II 研究デザインの設定

研究計画を立てていくうえで，PECOもしくはPICO（表2）を明確化していくことは臨床研究を行ううえでの基本となる．PECOを書き出して，実施可能性や倫理性など前述のFINERを勘案しながら，研究デザインを設定していく．

対象を一般住民とするのか，受療患者とするのか，曝露や介入を能動的に行うのか否か，対象の比較となるコントロールをどうとるか，アウトカムとして何を用いるのか，この4点で研究デザインが大きく変わってくる．一般には住民ベースの臨床研究で能動的な介入を行うことは不可能に近いので，曝露や介入の影響は生活歴，既往歴などから検討される．受療患者を対象とした場合には介入研究が可能であるが，対象が一般人口を代表しないので罹患率や有病率を調べたいという場合には不向きである．

例として「白内障術後眼内炎を眼内灌流液への抗菌剤添加によって減少させることができる」という研究仮説をもったとしよう．この場合，PECOのPは白内障手術患者になり，Eは手術時の眼内灌流液への抗菌剤添加，Cは白内障手術を受けるが眼内灌流液への抗菌剤添加をしない患者，Oのアウトカムは術後眼内炎の発症の有無ということになる．研究デザインとその実施可能性を考えるうえでは，サンプルサイズと割付けの有無（介入研究か観察研究か）が重要となる．

サンプルサイズの問題を考えてみよう．術後眼内炎のような発生頻度の低い事象をアウトカムとする場合，治療群と対照群の差を出すためには非常に大きな組み入れ症例数が必要となる．仮に，対照の眼内炎発症率を0.1%，治療群の発症率を0.05%と見積もると必要症例数は8万例以上と計算される．眼内灌流液への抗菌剤添加の例はESCRS Endophthalmitis Study Groupによりランダム化比較試験として実際に行われた臨床研究であるが，こうした大規模な臨床研究を実施するのは容易ではない．ESCRSの研究では対照群の眼内炎発症率が0.29%とかなり高かったこともあり，有意差が明らかとなった時点で中止されたが，それでも約1万6千例がエントリーされている．同じ研究テーマを症例の割付けを行わないコホート研究のデザインで検討した臨床研究もいくつかあるが，これらでは数万例から20万例以上を対象としている．コホート研究や症例対照研究はランダム化比較試験よりもエビデンスレベルは低いとされるが，後ろ向き研究が可能というメリットもある．また，介入研究による割付けが困難な場合や割付けの倫理性が問題になる場合には観察研究を選択したほうがよいし，介入研究の場合には健康被害に対する補償を考慮する必要がある．補償については後述する．

サンプルサイズを減らして研究の実施可能性を高めるためにはアウトカムの取り方を変えることも重要である．たとえば，ある眼圧降下剤の効果をみる場合に，眼圧が5mmHg以上下がったかどうか（2区分変数とよばれ，眼内炎の例も2区分変数である）をアウトカムとするよりも，投与前後の眼圧をアウトカム（連続変数）としたほうがサンプルサイズを小さくできるし，場合によってはサンプル比（介入群と対照群の比）を変えると実施可能性が高くなることがある．もう1つの戦略は，より頻度の高いアウトカムを用いることである．術後眼内炎の例では，手術終了時に前房水や結膜嚢ぬぐい液を採取して細菌培養検査に供し，その陽性率をアウトカムとする方法が考えられる．前房水から細菌が検出されることと眼内炎が発症することはイコールではなく，このようなアウトカムは代替アウトカムとよばれる．エビデンスレベルは低くなるが，限られた症例数で結果を出しやすいというメリットがある．

以上のようにPECO（PICO）を基にして研究デザインを考えることで，研究計画が固まっていくことになる．この部分は重要であり，研究グループの規模，研究資金なども考慮して実施可能性が高く，しかも研究仮説を実証できるようなデザインを選択する必要がある．

表2 研究デザイン設定のPECO（PICO）

P：Patient	対象
E（I）：Exposure（Intervention）	曝露（介入）
C：Comparison	比較対照
O：Outcome	結果・効果

III　研究計画書の作成

研究テーマとデザインが決まったら，次に研究計画書を書くことで具体的に細部を詰めていく．研究を始める前に研究計画書の作成を行うことは，研究の目的，プロセス，意義を自分のなかでまとめていく良い機会となる．研究計画書は将来の論文のひな型になるので決して無駄にはならない．

研究計画書では，研究の目的と背景，研究仮説（テーマ），研究デザイン，セッティング（研究・調査を行う場所），対象，観察項目とその測定方法，データの取得方法，データの管理方法，データの解析方法，倫理的事項などについて記載を行う．先行研究や類似の研究がある場合には，その対象と方法の項を参考にすることができる．学術論文は結果や考按，結論の部分がおもに読まれ，対象と方法の項目はざっと流し読みという場合が多いが，この場合は対象の取り方（選択基準や除外基準）や観察項目，データの取得方法，解析方法を読み込むことが必要となる．「批判的に」読むことがおそらく大切であり，先行研究をなぞって研究計画を立てるのではなく，先行研究の欠陥や見落としている点を吟味し，自らの研究計画に生かすという方針で行いたい．

研究計画書のドラフトができたら，今度は逆に，共同研究者や同僚に自分の計画書を批判的に吟味してもらうとよい．このことは多施設共同研究など大規模な研究を始める場合には特に重要であり，自分の施設での常識は他の施設では非常識ということもある．本格的に研究を開始する前に，少数の施設でパイロット研究を行って，研究の実施に問題がないかどうか確認し，修正点を洗い出すのも良い方法である．

IV　倫理的事項とインフォームド・コンセント

医学研究における倫理の一般原則には，人権尊重，最善，公正の3つがあげられる．人権尊重の原則には，インフォームド・コンセントを得ること，判断能力が損なわれた人を守ること，個人の秘密を守ることが含まれる．最善の原則は，対象者のリスクに見合うだけの価値ある成果が得られるように最大の努力を払うことである．この場合のリスクには検査や治療に伴う身体的リスクだけでなく，個人情報の漏洩，偏見・差別など心理社会的なリスクが含まれる．公正の原則は，研究に伴う恩恵とリスクに関して対象者間に不公正が生じないように配慮することである．こうした概念はヘルシンキ宣言に含まれている．

わが国における倫理指針としては，疫学研究に関する倫理指針，臨床研究に関する倫理指針が代表であり，この他にヒトゲノム・遺伝子解析研究に関する倫理指針，遺伝子治療臨床研究に関する指針，ヒト幹細胞を用いる臨床研究に関する指針などがある．指針の一覧と内容に関しては，厚労省 HP（http://www.mhlw.go.jp/general/seido/kousei/i-kenkyu/index.html）がわかりやすい．疫学研究，臨床研究に関する倫理指針の全文は http://www.lifescience.mext.go.jp/files/pdf/37_139.pdf と http://www.mhlw.go.jp/general/seido/kousei/i-kenkyu/rinsyo/dl/shishin.pdf から入手できる．

特殊な例を除くとほとんどの臨床研究は，疫学研究に関する倫理指針か臨床研究に関する倫理指針の適用を受ける．指針を遵守しなくとも法的に罰せられることはないが，公的研究費の応募資格が制限されることがある．また，倫理審査委員会で研究内容の承認を受ける際には指針の遵守が求められ，論文で研究内容を発表する場合にも倫理指針の遵守が求められるようになりつつある．

厚労省の指針では，疫学研究は人の疾病の成因及び病態の解明並びに予防及び治療の方法の確立を目的とする科学研究とされ，その対象となる範囲は広い（表3）．ある疾患について患者の診療情報を収集・集計し，解析する臨床研究，診断・治療などの医療行為について当該方法の有効性・安全性を評価するため診療録など診療情報を収集・集計して行う観察研究などはもちろんのこと，介入研究であっても介入の内容が医薬品でなく，食品（健康食品など）の場合には疫学研究指針の範囲となる．

インフォームド・コンセントの形式に関して整理してみると，文書説明，文書同意が必要かどうかは，適用される指針の種類，介入研究か観察研究か，人体からの試料採取の有無，採取の侵襲性の4つから区分される（表4）．いずれの場合でも拒否の自由が明記されていなければならない．これをみると観察研究の多くは必ずしも文書説明，文書同意が必要ではなく，研究情報公開（研究

表 3　疫学研究指針と臨床研究指針の適用範囲

	疫学研究指針の範囲	臨床研究指針の範囲
対象となる研究の概念	原則として，自ら異なる事象を作り出すのではなく，存在する異なる事象を分類し，比較する研究が対象． 〔例：実際に異なる治療（治療薬を含む）が行われている被験者などを対象として，複数の群に分けて評価を行う研究．〕	主として，能動的に人体に働きかけ，その効果を比較する研究であって，異なる事象（投薬群と非投薬群など）を作り出す研究が対象． 〔例：当該研究の目的のために，被験者に能動的医療介入（治療など）を行う研究．〕
治療等の医療行為の有無	能動的に医療介入（採血などのサンプリングに係る医療行為は除く）の効果を評価する研究は対象外．	主として，能動的に医療介入の効果を評価する研究が対象．
注意点	医療介入（治療）以外の能動的な人体への働きかけ（食品摂取比較など）を行い，その効果を比較する研究は疫学研究．	

表 4　疫学研究指針と臨床研究指針におけるインフォームド・コンセント要件

疫学研究指針			
	検　体	その他	インフォームド・コンセント要件
介入研究 （群間比較研究）	人体採取検体あり	採取に侵襲性あり	文書説明，文書同意
		採取に侵襲性なし	説明，同意の記録
	人体採取検体なし	個人単位の研究	説明，同意の記録
		集団単位の研究	IC 不要，研究情報公開，拒否機会付与
観察研究 （介入研究以外）	人体採取検体あり	採取に侵襲性あり	文書説明，文書同意
		採取に侵襲性なし	説明，同意の記録
	人体採取検体なし	既存資料以外の情報を用いる	IC 不要，研究情報公開，拒否機会付与
		既存資料のみを用いる	IC 不要，研究情報公開

