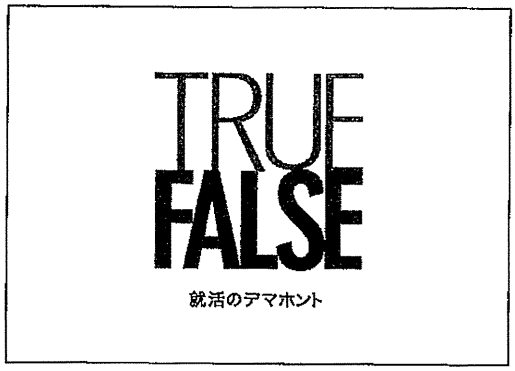

図３ クリエイティブ案『TRUE FALSE 就活のデマホント』

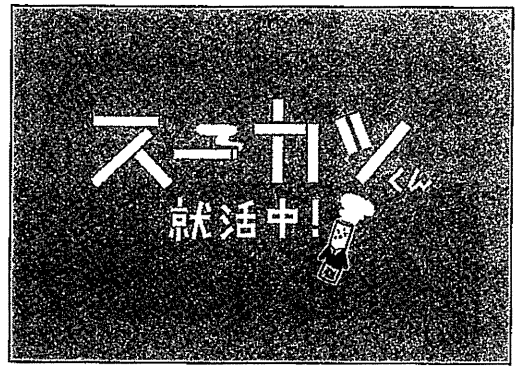

図４ クリエイティブ案『スーカツくん就活中！』

る．TRUE FALSEは複数のパターンを作成し，目を引くようTRUE FALSEのロゴを作成するとともに，全体に一貫して青（TRUE）と赤（FALSE）の象徴的な配色を用いている．『スーカツくん就活中！』は，スーカツくんという就職活動中の喫煙者である学生のキャラクターを作成し，主人公の就職活動の様子を実際の就職活動の時期にあわせて配信していくものである．主人公が喫煙者であるために，就職活動で困難に直面し，うまくいかないというストーリーになっている．

次に，介入対象となる大学生の喫煙者８名を対象に，2010年８月に個別面接によるクリエイティブの評価を行った．両案とも評価が高く，受け入れられていたが，『TRUE FALSE 就活のデマホント』の方がやや好まれる傾向にあった．そして，両案についてそれぞれ長所，改善点などが明らかになった．

調査の結果を受けて，内容の信頼性や目の引きやすさ，話題になりやすさ，キャンペーンの継続可能性，普及資材としての提供のしやすさ（利用者が利用目的・対象にあわせて内容の改変が行える）などから総合的に判断して，クリエイティブとして『TRUE FALSE 就活のデマホント』を採用した．

8）普及のための基本方針の策定

企業における商品をPRする際の基本である，「広告投下量を多くすると，認知度が上がる」×「認知度が上がると，買おうとする気持ちが強くなる」という関係に倣い，「“たばこを吸うと就職に不利”のコンセプトをメディアに載せて認知度を上げる」×「コンセプトを知ると，たばこを吸わない（禁煙・防煙）という気持ちが強くなる」という２つの柱を普及のための目標とし，その実現に向けた方策をとることとした．また，コンセプトを具体化していく際には，「たばこを吸うと就職できない」というようなfear appealではなく，「就職活動を機に，自分の人生や社会に出ることを考えるのと同様に，たばこを吸うことについても考えてみよう」というメッセージとした．

9）情報環境分析とメディアプランニング・PR（図２の調査⑦）

普及方法の検討のため，対象となる大学生の情報環境分析を行った．情報環境分析には，就職活動中の大学３年生と就職活動後の４年生の男性喫煙者各４人の計８人を対象に2010年８月に個別面接を実施した．

就職活動の際の情報源は，リクナビ，マスナビなどの就職情報サイト，みんなの就職活動日記などの掲示板，ブログ，ツイッター(twitter)，新聞などが多くあげられた．エントリーシートの受付などをウェブサイト上で行う企業が多いため，インターネットが就職活動の中心的なツールとなっていた．

続いて，情報環境分析の結果を検討し，インター

ネットを中心とするメディアプランニングを実施し，戦略的なPRを行った．禁煙・防煙プロジェクトの企画の段階からNHKとタイアップし，後述の喫煙と就職に関する調査結果やシンポジウムを番組で取り上げてもらうこととした．また後述のシンポジウムに関しては，共同通信PRワイヤーを利用しプレスリリースを行い，またインターネットを中心とするメディアや新聞記事，就職活動を行う大学生が多く登録する大手メーリングリスト，SNS，ブログ，ツイッターなどを使って告知を行った．

10)「喫煙と就職」のエビデンスの構築

非喫煙者であることを採用条件とする企業がいくつか存在していることは新聞などでも報道されていたが，非喫煙者を採用条件としている企業の分布や採用担当者の認識などは十分明らかになっていなかったため，普及に用いる「喫煙と就職」に関するエビデンスの構築を進めた．具体的には，喫煙と採用に関する3つの調査（インタビュー調査，郵送調査，インターネット調査）を実施した．

インターネット調査では，ネットリサーチ会社に登録されている20～69歳の男女のモニターから，職業として現在，企業の人事担当を行っている者を対象とし，調査は2010年6月に実施，838人から有効回答が得られた（有効回答割合37.3%）．結果として，新社会人や大学生が喫煙することに対して，回答者の半数以上が好感がもてないと感じていることが明らかとなった．また，所属企業がすでに現在，公式に喫煙の有無を採用基準としているのは回答者の3.7%，現在検討中が14.3%であったが，設定も検討もしていない687人についても，7.6%が今後採用基準としてもよいと思う，45.7%が採用基準ではないが考慮してもよいと思うと回答した．人事担当者個人の経験として，応募者の喫煙の有無が採用に影響を与えた可能性があると回答したのは30.0%，今後影響を与える可能性があると回答したのは48.7%であった[5]．

インタビュー調査，郵送調査，インターネット調査の結果を総合すると，非喫煙者を採用基準とする企業が少なからず存在すること，その傾向は今後強まっていく可能性が高いことなどが明らかになった．

11) 普及の実施（図2の調査⑧）

(1) 普及コンテンツの作成

クリエイティブ（普及資材）の開発・評価の結果から，最終的には「TRUE FALSE―就活と喫煙にまつわる不都合な真実―」を禁煙・防煙プロジェクトのテーマとした．このTRUE FALSEをもとに，研究班のウェブサイト（http://prev.ncc.go.jp/truefalse/index.html，URLは2012年1月12日現在）を立ち上げた．研究班のウェブサイトでは，TRUE FALSEだけでなく，喫煙と就職に関する調査結果の詳細やたばこによる健康被害，禁煙方法，禁煙に関するリンク集なども掲載している（図5）．

また，TRUE FALSEを中心に，音楽やロゴなどを組み合わせたムービーを作成した．ムービーは下記シンポジウムで公開したほか，研究班のウェブサイトやユーチューブ（You Tube）でも公開している（図6）．

(2) シンポジウムの開催

続いて，2011年1月に，本研究班の主催により，“大学生の就職活動と喫煙の関係を考えるシンポジウム「就活と喫煙にまつわる不都合な真実」”と題して，大学生を対象にシンポジウムを開催した．

シンポジウムは3部構成とし，第1部で喫煙と就職に関する調査結果の報告を行った．第2部では，就職と喫煙に関してさまざまな企業の人事担当責任者などによるパネルディスカッションを行った．パネルディスカッションには，本研究班の趣旨に賛同が得られた企業人事担当者6名が参加した（パネリスト：ファイザー株式会社，株式会社マッキャン・ワールドグループホールディングス，株式会社毎日新聞社，森永製菓株式会社，エスエムジー株式会社，株式会社電通パブリックリレーションズ）．第3部では，「大学生の就職活

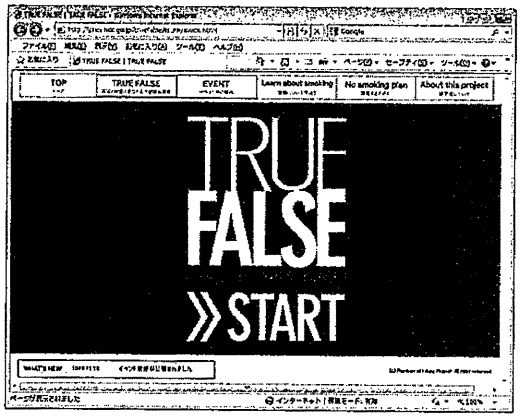

図5 研究班ウェブサイト

動のバイブル」といわれる漫画『銀のアンカー』や大学受験漫画として有名な『ドラゴン桜』の作者である三田紀房氏の協力を得て，「内定請負漫画『銀のアンカー』作者が指南する就活論～社会人になるということ～」と題した講演を行った．また，シンポジウムの終了後にはシンポジストや講演者，学生との懇親会を設けた．シンポジウムには約50人の大学生が参加した．

