回復期 25173 円）。

(2) ケースミックス区分の評価

今回の3年間にわたる研究の最大の目標は、慢性期（急性期後）の包括的な指標の開発であった。しかし平成18年度7月からの医療療養型の診療報酬は、医療区分、ADL区分をもとに支払われることが決まり、この区分が事実上の慢性期の日本における患者区分のスタンダードになることは間違いない。このような区分が導入されることが、昨年度の研究を始める前に予測できたので、昨年度より患者区分の焦点を亜急性と回復期に移した。

昨年度の研究で行なった亜急性病床の全国調査において、使用した亜急性区分が現場において十分使用可能なものであり、また集計結果より亜急性病床の把握に十分役立つものであることを確かめた。また、昨年度のちょうさにおいて亜急性病床の使われ方が、回復期リハビリテーション病床と近いものであることが確かめられた。

そこで今年度は、定量データを用いて、昨年用いたケースミックス区分の評価を試みた。以下、今回の調査より得たケースミックス区分に関する評価結果を示す。

- (i) 脳血管疾患の患者と脊髄損傷の患者は、入院期間が大きく異なるので、別の区分として扱うことが望ましい。
- (ii) 今回の集計結果より、各病棟の患者構成の違いをよく把握することができたので、亜急性・回復期病棟の患者像を把握するツールとしては有用だと思われる。ただし昨年の全国調査で入院患者が存在していたカテゴリーの多くが、今年後は該当患者がいなかった。
- (iii) 入院期間を予測するモデルとしての今回のケースミックス区分の評価は、 $R^2 = 0.296$ であり、有用であると評価できる。
- (iv) 診療報酬の包括部分医療費（入院費＋検査処置費）を予測するモデルとしての今回のケースミックス区分の評価は、 $R^2 = 0.017$ であり、医療資源消費も出るとしては、役立たないという結果となった。

今回の2病院の回復期リハビリテーション病棟に関しては、包括部分医療費（入院費＋検査処置費）のばらつきは小さいので、現行の一患者あたり定額という診療報酬の設定は妥当と考えられる結果といえよう。一方亜急性は、検査処置費のばらつきが大きく、特に非リハビリ目的の患者のばらつきが大きい。今後、急性期病院からの高密度医療の継続的治療目的の患者を亜急性病床で受け入れる、あるいは在宅からの緊急入院患者を在宅復帰させる目的で亜急性病床に受け入れることを、亜急性病床の使命とするならば、なんらかのケースミックスによる支払い体制が必要と思われる。

今回の調査は、2病院にとどまったが、今後より多くの病院からデータを収集し、今回2病院で得た結果が、どの程度普遍性があるのかを検証し、今回開発した亜急性・回復期のケースミックス区分を発展させていきたい。

E. 結論

今回の調査対象病院において、回復期リハビリテーション病棟は、亜急性病棟と比べ、リハビリ目的の寝たきり度がより低い患者が入院し、より長期かつ密度の高いリハビリを受け、退院時の寝たきり度の回復の度合いも高いことが確かめられた。

今回開発した亜急性・回復期のケースミックス区分は、入院期間を予測するモデルとしての R^2 が 0.296 であり有用であるが、診療報酬の包括部分医療費（入院費＋検査処置費）を予測するモデルとしては、 $R^2=0.017$ であり、医療資源消費も出るとしては、役立たないという結果となった。

G. 研究発表

1. 論文発表

1) 高橋泰、猪口雄二、安藤高郎

亜急性病床の現状と今後のあるべき姿

日経ヘルスケア 特別寄稿（2005年5月15日号）

2. 学会発表

1) 第43回日本病院管理学会（東京）

「亜急性病床の現状調査と評価」

II. 研究成果の刊行に関する一覧表

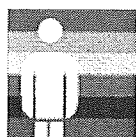
研雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Jiro Okochi, Sakiko Utsunomiya Tai Takahashi	Health measurement using the ICF: Test-re test reliability study of ICF codes and qualif iers in geriatric care	<i>Health and Quality of Li fe Outcome s</i>	3	46 http://www. hqlo.com/c ontent/3/1/ 46	2005
高橋泰、 猪口雄二、 安藤高郎	亜急性病床の現状と 今後のあるべき姿 (特別寄稿)	日経 ヘルスケア	2005年 5月15日号	8-14	2005

Ⅲ. 研究成果の刊行物・別刷

**Health measurement using the ICF: Test-retest reliability study of ICF codes
and qualifiers in geriatric care**

Jiro Okochi, Sakiko Utsunomiya, Tai Takahashi
Health and Quality of Life Outcomes 2005 3:46


**Health and Quality of Life
Outcomes**

Volume 3

Viewing options:

- Abstract
- **Full text**
- PDF (434KB)

Associated material:

- Readers' comments
- PubMed record

Related literature:

- Articles citing this article on Google Scholar
- ⊕ Other articles by authors
- Similar articles (PubMed)

Tools:

- E-mail to a friend
- Download references
- Post a comment

Key:

- ✉ E-mail
- ✉ Corresponding author
- * Contributed equally

Research

Health measurement using the ICF: Test-retest reliability study of ICF codes and qualifiers in geriatric care

Jiro Okochi* 1 ✉, **Sakiko Utsunomiya* 2** ✉ and **Tai Takahashi* 2** ✉

¹Department of Health Services Coordination, Graduate School of Medical Sciences, Kyu University, Maedashi 3-1-1 Higashiku, Fukuoka, 812-8586, Japan

²Department of Health Service Management, International University of Health and Welfare, 2600-1 Kita-Kanamaru Ohtawara, Tochigi, 324-0011, Japan

Health and Quality of Life Outcomes 2005, **3**:46 doi:10.1186/1477-7525-3-46

The electronic version of this article is the complete one and can be found online at:
<http://www.hqlo.com/content/3/1/46>

Received 23 February 2005**Accepted** 29 July 2005**Published** 29 July 2005

© 2005 Okochi et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Outline

- Abstract
- Background
- Methods
- Results
- Discussion
- Conclusion
- Authors' contributions
- Acknowledgements
- References

Abstract**Background**

The International Classification of Functioning, Disability and Health (ICF) published by the World Health Organization (WHO) to standardize descriptive health and disability. Little is known about the reliability and clinical relevance of measurements using the ICF and its qualifiers. This study examines the test-retest reliability of ICF codes, and the rate of immeasurability in long care settings of the elderly to evaluate the clinical applicability of the ICF qualifiers, and the ICF checklist.

Methods

Reliability of 85 body function (BF) items and 152 activity and participation items of the ICF was studied using a test-retest procedure with a sample of elderly persons from 59 institutional and at home care service centers. Test-retest reliability was estimated using the weighted kappa statistic. The clinical relevance of the ICF was estimated by calculating immeasurability. The effect of the measurement settings and evaluators' experience was analyzed by stratification of these variables. The properties of each item were evaluated using both the kappa statistic and immeasurability rate to assess the clinical applicability of WHO's ICF checklist in the elderly care setting.

Results

The median of the weighted kappa statistics of 85 BF and 152 AP items was

and 0.55 respectively. The reproducibility statistics improved when the measurements were performed by experienced evaluators. Some chapter as genitourinary and reproductive functions in the BF domain and major li in the AP domain contained more items with lower test-retest reliability r and rated as immeasurable than in the other chapters. Some items in the checklist were rated as unreliable and immeasurable.

Conclusion

The reliability of the ICF codes when measured with the current ICF qualif relatively low. The result in increase in reliability according to evaluators' experience suggests proper education will have positive effects to raise th reliability. The ICF checklist contains some items that are difficult to be ap the geriatric care settings. The improvements should be achieved by selec most relevant items for each measurement and by developing appropriate qualifiers for each code according to the interest of the users.

Outline Background

Abstract
Background
Methods
Results
Discussion
Conclusion
Authors' contributions
Acknowledgements
References

The International Classification of Functioning, Disability and Health (ICF) published by the World Health Organization (WHO) in 2001 to standardize descriptions of health and disability[1]. Not only has the ICF provided a conceptual framework for description of functioning and disability, health professionals can use it as a tool to describe necessary information conce people with disabilities[2]. Whence the standardization of the language is achieved with the ICF, areas of potential application include description of disability cases[3]; standardization of clinical recording systems, and com of disability statistics between countries[4]. As WHO has provided the ICF qualifier, existing health measures can be mapped to the ICF [5,6]. It ma be possible to develop measurement scales from the ICF codes[7]. The IC consists of four domains: *body structures*, *body functions* (BF), *activities & participation* (AP), and *environment*. The term "disability" is further define "impairment" (dysfunction or loss of "body functions or structure"), "limita (the difficulty an individual may experience in executing a particular activi "restriction" (problems an individual may experience in involvement in life situations). Every domain of the ICF has hierarchical structure, with increa code values (higher digit items) corresponding to more specific functions activities. These are the characteristics of the ICF as taxonomy. In additio applied the ICF to describe the level of disability. For this purpose, WHO developed the qualifiers relevant to each domain, and they were added to codes. For example, using the *Performance* qualifier, mild restriction of "d walking short distance" is coded d4500.1. According to WHO, the ICF cod without qualifier does not have an inherent meaning when used for indivi cases[8]; thus the qualifier is indispensable to denote the level of health.