臨床研究指針			
	検　体	その他	インフォームド・コンセント要件
介入研究 （能動的医療介入研究）			文書説明，文書同意
観察研究 （介入研究以外）	人体採取検体あり	採取に侵襲性あり	文書説明，文書同意
		採取に侵襲性なし	説明，同意の記録
	人体採取検体なし		IC 不要，研究情報公開

機関の HP や院内掲示板での告知）で良い場合も少なくない．

表 4 の人体採取検体の定義はむずかしく，たとえば感染性角膜炎から分離された微生物だけを扱う研究では人体からの試料とみなされないが，症例の臨床情報と結びつける場合（実際上はほとんどこちらに該当する）には人体採取検体として扱うことになる．採取の侵襲性に関しては，採血は侵襲性あり，尿や便は侵襲性なしに分類するとされている．なお，手術中に治療を目的として切除された組織（角膜移植時の摘出角膜や線維柱帯切除術での虹彩・隅角組織など）は採取に侵襲性なしと判断される．

以前の手術時に得られた病理標本を用いて新たな研究を行おうという場合など，臨床研究開始前に得られた既存の試料を用いる研究は少なくない．この場合には被験者の同意を得るのが原則であるが，古い試料で被験者の所在が不明，死亡などの事情により同意を得ることが困難な場合もある．このような場合には，試料が匿名化さ

れているか，同意がなされているかによって多少の手続きの違いはあるが，倫理審査委員会と研究機関の長の承認が得られれば，研究を行うことができる．

V　倫理審査委員会

研究計画書が整ったら，説明文書，同意文書など各種の書類を添付して倫理審査委員会の承認を得る必要がある．臨床研究の倫理指針には，「当該臨床研究機関の長が設置した倫理審査委員会以外の倫理審査委員会に審議を依頼することが出来る」という文章があるが，基本的には自施設の倫理審査委員会の承認が原則である．ただし，多施設共同研究の場合には主施設の倫理審査委員会で承認されれば，他の施設では無審査で承認または迅速審査に付される可能性がある．倫理審査委員会が設置されていない病院やクリニックで臨床研究を行う場合にはこのような外部付託の仕組みを活用して外部機関に審査を付託していく必要がありそうである．

VI　健康被害に対する対応

介入研究を行う場合に問題になる点の1つが，臨床研究に関する倫理指針にある「医薬品・医療機器を用いた介入研究では，健康被害に対する補償のための保険その他の必要な措置を講じなければならない」という項である．補償は賠償とは異なり，「過失の有無とは関係なく，有害事象に対する救済措置を行う」という概念である．さらにここでの有害事象とは「医療介入がなされた際に起こる，あらゆる好ましくない，あるいは意図しない徴候（臨床検査値の異常を含む），症状，または病気のことであり，当該医療介入との因果関係の有無は問わない」と定義されている．素直に読むと，因果関係がなく過失のない健康被害についても臨床研究にエントリーされている限り，研究者に補償の義務があることになる．

臨床研究の補償保険は，東京海上日動，日本興亜損害保険，損害保険ジャパンなど数社が取り扱っているが，このような補償の概念のもとでは介入を伴う臨床研究の実施が困難となる懸念がある．ただし，抗がん剤や免疫抑制剤など重篤な副作用が高頻度で発現することが予想される場合には補償保険の必要はなく，医療費，医療手当などの手段で補完してもよいとされている．補償保険に関しては，研究者，保険会社のいずれもが手探りという状態であり，倫理審査委員会での扱いも施設による幅が大きいようである．現実的な対応として，少なくとも研究計画書や同意説明文書に補償に係る方針や金銭的事項について明記することが推奨される．研究の内容やリスクに応じて「健康被害が生じた場合には回復のために必要な治療などの措置を速やかに講ずるが，通常の健康保険を適用し，自己負担分は患者に負担してもらう」「健康被害が生じた場合の補償はしない」といった内容でもよく，倫理審査委員会の判断を仰ぐことになる．

なお，介入研究でも医薬品・医療機器を用いない研究（たとえば介入内容が食事指導である場合など）や観察研究では補償のための措置，補償保険への加入は必要ないが，補償の有無の説明は必要とされており，説明文書に明記しておくとよい．

臨床研究の倫理指針によれば，臨床研究に関連する重篤な有害事象，不具合が発生した場合には，研究責任者は必要な措置を講じるとともに臨床研究機関の長に報告する義務がある．臨床研究機関の長は，倫理審査委員会等に報告し，多施設共同研究の場合には共同臨床研究機関への周知などを行わなければならない．有害事象が予期しない重篤なものの場合には厚労省への報告も必要となる．最近話題となったがんペプチドワクチンの臨床研究では，この部分が問題とされている．

VII　研究計画の事前登録

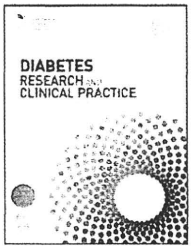
臨床研究に関する倫理指針には「侵襲性を有する介入研究では，予め公開データベースに臨床研究計画を登録しなければならない」という項もある．出版バイアス（ネガティブデータは論文として公表されにくい傾向がある）を防ぐための措置であり，すでに欧米の一流学術雑誌では研究計画の事前登録が求められている．日本では3つの登録サイトがあり，UMINの臨床試験登録システム（http://www.umin.ac.jp/ctr/index-j.htm），日本医薬情報センターの臨床試験情報（http://www.clinicaltrials.jp/user/cte_main.jsp），日本医師会臨床試験登録システム（https：//dbcentre3.jmacct.med.or.jp/jmactr/）のいずれかを用いることになる．

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.elsevier.com/locate/diabres

International Diabetes Federation



Cost-effectiveness of administering oral adsorbent AST-120 to patients with diabetes and advance-stage chronic kidney disease

Yasuaki Hayashino[a,*], *Shunichi Fukuhara*[a], *Tadao Akizawa*[b], *Yasushi Asano*[c], *Takafumi Wakita*[d,e], *Yoshihiro Onishi*[d], *Kiyoshi Kurokawa*[f]

for the CAP-KD study group

[a] *Department of Epidemiology and Healthcare Research, Kyoto University Graduate School of Medicine, Konoe-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan*

[b] *Department of Nephrology, Showa University School of Medicine, Tokyo, Japan*

[c] *Department of Internal Medicine, Koga Red-Cross Hospital, Ibaraki, Japan*

[d] *Institute for Health Outcomes and Process Evaluation Research, Kyoto, Japan*

[e] *Department of Psychology, Faculty of Sociology, Kansai University, Osaka, Japan*

[f] *National Graduate Institute for Policy Studies, Tokyo, Japan*

ARTICLE INFO

Article history:
Received 12 March 2010
Received in revised form
12 July 2010
Accepted 15 July 2010
Published on line 13 August 2010

Keywords:
Diabetes
Chronic kidney disease
Oral absorbent
Cost-effectiveness

ABSTRACT

Aims: AST-120, an oral adsorbent currently on-label only in Asian countries with phase III trials ongoing in the US, slows renal disease progression in patients with diabetes and advanced-stage chronic kidney disease (CKD). The objective of this study is to evaluate the cost-effectiveness of using AST-120 to treat patients with type 2 diabetes and advanced-stage CKD.
Methods: We used Markov model simulating the progression of diabetic nephropathy. Data were obtained from randomized trials estimating the progression of diabetic nephropathy with and without AST-120, and published literature. The base population was patients 60 years of age with type 2 diabetes and Stages 3 and 4 CKD.
Results: Treating patients with diabetes and advanced-stage CKD was found to be a dominant strategy, and quality of life improved further and more money was saved (0.22 quality-adjusted life years [QALYs] and $15,019 per patient) using AST-120 than the control strategy. Sensitivity analysis results were robust with regard to cost, adherence, and quality of life associated with AST-120 therapy, as well as age at diagnosis. The model was relatively sensitive to the effectiveness of AST-120.
Conclusions: Treating patients with type 2 diabetes and advanced-stage CKD with AST-120 appears to extend life and reduce costs.

1. Introduction

Although tight control of blood sugar and lowering blood pressure using angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) may slow the development and progression of diabetic complications [1–4], the prevalence of diabetes and its complications nevertheless continues to increase substantially in the US and Japan [5–9]. Diabetes mellitus is the most common cause of end-stage renal disease (ESRD), and two-thirds of ESRD cases occur in

* *Corresponding author.* Tel.: +81 75 753 9467; fax: +81 75 753 4644.
E-mail address: hayasino-y@umin.net (Y. Hayashino).

0168-8227/$ – see front matter
doi:10.1016/j.diabres.2010.07.007

patients with type 2 diabetes [10,11]. Given the increasing prevalence of renal disease in these patients, a cost-effective yet still appropriate treatment method is necessary.

AST-120 (Kremezin; Kureha Corporation, Tokyo, Japan) is an oral adsorbent effective in removing uremic toxins such as indoxyl sulfate from the gastrointestinal tract. Several animal studies have shown AST-120 effective in ameliorating the development of renal fibrosis by reducing the gene expression of transforming growth factor β1 [12,13]. At present, this compound is on-label and widely used in Asian countries to treat chronic kidney disease (CKD), and phase III trials are ongoing in the U.S. [14]. A recently published randomized controlled trial evaluating the effectiveness of AST-120 in patients with CKD revealed that this agent was indeed effective in slowing creatinine clearance (CCr) deterioration [15].

Before implementation, steps to avoid development of ESRD in a few patients must be weighed against the cost of treating many patients with AST-120. Unfortunately, given the large number of patients and long follow-up required to accurately assess these items, this issue is unlikely to be settled in clinical trials. We therefore used a decision model to simulate costs and clinical outcomes associated with either AST-120 or a control strategy for preserving renal function in patients with type 2 diabetes and advanced-stage CKD.

2. Materials and methods

2.1. *Decision analytic model*

We developed a Markov model simulating the natural history of renal complications in diabetes and risk reduction due to AST-120 (Fig. 1). Model outcomes included renal disease progression, quality-adjusted life expectancy, lifetime costs, and incremental cost-effectiveness ratios. Our model builds on previous models of ACE inhibition for diabetic nephropathy [16], with a cohort of individuals 60 years of age with diabetes and CKD entering the model and transitioning through renal disease states with rate of disease progression modified by the use of AST-120. In our model, the time horizon of the analysis is divided into 1-year cycles, and the cohort is followed over its lifetime.