研究班で実施したシンポジウムの様子は，当日のNHK「ニュースウォッチ9」で大きく取り上げられ，7分20秒にわたって放送された．

12) 普及方法の評価

ソーシャルマーケティング手法の公衆衛生分野への応用の評価に関しては，現段階では確立した評価方法が存在せず，開発途上にある．そのため，今回の評価については，PR効果の評価やマーケティングにおける評価手法を参考とした．

普及にあたっては，前述した通り「広告投下量を多くすると，認知度が上がる」×「認知度が上がると，買おうとする気持ちが強くなる」という関係に倣い，「“たばこを吸うと就職に不利”のコンセプトをメディアに載せて認知度を上げる」×「コンセプトを知ると，たばこを吸わない（禁煙・防煙）という気持ちが強くなる」という2つの柱を普及のための目標として「禁煙・防煙」キャンペーンを実施した．そのため，評価についても，2つの柱のそれぞれについて，情報流通経路やメディア露出の調査，広告換算，シンポジウム参加者アンケートなどによって効果の測定を行った．

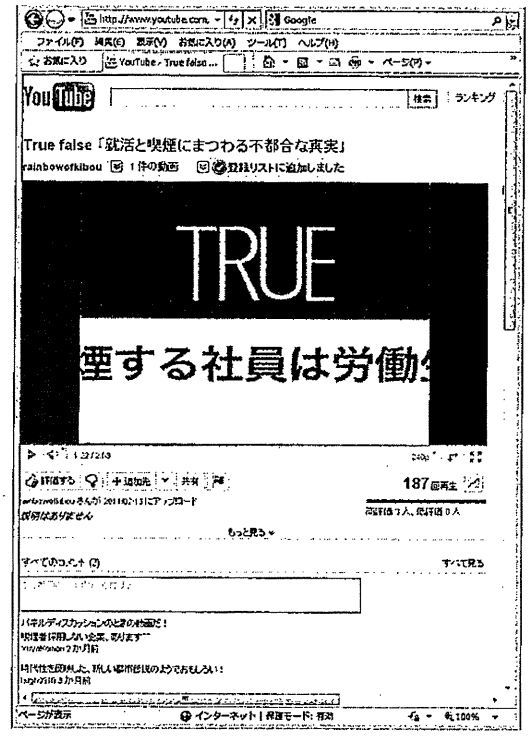

図6 ユーチューブでの公開ムービー

「“たばこと就活”のコンセプトをメディアに載せて認知度を上げる」に関しては，シンポジウムなど本研究の取り組みや調査結果がNHK報道番組で3回，日本経済新聞など新聞6紙，Yahoo！ニュースなど30以上のポータルサイト，4,000以上のブログなどで紹介された．

「コンセプトを知ると，たばこを吸わない（禁煙・防煙）という気持ちが強くなる」に関しては，シンポジウム参加者に対する会場アンケートによる評価を行った．シンポジウムの前後による比較では，「喫煙で就職が不利になる可能性」が「ある」と思っていたのはシンポジウム前では25.6％だったものが，シンポジウム後では82.1％に増加し，また，「今後，喫煙と就職の関係は強くなっていく」と思うと回答したのは76.9％，「喫煙で就職が不利になる可能性について周囲の人に教えてあげたい」と答えたのは84.6％と，コンセプトが信頼され，また口コミ効果も期待できることが明らかになった．研究班の

取り組みであるため，単純な評価はできないが，テレビ，新聞，ポータルサイトによる報道を広告換算してみると，少なくとも8,421万円以上の広告効果があり，テレビと新聞の報道は，推計で2,400万人以上の人の目に触れたことが期待される．

本研究における評価を通じ，より適切な評価方法を開発していく予定であるが，がん予防に関する新しい規範を形成し，メディア等を戦略的に活用することで，より広い普及と社会規範としての醸成を目指すという本研究の目標にとっても，「喫煙と就職」というコンセプトによる今回の普及は，十分な成果が得られたものと考えられる．

6. 今後に向けて

今回，ソーシャルマーケティング手法を活用し，ターゲットを決め，ターゲットのインサイトに狙いを絞って「禁煙・防煙」の普及を行った．普及方法は，認知度を上げる点，たばこを吸わないという気持ちを強める点で，効果があったことが期待され，「新しい社会規範」の醸成を前進させることができたと考えている．「禁煙・防煙」に関しては，今後もイベントの開催やウェブサイトの充実などを継続的に行い，より広い展開を行っていく予定である．

2011年より，「身体活動増加」と「野菜摂取量増加」を研究班の活動の中心とし，ソーシャルマーケティング手法を活用した普及方法の開発を進めているところである．

ソーシャルマーケティング手法は，欧米では国の施策として活用されはじめているが，日本ではまだ活用がはじまったばかりである．費用効果の検討やメディア等の積極的な活用などを取り入れた戦略的な普及の取り組みは，日本において，今後，行政などにおいてもさらに重要性が増していくと考えられる．われわれはソーシャルマーケティング手法のがん予防普及における導入を進めているが，今後，さまざまな分野において事例を蓄積し，理論や方法，経験を公開し，提供し合うことによって，日本における効率的な普及方法の理論と実践の拡大につながることが期待される．

［文　献］

1) 独立行政法人国立がん研究センターがん対策情報センター：がん情報サービス 統計（http://ganjoho.jp/professional/statistics/index.html, http://ganjoho.jp/public/statistics/index.html）．
2) Inoue M et al.: Attributable causes of cancer in Japan in 2005-systematic assessment to estimate current burden of cancer attributable to known preventable risk factors in Japan. Annals of Oncology, 2011 Nov 2.［Epub ahead of print］.
3) 独立行政法人国立がん研究センターがん対策情報センター：がん情報サービス 日本人のためのがん予防法（http://ganjoho.jp/public/pre_scr/prevention/evidence_based.html）．
4) Farrelly MC et al.: Evidence of a dose-response relationship between "truth" antismoking ads and youth smoking prevalence. Am J Public Health, 95: 425－431, 2005.
5) 溝田友里ほか：人事担当者における就職応募者の喫煙に関する認識．日本衛生学雑誌，67（1），2012（in press）．

JJCO Japanese Journal of Clinical Oncology

Clinical Trial Notes

Jpn J Clin Oncol 2011;41(2)275–277
doi:10.1093/jjco/hyq214
Advance Access Publication 2 December 2010

Randomized Controlled Trial on Effectiveness of Ultrasonography Screening for Breast Cancer in Women Aged 40–49 (J-START): Research Design

Noriaki Ohuchi[1,*], Takanori Ishida[1], Masaaki Kawai[1], Yoko Narikawa[1], Seiichiro Yamamoto[2] and Tomotaka Sobue[2]

[1]Department of Surgical Oncology, Graduate School of Medicine, Tohoku University, Miyagi and [2]Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan

*For reprints and all correspondence: Noriaki Ohuchi, Department of Surgical Oncology, Graduate School of Medicine, Tohoku University, 1-1, Seiryo-machi, Aoba-ku, Sendai-shi, Miyagi 980-8574, Japan. E-mail: noriaki-ohuchi@med.tohoku.ac.jp

Received October 22, 2010; accepted October 22, 2010

In cancer screening, it is essential to undertake effective screening with appropriate methodology, which should be supported by evidence of a reduced mortality rate. At present, mammography is the only method for breast cancer screening with such evidence. However, mammography does not achieve sufficient accuracy in breasts with high density at ages below 50. Although ultrasonography achieves better accuracy in Breast Cancer detection even in dense breasts, the effectiveness has not been verified. We have planned a randomized controlled trial to assess the effectiveness of ultrasonography in women aged 40–49, with a design to study 50 000 women with mammography and ultrasonography (intervention group), and 50 000 controls with mammography only (control group). The participants are scheduled to take second round screening with the same modality 2 years on. The primary endpoints are sensitivity and specificity, and the secondary endpoint is the rate of advanced breast cancers.