The ICF in its current version consists of 1424 codes. Therefore, it is nece select a subset of the codes as needed for any given purpose. One of such activities is the development of the ICF checklist, which is composed of m three digit ICF items, as a practical tool to elicit and record information at individual's functioning and disability[9,10]. Other such studies involve th development of the ICF core-sets[11]. They are developed to standardize measure for each chronic condition.

In addition to these studies, which are aimed at determining what to mea is necessary to consider how to describe health and its related status usin ICF codes. One possible approach, as taken in this study, is to apply as m items as possible, as measures to describe health conditions, then to disc reliability and applicability of the ICF codes in a health domain such as ge

care.

In Japan, after the implementation of the long-term care insurance (LTCI) 2000[12], accurate assessment of the needs of elderly clients using LTCI became necessary. In addition, the Ministry of Health, Labor and Welfare recommends the use of the ICF in rehabilitation care planning. Therefore, imperative to assess the accuracy of the application of the ICF codes in geriatric care and rehabilitation settings.

This study examines test-retest reproducibility and the clinical relevance of ICF codes related to the geriatric care setting in the context of this background. The study also aims at evaluating the content validity of the ICF checklist. The research targets the BF and AP domains only, since they contain more easily measurable items and also it was reasonable to lessen the burden of the evaluators.

Outline **Methods**

Abstract
Background
Methods
Results
Discussion
Conclusion
Authors' contributions
Acknowledgements
References

788 elderly patients (age >= 65 years) using LTCI services were selected from 29 hospitals, 29 long-term care institutions, 11 day-care centers, and 14 visiting nursing service centers. Candidates were selected without regard to age, sex, or level of function. However the participants were selected on the basis of functional stability during the one week test-retest period, and ability to give informed consent to study participation. Two independent evaluators judged the stability of candidates. Therefore, a randomization approach was not used to select study participants. Written consent was obtained from all study participants except when it was obtained from a family member by proxy in cases where the subject was unable to provide written consent by him/herself. Subjects who exhibited an acute decline in function during the course of the study were excluded from the final analysis.

Figures



Figure 1
Questionnaire used in this study

Between May and October 2003, each subject was independently evaluated using numerous ICF codes by two health care professionals. The two evaluations were performed within a week of each other. In addition, all evaluators concurrently administered the Typology of the Aged with Illustrations (TAI) questionnaire, a simple illustrative assessment tool developed for care-management of the long-term care insurance to assess the reproducibility of ICF items[13]. The TAI is a four-scale instrument whose reliability and validity have previously been established [14,15]. Evaluators also documented subjects' chronic medical conditions, health behaviours, and living status. Each evaluator also reported on her own professional background and years of work experience on the questionnaire. All evaluators were provided with a comprehensive guide to ICF codes and qualifiers and to the TAI in advance of the assessment.

Since the ICF checklist is composed of three-digit code items (31 items in BF and 48 items in AP domain), every three digit item in the BF and AP domain was initially selected. For more detailed analysis, additional four-digit items were included, by the consensus recommendations of a panel of physiotherapists, occupational therapists, speech therapists, nurses, social workers and care-managers consulted by the authors. 85 BF items (79 three-digit and four-digit) and 152 AP items (81 three-digit and 71 four-digit) ultimately composed the study instrument.

Eight BF domain chapters included were: 1. "Mental Functions" (21 codes); 2. "Sensory Functions and Pain" (15 codes); 3. "Voice and Speech Functions" (10 codes); 4. "Cardiovascular, Haematological, Immunological and Respiratory Systems" (10 codes); 5. "Function of the Digestive, Metabolic and Endocrine Systems" (10 codes); 6. "Genitourinary and Reproductive Functions" (7 codes); 7. "Neuromusculoskeletal and Movement-Related Functions" (12 codes); 8. "Functions of the Skin and Related Structures" (6 codes).

Nine AP domain chapters included were: 1. "Learning and Applying Knowl (15 codes); 2. "General Tasks and Demands" (4 codes); 3. "Communicati codes); 4. "Mobility" (64 codes); 5. "Self-care" (20 codes); 6. "Domestic I codes); 7. "Interpersonal Interactions and Relationships" (7 codes); 8. "M Areas" (12 codes) and 9. "Community, Social and Civic Life" (5 codes). So the three-digit ICF codes not applicable to Japanese geriatric population, "riding animals for transportation" (d460) were nevertheless intentionally included in order to test whether they were correctly identified as irrelevant items.

In response to concerns raised by evaluators, illustrations to each item or study questionnaire were added to promote efficient comprehension of each item [16]. These illustrations are available on the authors' website [17].