2.2. *Initial population distribution*

With regard to population distribution, because the clinical trial from which we obtained the data for the current analysis

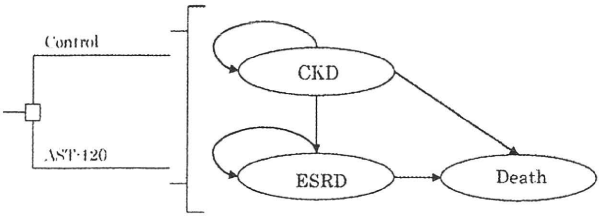


Fig. 1 – Health status and clinical strategies in the Markov model. CKD, chronic kidney disease and ESRD, end-stage renal disease.

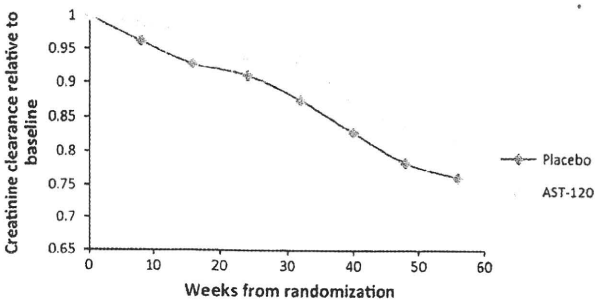


Fig. 2 – Creatinine clearance relative to baseline during actual follow-up of a randomized controlled trial of AST-120 in a subgroup of patients with diabetes. The black diamond indicates changes in creatinine clearance from baseline of placebo; gray square, AST-120.

examined patients with diabetes and Stage 3 or 4 CKD, we chose to use a cohort with the same makeup.

2.3. *Likelihood of end-stage renal disease*

To estimate the risk for developing and progressing through the stages of nephropathy (CKD and ESRD) and risk reductions from AST-120 treatment, we assessed data obtained from a clinical trial evaluating the effectiveness of AST-120 in diabetes patients with CKD over 56 weeks. Subject and method details have been previously reported elsewhere [15]. Briefly, the present study was a randomized controlled trial involving 460 patients at 75 medical facilities with CKD and who were not undergoing dialysis. Patients were randomly assigned to receive a low-protein diet with either antihypertension medication alone (either or both an ACE inhibitor or ARB) or combined with oral AST-120 (6 g/day).

We obtained data from a subgroup of 95 patients with diabetes and Stage 3 or 4 CKD who participated in a randomized controlled trial of AST-120 (mean age, 66.0 years: mean HbA1c, 6.2%: mean CCr, 27.1). In this particular subgroup, CCr declined by an average of 18.4% per year in the AST-120 group and 25.3% per year in the control group among patients with diabetic nephropathy (Fig. 2). The magnitude of CCr decline varied depending on baseline CKD stage; CCr declined by an average of 7.0% per year in the AST-120 group and 16.3% per year in the control group among patients whose baseline CKD stage was 3 and by an average of 21.8% per year in the AST-120 group and 26.7% per year in the control group among patients whose baseline CKD stage was 4. Based on these results, we estimated the CCr after 1 year by extrapolating the CCr yearly decline rate to the baseline CCr of participants of this trial; we then developed the following formula using the declining exponential approximation to life expectancy method [17–19]:

$$CCry_i = CCrb_i \times \exp(-YDR_i \times y)$$

where $CCry_i$ indicates the CCr for Participant i at y years from baseline, $CCrb_i$ indicates the CCr for Participant i at baseline, and YDR_j indicates the yearly decline rate of CCr in Group j (AST-120 or placebo).

YDR_j values varied depending on the CKD stage (3 or 4) to which patients belonged at y years. Assuming that patients with an estimated CCr < 10 suffer from ESRD and thus receive hemodialysis in accordance with renal disease guidelines [20], we then calculated the yearly rate of progression to ESRD. Estimated cumulative progression rates to ESRD in AST-120 and placebo groups are shown in Fig. 3.

2.4. Mortality

We based the age-dependent probability of death from non-ESRD-related or ESRD-related complications on Japan Life Table data from 2007 [21] multiplied by a standardized mortality ratio for diabetes in the elderly [5]. Standard mortality ratio for ESRD was calculated by dividing the mortality of Japanese ESRD patients (9.6%) registered in the Japanese Society for Dialysis Therapy database (average age: 67.2 years) [22] by mortality at 67 years of age from Japan Life Table data from 2007 [21].

2.5. Utilities

Health state utilities, measures of value for given health states, can be thought of as quality of life weights that are bounded by 1 for perfect health and 0 for death. When utilities are multiplied by the length of time individuals spend in their respective health states, the resultant metric is a quality-adjusted life expectancy (measured in quality-adjusted life years [QALYs]), which reflects both the quantity and quality of remaining years of life [23]. We obtained utilities for the present study from previously published studies [24–26].

2.6. Treatment discontinuation

Because noncompliance with medications can be high, we allowed for discontinuation of AST-120 in our model. We assumed that most discontinuation of therapy would appear during the first three months of treatment, and in our base-case analysis, we allowed that 15% of patients starting AST-120 therapy would discontinue treatment every year based on results from clinical trials [15]. We also assumed that patients who discontinued AST-120 inhibitor therapy for any reason (side effects or noncompliance) would not restart treatment and that patients who discontinued therapy during the first year of treatment would not receive any benefit from the medication (that is, they experience the same rate of disease progression as untreated patients).

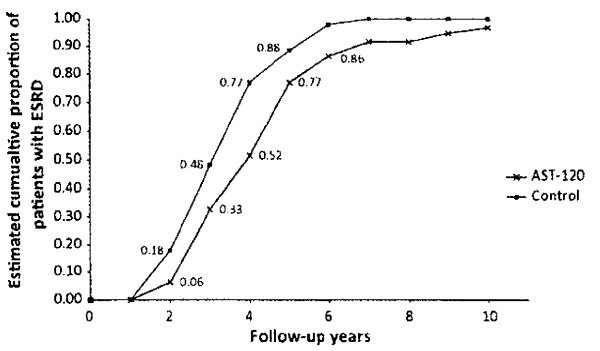


Fig. 3 – Cumulative proportion of patients with ESRD during follow-up in the AST-120 and control groups, estimated based on data from the diabetes subgroup of participants in a clinical trial. The black circle indicates control group; × indicates AST-120.

2.7. Costs

We conducted our analysis from a societal perspective. For our base-case analysis, we considered health care costs associated with either AST-120 therapy or treatment of ESRD. We obtained ongoing costs of care, including medical costs of future years of life added, for years in which no discrete event occurred in the published literature [27]. The cost of AST-120 was adjusted based on the consumer price index in the medical sector in Japan in 2008 and converted to U.S. dollars using the average exchange rate in 2008 (102.16 yen/dollar) [28]. The baseline cost estimates and the ranges tested in sensitivity analyses are shown in Table 1.

2.8. Base-case analyses

We performed base-case analyses from the societal perspective, and data are reported for a hypothetical 60-year-old beneficiary with diabetes and CKD. We discounted future costs and QALYs at an annual rate of 3% (48), and analyses were conducted using the TreeAge Pro Healthcare Module 2009 (TreeAge Software, Inc., Williamstown, MS, USA).

2.9. Sensitivity analyses

To assess the robustness of our findings, we performed extensive one-way sensitivity analyses (ranges in Table 1), derived by adding or subtracting 30% to or from the baseline estimate. In addition, we also performed a sensitivity analysis by varying the effectiveness of AST-120 (yearly decline rate of CCr of AST-120 divided by that of control) from 0.75 to 0.99 (0.73 for base-cases).

3. Results

3.1. Base-case analysis

In a scenario using AST-120, the total lifetime cost per 60-year-old patient with diabetes and Stage 3 or 4 CKD was $434,765. Adopting this strategy resulted in a quality-adjusted life expectancy of 7.31 QALYs and corresponding life expectancy (without quality adjustment) of 10.09 life years. In a scenario without using AST-120, lifetime costs increased to $449, 784 and quality-adjusted life expectancy decreased to 7.09 QALYs. Given these results, using AST-120 was found to be a dominant strategy, meaning that using AST-120 saved both lives and money (0.22 QALYs and $15,019 per patient).

In one-way sensitivity analyses, our results were robust to a wide range of plausible estimates of costs, utilities, and discount rate, and the AST-120 strategy remained dominant in all univariate sensitivity analyses. Our results were also sensitive to the effectiveness of AST-120. Results of sensitivity analysis by varying the effectiveness of AST-120 (yearly

Table 1 – Baseline values in the decision analysis model and their range in a sensitivity analysis.

	Baseline value	Sensitivity analysis
Annual decline rate ratio of Ccr in AST-120 group vs. control	0.73	0.70–0.95
Non-ESRD mortality	Age-based	
ESRD mortality	Age-based	
AST-120 discontinuation, %		
Each follow-up year	15	0–50
Utility		
ESRD	0.61	0.43–0.79
Diabetes mellitus	0.88	0.62–1 0
Annual cost, $		
AST-120 (6 g/day)	710	497–922
Outpatient treatment for CKD	299	210–389
Dialysis	12,692	8,885–16,500
Hospitalization for onset of dialysis	4,316	3,021–5,611
Surgery for vascular access	104	73–136
Hospitalization for CKD	1,991	1,394–2,589
Death	8,313	5,819–10,807

ESRD, end-stage renal disease; Ccr, creatinine clearance; CKD, chronic kidney disease.

decline rate of CCr of AST-120 divided by that of control) are shown in Fig. 4. The incremental cost-effectiveness ratio of AST-120 increased compared to the control as the annual decline ratio of Ccr increased. AST-120 remained dominant even at a rate ratio of 0.95 (0.04 QALYs and $999 per patient), but the incremental cost-effectiveness ratio increased to $185,438 per QALY if the rate ratio was 0.99.

4. Discussion

Diabetes is a major cause of ESRD and cardiovascular disease, and recent studies have shown that the oral adsorbent AST-120, which is currently on-label in Japan, Taiwan, and Korea is effective in reducing complications associated with these diseases. Our findings suggest that a strategy using AST-120 to treat patients with type 2 diabetes and Stage 3 or 4 CKD would not only slow the progression to ESRD but actually decrease total costs associated with patient care.