Key words: breast cancer screening – mammography – ultrasonography – randomized controlled trial

INTRODUCTION

Breast cancer is one of the most common cancers worldwide (1). The age-standardized incidence rate is the first among all female cancers, and it is continuously increasing in Japan (2,3), although Japan has a lower risk of breast cancer in comparison with Western countries. The incidence peaks at ages 45–49, and the mortality peaks at ages 55–59 in Japan (2). In breast cancer screening, it is essential to undertake effective screening with appropriate methodology. Effective screening should be supported by evidence of a reduced mortality rate. At present, mammography (MG) is the only method for breast cancer screening that has such evidence. However, MG does not achieve sufficient screening accuracy in breasts with high mammary gland density. Dense breasts are common at ages below 50 and are more common in Japanese populations than in Western populations (4). As the US Preventive Services Task Force (USPSTF) recommends against routine screen MG in women aged 40–49 years, the issue of breast imaging to screen women aged 40–49 still remains unclear (5).

Since ultrasonography (US) achieves better accuracy in breast cancer detection even in dense breasts (6) and supplemental screening US has the potential to depict early breast cancers not seen on MG (6–8), several single-institution observational studies in screening setting began. As mentioned in the WHO guidelines, 'population-based cancer screening' conducted as a public health program should be undertaken only when there is evidence of a

reduced mortality rate (9). Before introducing any new technology in population-based breast cancer screening, it is essential to evaluate the effectiveness. However, randomized controlled trials (RCTs), cohort studies or case–control studies have not been completed to assess the efficacy of screening US to reduce breast cancer mortality, and the effectiveness has not been verified.

Therefore, we have planned an RCT to assess effectiveness of screening US for breast cancer, the Japan Strategic Anti-cancer Randomized Trial (J-START) in 2006. The defined study population is women aged 40–49 years, because this is the age range at which breast cancer peaks in Japan (2) and because a high percentage of Japanese women aged 40s have dense breast. This is a large-scale controlled trial, designed to study 50 000 women with MG and US (intervention group) and 50 000 controls with MG only (control group).

The primary endpoints of this trial are the inter-group comparisons of the sensitivity and specificity, and the secondary endpoint is the inter-group comparison of the accumulated incidence rate of advanced breast cancer during the follow-up period. The most important index in the evaluation of the effectiveness of cancer screening is the mortality rate from the cancer in question in the target population. However, in view of the natural history of breast cancer, the 4-year period scheduled in the strategic study grant is too short to observe a significant inter-group difference. Although the rate of advanced breast cancer could be a surrogate for mortality reduction, it is necessary to have a system that has the long-term follow-up of the survival status of individuals even after the completion of the strategic study, J-START.

This study may have several limitations. First, the screening interval is 2 years, despite evidence that screening MG at age 40–49 years is more effective with annual screening. The recent USPSTF, however, recommends biannual MG screening in view of reducing 'harm', i.e. higher recall rate at age 40–49 years (5). Secondly, the study population, which is so different from that in Western countries, may limit the generalization of study outcomes. Most countries in Asia, however, demonstrate the similar trend of breast cancer incidence as observed in Japan; therefore, this trial may influence their health strategy against breast cancer. Nevertheless, for women aged 40–49 years even in Western countries, there is a limitation of MG screening as the USPSTF recommends against the routine use of screening MG for this age group. Thirdly, the study may be underpowered to provide follow-up data on breast cancer deaths because of the low breast cancer risk of native Japanese women. In this context, as much as 100 000 women are targeted in this trial to ensure the statistical power be sufficient enough in comparison between the two groups.

PROTOCOL DIGEST OF THE STUDY

Purpose

The aim of this study is to assess the effectiveness of screening US for breast cancer in women aged 40–49 (Fig. 1).

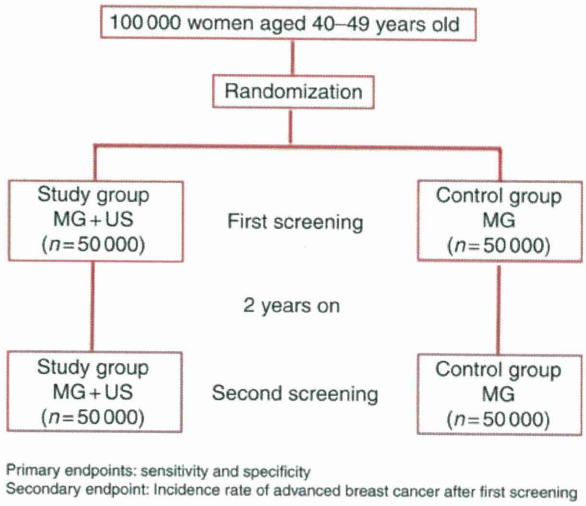

Figure 1. J-START study design. MG, mammography; US, ultrasonography.

Study Setting

This study is a multi-institutional prospective RCT, with 42 participating centers in 23 prefectures in Japan as of 31 March 2011.

Endpoints

The primary endpoints of this trial are sensitivity and specificity, based on the data of each incremental cancer detection rate, false-positives and false-negatives should be forthcoming in 2 years. The secondary endpoint is the rate of advanced breast cancers, as this has been demonstrated in the screening MG RCTs to be a surrogate for mortality reduction (10).

Eligibility Criteria

Inclusion criteria are as follows:

(i) women aged 40–49 years when registered;
(ii) women signed the informed consent to participate in the study.

Exclusion criteria are as follows:

(i) women with a history of breast cancer;
(ii) women with a history of malignant disease other than breast cancer within 5 years;
(iii) women in severe condition, who are not expected to live for 5 years.

Treatment Methods

Patient Assignments

Each participating center confirms the participants' eligibility and screening methods are assigned according to the random

number provided by the Japan Clinical Research Supporting Unit (J-CRSU) Data Center. Cluster randomization is also used in some institutions.

Screening Method and Assessment

For the intervention arm, US and MG are performed at the same time. For the control arm, MG is performed. The technologists and the physicians involved in this trial are asked to finish 2-day, 16-h education program for the standardization of US screening for breast cancer. Regarding the procedure in screening with US, the handheld US is performed by a technologist or by a physician, and later, the US image is interpreted by a physician. An interpretation of MG is performed by a physician who is not regulated to be the same doctor interpreting US image or not, although the categorization of the two modalities are defined separately in the protocol. The findings of MG and/or US are subsequently evaluated by authorized screeners and are classified into five categories as follows: Category 1, negative; Category 2, benign finding(s); Category 3, probably benign finding(s); Category 4, suspicious abnormality; and Category 5, malignancy. The women who are rated in Category 3 or higher by the MG and/or US are referred for further diagnostic examinations.

Statistical Analysis

The sample size was calculated on the hypothesis that adjunct US is expected to improve sensitivity of the intervention group compared with the control group. Our previous data demonstrated the lower sensitivity of MG screening, 71% in women aged 40–49, when compared with those in women aged 50–59 and 60–69, 85 and 86%, respectively (11). Assuming that the sensitivity increases from 71 to 86% by adding US to MG, 42 500 subjects for each arm is needed to make it 5% statistical significance (two-sided) with 80% power. Thus, the number of 100 000 subjects (two arms combined) is set to be a targeted sample size to verify the primary endpoint, a sensitivity improvement in the intervention group when compared with the control group.

Follow-up Period

The participants are invited to be screened 2 years after the first recruitment or asked to answer questionnaires of health status, history of receiving other screening program, incidence of breast cancer, and history of hospital consultation with any breast symptoms within 2 years. For evaluating the actual evidence of a reduced mortality rate of the intervention group compared with the control group, there must be needed to establish follow-up strategies for a long time period and systematic, nationwide population-based cancer registries.

Registration of the Protocol

The J-START was registered on the University Hospital Medical Information Network Clinical Trial Registration (UMIN-CTR), Japan (registration number: UMIN000000757), on 2007. Details are available at the following address: https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?function=brows&action=brows&type=summary&recptno=R000000910&language=E.

Funding

This work was supported by the 3rd term comprehensive control research for cancer (grant number: H-18-Senryaku-001), the Ministry of Health, Labour and Welfare of Japan.

Conflict of interest statement

None declared.