The *body functions* qualifier is used in BF measurement in this study. Two qualifiers were used in AP measurement (the performance qualifier and the capacity qualifier). The performance qualifier describes what an individual performs in his or her current environment, while the capacity qualifier describes an individual's ability to execute a task or an action. In this study, the performance qualifier was used to evaluate the AP limitation or restriction. According to the WHO definition, the qualifiers were graded as follows: Level 0 indicates "no problem" (0–4% limitation or restriction); Level 1 "mild problem" (5–24% limitation or restriction); Level 2 "moderate problem" (25–49% limitation or restriction); Level 3 "severe problem" (50–95% limitation or restriction) and Level 4 "complete problem" (96–100% limitation or restriction). Levels 8 were used to describe conditions "not specific" meaning the available information does not suffice to quantify the severity of the problem.

Level 9 were used to describe conditions that were "not applicable" For example, the category *d760* (Family relationships) is not applicable to a patient without living family members.

Figure 1 shows the format of the questionnaire used in this study. All the descriptions used in the questionnaire were identical to the WHO publication of ICF.

The evaluators were given instructions by the authors, using a manual comparable to the ICF checklist [9]. The evaluators are asked to evaluate all possible information available, including interviews, proxy, and medical records.

To evaluate the reliability of each ICF item, the weighted kappa of each item was calculated using the data obtained by the two independent evaluators [18]. To estimate the reliability between the two, the required number of pairs is $\frac{1}{\alpha \beta}$ where α and β are the alpha and beta error level of 0.05 and 0.20 respectively, with a minimum required kappa level of 0.4 and acceptable kappa level of 0.6 [19]. Stratification was done so that there were more than 86 relevant data both in terms of measurement settings and evaluators' experience. In this study, the weighted kappa was classified according to Landis *et al* to moderate (0.41–0.60), substantial (0.61 to 0.8), and almost perfect agreement (above 0.8) [20].

The kappa value was further evaluated by stratifying the experience of the evaluators and care settings. In evaluation by box plot, the kappa statistics showing negative values were replaced to 0.

As an index of irrelevance of the ICF codes, the immeasurability rate of each item was calculated using the sum of samples judged "non specific" or "not applicable" as the numerator and the total number of evaluations as the denominator. The properties of each item were evaluated by using the values of the weighted kappa and the immeasurability rate of each item. The analyses were performed using STATA (version 8.17).

Outline **Results**

Abstract
Background
Methods
Results
Discussion
Conclusion
Authors' contributions
Acknowledgements
References

Figures

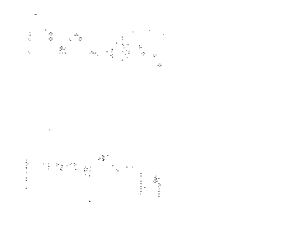


Figure 2
Box plot of weighted kappa of the ICF BF domain by chapter

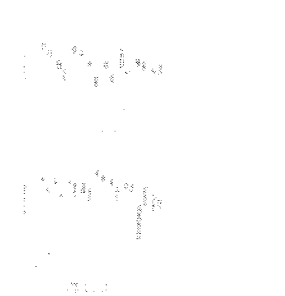


Figure 3
Box plot of weighted kappa of the ICF AP domain by chapter



Figure 4
Box plot of immeasurability rate of the ICF BF domain by chapter

Evaluations were performed on a total of 788 participants. Among these evaluations, one of two evaluations was not complete in 46 persons (6%) cases were excluded from the analysis. Thus, two sets of data were independently obtained from 742 geriatric subjects yielding a total of 1484 data sets. 25 subjects were male (mean age 78.8 years; SD = 9.2 years) and 75% were female (mean age 84.1; SD = 7.6 years); 593 were institutionalized (evaluated at the residential institution) and 149 lived at home (75 evaluated at day-services and 74 evaluated at home).

289 experienced care professionals served as evaluators: nurses (24%), therapists (26%), care managers (22%) and social workers/caregivers (22%). The average amount of work experience as health professional was 10 years (SD8) with a median of 8 years. Among the measurements, 227 pairs were performed by evaluators who both had 8 years or more experience. Conversely, 205 pairs of measurements were performed by evaluators who both had less than 8 years of experience.

Test-retest reliability of the ICF items

The result distribution and rating as non specific (n.s), not applicable (n.a) weighted kappa of 85 BF items (79 three-digit items and 6 four-digit items) and 152 AP items (81 three-digit items and 71 four-digit items) are shown in [Additional files 1](#) (BF domain) and [2](#) (AP domain).