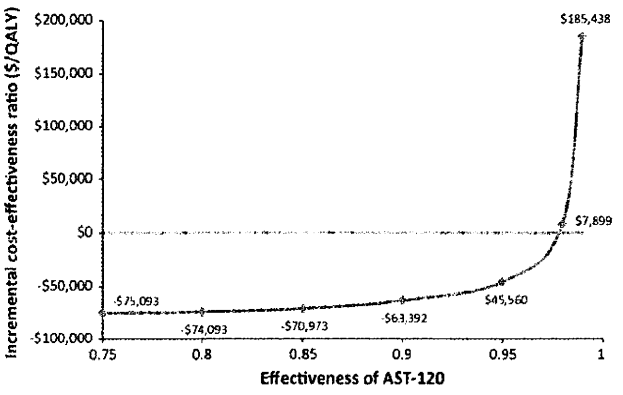


Fig. 4 – Incremental cost-effectiveness ratio of AST-120 compared with control group by varying the effectiveness of AST-120. Effectiveness of AST-120 indicates the ratio of yearly decline rate of CCr of AST-120 to that of control.

Cost savings remained even on sensitivity analysis, and our findings were robust to wide variation in model estimates. The AST-120 strategy remained dominant in all univariate sensitivity analyses, and our results were sensitive to the effectiveness of AST-120 on declining ratio of Ccr. Further, cost savings remained if the relative effectiveness of ATS-120 (yearly decline rate of CCr of AST-120 divided by that of control) was 0.95, but if the effectiveness rose to .99, using ATS-120 became cost-ineffective. Sensitivity analysis revealed that AST-120 is cost-effective if the effect of AST-120 is extremely small.

Cost-saving interventions are rarely implemented in medical practice [29], with examples including folic acid fortification of cereal grains [23] and pneumococcal vaccination in the aged 65+ population [30]. However, results from the present study demonstrate that considerable gains in overall health can be realized while simultaneously saving substantial resources. In this era of growing concerns over future medical expenses, rare opportunities to improve quality of life while also saving money should not be treated lightly.

Several notable strengths to our analysis merit mention. First, the point estimates for efficacy are based on the data of a randomized trial rather than on observational data. Second, a thorough review of the biomedical literature provided reliable ranges for our sensitivity analysis that are appropriate for our study population. Finally, Markov models allowed us to incorporate changes in disease progression following CKD in advanced stages, although additional controlled trials with longitudinal follow-up would likely provide useful prognostic information regarding important complications related to kidney disease.

The mechanisms by which AST-120 benefits renal diseases may be derived from findings in several animal studies. For example, diabetic rats fed AST-120 from an early stage of disease progression had higher creatinine clearance and serum albumin levels and lower urinary protein levels than diabetic control rats [31]. Further, glomerulosclerosis, mesangial matrix accumulation, tubular dilatation, and incidence of proteinaceous casts were also reduced. AST-120 has also been

shown to reduce the accumulation of indoxyl sulfate, a well-proven uremic toxin, in tubular cells, and to prevent decline in renal clearance of indoxyl sulfate, suggesting that AST-120 plays a role in preserving tubular function [32,33].

Several limitations to the present study also warrant mention, particularly with regard to our input data. First, only one study thus far has assessed the ability of AST-120 to slow the reduction in glomerular filtration rate in a randomized controlled trial. If the point estimate of benefit obtained from this study is wrong, the conclusions of our analysis could change drastically. Second, we extrapolated the data for 1 year of CCr reduction from a single trial to several years of follow-up, which may be an inaccurate assumption. Third, we must emphasize that the effect of AST-120 in slowing or preventing ESRD in patients with type 2 diabetes and advanced-stage CKD was inferred from the above results, as no single study has directly related the use of AST-120 in patients with type 2 diabetes to the development of ESRD.

End-stage renal disease is associated with an extremely high mortality rate, reduced quality of life, and high treatment cost, and thus any strategy that might slow or prevent the development of ESRD is attractive from the perspective of both the patient and society. In the present study, AST-120 therapy in patients with diabetes in advanced stages of CKD appears to extend life and reduce Medicare program costs. A reduction in costs from such an intervention may mean more money is available to spend on other health care needs of the elderly at a time when medical expense represent a national health policy concern.

Acknowledgement

We thank Dr. Kawaguchi for his invaluable support with this study.

Conflict of interest

The authors have a competing interest to declare. Dr. Hayashino is an endowed faculty of collaborative research program at Kyoto University funded by Kureha Corporation, Tokyo, Japan. The funding source had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

REFERENCES

[1] Gaster B, Hirsch IB. The effects of improved glycemic control on complications in type 2 diabetes. Arch Intern Med 1998;158(2):134–40.
[2] Parving HH. Clinical experience in the treatment of diabetic renal disease (type 1 and 2): summary and concluding remarks. Kidney Int 1994;45(Suppl.):165–6.
[3] Parving HH. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. N Engl J Med 1994;330(13):937–8.
[4] Brenner BM, Cooper ME, de Zeeuw D, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med 2001;345(12):861–9.
[5] Bertoni AG, Krop JS, Anderson GF, Brancati FL. Diabetes-related morbidity and mortality in a national sample of U.S. elders. Diabetes Care 2002;25(3):471–5.
[6] Mokdad AH, Bowman BA, Ford ES, Vinicor F, Marks JS, Koplan JP. The continuing epidemics of obesity and diabetes in the United States. JAMA 2001;286(10):1195–200.
[7] Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 2003;289(1):76–9.
[8] Mokdad AH, Ford ES, Bowman BA, et al. Diabetes trends in the U.S.: 1990–1998. Diabetes Care 2000;23(9):1278–83.
[9] Nakai S, Masakane I, Akiba T, et al. Overview of regular dialysis treatment in Japan as of 31 December 2006. Ther Apher Dial 2008;12(6):428–56.
[10] United States Renal Data System. Patient mortality and survival. Am J Kidney Dis 1998;32(2 (Suppl. 1)):S69–80.
[11] Ritz E, Stefanski A. Diabetic nephropathy in type II diabetes. Am J Kidney Dis 1996;27(2):167–94.
[12] Aoyama I, Shimokata K, Niwa T. Combination therapy with benazepril and oral adsorbent ameliorates progressive renal fibrosis in uremic rats. Nephron 2002;90(3):297–312.
[13] Miyazaki T, Aoyama I, Ise M, Seo H, Niwa T. An oral sorbent reduces overload of indoxyl sulphate and gene expression of TGF-beta1 in uraemic rat kidneys. Nephrol Dial Transplant 2000;15(11):1773–81.
[14] A study of AST-120 for evaluating prevention of progression in chronic kidney disease (EPPIC-1). ClinicalTrials.gov; 2008.
[15] Akizawa T, Asano Y, Morita S, et al. Effect of a carbonaceous oral adsorbent on the progression of CKD: a multicenter, randomized, controlled trial. Am J Kidney Dis 2009;54(3):459–67.
[16] Golan L, Birkmeyer JD, Welch HG. The cost-effectiveness of treating all patients with type 2 diabetes with angiotensin-converting enzyme inhibitors. Ann Intern Med 1999;131(9):660–7.
[17] Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the "DEALE"). I. Validation of the method. Am J Med 1982;73(6):883–8.
[18] Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"). II. Use in medical decision-making. Am J Med 1982;73(6):889–97.
[19] Kunz K, Weinstein M. Modelling in economic evaluation. In: Drummond M, McGuire M, editors. Economic evaluation in health care: merging theory with practice. New York, NY: Oxford University Press; 2001.
[20] Clinical practice guidelines for hemodialysis adequacy, update 2006. Am J Kidney Dis 2006;48 (Suppl. 1):S2–90.
[21] Arias E. United States life tables, 2004. Natl Vital Stat Rep 2007;56(9):1–39.
[22] The Japanese Society for Dialysis Therapy. The crude 1-yr mortality rates among hemodialysis patients; 2008, Available at http://docs.jsdt.or.jp/overview/pdf2009/p021.pdf (accessed 14.06.10).
[23] Gold M, Siegel J, Russel L, Weinstein MC. Cost-effectiveness in health and medicine. Philadelphia: Saunders; 1996.
[24] The Diabetes Control and Complications Trial Research Group. Lifetime benefits and costs of intensive therapy as practiced in the diabetes control and complications trial. JAMA 1996;276(17):1409–15.
[25] Lawrence WF, Grist TM, Brazy PC, Fryback DG. Magnetic resonance angiography in progressive renal failure: a technology assessment. Am J Kidney Dis 1995;25(5):701–9.

[26] Brown GC, Brown MM, Sharma S, Brown H, Gozum M, Denton P. Quality of life associated with diabetes mellitus in an adult population. J Diabetes Complications 2000;14(1):18–24.
[27] Takahashi T, Reed SD, Schulman KA. Cost-effectiveness of the oral adsorbent AST-120 versus placebo for chronic kidney disease. Nephrology (Carlton) 2008;13(5):419–27.
[28] Japanese Ministry of Health Labour and Welfare. Standard prices for medicines in 2008. Tokyo: Shakai-hoken Kenkyujyo; 2006.
[29] Rosen AB, Hamel MB, Weinstein MC, Cutler DM, Fendrick AM, Vijan S. Cost-effectiveness of full medicare coverage of angiotensin-converting enzyme inhibitors for beneficiaries with diabetes. Ann Intern Med 2005;143(2):89–99.
[30] Ellis JJ, Erickson SR, Stevenson JG, Bernstein SJ, Stiles RA, Fendrick AM. Suboptimal statin adherence and discontinuation in primary and secondary prevention populations. J Gen Intern Med 2004;19(6):638–45.
[31] Okada K, Shibahara H, Takahashi S. Effect of oral adsorbent on the progression of diabetic nephropathy. Nephron 1997;76(4):489–90.
[32] Aoyama I, Shimokata K, Niwa T. Oral adsorbent AST-120 ameliorates interstitial fibrosis and transforming growth factor-beta(1) expression in spontaneously diabetic (OLETF) rats. Am J Nephrol 2000;20(3):232–41.
[33] I. NKF-K/DOQI clinical practice guidelines for hemodialysis adequacy: update 2000. Am J Kidney Dis 2001;37(1 (Suppl. 1)):S7–64.

ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 301–308

Journal of Clinical Epidemiology

Validation testing of a three-component model of Short Form-36 scores

Yoshimi Suzukamo[a,b,*], Shunichi Fukuhara[a,c], Joseph Green[a,d], Mark Kosinski[e], Barbara Gandek[e], John E. Ware[f]

[a]*Institute for Health Outcomes and Process Evaluation Research, Kyoto and Tokyo, Japan*
[b]*Department of Physical Medicine and Rehabilitation, Tohoku University, Sendai, Japan*
[c]*Department of Epidemiology and Healthcare Research, Kyoto University, Kyoto, Japan*
[d]*Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*
[e]*QualityMetric Inc., Lincoln, RI, USA*
[f]*University of Massachusetts-Boston, Boston, MA, USA*

Accepted 13 April 2010

Abstract

Objective: The two-component factor structure underlying Short Form-36 (SF-36) summary scores may not be valid worldwide. We studied a three-component model of SF-36 scores in Japan.

Study Design and Setting: The SF-36 scores came from representative samples of the population of Japan. Factor analysis and structural equation modeling were used. The two-component model gave physical component summary (PCS) scores and mental component summary (MCS) scores. The three-component model gave scores on the PCS, the MCS, and also on the third component, which we call the role component summary (RCS) score. These were evaluated with external criteria.

Results: In the three-component model, the RCS was strongly associated with the role-physical, social functioning, and role-emotional subscales, whereas the PCS and MCS were associated with the physical functioning and mental health subscales, as expected. The goodness-of-fit index was 0.945 for the three-component model and 0.935 for the two-component model. The PCS discriminated between groups stratified by comorbid conditions, and the MCS discriminated between groups stratified by psychological depression. Absence from work was associated with both PCS and RCS.

Conclusion: The three-component model is better than the two-component model, and it provides more useful PCS and MCS scores. Criteria for validation testing of the RCS are needed.

Keywords: Quality of life; Structure; SF-36; Summary score; Role/social; Validity

1. Introduction

Measures of health-related quality of life are often based on explicit conceptual models. The model associated with the original, US-English version of the Short Form-36 (SF-36) has eight subscales [1,2]. The four subscales measuring physical functioning (PF), limitations on role functioning because of physical health (RP), bodily pain (BP), and general health (GH) are often given the greatest weights to form a "physical component." The other four, which measure mental health (MH), limitations on role functioning because of emotional problems (RE), social functioning (SF), and vitality (VT) are the greatest contributors to a "mental component" [3]. "The physical and mental components are conceived as being parts of a higher-order concept: health-related quality of life" [1,4]. Through the International Quality of Life Assessment project [5], translated and adapted versions of the SF-36 were made available for use in many countries and language-based cohorts. The SF-36 has been used worldwide, and results from studies done in East Asia have now brought its original conceptual model into question.

Factor structures differing from that found in Western countries have been found in Japan [6], China [7,8], Singapore [9], and Taiwan [10], although not in Hong Kong [11]. In Japan, Fukuhara et al. [6] found that the factor structure of the SF-36 differed from that in other countries in three ways: (1) Scores on the role-emotional (RE) subscale loaded strongly on the "mental" component in the United States and in Western European countries, but they loaded strongly on the "physical" component in Japan. (2) Scores on the VT subscale loaded strongly on the "physical" component in the other countries, but they loaded strongly on the

* Corresponding author. Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, 2-1 Seiryo-cho, Aoba-ku, Sendai 980-8575, Japan. Tel.: +81-22-717-7338; fax: +81-22-717-7340.

E-mail address: suzukamo@med.tohoku.ac.jp (Y. Suzukamo).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.04.017

What is new?

For SF-36 scores in Japan, using a model that includes three components (role-social, physical, and mental) solves problems that arise when a simpler model is used.

"mental" component in Japan. (3) Scores on the BP subscale loaded strongly on the "physical" component in the other countries, but they loaded strongly on both the "mental" and the "physical" components in Japan.

The original SF-36 measurement model has also been reconsidered in the light of findings from Western countries. Using structural equation modeling and data from the United States and Europe, Keller et al. [4] found support for explanations invoking a "general well-being" factor at the same level as the "physical" and "mental" factors, and also a single higher-level factor (interpreted as "health"). Five years later, Ware [12] proposed a model including a "participation (role, social)" factor in addition to the "physical" and "mental" factors.

In addition to the eight subscale scores, physical component summary (PCS) scores and mental component summary (MCS) scores are often computed, at least in Western countries. As convenient as these summary scores may be, their validity depends on the appropriateness of the two-factor model. The differences in factor structure between Western and Asian countries, and appreciation of the value of role and social participation as an independent domain, have led to proposals for three-component summaries of SF-36 scores. For example, using structural equation modeling, Huang et al. [13] found a good fit between SF-36 data from the general population of Taiwan and a model with three second-order components (which they interpreted as physical, mental, and social) and one third-order component (which they interpreted as health).

To develop a more appropriate way to summarize SF-36 scores, we compared a model with two components (mental and physical) to a model with three components (mental, physical, and role-social).

2. Methods

2.1. Sample

To study the factor structure of the SF-36 scores and to determine how to compute the three summary scores, we used data from the 2002 national norm survey done in Japan (sample 1, $n = 2,966$) [14]. For validation tests of the three-component model, we used data from the 1995 national norm survey [6] (sample 2, $n = 3,395$; some data used for validation tests were missing in 308 cases; hence, the total number used was 3,087). All residents of Japan who were 20 through 80 years old were potential subjects of these surveys, and participants were selected by two-stage stratified random sampling. All participants completed the SF-36 self-report form.

2.2. Data collection

The 2002 national norm survey was done with the Japanese SF-36 version 2, and the 1995 survey was done with version 1.2. For six of eight subscales, the difference in scores between versions was less than 1.0 point [14]. For the scales measuring limitations on role functioning because of physical health and emotional problems, the differences were somewhat greater, as in the United States [15]. The SF-36 comprises one question item about recent changes in health (the data from which were not used in this study) and 35 question items that are scored in eight subscales: PF, RP, BP, GH, VT, SF, RE, and MH.

Together with the SF-36, the respondents were given a list of 19 medical conditions and, for each condition, were asked to indicate whether they had the condition or not. The 19 medical conditions were hypertension, diabetes, stroke, myocardial infarction, angina, congestive heart failure, eye disease, respiratory disease, gastrointestinal disease, blood disease, kidney disease, urological disease, bone or muscle disease, skin disease, neurological disease, depression, hormonal disease, gynecological disease, and any other chronic condition.

Respondents were also asked to complete the Zung Self-Rating Depression Scale (ZSDS) [16]. Reliability and validity studies of the Japanese version of the ZSDS have been reported [17], and scores on the ZSDS (lowest possible score, 20; highest possible score, 80) were taken as indicators of the severity of depression, as follows: greater than 55, severe; 48 through 55, moderate; and 40 through 47, mild.

In addition, the respondents were asked to indicate the number of days they had missed work, school, or housework for health-related reasons during the previous year.

2.3. Analysis

The original measurement model is a two-component subscale-level model [1–3]. To test the hypothesis of a three-component structure, we first used exploratory factor analysis. With the limit set at three, factors were extracted using the principal components method, followed by varimax rotation, which resulted in the three-component subscale-level model. The results were examined to see whether they fulfilled the following criteria:

1. At least 70% of the variance should be explained by the three factors.
2. The subscale that correlated most strongly with the physical component should be the PF subscale.
3. The subscale that correlated most strongly with the mental component should be the MH subscale.
4. The subscales that correlated most strongly with the role-social component should be the RP, RE, and SF subscales.

5. The correlations between the mental component and PF scores should be weak ($r < 0.2$).
6. RP scores should correlate more strongly with the role-social component than with the physical component.
7. RE scores should correlate more strongly with the role-social component than with the mental component.
8. SF scores should correlate more strongly with the role-social component than with the mental component.

Next, we used confirmatory factor analysis to study four different models: the two-component and three-component subscale-level models, which were based on the eight subscales of the SF-36, and the two-component and three-component item-level models, which had eight first-order factors and were based on the 35 question items of the SF-36 (as noted above, the one question item asking about health transition was not used).

As indices of goodness of fit, we used the chi-squared test, the comparative fit index (CFI) [18], and the root mean squared error of approximation (RMSEA) [19]. The criteria for a good fit between data and model were a CFI of at least 0.9, and an RMSEA of less than 0.1. Confirmatory factor analysis was done with Amos 17.0 software (Amos Development Co., Crawfordville, FL, USA).

For validation testing of the subscale-level models, we assigned respondents to one of three groups and compared the three summary scores between groups. The three groups were defined by criteria external to the SF-36, as follows:

Group 1. Healthy: respondents who did not indicate that they had any of the 19 medical conditions, and whose ZSDS score was less than 48.

Group 2. Physical problem only: respondents who indicated that they had at least one of five medical conditions previously found to be strongly associated with SF-36 scores (bone disease, cerebrovascular disease, myocardial infarction, angina, or congestive heart disease) [1], and whose ZSDS score was less than 48.

Group 3. Mental problem only: respondents who either had a ZSDS score of at least 48 or indicated that they had a mental disorder. Those who indicated that they had any of the remaining 18 (physical) conditions in the checklist were not included.

Data from respondents who did not meet any of the above criteria were not used in these analyses. That would include, for example, respondents who indicated that they had a dermatologic condition only or an ophthalmologic condition only.

We used a general linear model that included adjustments for sex and age to estimate mean values for each summary score in each of the three groups, and then compared those summary scores between the pairs of groups. To compute the relative validity (RV) of each component, the F statistic of each summary scale was divided by the highest F statistic of all summary scales (thus, the resulting values were positive and had a maximum of 1.0). This ratio was referred to as the RV coefficient, as in previous studies [3,20].

Comparing groups 2 and 3 with the healthy group, we tested three hypotheses:

1. In respondents who report only physical problems (i.e., comparing group 2 with group 1), RV should be rank ordered from highest to lowest as follows: PCS score, role-social summary score, MCS score.
2. In respondents who report only mental problems (i.e., comparing group 3 with group 1), RV should be rank ordered from highest to lowest as follows: MCS score, role-social summary score, PCS score.
3. In respondents who report only mental problems (again, comparing group 3 with group 1), the RV value of the physical component in the three-component model should be lower than that of the physical component in the two-component model.