References

1. Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, et al. *Cancer Incidence in Five Continents, Vol. IX*. IARC Scientific Publications No. 160. Lyon: IARC 2007.
2. Matsuda T, Marugame T, Kamo KI, Katanoda K, Ajiki W, Sobue T. Cancer incidence and incidence rates in Japan in 2005: based on data from 12 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. *Jpn J Clin Oncol* 2010 doi:10.1093/jjco/hyq169 (Epub ahead of print).
3. Minami Y, Tsubono Y, Nishino Y, Ohuchi N, Shibuya D, Hisamichi S. The increase of female breast cancer incidence in Japan: emergence of birth cohort effect. *Int J Cancer* 2004;108:901–6.
4. Hasegawa S, Ohnuki K, Nagakubo J, Mizukubo J, Kitami J, Oyama K, et al. Evaluation of breast cancer visualization ability of mammography by age group and by breast constitution (mammary gland/fat ratio). *J Jpn Assoc Breast Cancer Screen* 2003;12:101–7 (in Japanese).
5. Screening for Breast Cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151:716–26, W-236.
6. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. *Am J Roentgenol* 2003;181:177–82.
7. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002;225:165–75.
8. Leconte I, Feger C, Galant C, Berliere M, Berg BV, D'Hoore W, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *Am J Roentgenol* 2003;180:1675–9.
9. WHO [Internet]. *National Cancer Control Programmes: Policies and Managerial Guidelines*. 2nd edn. 2002 [cited 21 October 2010]. http://www.who.int/cancer/media/en/408.pdf.
10. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42:793–806.
11. Suzuki A, Kuriyama S, Kawai M, Amari M, Takeda M, Ishida T, et al. Age-specific interval breast cancers in Japan: estimation of the proper sensitivity of screening using a population-based cancer registry. *Cancer Sci* 2008;99:2264–7.

Diabetol Int (2011) 2:47–54
DOI 10.1007/s13340-011-0025-z

ORIGINAL ARTICLE

Validity and applicability of a simple questionnaire for the estimation of total and domain-specific physical activity

Hitomi Fujii · Seiichiro Yamamoto · Fujimi Takeda-Imai · Manami Inoue · Shoichiro Tsugane · Takashi Kadowaki · Mitsuhiko Noda

Received: 9 November 2010 / Accepted: 15 March 2011 / Published online: 3 May 2011

Abstract
Purpose We developed and evaluated a simple, robust and valid self-administered questionnaire for the estimation of physical activity (PA). Here, we examined the validity of this questionnaire in subjects with differing sex, ages, occupations and living circumstances.
Methods The questionnaire consists of four domains, namely occupational activity, including housekeeping and commuting; leisure time activity; sleeping; and other activities. It was validated with 8-day, 24-h physical activity records (24 h-R) as the gold standard in 110 volunteers.
Results Total PA estimated by the questionnaire and the 24 h-R showed a moderately strong correlation ($r = 0.69$). Correlations between total PA by the PAQ and the 24 h-R for various subgroups, such as sex, age, area, occupation and BMI, were moderate to strong (0.55–0.80). Validity of domain-specific PA calculated by the questionnaire was also moderate to high.
Conclusion This simple questionnaire produces valid estimates of total and domain-specific PA and can be applied to a broad population.

Keywords Physical activity questionnaire · Physical activity · Epidemiology · Validity · 24-h Physical activity record

On behalf of the JPHC Study Group.

H. Fujii
Tama Center Clinic Mirai, Tokyo, Japan

S. Yamamoto (✉)
Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
e-mail: siyamamo@ncc.go.jp

F. Takeda-Imai · M. Inoue · S. Tsugane
Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan

F. Takeda-Imai
Department of Environmental and Occupational Health, Toho University School of Medicine, Tokyo, Japan

T. Kadowaki
Department of Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

M. Noda
Department of Diabetes and Metabolic Medicine, National Center for Global Health and Medicine, Tokyo, Japan

Introduction

Interest in the association between daily physical activity (PA) and the development of chronic diseases, such as diabetes [1–4], ischemic heart disease [5, 6] and cancers [7], has increased. This in turn has led to a need to quantify the effective threshold of activity [8], qualify varying physical activities and determine whether disease prevention is better discussed in terms of vigorous or moderate activity, or leisure time activity or occupational activity.

For practical reasons, large-scale epidemiological studies usually measure physical activity using a physical activity questionnaire (PAQ) not using more precise measurements, such as doubly labeled water or physical activity records. Moreover, because whole questionnaires used in epidemiological studies devote relatively little time and space to physical activity, the PAQ component must be short and accurate. This in turn hampers the development of PAQs that can estimate PA in different target

populations (e.g., by gender, age, occupation, geographical location, environment, etc.). Accordingly, many of the PAQs reported to date were developed for use in specific populations, such as nurses [5], physicians [9], office workers [10–12] and college alumni [7, 13], which tend to be homogenous in terms of both background and physical activity pattern. Given the difficulty in asking about overall activity, previous PAQs focused mostly on specific activities, such as leisure time, or moderate or vigorous activity, rather than total PA.

We therefore sought to develop a simple, robust and valid PAQ capable of measuring physical activity in a range of subjects, such as people living in both urban and rural areas, or subtropical to cold or snowy places. The aim of the PAQ is to be used in epidemiological studies for ranking individuals by PA adjusted with body size rather than estimating energy expenditure. Further, we sought to measure total PA as well as PAs for specific activities, such as at leisure and occupation, and for specific intensity, such as moderate or vigorous activities. In addition, because physical activity may differ among seasons, validity may be affected if the answer to the PAQ depends on when it is administered. We therefore aimed at measuring year-round average physical activity regardless of when the PAQ is administered.

Here, we developed and validated a PAQ in a Japanese cohort study against 24-h physical activity records (24 h-R).

Methods

The JPHC Study

The questionnaire was originally developed and used in the Japan Public Health Center-based prospective Study (JPHC Study) [14], a prospective follow-up study conducted widely throughout Japan that investigated mainly cancer and cardiovascular diseases. Cohort I was begun in 1990 and enrolled 61,595 subjects (29,980 men and 31,615 women) in four population cohorts and one health-checkup cohort, while Cohort II was begun in 1993 and enrolled an additional 78,825 residents (38,740 men and 40,085 women) in five population-based and one health-checkup cohort.

Subjects of the validation study

These subjects represented a subsample of participants in specific geographic areas covered by four public health centers (PHC) of the JPHC Study.

A total of 110 subjects (54 males and 56 females; 20 from the Katsushika PHC area of Tokyo, 38 from the Miyako PHC area of Okinawa Prefecture, 30 from the Saku PHC area of Nagano Prefecture and 22 from the Kashiwazaki PHC area of Niigata Prefecture) were recruited from both cohorts in an attempt to include urban areas as well as rural agricultural areas that cultivated rice, apples, sugarcane, etc. We also wanted to include subjects from both the western and eastern parts of Japan, and both cold and subtropical areas.

We selected married couples in their 50s or 60s for convenience and requested their voluntary participation. Approval for the study was obtained from the Institutional Review Board of the National Cancer Center, and written informed consent was obtained from the subjects.

The PAQ in the JPHC Study

The PAQ questionnaire, which was originally developed as part of the 10-year follow-up questionnaire survey used in the JPHC Study, was constructed based on the simple notion that total daily energy expenditure consists of four domains, namely occupational activity, including housekeeping and commuting; leisure time activity; sleeping; and other activities (Appendix). The questionnaire is intended to estimate habitual total and domain-specific physical activity in metabolic equivalent (MET) (kcal/kg/h)-hours per day averaged over 1 year. Subjects were asked about their activities in each of these domains during the preceding year. With regard to occupational activity, they were asked about the number of hours spent at different levels of intensity (sitting, standing, walking, strenuous work); and for leisure time activity about the frequency and number of hours spent at different levels of intensity (walking slowly, walking quickly, light to moderate and strenuous exercise). A question about sleeping hours was also asked, and then hours for other activities were calculated as the difference between the sum of the other three activities and 24 h. Total PA was calculated as the sum of hours spent for the respective activity multiplied by METs, namely 1.5, 2, 2, 4.5 and 0.9 for sitting, standing, walking, strenuous work and sleeping, respectively, for occupational activities and sleeping; and 3, 4, 4 and 4.5 for walking slowly, walking quickly, light to moderate exercise and strenuous exercise, respectively, for leisure time activities. Intensity for "other activities" was considered the same as that for sitting, at 1.5. This estimation method is referred to below as "PA without seasonal variation."

To take account of seasonal differences in activities, the PAQ also asked several questions concerning working hours and duration in months in busier seasons. The seasonally adjusted total PA was calculated as the weighted mean of the total PA for the normal season and that for the busy season, for which PA for occupational activities was multiplied by the ratio of working hours in the busy season to the normal season. This estimation method is referred to below as "PA with seasonal variation."

Study design

The subjects were asked twice about their physical activities in different seasons, one time during the main harvest season (if present) and the other time in a different season for each study area (Fig. 1). The PAQ and 24 h-R were used in each season over a study period of 9 days. On the first day, each participant was asked to complete the PAQ, then given instructions on the study procedure and filled out the 24 h-R on 4 days. On day 9, the 24 h-Rs were collected and checked. Below, the PAQs conducted in the first survey are denoted as PAQ1 and those in the second survey as PAQ2.