Weighted kappa values of BF domain items ranged from 0.13 to 0.72, with an average of 0.46 and a median of 0.44, while that of AP domain items ranged from -0.17 to 0.79, with an average of 0.55 and a median of 0.59.

[Table 1](#) shows the weighted kappa for the BF and AP domains stratified by evaluators' years of experience, and by care settings of the samples, along with the referential weighted kappa value of the TAI scales. The institutionalized setting showed higher average and median kappa values, compared to those of living at home. Average and median of the evaluation performed by a more experienced pair of evaluators exceeded those performed by a less experienced pair. The kappa values of the four TAI scales concurrently measured with ICF items were as follows: "Mobility" 0.80 (95% C.I. 0.75–0.84); "Mental functioning" 0.75 (0.70–0.80); "Toileting" 0.76 (0.71–0.82); "Eating" 0.78 (0.73–0.83). The weighted kappa value of the TAI scales did not show marked differences between care settings and evaluators' experience.

The higher average kappa value of measurement in institution is not likely due to the experience of the evaluator, since the measurement performed at home contained more pairs of evaluation by experienced evaluators (69%).

[Figure 2](#) shows the box plot of weighted kappa statistics by chapters of the BF and AP domains respectively. Chapters 2,4,5,6 and 7 in the BF domain and Chapters 8 and 9 in the AP domain showed relatively low reliability. In the BF domain the weighted kappa result by the pair of experienced evaluators showed better measurement reproducibility for all chapters. In the AP domain, experienced evaluators showed better reproducibility in chapters 1,5,6,7,

[Figure 3](#) shows the box plot of weighted kappa stratified by care setting. In all chapters, except for the chapter 8 and 9 of the AP domain, showed high reliability. However, caution must be paid when interpreting the weighted kappa result in chapter 8, because kappa values were far less accurate when the results were stratified.

Immeasurability rate



Figure 5
Box plot of immeasurability rate of the ICF AP domain by chapter

Tables

Chapter	Average	Median
1		
2		
3		
4		
5		
6		
7		
8		

Table 1
Average and median weighted kappa values, by care settings and evaluator experience

Additional Material

Additional File 1

Additional File 2

Additional File 3

Additional File 4

The immeasurability rate of BF domain items ranged from 0.00 to 0.96, with an average of 0.06 and a median of 0.03. That of the AP domain items range from 0.00 to 0.90, with an average of 0.13 and a median of 0.02.

Figure 4 shows the box plot of the immeasurability stratified by the evaluators' years of experience. In the BF domain, the highest immeasurability item was "sexual functions" (b640): 0.96. Because this is an exceptionally large figure, it was not plotted on Figure 4. Other top five items rated as immeasurable in the BF domain were "menstruation functions" (b650): 0.28, "sensations associated with genital and reproductive functions" (b670): 0.26, "endocrine gland functions" (b555): 0.18, "procreation functions" (b660): 0.18. Except for the item "endocrine gland functions" (b555), all fell within chapter 6 of the BF domain, "Genitourinary and Reproductive Functions". The top 5 items rated as immeasurable in the AP domain were "preschool education" (d815): 0.90; "school education" (d820): 0.90; "higher education" (d830): 0.89; "producing meaning in formal sign language" (d340): 0.89; and "vocational training" (d825): 0.89. Except for the item "communicating in formal sign language" (d340), all fell within chapter 8 of the ICF AP domain, "Major Life Areas". The pattern of immeasurability by chapter did not differ according to the evaluators' experience.

Figure 5 shows the immeasurability rate by care settings of elderly persons. Except for the immeasurability rate of BF domain chapter 6 (domestic life settings), there were no marked differences in immeasurability rate by care setting.

Properties of the ICF items

The weighted kappa statistics and the immeasurability rate of three-digit BF domain codes are categorized as shown in Additional file 3 and 4 respectively. Items were classified into 3 categories: high reliability (weighted kappa \geq 0.6); intermediate reliability ($0.4 \leq$ weighted kappa < 0.6); and low reliability (weighted kappa < 0.4) using all data. Items were secondly categorized by the median value of the immeasurability rate (immeasurability ≥ 0.3 for the BF domain; immeasurability ≥ 0.2 for the AP domain, the median score) or of low reliability (immeasurability < 0.3 for the AP domain; immeasurability < 0.2 for the AP domain). Additionally, each item was flagged as to whether it was included in the ICF checklist.