We also used three other external measures: the number of medical conditions, scores on the ZSDS, and the number of days of work missed per year for health-related reasons. The hypotheses based on those measures were as follows:

1. That the PCS scores would be inversely related to the number of medical conditions, excluding mental disorders (each respondent was assigned to one of three groups: 0, 1, and 2 or more conditions)
2. That the MCS scores would be inversely related to the number and frequency of symptoms of depression (each respondent was assigned to one of two groups defined by ZSDS score: less than 48, and 48 or greater).
3. That the role-social component scores would be inversely related to the number of days of work missed per year for health-related reasons (each respondent was assigned to one of two groups: 10 or fewer days, and more than 10 days).

To test those hypotheses, we examined means, F statistics, and RV values of the summary scores.

3. Results

3.1. Exploratory factor analysis

The three-component subscale-level model explained a total of 76.3% of the variance (Table 1). The PF subscale loaded most strongly on the physical factor and the MH subscale loaded most strongly on the mental factor. The RP, RE, and SF subscales loaded most strongly on the role-social factor. The loading of the MH subscale on the physical factor was very weak (0.023), as was the loading of the PF subscale on the mental factor (0.095).

On the other hand, the two-component subscale-level model explained a total of 60.0% of the variance. Unlike the hypothesized associations, the RE subscale loaded strongly on the physical factor, the VT subscale loaded strongly on the mental factor, and the BP subscale loaded about equally strongly on the mental and physical factors.

Table 1
Hypothesized associations and observed factor loadings for the two-component and three-component subscale-level models (2002 Japanese norm data, $n = 2,966$)

	2-Component subscale-level model					3-Component subscale-level model						
	Hypothesized associations		Factor loadings Varimax rotation			Hypothesized associations			Factor loadings Varimax rotation			
	Physical	Mental	Physical	Mental	Communality[a]	Physical	Mental	Role-social	Physical	Mental	Role-social	Communality[a]
PF	●	○	**0.610**	0.246	0.61	●	○	○	**0.832**	0.095	0.300	0.79
RP	●	○	**0.935**	0.219	0.92	◑	○	●	0.584	0.117	**0.701**	0.84
BP	●	○	0.436	0.466	0.41	◑	◑	○	0.566	0.480	0.213	0.60
GH	◑	◑	0.357	**0.604**	0.49	◑	◑	○	0.535	**0.680**	0.072	0.76
VT	◑	◑	0.258	**0.824**	0.75	◑	◑	○	0.204	**0.843**	0.240	0.81
SF	◑	●	0.430	0.502	0.44	○	◑	●	0.116	0.453	**0.671**	0.67
RE	○	●	**0.687**	0.399	0.63	○	◑	●	0.306	0.255	**0.811**	0.82
MH	○	●	0.215	**0.822**	0.72	○	●	○	0.023	**0.827**	0.374	0.83

Abbreviations: PF, physical functioning; RP, physical interference in role functioning; BP, bodily pain; GH, general health perceptions; VT, vitality; SF, social functioning; RE, emotional interference in role functioning; MH, mental health.

● Strong association (factor loading ≥ 0.6).

◑ Moderate association (0.3 < factor loading < 0.6).

○ Weak association (factor loading ≤ 0.3).

The bold values indicate factor loadings that are "strong," as defined in the table footnote (i.e., ≥0.6). The factors in the two-component model explained a total of 60.0% of the variance and those in the three-component model explained a total of 76.3% of the variance. The observed factor loadings for the three-component model agreed closely with the hypothesized associations, whereas those for the two-component model did not.

[a] Proportion of variance in each scale explained by the two components or three components.

3.2. Confirmatory factor analysis

The CFIs were greater than 0.9 for all four models, and they were higher for each of the three-component models than for the corresponding two-component models (i.e., higher for the three-component subscale-level model than for the two-component subscale-level model, and higher for the three-component item-level model than for the two-component item-level model). The subscale-level models did not meet the preestablished criterion for RMSEA, but the item-level models did (Table 2). For the item-level models, the completely standardized solutions are shown in Figs. 1 and 2.

3.3. Validation tests using external criteria

The group comprising respondents who reported no medical condition had the highest scores on all three summary measures. In the "physical condition only" group, the lowest score was on the physical component, and in the "mental condition only" group, the lowest score was on the mental component (Table 3).

The PCS of the "physical condition only" group was lower than that of the "healthy" group, and the F-value for that difference in scores was very high. For the comparison between the "healthy" and the "physical condition only" groups, both the mental component and the role-social component had very low values of RV (Table 3).

The MCS of the "mental condition only" group was lower than that of the "healthy" group, and the F-value for that difference in scores was very high. For the comparison between the "healthy" and the "mental condition only" groups, both the physical component and the role-social component had low values of RV (Table 3).

With both the two-component and the three-component models, the mean scores on the physical and mental components varied as expected between the groups known to vary with regard to physical and mental conditions (Table 3).

For the groups defined by the presence or absence of physical and mental conditions, the results were consistent with the three hypotheses. First, comparing group 2 with group 1, the rank order of RV values was physical, role-social, mental (1.00, 0.04, and 0.03, respectively). Second, comparing group 3 with group 1, the rank order of RV values was mental, role-social, physical (1.00, 0.21, and 0.01, respectively). Third, when the respondents with a mental condition only (group 3) were compared with the healthy respondents (group 1), the RV value of the physical component was much lower in the three-component model than in the two-component model (0.01 for group 3 vs. 0.20 for group 1, Table 3).

For the groups defined by the number of comorbid conditions, by ZSDS score, and by the number of work days missed for health reasons, the three summary scores are shown in Table 4.

PCS scores were lower in respondents who reported more comorbid conditions, and the RV values for both the mental component and the role-social component were low (Table 4).

Table 2
Goodness of fit indices of four models

Models tested	Factor structure	χ^2	Df	CFI	RMSEA
Subscale-level	2-component	803.9	14	0.935	0.138
	3-component	712.0	11	0.945	0.147
Item-level	2-component	6,815.8	542	0.904	0.063
	3-component	6,278.6	537	0.912	0.060

Abbreviations: Df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean squared error of approximation.

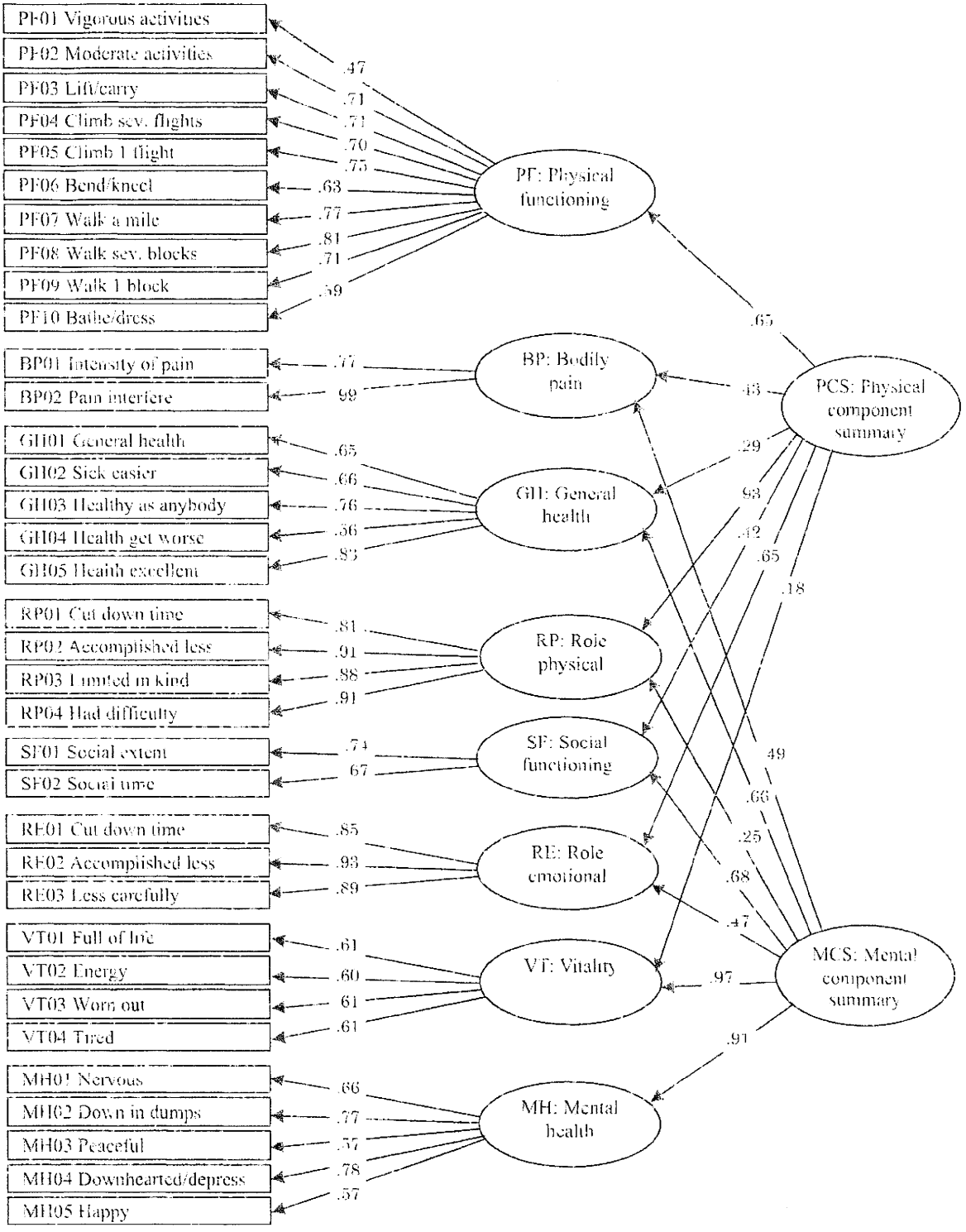


Fig. 1. Standardized solution for the two-component item-level model, based on confirmatory factor analysis.