The 24-h physical activity record

The participants were asked to complete a 24 h-R based on the form of Bouchard [15], modified by Naito et al. [12], by recording activity for each 15-min interval. They were instructed to record their activities in as much detail as possible for 4 days each for two seasons, which were to include one Sunday or other day off plus three ordinary work days.

The data were converted into MET-hour values based on a compendium of physical activities [16] and simply averaged over 8 days to estimate habitual daily physical activity. These MET-hour values were used as the gold standard to assess the validity of the PAQ estimates [13, 17].

Statistical analysis

Validity of the PAQ was determined by comparing total PA estimated by the PAQ (PA with and without seasonal variation) to that by the 24 h-R as gold standard. Comparisons were made using Spearman's correlation coefficients with 95% confidence intervals (CIs). To examine reproducibility, the MET-hours estimated by the two PAQs administered in the different seasons were compared. In addition, to evaluate the broad applicability of the PAQ, correlation coefficients with the 24 h-R were calculated by sex, area, age group, occupation and body mass index (BMI). Four domain-specific PAs were also evaluated for validity.

All analyses were performed using JMP Software (version 6. SAS Institute Inc., 2005).

Results

Characteristics and daily total PA of the study participants

Data were analyzed for 110 subjects who participated in two surveys. Mean age was 60.7 years, and mean BMI was 24.2. Although we recruited people in their 50s and 60s, two spouses in their 40s and nine in their 70s were included in the analyses.

Daily total PA calculated with the 24 h-R by subgroup is shown in Table 1. There were statistically significant differences in occupation, but not in sex, age group, area or BMI. For descriptive purposes, seasonal difference was also examined. There were seasonal variations between two surveys, which mainly came from the difference in agricultural area, such as Saku and Kashiwazaki.

Validity and reproducibility of the PAQ

With regard to the 24 h-R, Spearman's correlation coefficients were 0.69 (95% CI 0.57–0.77) for PAQ1 without seasonal variation, 0.68 (95% CI 0.57–0.77) for PAQ1 with seasonal variation, 0.55 (95% CI 0.43–0.70) for the PAQ2 without seasonal variation, and 0.58 (95% CI 0.43–0.70) for the PAQ2 with seasonal variation.

Total PA calculated with the PAQ tended to underestimate those with the 24 h-R. While correlation coefficients with the 24 h-R were almost the same for the PAQ with and without seasonal variation, calculated METs tended to be a little larger for the PAQ with seasonal variation than without seasonal variation (Table 2). As for reproducibility, correlations between the two PAQs were 0.68 (95% CI 0.56–0.77) without and 0.68 (95% CI 0.57–0.78) with seasonal variation.

Validity by different subgroups

Physical activity estimated by the 24 h-R for various subgroups is shown in Table 3. Variations were observed by

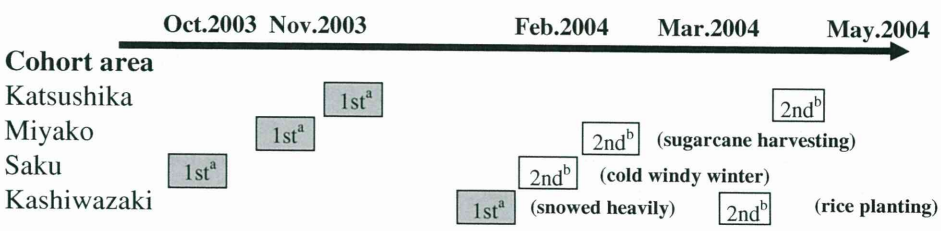

Fig. 1 Design of the validation study. *a* First survey, *b* second survey

Table 1 Daily total physical activity in MET-hours calculated from the 24 h-R by subgroup

Characteristics	Mean daily PA for 1st and 2nd 24 h-R surveys							Mean daily PA for 24 h-R1			Mean daily PA for 24 h-R2			
	n	Min	Med	Max	Mean	SD	p^a	*n*	Mean	SD	*n*	Mean	SD	p^b
All subjects	110	31.1	40.1	63.6	40.3	4.87	–	107	39.9	4.75	103	40.6	5.99	0.04
Sex														
Male	54	31.1	39.6	63.6	39.7	5.63	0.22	52	39.2	5.53	51	40.0	6.68	0.09
Female	56	33.8	40.5	57.3	40.8	3.97		55	40.5	3.81	52	41.1	5.24	0.25
Age group														
50	46	31.1	39.9	50.6	39.5	4.39	0.19	43	39.2	4.61	42	39.8	5.18	0.22
60	55	32.2	40.3	57.3	40.5	4.24		55	40.0	4.18	52	40.8	5.29	0.12
70	9	31.7	41.3	63.6	42.6	9.12		9	42.0	7.85	9	43.2	11.41	0.63
Area														
Katsushika	20	32.2	40.2	46.3	39.4	3.82	0.52	20	39.0	4.25	20	39.8	4.18	0.35
Miyako	38	31.1	39.7	63.6	41.1	6.50		37	40.5	5.86	31	41.4	8.42	0.17
Saku	30	31.5	40.0	46.9	39.6	3.51		28	40.8	3.78	30	38.6	3.50	0.00
Kashiwazaki	22	31.7	41.3	45.7	40.5	3.97		22	38.2	3.83	22	42.8	5.32	0.00
Occupation														
Clerk	17	31.1	37.4	50.6	37.7	4.80	0.00	17	37.5	5.09	14	36.9	3.74	0.59
Shopkeeper, sales	43	31.5	39.6	47.9	39.6	3.42		40	38.7	3.63	43	40.6	4.55	0.01
Housewife, light activity	20	32.6	40.0	47.6	39.6	4.35		20	39.8	4.45	18	39.5	4.98	0.57
Agriculture	30	36.1	41.9	63.6	43.1	5.80		30	42.8	4.85	28	43.1	8.16	0.56
BMI														
Less than 22	25	31.7	42.1	50.6	41.2	4.78	0.53	25	40.6	5.09	22	40.8	5.02	0.20
22–25	47	31.5	39.6	57.3	39.8	4.57		44	39.7	4.37	45	40.2	5.75	0.40
More than 25	38	31.1	39.8	63.6	40.2	5.30		38	39.5	5.01	36	41.0	6.90	0.16

METs Metabolic equivalents, *24 h-R1* 24-h physical activity record 1st survey, *24 h-R2* 24-h physical activity record 2nd survey

[a] *p* Values by ANOVA comparing mean daily PA within each characteristics

[b] *p* Values by paired t test comparing mean daily PA between 24 h-R1 and 24 h-R2

Table 2 Daily total physical activity (MET-hours) calculated from 24-h physical activity record and physical activity questionnaire

All subjects	*n*	Min	Med	Max	Mean	SD
24-h Physical activity record	110	31.1	40.1	63.6	40.3	4.87
PAQ1 (without[a])	109	27.3	37.0	57.4	38.2	6.07
PAQ1 (with[b])	109	27.3	37.4	57.6	38.4	6.16
PAQ2 (without[a])	105	26.9	36.2	56.4	37.3	6.00
PAQ2 (with[b])	105	26.9	36.3	59.6	37.6	6.31

PAQ Physical activity questionnaire, *PAQ1* first survey, *PAQ2* second survey

[a] Without seasonal variation

[b] With seasonal variation

sex, age group, area, occupation and BMI, although not necessarily with significance. Consistently high correlations were observed between the 24 h-R and PAQ both with and without seasonal variation for most of the subgroups (Table 3). Correlation coefficients by the PAQ with and without seasonal variation are virtually same (data not shown).

Distribution of PA among domains

The four domain-specific PAs are shown in Table 4. For total PA, occupational activity and other activities each contributed more than one-third of the total. With regard to the correlations between domain-specific PA from PAQ1 without seasonal variation and those from 24 h-R, high correlation was observed for sleeping, and moderate correlations were observed for occupational activity and leisure time physical activity. In contrast, a low correlation was observed for other activities. Further, domain-specific correlation was not improved when seasonal variation was taken into account. The results for PAQ2 were similar to those for PAQ1 (data not shown).