High reliability and measurable items included in the study instrument, but not found in the ICF Checklist were: "global psychosocial functions" (b122); "temperament and personality function" (b126); "calculation functions" (b176); "mental function of sequencing complex movements" (b176); "articulation functions" (b320) and "gait pattern functions" (b770) in BF domain, and "focused attention" (d160); "making decisions" (d177); "transferring oneself" (d420); "Moving around in different location" (d460) in the AP domain. On the other hand, items evaluated as low reliability and immeasurable in the ICF Checklist were: "blood pressure functions" (b420); "Haematological system functions" (b430); "immunological system functions" (b435); "respiration functions" (b440); "digestive functions" (b515); endocrine gland functions (b555) and "sexual functions" (b640) in BF domain, and "school education" (d820); "apprenticeship" (d840); "religion and spirituality" (d930) and "human rights" (d940) in the AP domain.

Outline

Abstract
Background

Discussion

The clinical application of ICF codes to diverse populations remains an act of discussion [21-27], with little consensus as to how each code and quality must be utilized for specific populations. There are related previous studies

Methods	deal with the concept of ICF model using different existing scales [28-30]
Results	studies dealt with the ICF reproducibility to assign ICF categories to extar
Discussion	measures[3]. In geriatric care research, Jette <i>et al.</i> have identified distinc
Conclusion	concepts shared by activity and participation[31].
Authors' contributions	However, still to date, to the best of the authors' knowledge, there is no s
Acknowledgements	that has shown the test-retest reproducibility of the ICF as a scale to eval
References	functioning in a specific population.

The ICF is based on a universal model that theoretically can be applied re of cultures, age groups or care settings [7,27,32]. However, various code have different implications for various care settings in practical terms, and individual ICF items requires validity and reliability studies in application t diverse populations. Such efforts are already underway in the form of development of ICF core-sets for specific medical conditions[11]. Concept applications of the ICF to National surveys have also been undertaken[4,3

This study differs from both these approaches, as it does not rely on the e opinions to assure face and content validity, but applies the ICF directly a instrument of geriatric assessment to select more adequate items, while a to develop new scales using ICF taxonomy.

It requires a certain level of test-retest reproducibility and measurability, discard of items which are not appropriate to create new scales.

The authors are now developing the elderly communication performance s according to the result of test-retest reliability statistics, because AP item to communications have acceptable level of test-retest reliability.

Items such as d320 and d340, which are related to communication using sign language, showed low measurability. These items are not always app in the general geriatric care setting, but are pertinent for individuals with loss. Thus, the scale developer can select ICF items with certain reliability measurability according to the scope of each scale.

The other rationale of testing such a wide range of the ICF codes is that e persons hold problems that cover multiple disciplines. This contrasts with core-set project which is relatively disease focused.

Reliability of the ICF qualifiers

Our findings raise concerns about the low reliability of the ICF items using qualifiers.

Although overall reliability of the ICF items was low, it had improved considerably, when the weighted kappa statistics were stratified by the w experience of the evaluators. As shown in Table 1, the weighted kappa of scales did not show marked differences compared to the ICF items. A pre study on the TAI scales also indicated that the reliability was not depende the experience of the evaluators [15]. It indicates the ICF items and its q may be too difficult to quantify in some cases.

By stratifying the results by care-settings, it was possible to get better te: reproducibility in the institutional setting. This may be because more infor including medical records, are available in the setting.

The result of reliability differs depending on the chapter. As shown in Figu and 3, the low weighted kappa value of chapters 4, 5 and 8 of the BF domr and chapters 8 and 9 of the AP domain contribute to the overall low reliat the ICF.

In BF domain, chapter 4("Functions of the Cardiovascular, Haematologica Immunological and Respiratory Systems"), 5 ("Functions of the Digestive,

Metabolic and Endocrine Systems") and 8 (Functions of the Skin and Related Structures") are composed of items that can be described with specific medical examination.

For example, "blood pressure functions" (b420) can be described much more easily with blood pressure level measurable with arm cuff than using qualitative levels from 0 to 4.

Immeasurability of the ICF items

What we call immeasurable in this study include level 8 – not specified (a) information does not suffice to quantify the severity of the problem) and 1 not applicable (e.g., *d760, Family relationships* is not applicable to an elderly person without family).

For example, in case of the global psychosocial functions (b122:immeasurability rate 2.6%), 38 evaluators could not quantify it because the sufficient information was not available and one evaluator rated it as not applicable as shown in [additional file 1](#). This indicates that items with low immeasurability rate can be easily evaluated.

In contrast, 96% of the measurement was rated as immeasurable in sexual function (b640), and most of them were rated as not applicable, as expected in the target sample of this study. Overall, most of the ratings as immeasurable by level 8 (not specific), although some items such as chapter 8 ("Major life area") of the AP domain showed more level 9 than level 8.