MCS scores were lower in respondents who had high ZSDS scores, and the RV value for the physical component was low. The RV value for the role-social component was somewhat higher: 0.39 (Table 4). In the two-component model, not only MCS scores but also PCS scores were lower in respondents who had high ZSDS scores, and the RV value of physical component was relatively high (0.35).

For the groups defined by the number of days of work missed for health reasons, the highest F-value was associated with the difference in scores on the physical component. For the number of work days missed, the RV value of the mental component was low, whereas that of the role-social component was higher: 0.49 (Table 4).

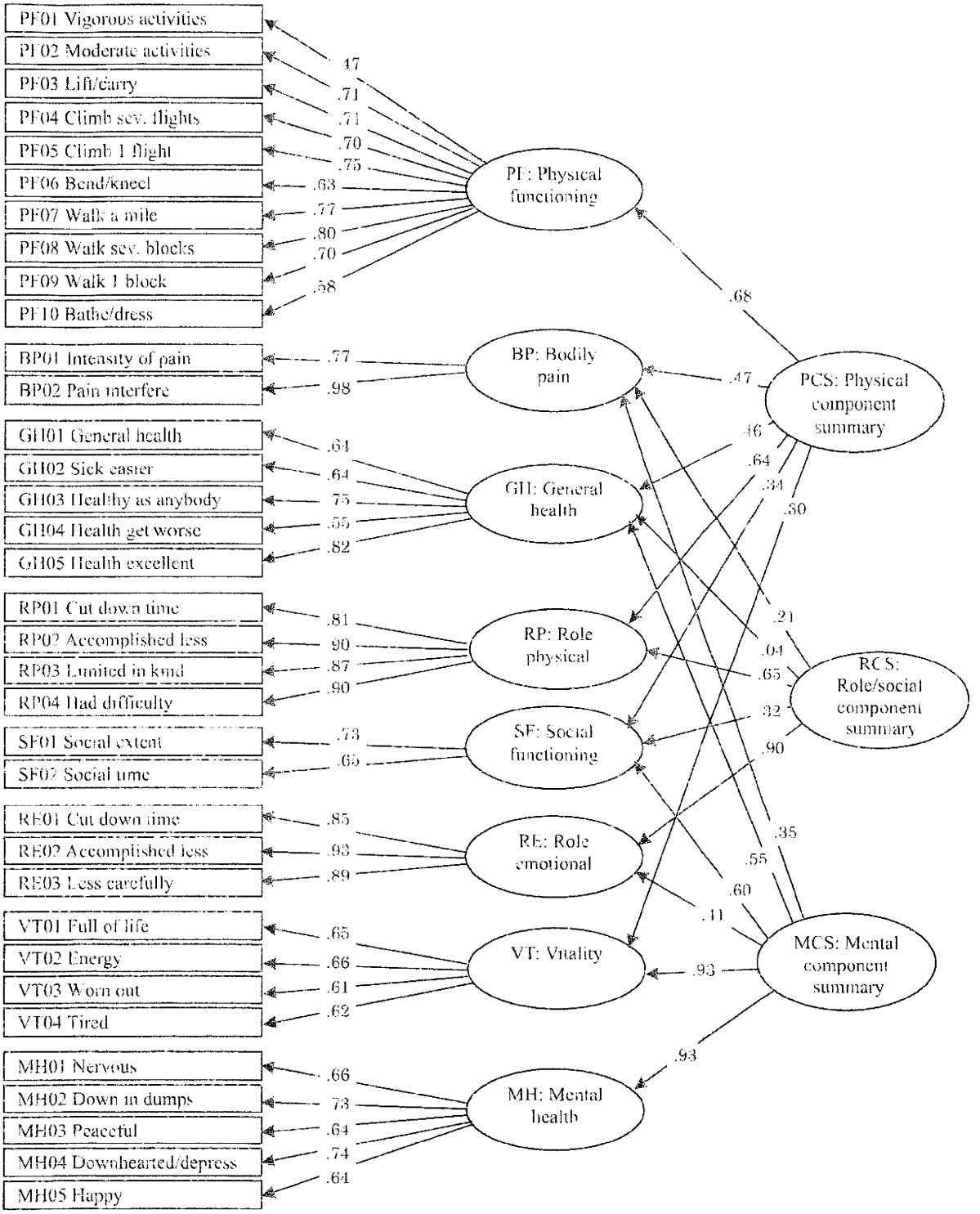


Fig. 2. Standardized solution for the three-component item-level model, based on confirmatory factor analysis.

4. Discussion

4.1. Factor structure

Overall, these results from Japan support the three-component model over the two-component model. The results of exploratory factor analysis showed that the three-component subscale-level model explained about 15% more of the total variance than did the two-component subscale-level model (76% vs. 60%). The three-component subscale-level model also met seven of the eight criteria that were set at the start of the study. The only exception was criterion 3. Specifically, the subscale correlated most

Table 3
Validation tests based on physical and mental health

	Group 1: Healthy (n=1,916)	Group 2: Physical condition only (n=128)	Group 3: Mental condition only (n=255)	Group 2 vs. group 1		Group 3 vs. group 1	
	Mean (SD)	Mean (SD)	Mean (SD)	F	RV	F	RV
3-component							
PCS_3	52.7 (7.3)	39.3 (10.8)	51.8 (11.1)	376.7	1.00	3.0	0.01
MCS_3	52.3 (8.7)	49.7 (9.2)	41.3 (10.7)	11.2	0.03	344.2	1.00
RCS_3	51.0 (7.9)	48.1 (13.9)	46.2 (12.1)	14.6	0.04	71.0	0.21
2-component							
PCS_2	52.9 (6.1)	40.9 (9.6)	49.0 (9.2)	461.3	1.00	79.2	0.20
MCS_2	52.0 (7.7)	50.1 (9.7)	41.0 (11.5)	7.0	0.02	402.6	1.00

Abbreviations: SD, standard deviation; RV, relative validity; PCS, physical component summary, MCS, mental component summary; RCS, role-social component summary.

F-ratios, RV coefficients, and mean (and SD) of scores computed using three-component and two-component models (1995 Japanese norm data).

The "healthy" group had the highest scores on all components. The lowest score in the "physical condition only" group was on the physical component, and the lowest score in the "mental condition only" group was on the mental component.

The F-value for the difference in physical component scores between the "physical condition only" group and the "healthy" group was very high, and both the mental component and the role-social component had very low values of RV. Similar results were seen for the difference between the "mental condition only" group and the "healthy" group.

strongly with the mental component was the VT subscale (not the MH subscale), and this was true for both two-component and three-component subscale-level models. As shown in Table 1, substituting the three-component model for the two-component model would give "purer" factor-loading patterns in the physical and mental components, with the mental component showing the greater improvement.

The results of confirmatory factor analysis showed that a three-component model fits the data better than did a two-component model. In particular, we note that only three subscales had strong loadings on the role-social factor: RP, RE, and SF.

The reasons for including a role-social factor are now both theoretical and empirical. Doing so helps to make the interpretation of SF-36 scores consistent with the concept that role and social participation are important to health-related quality of life [12], and it resolves the problems in factor structure previously found with the two-component model in Japan.

4.2. Criterion-related validation testing

The PCS score varied as expected among subgroups defined by the number of medical conditions, and the MCS score varied as expected among subgroups defined by scores on an index of psychological depression.

Because the survey was not originally designed to collect data for validation testing of a three-component model, our inferences regarding the role-social summary score are less direct. We expected that role-social summary scores would vary with the number of days of work, school, or housework missed for health-related reasons. That expectation was met to some degree by the high value of RV (0.49, Table 4). The strength of the association between the

Table 4
Results of validation tests based on the number of comorbid conditions, depression-scale scores, and number of days of work missed

	Number of comorbid conditions					ZSDS score				Number of days of work missed for health-related reasons per year			
	0	1	≥2	F	RV	<48	≥48	F	RV	≤10 days	≥11 days	F	RV
3-component													
PCS_3	52.9 (7.6)	45.9 (10.8)	40.5 (12.7)	336.3	1.00	50.8 (8.9)	46.8 (13.9)	62.9	0.12	50.6 (9.2)	42.7 (13.3)	106.0	1.00
MCS_3	51.4 (9.6)	48.3 (10.5)	45.4 (10.9)	59.1	0.18	51.6 (9.0)	40.5 (11.1)	526.8	1.00	47.2 (9.4)	44.7 (10.7)	12.9	0.12
RCS_3	51.4 (8.4)	49.2 (11.9)	47.8 (10.1)	11.0	0.03	51.0 (8.9)	44.0 (13.2)	204.2	0.39	49.8 (10.3)	43.8 (14.2)	52.3	0.49
2-component													
PCS_2	52.7 (6.4)	46.1 (9.6)	41.1 (11.2)	390.8	1.00	51.2 (7.6)	44.3 (11.9)	254.6	0.35	49.8 (7.9)	41.3 (11.9)	164.7	1.00
MCS_2	51.0 (8.7)	48.7 (10.8)	46.1 (11.6)	40.4	0.10	51.7 (8.2)	39.7 (11.5)	720.7	1.00	47.9 (9.8)	43.6 (11.9)	32.6	0.20

Abbreviations: RV, relative validity; ZSDS, Zung Self-Rating Depression scale; PCS, physical component summary; MCS, mental component summary; RCS, role-social component summary.

F-ratios, RV coefficients, and mean (and standard deviation) of scores computed using three-component and two-component models (1995 Japanese norm data).

The physical component scores were lower in respondents who reported more comorbid conditions, and mental component scores were lower in respondents who had high ZSDS scores. For the groups defined by the number of days of work missed for health reasons, the difference in role-social component scores was larger than the difference in mental component scores between the two groups; and the RV value of the mental component was lower than that of the role-social component.

number of days missed and the PCS score is also not surprising. Further validation studies should now be designed to find the most appropriate interpretation of role-social summary scores.

A three-component model of SF-36 scores might be useful in countries and regions other than Japan, but the analyses needed to understand how such models can best be constructed and applied remain to be done. In particular, we note that one open question is whether (and, if so, how) algorithms for computing the three summary scores should differ among populations.

4.3. Models for SF-12 and SF-8 scores

These findings may help resolve problems in using the SF-36v2. Further validation studies will be done on three-component summary scores estimated from the SF-12 and SF-8.