Table 3 Correlation coefficients of the physical activity questionnaire and 24-h physical activity record by subgroup

Characteristics	PAQ1			
	Without[a]		With[b]	
	r^c	95%CI	r^c	95%CI
All subjects	0.69	0.57–0.77	0.68	0.57–0.77
Sex				
Male	0.80	0.68–0.88	0.80	0.68–0.88
Female	0.58	0.37–0.73	0.57	0.36–0.73
Age group				
50	0.76	0.61–0.86	0.77	0.62–0.87
60	0.59	0.39–0.74	0.60	0.40–0.74
70	0.80	0.29–0.96	0.78	0.25–0.95
Area				
Katsushika	0.76	0.47–0.90	0.76	0.47–0.90
Miyako	0.72	0.52–0.85	0.73	0.53–0.85
Saku	0.55	0.22–0.76	0.58	0.27–0.78
Kashiwazaki	0.68	0.36–0.86	0.60	0.24–0.82
Occupation				
Clerk	0.73	0.39–0.90	0.65	0.25–0.86
Shopkeeper, sales	0.65	0.43–0.79	0.65	0.44–0.80
Housewife, light activity	0.57	0.18–0.81	0.58	0.19–0.82
Agriculture	0.74	0.52–0.87	0.75	0.53–0.88
BMI				
Less than 22	0.72	0.46–0.87	0.71	0.43–0.86
22–25	0.58	0.35–0.75	0.58	0.35–0.75
More than 25	0.69	0.48–0.83	0.70	0.49–0.83

PAQ Physical activity questionnaire, *PAQ1* first survey

[a] Without seasonal variation

[b] With seasonal variation

[c] Spearman's correlation coefficient

Discussion

This study demonstrates the validity and reproducibility of the PAQ used in the JPHC study. Further, the validity of the PAQ answers for the estimation of average physical activity over 1 year was confirmed against the possibe seasonal variation in PA even within rural agricultural areas. In addition, the validity of the PAQ was also confirmed for various subgroups defined by sex, age, area of residence, occupation and BMI.

Reflecting the subjects of the JPHC study, our present subjects came from widely varying areas of Japan and occupations. The age category of the 50s and above was selected to reflect the JPHC study's purpose, namely to study the onset of diseases such as cancer, diabetes and cardiovascular disease. Despite the wide variation among subjects, our simple PAQ showed good validity, with a higher correlation with 24 h-R PA values ($r = 0.55$–0.80) than in a study on female alumnae ($r = 0.15$–0.52) (Historical Leisure Activity Questionnaire) [13].

The results from the 24 h-R in this study showed the presence of a seasonal difference in PA, especially in agricultural areas and among farm laborers. An increase in PA during the 'busy season' came from farm work such as harvesting, while a relative decrease in PA during other seasons resulted from environmental factors such as hot summers and cold winters, including snowfall-related sedentariness, resulting in seasonal differences. Thirty of our subjects described themselves as farmers, but a number of others also farmed as a side job, and some housewives in rural areas also farmed. The first survey in the Kashiwazaki area was conducted

Table 4 Domain-specific physical activity and their correlations between 24-h physical activity record and physical activity questionnaire

	Total PA (METs) by 24 h-R		Domain-specific PA by PAQ1 (without[a]) versus domain-specific PA by 24 h-R		Domain-specific PA by PAQ1 (with[b]) versus domain-specific PA by 24 h-R	
	METs	%	r^c	95% CI	r^c	95% CI
24-h Physical activity record						
Sleeping	6.8	16.9	0.51	0.36 to 0.64	–	–
Occupational activity	14.8	36.7	0.31	0.12 to 0.47	0.12	−0.07 to 0.30
Leisure time physical activity	4.6	11.4	0.31	0.13 to 0.47	–	–
Other activity	14.1	35.0	0.09	−0.1 to 0.27	−0.07	−0.26 to 0.12
Total activity	40.3	100.0	0.69	0.57 to 0.77	0.68	0.57 to 0.77

PA Physical activity, *METs* metabolic equivalents, *24 h-R* 24-h physical activity record, *PAQ* physical activity questionnaire, *PAQ1* first survey

[a] Without seasonal variation

[b] With seasonal variation

[c] Spearman's correlation coefficient

during a period of heavy snowfall, and although the subjects worked hard to remove snow, they seldom ventured outdoors. At the time of the second survey, in contrast, they were engaged in planting rice. The second survey in the Saku area was conducted during a cold winter when the subjects may have engaged in less outside farm work. These findings may indicate that physical activity questions should ask about "an average over the year;" otherwise, estimated physical activity depends on the season when the survey was conducted.

In our study, adding the extra working hours of the 'busy season' did not improve the correlation with the 24 h-R estimates, but tended to mitigate the underestimation of total PA calculated with PAQ. Although this can be interpreted as questions about seasonal variation in physical activity not being necessary to estimate average MET-hours over a year, even when such variation exists, the possibility remains that seasonal variation could be more accurately considered using a more sophisticated questionnaire. For example, subjects were asked about the length of the busier season for 1 year and the length of working in the busier season in our questionnaire. This question considered the variability between subjects, but the difference in intensity of working in busier season was not considered.

The limitation of our study was using the 24 h-R as a gold standard. In the 24 h-R, physical activity is calculated using the compendium [16]. In an aging society like in rural Japan, farming or housework may be more labor-saving than previously, and use of the compendium may have overestimated PAs. The higher correlation between PA estimates based on the 24 h-R and PAQ may be attributable to the fact that, because the same conversion table (compendium) was used for physical activity to METs, they had the same direction of measurement error, although the 24 h-R was used as a gold standard in a previous study [13] and its validity to doubly labeled water was confirmed [17].

The PAQ is based on the simple notion that total daily PA consists of four domains, namely occupational activities, including housekeeping and commuting, leisure time activities, sleeping and other activities. This may explain the high validity and reproducibility of this instrument, but may also be the reason for the lack of increase in validity when seasonal variation is considered. Although the domain "occupational activities" includes housekeeping, PA is estimated from hours spent for activities in different levels of intensity, such as sitting, standing, walking and strenuous work, and these categories may not sufficiently represent housekeeping activities. It thus appears possible to improve the PAQ to more precisely determine intensity levels, especially for housekeeping.

In summary, this study shows that our simple PAQ provides a valid estimation of total as well as domain-specific PA in different populations without considering seasonal variation, not only in our JPHC cohort study, but also in broader populations differing by age, sex, residential area, occupation and BMI in epidemiological study.

Acknowledgments The authors wish to thank Prof. Hirotsugu Ueshima and Dr. Yoshikuni Kita of the Department of Health Science, Shiga University of Medical Science, for their cooperation and contribution to pilot studies. We also wish to thank Prof. Yasuo Ohashi and Dr. Akiko Harada of the Department of Biostatistics/Epidemiology and Preventive Health Science, Graduate School of Medicine, University of Tokyo, for their advice regarding accelerometer. We also wish to thank Dr. Hiroji Kitazato of the Department of Endocrinology and Metabolism, The Institute for Adult Diseases, Asahi Life Foundation and Dr. Masayuki Kato of the Japan Foundation for the Promotion of International Medical Research Cooperation, for their cooperation in this research. This study was supported by Health Sciences Research grants (Clinical Research for Evidence-based Medicine H15-006, Comprehensive Research on Cardiovascular Diseases H16-019, H17-019, H18-028, H19-019) and a grant-in-aid for Cancer Research and for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan. The authors also thank all staff members in each study area especially for their efforts with PAQ validation study.

Appendix

JPHC Physical Activity Questionnaire

Qa, How many hours do you usually sleep?

( ) hrs

Qb, Do you have 'busier season' than usual during the year?
If you have, please answer the duration of it.

No	less than 1 month	more than 1 month to 2 months	2 to 3 months	3 to 4 months	4-5 months	5-6 months
⬭	⬭	⬭	⬭	⬭	⬭	⬭
⬭	⬭	⬭	⬭	⬭	⬭	⬭

Qc, How long did you work on average? Including comuting and household chores.
If you have a 'busier season', please also answer about it.

work hours	less than 1 hr	more than 1hr to 3 hrs	3 to 5 hrs	5 to7 hrs	7 to 9 hrs	9 to 11 hrs	more than 11hrs
1 on average	⬭	⬭	⬭	⬭	⬭	⬭	⬭
2 in busier season	⬭	⬭	⬭	⬭	⬭	⬭	⬭

Qd, Please answer the hours spent for each activities about on the average day in the last year.
Including comuting and household chores.

	none	less than 1 hr	more than 1hr to 3 hrs	3 to 5 hrs	5 to7 hrs	7 to 9 hrs	9 to 11 hrs	more than 11hrs
sitting	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
standing	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
walking	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
strenous work	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

Qe, Please answer of the leisure time physical activity frequencies and hours spent per one opportunity last year.

exercise	frequency					time spent for one opportunity					
	less than once a month	once to 3 times a month	once to twice a week	3 to 4 times a week	almost every day	less than 30 minutes	30 to 59 mins	1 to 2 hrs	2 to 3 hrs	3 to 4 hrs	more than 4hrs
walking slowly such as taking walk	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
walking fast	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
light to moderate exercise such as golf, gardening.	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
vigorous exercise such as tennis, jogging, aerobics, and swimming	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

References

1. Eriksson KF, Lindgarde F. No excess 12-year mortality in men with impaired glucose tolerance who participated in the Malmo Preventive Trial with diet and exercise. Diabetologia. 1998;41(9):1010–6.
2. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002;346(6):393–403.