Chapter 8 ("Major life area") is comprised of the categories "education" (d810-d839), "work and employment" (d840-d859), and "economic life" (d860-d879), while Chapter 9 (Community, social and civic life) includes "community life" (d910), "recreation and leisure" (d920), "religion and spirituality" (d930), "human rights" (d940) and "political life and citizenship" (d950). To accurately assign scores in the sub-domains of education, work and employment, community life, and political life in a population of institutionalized elderly patients may be difficult, or even inappropriate. Thus, the large proportion of institutionalized geriatric patients in our study sample may have affected high immeasurability scores in these two chapters. The measurement of "education and spirituality" and "human rights" requires multidimensional and subjective assessment. Thus it is difficult to assign either of them into a single code.

The low reliability shown in this study indicates the difficulty of using the measurement tool and is also attributable to the ambiguous nature of the qualifiers. For example, when an evaluator judges the performance level of education, he or she may assess the subject as level 4 ("complete difficulty") because of the subject's inability to obtain further education or to attend an institution for learning. However, this item may also be regarded as "not applicable" or "not specified," especially in the context of institutionalized geriatric patient for whom school attendance is not an expected component of daily life.

In contrast, frequently assessed items in the LTCI assessment appeared to have high reliability. Presumably because items such as toileting and self-dressing constitute a part of a standard self-care assessment already widely used by healthcare professionals [37]. This similarity may explain the high reproducibility of self-care item assessments between independent evaluators in our study.

Validity of the ICF Checklist

An additional purpose of this study was to evaluate the validity of the ICF Checklist in geriatric assessment. We have also used the checklist as a training tool for evaluators, because it was the sole available material at the commencement of this study for official training of the ICF. We have found that the existing ICF Checklist lacks several items which we found scored high

reliability and low in immeasurability rate. These items include, "global psychosocial functions" (b122); "temperament and personality function" ("calculation functions" (b172); "mental function of sequencing complex movements" (b176); "articulation functions" (b320) and "gait pattern function" (b770) in the BF domain, and "focusing attention" (d160); "making decisions" (d177); "transferring oneself" (d420); "Moving around in different locations" (d460) in the AP domain.

The ICF checklist includes less reliable and immeasurable items, e.g. "blood pressure functions" (b420); "haematological system functions" (b430); "immunological system functions" (b435); "respiration functions" (b440) "digestive functions" (b515); "endocrine gland functions" (b555) and "sexual functions" (b640) in the BF domain, and "school education" (d820); "apprenticeship" (d840); "religion and spirituality" (d930) and "human rights" (d940) in the AP domain. Some of the body function related items could be described with chronic disease, such as high blood pressure, anemia, and diabetes. Items not relevant to the elderly care settings such as school education and apprenticeship might be just omitted when applying the scheme to those settings.

Importance of participation in religions and spirituality might vary depending on cultural settings. Also, human rights (d940) may play a pivotal role on understanding geriatric domestic violence.

This result should help selecting more useful sets of the ICF items that will reflect evaluators' needs and reliability of items. Some modification to the checklist may also facilitate the use of the ICF.

Study Limitations

There are a few limitations in this study. The samples were selected from service providers based on the stability of the function during the test-retest period. The kappa statistic is dependent on the samples. Therefore these samples might not fully represent the target population, namely the elderly using long-term care services in Japan. However, the use of a large sample obtained from multiple centers is nevertheless indicative of relatively low reliability of ICF items measured with the qualifiers.

Also, other possible confounders such as the cultural settings and evaluator professional backgrounds may influence the ICF measurement values. It is possible that some of the ICF items show different item functioning (DIF) depending on these confounders. The Rasch measurement technique is applied to answer this question, which remains to be studied[38]. The illustrations provided by the authors to clarify the definition of each item could have biased the results. However, our intention in incorporating illustrations was to standardize evaluations. Previous studies have shown that illustrations increase the reliability of assessment instruments[39].

Lastly, the authors used the sum of qualifiers 8 and 9 as a simple index of immeasurability. Items with a high prevalence of level 8 suggested that it was difficult for the evaluator to ask the question or obtain the information from the medical chart. In contrast, assignment of a qualifier of 9, which was more prevalent in chapter 8 of AP domain, suggested these items were not applicable. However, these two qualifiers may convey quite different information, and the study design made it difficult to compare the differences between these two qualifiers. In addition, it was difficult to analyze inter-rater reliability of qualifiers 8 and 9 because of the skewed distribution of the result between these two levels. However, the prevalence of these qualifiers, as shown in Additional File 1, should help in selecting ICF items for future research.

Conclusion

The reliability of the ICF codes as measured with qualifiers is relatively low. The ICF Checklist requires modification. Improvements should be achieved by selecting the most relevant items for each measurement and constructing appropriate qualifiers for each code according to the interest of users.