5. Conclusion

These findings support the use of a three-component model of SF-36 scores in Japan. They can be interpreted as successful validation testing of the PCS and MCS scores that are based on this three-component model. When the role-social concept is used as described here, the problems associated with using PCS and MCS scores are solved, and we expect that the PCS and MCS scores computed with the algorithms developed in this work will be useful in many contexts in Japan.

One important topic remaining to be studied in detail concerns the meaning of the RCS score. Our methods for validation testing of the RCS score were insufficient, so the next tasks are to clarify the concepts of role and social functioning, and to determine how to measure them. In addition, we hope that further work in other populations will eventually answer questions about whether these findings are generalizable.

References

[1] Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston, MA: New England Medical Center, The Health Institute; 1993.

[2] Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

[3] McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.

[4] Keller SD, Ware JE, Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. Use of structural equation modeling to test the construct validity of the SF-36 health survey in ten countries: results from the IQOLA project. J Clin Epidemiol 1998;51:1179–88.

[5] Ware JE, Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. Int J Technol Assess Health Care 1995,11:525–51.

[6] Fukuhara S, Ware JE, Kosinski M, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. J Clin Epidemiol 1998;51:1045–53.

[7] Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalization. J Epidemiol Community Health 2003;57:259–63.

[8] Yu J, Coons SJ, Draugalis JR, Ren XS, Hays RD. Equivalence of Chinese and US-English versions of the SF-36 health survey. Qual Life Res 2003;12:449–57.

[9] Thumboo J, Fong KY, Machin D, Chan SP, Leon KH, Feng PH, et al. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. Qual Life Res 2001;10:175–88.

[10] Fuh JL, Wang SJ, Lu SR, Juang KD, Lee SJ. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community. Qual Life Res 2000;9: 675–83.

[11] Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998;51:1139–47.

[12] Ware JE. Conceptualization and measurement of health-related quality of life: comments on an evolving field. Arch Phys Med Rehabil 2004;84:S43–51.

[13] Huang IC, Frangakis C, Wu A. Are there different components for the Taiwan SF-36 compared to western versions—physical, mental and social? Qual Life Res 2004;13:1495.

[14] Fukuhara S, Suzukamo Y. Manual of SF-36v2 Japanese version. Kyoto, Japan: Institute for Health Outcomes & Process Evaluation Research; 2004.

[15] Ware JE, Kosinski M, Dewey JE. How to score version 2 of the SF-36 health survey. Lincoln, RI: QualityMetric Incorporated; 2000.

[16] Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63–70.

[17] Fukuda K, Kobayashi S. A study on a self-rating depression scale. Seishin Shinkeigaku Zasshi 1973;75:673–9.

[18] Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1995;107:238–46.

[19] Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen K, Long JS, editors. Testing structural equation models. Newbury Park, CA: SAGE Publications; 1993. p. 147–62.

[20] Keller SD, Ware JE Jr, Hatoum HT, Kong SX. The SF-36 Arthritis-Specific Health Index (ASHI): II. Tests of validity in four clinical trials. Med Care 1999;37(Suppl):MS51–60.

Archives of Environmental & Occupational Health, Vol. 65, No. 4, 2010

The Association Between Socioeconomic Status and Prevalence of Diabetes Mellitus in Rural Japan

Yasuaki Hayashino, MD; Shin Yamazaki, PhD; Takeo Nakayama, MD; Shigeru Sokejima, MD; Shunich Fukuhara, MD

ABSTRACT. The objective of this study is to investigate the association between socioeconomic status and diabetes prevalence. A population-based cross-sectional survey was conducted in Japan. The association between household income tertile, duration of education (<12, 12, >12 years), or occupation (blue collar, white collar) and diabetes prevalence were assessed in 6,197 participants using multivariable-adjusted logistic regression models. Blue collar occupation and middle household income were found to be associated with high prevalence of diabetes mellitus.

KEYWORDS: diabetes, education, income, occupation, socioeconomic status

The "economic inequality theory" says that disparities in income among members of a community affect their health and, specifically, that economically egalitarian communities or societies have better health outcomes than more unequal communities.[1,2] Some proponents argue that inequality in incomes is a stronger determinant of health than the income of individual or families.

Type 2 diabetes mellitus imposes a major public health burden across populations. Over 150 million people suffer from diabetes worldwide,[3] including more than 18 million in the United States, and this disease is a major source of morbidity and mortality causing significant medical complications. The prevalence of type 2 diabetes is strongly influenced by socioeconomic status (SES). Persons of lower SES have limited income, poorer occupational opportunity, and reduced access to health care services and information, and these factors may contribute to both diabetes and complication risks. Persons with a lower SES consistently have a higher prevalence of diabetes and an excess burden of morbidity and mortality compared with persons with higher SES.[4–6] This association is evident because known diabetes risk factors such as obesity, large waist circumferences, and physical inactivity are patterned by SES.

In contrast with the clear evidence of the relationship between SES and diabetes mellitus in the Western countries, we have few data where we may observe this association in other regions of the world. The objective of this study is to investigate the association between SES and prevalence of diabetes mellitus in a population-based sample of rural Japanese population.

METHODS

Sample and data collection

We used the baseline survey from a population-based prospective study in the town of Naie in Japan for the current analysis. The town of Naie is an agricultural and rural area

Yasuaki Hayashino, **Shin Yamazaki**, *and* **Shunich Fukuhara** *are with the Department of Epidemiology and Healthcare Research at Kyoto University Graduate School of Medicine in Kyoto, Japan.* **Takeo Nakayama** *is with the Department of Health Informatics at Kyoto University School of Public Health in Kyoto Japan.* **Shigeru Sokejima** *is with the Section of Health Policy Development, Department of Public Health Policy, at the National Institute of Public Health in Wako, Japan.*

and is situated at the southern part of Hokkaido, which is the northernmost prefecture in Japan. The cohort consisted of all residents who were 20 years old or older and lived in Naie town on August 31, 2000. Fiscal year 2000 was at the bottom of over 10-year long economic depression in Japan, and economics in rural areas were more damaged than those in city area. Thus we expected to observe a wider disparity of SES if we choose this rural area of Japan. The baseline data were obtained in October–December 2000. The baseline survey included age, sex, height, weight, physical activity, smoking status, alcohol drinking, household income, the number of family members, duration of education, current occupation, health-related quality of life, and comorbidities. Comorbidities were obtained by self-report for hypertension, diabetes mellitus, coronary heart disease, chronic heart failure, cerebrovascular disease, and cancer. For physical activities, a metabolic equivalent task (MET) score was assigned based on the energy cost of each activity,[7] and then the equivalent energy expenditure in hours per week (MET-h/week) for each participant was estimated. The frequency of alcohol consumption during a typical week and the total alcohol intake on each occasion was determined and used to calculate the alcohol intake per day (g/day). Body mass index (BMI) was calculated as weight (kilograms) divided by height (meters) squared. Mental health was measured using a Japanese version of SF-36, which is the validated and widely used measure for quality of life.[8] Mental health subdomain score was estimated on a scale from 0 to 100 using SF-36 measures.

The trained research assistants distributed self-administered questionnaires and collected them by visiting each family during the study period. The Public Health Research Foundation organized this study and the Institutional Review Board of this foundation approved it. Informed consent was obtained from each participant.

Socioeconomic status measures

We collected data on income, education, and occupation as measures of socioeconomic status. We collected these data by using self-administered questionnaire. Income at the individual level was measured as yearly household income expressed in Japanese Yen, which included earning from work, transfer income, and other sources. Then income data were converted into US dollars at the rate of 0.93 dollar/100 yen, the average exchange rate in 2000. Self-reported current occupation was coded based on the Standard Industrial Classification for Japan, which is used to classify Japanese census data. We used following classifications: (1) professional and technical workers, (2) managers and officials; (3) clericals and related workers; (4) sales workers; (5) protective service and other service workers; (6) agricultural, forestry, and fishery workers; (7) workers in transport and communication; (8) craftsmen and manufacturing and construction workers; (9) housewife, (10) unclassified. Education was measured by the total years of education since the beginning of elementary school. We did not record health insurance status, because in Japan universal health insurance has been achieved.[9]

Statistical analysis

We used the baseline data of participants who answered questions on household income, years of education, or current occupation for the purpose of this analysis. We collapsed household income into 3 categories; low, middle, and high categories were created using tertiles of the income distribution. In the sensitivity analysis, we substituted income adjusted for family size; household income was divided by the square root of the number of families living with the subject. We sorted occupation classification into 3 categories: blue collar, white collar, housewife or unclassified. Because of the vague "unclassified" category, results were limited to white collar and blue collar categories. Total years of education were collapsed into 3 categories: less than 12 years, 12 years, more than 12 years. Using these categories, we investigated the association between income, total years of education, or occupational status and prevalent diabetes prevalence. For the analysis of occupation, we investigated the interaction between gender and occupational status in addition to the main analysis. Likelihood-ratio tests were used to test statistical interactions by comparing the -2 log(likelihood) between 2 nested models, one with only the main effects, and the other with main effects and interaction terms.

We used direct standardization to compare categorical variables adjusted for age (5-year increments), and the generalized linear model to compare age-adjusted continuous measurements. We used multivariable logistic regression model to analyze the association between incomes, total years of education, or occupational status and diabetes prevalence. We used age-adjusted model and 3 multivariable models. The first multivariable model (model 1) controlled for age (continuous), BMI (continuous), and gender. The second multivariable model (model 2) controlled for variables considered to be potential confounders of the association between the aimed categories and prevalent diabetes cases. This model controlled for age (continuous), BMI (continuous), gender, history of smoking (never, past, current), history of hypertension, history of high cholesterol, alcohol use (continuous), SF-36 mental health subdomain score (continuous), and physical activity (continuous). We calculated age- and multivariable-adjusted odds ratios (ORs) and the corresponding 95% confidence intervals (CIs). All analyses were performed using commercially available statistical software packages (Intercooled Stata 8.2; Stata, College Station, TX).

RESULTS

Baseline characteristics according to household income, education, and occupation were shown in Table 1. Of the 6,197 participants enrolled in this study, 2,771 reported household income. Median household income for