3. Laaksonen DE, Lindstrom J, Lakka TA, Eriksson JG, Niskanen L, Wikstrom K, Aunola S, Keinanen-Kiukaanniemi S, et al. Physical activity in the prevention of type 2 diabetes: the Finnish diabetes prevention study. Diabetes. 2005;54(1):158–65.
4. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, Hu ZX, Lin J, Xiao JZ, Cao HB, Liu PA, Jiang XG, Jiang YY, Wang JP, Zheng H, Zhang H, Bennett PH, Howard BV. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. Diabetes Care. 1997;20(4):537–44.
5. Hu FB, Stampfer MJ, Solomon C, Liu S, Colditz GA, Speizer FE, Willett WC, Manson JE. Physical activity and risk for cardiovascular events in diabetic women. Ann Intern Med. 2001;134(2):96–105.
6. Kannel WB, Belanger A, D'Agostino R, Israel I. Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham study. Am Heart J. 1986;112(4):820–5.
7. Lee IM, Paffenbarger R, Hsieh C. Physical activity and risk of developing colorectal cancer among college almuni. J Natl Cancer Inst. 1991;83(18):1324–9.
8. Pate RP, Steven MP, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA. 1995;273(5):402–7.
9. Lee IM, Manson JE, Ajani U, Paffenbarger RS, Buring JE. Physical activity and risk of colon cancer: the Physicians' Health Study. Cancer Causes Control. 1997;8(4):568–74.
10. Nakanishi N, Nishina K, Yoshida H, Matsuo Y, Nagano K, Nakamura K, Suzuki K, Tatara K. Hours of work and the risk of developing impaired fasting glucose or type 2 diabetes mellitus in Japanese male office workers. Occup Environ Med. 2001;58(9):569–74.
11. Okada K, Hayashi T, Tsumura K, Suematsu C, Endo G, Fujii S. Leisure-time physical activity at weekends and the risk of type 2 diabetes mellitus un Japanese men: the Osaka Health Survey. Diabet Med. 2000;17(1):53–8.
12. Naito Y. Relationship between physical activity and health examination variables in male workers—new methods to assess physical activity and their applications to epidemiologic research. Nippon Koshu Eisei Zasshi. 1994;41(8):706–19.
13. Chasan-Taber L, Erickson JB, Philip, Nasca PC, Chasan-Taber S, Freedson PS. Validity and reproducibility of a physical activity questionnaire in women. Med Sci Sports Exerc. 2002;34(6):987–92.
14. Baba S, Mnnami T, Konishi M, Sasaki S, for the JPHC Study Group. Anthropometric measures, blood pressure and major laboratory examination results in the health check-up examination among the JPHC Study participants at baseline survery. J Epidemiol. 2001;11(6 Suppl):S87–93.
15. Bouchard C, Tremblay A, Leblanc C, Lortie G, Savard R, Theriault G. A method to assess energy expenditure in children and adults. Am J Clin Nutr. 1983;37(3):461–7.
16. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swarts AM, Strath SJ, O'brien WL, Bassett DR Jr, Schmits KH, Emplaincourt PO, Jacobs DR Jr, Leon AS. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc. 2000;32(9 Suppl):S498–504.
17. Rafamantanantsoa HH, Ebine N, Yoshioka M, Higuchi H, Yoshitake Y, Tanaka H, et al. Validation of three alternative methods to measure total energy expenditure against the doubly labeled water method for older Japanese men. J Nutr Sci Vitaminol (Tokyo). 2002;48(6):517–23.

Characteristics and Outcomes of Patients With Advanced Gastric Cancer Who Declined to Participate in a Randomized Clinical Chemotherapy Trial

By Chiharu Tanai, MD, Takako Eguchi Nakajima, MD, PhD, Kengo Nagashima, Ken Kato, MD, PhD, Tetsuya Hamaguchi, MD, PhD, Yasuhide Yamada, MD, PhD, FACP, Kei Muro, MD, Kuniaki Shirao, MD, PhD, Hideo Kunitoh, MD, Yasuhiro Matsumura, MD, PhD, Seiichiro Yamamoto, PhD, and Yasuhiro Shimada, MD

Medical Oncology Division, National Cancer Center Hospital, Tokyo; Faculty of Pharmaceutical Science, Josai University, Saitama; Medical Oncology Division, Aichi Cancer Central, Aichi; Medical Oncology Division, Oita University, Oita; Investigative Treatment Division, Research Center for Innovative Oncology, National Cancer Center Hospital East, Chiba; Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan

Abstract

Purpose: There is insufficient data to verify whether participation in clinical trials in itself can lead to better clinical outcomes. We have analyzed the characteristics and outcomes of patients who declined to participate in a randomized trial in comparison with those who participated in the trial.

Patients and Methods: A randomized trial for naive advanced gastric cancer was offered to 286 patients. The trial investigated the superiority of irinotecan plus cisplatin and the noninferiority of S-1 compared with continuous fluorouracil infusion. We retrospectively reviewed the characteristics and outcomes for both participants and nonparticipants in this trial.

Results: Of the 286 patients, 98 (34%) declined to participate in the trial. The rate of declining was significantly higher among younger patients ($P = .003$), and it varied significantly between attending physicians (range, 23% to 58%; $P = .004$). There were no other significant correlations between rate of declining and patient characteristics. No significant differences were observed in the clinical outcomes between the participants and nonparticipants, for whom the median survival times were 367 versus 347 days, respectively. The hazard ratio for overall survival, adjusted for other confounding variables, was 1.21 (95% CI, 0.91 to 1.60). No interaction was observed between participation and the various regimens.

Conclusion: There was no difference in clinical outcomes between participants and nonparticipants. However, the patient's age and the doctor-patient relationship may have an effect on patient accrual to randomized trials.

Introduction

A randomized clinical trial (RCT) is the definitive method for comparing the efficacy of treatments, and an RCT is a crucial step in the development of any new cancer treatment. Nevertheless, there is a consistent problem in that low accrual rates limit the progress of RCTs.[1-3]

Several factors that act as barriers to participation in trials have been documented,[1-6] and some have been successfully targeted for improvement.[4-5] Major barriers include a lack of availability of appropriate trials, limitations of eligibility criteria, socioeconomic factors (including insufficient awareness of clinical trials, lack of medical insurance, and geographical limitations), physician triage, and patient decision making. Insufficient data are available on the actual outcomes for nonparticipants in RCTs in comparison with those for participants.[7-11]

We have previously analyzed the characteristics and outcomes of patients who had been referred and were eligible for, but declined to participate in, RCTs and compared them with those of participants, with the aim of developing an approach to improve patient accrual to RCTs.[12] We found no evidence to suggest any significant differences in the characteristics or clinical outcomes between participants and nonparticipants. We also reported that the trial design and the doctor-patient relationship might have an effect on patient accrual to RCTs.

In the present study, we reviewed the characteristics and clinical outcomes of patients who met the eligibility criteria of an RCT designed to compare three different types of therapy, including both injection and oral agents, the levels of toxicity of which were estimated to be quite different. Our analysis was designed to test our previous findings. We also analyzed whether participation in an RCT that compared several arms with different efficacies affected patient outcomes.

Patients and Methods

All the patients who were recruited into this study fulfilled the entry criteria for the Japan Clinical Oncology Group RCT on unresectable or recurrent gastric cancer (JCOG 9912). The patients were informed of all aspects of the trial and were invited to participate. Irrespective of their participation or nonparticipation in the RCT, all received first-line chemotherapy at the National Cancer Center Hospital, Tokyo, Japan, between November 2000 and January 2006. Signed informed consent was obtained from the patients to permit future statistical analysis of data from their clinical courses and outcomes, even if they were treated outside the clinical trials.