Authors' contributions

JO, SU and TT carried out the study design, data collection, statistical analysis and preparation of this manuscript.

Acknowledgements

We would like to appreciate the great support of Professor Takashi Hashirama, Kyushu Rehabilitation College, Dr. Yasuko Arase, Fukuoka City Government, and all healthcare professionals who have cooperated in this study. This research was funded by Health Labor Science Research Grant; 1-15-Seisaku-018.

Outline

Abstract
Background
Methods
Results
Discussion
Conclusion
Authors' contributions
Acknowledgements
References

References

1. World Health Organization.: *International classification of functioning, disability and health : ICF*. Geneva , World Health Organization; 2001.
Return to citation in text: [1]
2. Stucki G, Ewert T, Cieza A: **Value and application of the ICF in rehabilitation medicine.** *Disabil Rehabil* 2002, **24**(17):932-938. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1]
3. Ogonowski J, Kronk R, Rice C, Feldman H: **Inter-rater reliability in assigning ICF codes to children with disabilities.** *Disabil Rehabil* 2004, **26**(6):353-361. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1] [2]
4. Madden R, Choi C, Sykes C: **The ICF as a framework for national data: introduction of ICF into Australian data dictionaries.** *Disabil Rehabil* 2003, **25**(11-12):676-682. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1] [2]
5. Perenboom RJ, Chorus AM: **Measuring participation according to the International Classification of Functioning, Disability and Health (ICF).** *Disabil Rehabil* 2003, **25**(11-12):577-587. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1]
6. Brockow T, Wohlfahrt K, Hillert A, Geyh S, Weigl M, Franke T, Resch KL, Cieza A: **Identifying the concepts contained in outcome measures of clinical depression and depressive disorders using the international classification of functioning, disability and health as a reference.** *J Rehabil Med* 2004, 49-55. [PubMed Abstract] [Publisher Full Text]

Return to citation in text: [1]

7. Simeonsson RJ, Leonardi M, Lollar D, Bjorck-Akesson E, Hollenweger J, Mart
Applying the International Classification of Functioning, Disability and Health (ICF) to measure childhood disability.
Disabil Rehabil 2003, **25**(11-12):602-610. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1] [2]
8. Ustun B: **The international classification of functioning, disability and health-a common framework for describing health states.**
[http://whqlibdoc.who.int/publications/2002/9241545518_Chap7.3.pdf]
Return to citation in text: [1]
9. World Health Organization.: **The ICF checklist.**
[<http://www3.who.int/icf/checklist/icf-checklist.pdf>]
Return to citation in text: [1] [2]
10. Ewert T, Fuessl M, Cieza A, Andersen C, Chatterji S, Kostanjsek N, Stucki G:
Identification of the most common patient problems in patients with chronic conditions using the ICF checklist.
J Rehabil Med 2004, **22**:22-29. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1]
11. Stucki G, Grimby G, (Eds): **ICF core sets for chronic conditions.**
J Rehabil Med 2004, **44**:5-141. [PubMed Abstract] [Publisher Full Text]
Return to citation in text: [1] [2]
12. Matsuda S: **The health and social system for the aged in Japan.**
Aging Clin Exp Res 2002, **14**(4):265-270. [PubMed Abstract]
Return to citation in text: [1]
13. Takahashi T, Okochi J, Takamuku K, Matsuda S: **The introduction of typographic illustrations for the aged with illustrations.**
Casemix Quarterly 2001, **3**(1):3-14.
Return to citation in text: [1]
14. Okochi J: **Increase of mild disability in Japanese elders: A seven year follow-up cohort study.**
BMC Public Health 2005, **5**(1):55. [PubMed Abstract] [BioMed Central Full Text] [PubMed Central Full Text]
Return to citation in text: [1]
15. Okochi J, Takahashi T, Takamuku K, Matsuda S, Takagi Y: **Reliability of a new assessment instrument with illustrations.**
Geriatr Gerontol Int 2005, **5**(1):37-47. [Publisher Full Text]
Return to citation in text: [1] [2]
16. Sutch S: **ICF illustration library.**
Bulletin of the World Health Organization 2004, **82**(7):550-551.
Return to citation in text: [1]
17. Takahashi T: **ICF illustration library.**
[<http://www.icfillustration.com>]
Return to citation in text: [1]
18. Fleiss J, Levine B, Cho Paik M: *Statistical Method for Rates and Proportion.*
New Jersey , Wiley and Sons,Inc; 2003.
Return to citation in text: [1]
19. Walter SD, Eliasziw M, Donner A: **Sample size and optimal designs for randomized studies.**
Stat Med 1998, **17**(1):101-110. [PubMed Abstract] [Publisher Full Text]