The RCT was a three-arm, phase III trial conducted by JCOG to investigate the superiority of irinotecan (CPT-11) plus cisplatin (CDDP) combination chemotherapy and the

noninferiority of *S*-1 compared with continuous fluorouracil (FU) infusion.[13]

The criteria for inclusion of patients were as follows: histologically documented unresectable or recurrent gastric cancer; no prior systematic chemotherapy or radiation therapy except for adjuvant chemotherapy with one oral fluoropyrimidine agent other than *S*-1, completed 6 months earlier; age 20 to 75 years; Eastern Cooperative Oncology Group performance status (PS) of 0 to 2; no history of chemotherapy or radiation therapy for malignant disease other than gastric cancer; and adequate hematologic, hepatic, and renal functions. Those with severe peritoneal dissemination resulting in impaired bowel passage, ascites beyond the pelvic cavity, or wall deformity detected by barium enema were excluded. A measurable lesion was not mandatory. Each patient was required to submit written informed consent before taking part in the RCT.

The treatment schedule of each arm was as follows: (1) Continuous FU infusion: FU was infused continuously over 120 hours; this required hospitalization for 7 days. (2) CPT-11 plus CDDP combination chemotherapy: CPT-11 was infused on days 1 and 15, and CDDP was infused on day 1; this required hospitalization for 5 days. (3) *S*-1 monotherapy: *S*-1 was administered orally on days 1 through 28 and repeated every 6 weeks. Patients were required to undergo a medical examination every 2 weeks. Patients who declined to participate ultimately selected their treatment regimen after discussions with their families and physicians. We provided the selected therapies after confirming that patients fully understood that the standard therapy at that time was FU infusion and that the CPT-11 plus CDDP combination therapy and the *S*-1 monotherapy were still under evaluation.

In the RCT, CPT-11 plus CDDP therapy resulted in a longer survival rate, and *S*-1 showed significant noninferiority compared with FU.[13] The hazard ratio (HR) of CPT-11 plus CDDP versus FU was 0.82 (95% CI, 0.68 to 0.99; $P = .019$). The HR of *S*-1 versus FU was 0.83 (95% CI, 0.68 to 1.00, *P* for noninferiority $< .001$). AQ: A

Six male physicians participated in the trial. At the start of the trial (November 2000), physicians A, B, C, D, E, and F had 8, 10, 11, 17, 19, and 19 years of experience, respectively, as gastrointestinal oncologists. One of these six attending staff physicians, together with one, two, or three residents or trainees, attended each consultation. They explained to the patients that this was a JCOG study, that standard therapy was continuous FU infusion, and that we could not tell which arm was superior, but the treatment schedule, toxicities, and required lengths of hospitalization were expected to be different among the various arms. If patients chose not to participate in the study, we recommended the standard therapy, but they could choose other, off-protocol regimens.

We reviewed all the medical records of patients who underwent chemotherapy for unresectable or recurrent gastric cancer between November 2000 and January 2006, and we selected 286 patients who were documented as having been offered the opportunity to participate in the RCT. During the study period, some other patients were judged to be ineligible for the study and were offered other treatments, as clinically indicated, but the number of such patients is not available. Paper and/or electronic medical records from the initial visit to our center to the end of follow-up were reviewed retrospectively. Demographic data (age and sex), medical information (tumor histology, clinical stage, PS, peritoneal dissemination, and therapy characteristics), and clinical outcomes (response rate [RR], follow-up time, overall survival [OS] time, and 1- and 2-year survival rates) were abstracted and analyzed. Response was evaluated by the attending physicians according to the Response Evaluation Criteria in Solid Tumors (RECIST).[14] It is our policy to assess clinical responses to RECIST, even in routine practice. The follow-up time at our institution was defined as the period from the first day of initial therapy to the last visit or the last day of hospitalization at our institution (including death during follow-up). Data on the survival of the patients who left our institution were collected through inquiries to the Japanese official agency for family registries.

The χ^2 test and logistic regression analysis were used to assess associations between patient characteristics and the rate of declining to participate. OS curves were prepared by using the Kaplan-Meier method and were compared with the results of the log-rank test. All participants (those who agreed to be enrolled onto the RCT), including two who were later found to be ineligible after random assignment, and all nonparticipants (those who declined to participate in the RCT), including one who was lost to follow-up, were included in the OS analysis. A Cox proportional hazards model was used to adjust for other potential confounding factors (ie, age, sex, histology, clinical stage, PS, nonsevere peritoneal dissemination, and treatment regimen) in comparing the OS of participants and nonparticipants. Interaction between participation and regimen was tested with an α level of 0.2; $P < .05$ was regarded as significant. Collected data were analyzed by using an SPSS II statistical package (SPSS, Chicago, IL). This study was approved by the institutional review board at the National Cancer Center and was conducted in accordance with the ethical principles stated in Japanese ethics guidelines for clinical and epidemiological studies. No patient explicitly refused to be analyzed for his or her outcome during this study period. The institutional review board approved the use of such clinical data for the study objective.

Results

Table 1 shows the patient characteristics and the rates of declining. A total of 190 patients accepted, and 96 patients (34%) declined, entry into the RCT. There was no significant correlation between the declining rate and sex, clinical stage, PS, tumor histology, or peritoneal dissemination. Patients younger than 60 years declined to participate at a significantly higher frequency ($P = .003$). There were also significant differences in the declining rates between the various attending physicians who informed the patients about the trial and asked for their participation ($P = .004$). The patients were divided equally among the offering physicians by characteristic. T1

Table 1. Patient Characteristics and Rate of Declining to Participate in Randomized Clinical Trials

Characteristic	Participants		Non-participants			Participants			Nonparticipants		
	No.	%	No.	%	ROD (%)	OR	95% CI	P	OR	95% CI	P
No.	190		96		16						
Sex											
Male	146	77	64	67	30	1.66	0.97 to 2.85	.07	0.49	0.26 to 0.90	.02
Female	44	23	32	33	42						
Age, years											
< 60	48	25	41	43	46	0.45	0.27 to 0.76	.003	2.54	1.44 to 4.47	.01
≥ 60	142	75	55	57	28						
Clinical stage											
III	1	1	0	0	0						
IV	146	77	70	73	32	1.30	0.74 to 2.26	.36	0.55	0.29 to 1.04	.06
Recurrent	43	23	26	27	38						
PS											
0	104	55	51	53	30						
1	84	44	44	46	34	0.96	0.58 to 1.58	.87	0.85	0.49 to 1.47	.56
2	2	1	1	1	33	0.98	0.09 to 11.07	.99	0.51	0.03 to 7.04	.61
Histology											
Well differentiated	75	39	34	35	31	0.85	0.51 to 1.42	.55	1.05	0.59 to 1.89	.86
Poorly differentiated	115	61	61	64	35						
Undifferentiated	0	0	1	1	100						
Peritoneal dissemination											
Yes	85	45	51	53	38	0.71	0.44 to 1.17	.18	1.54	0.89 to 2.69	.13
No	105	55	45	47	30						
Physicians											
A	31	16	19	20	38			.04			< .01
B	27	14	10	10	27						
C	35	18	13	14	27						
D	25	13	27	28	52						
E	67	35	20	21	23						
F	5	3	7	7	58						

NOTE. Univariate analysis was performed with Pearson's χ^2 test; multivariate analysis was logistic regression analysis.
Abbreviations: ROD, rate of declining; OR, odds ratio; PS, performance status.

T2 Table 2 shows the treatment options of patients who declined to participate in the RCT. Nearly 60% of all those who declined to participate selected *S*-1 monotherapy. Moreover, approximately 70% of nonparticipants who were under 60 years of age selected *S*-1 monotherapy. The proportion of those who selected CPT-11 plus CDDP, which was expected to be more beneficial but showed more severe toxicity and required hospitalization, was not necessarily higher among nonparticipants younger than 60 years than among older nonparticipants. No specific tendency was shown in selection of regimen in relation to the attending physician.

AQ: B Post-therapy was analyzed in 188 of the participants. This group excluded all 96 nonparticipants, as well as two patients found to be ineligible after random assignment: one patient who developed gastrointestinal bleeding several hours after entry, and another who was later diagnosed with adenosquamous cell carcinoma. Survival was analyzed in the 190 participants and the 96 nonparticipants. There were no treatment-related deaths among either the participants or the nonparticipants. There was no difference in the number of cycles of first-line chemotherapy received by participants or nonparticipants: 53% of the participants and 58% of the nonparticipants received fewer than three cycles ($P = .406$). A total of 85% of the participants and 70% of the nonparticipants were given more than two chemotherapy regimens. Statistically, more participants than nonparticipants were given chemotherapy after the initial therapy ($P = .003$). A total of 14 (7%) of the participants and 6 (6%) of the nonparticipants in the RCT participated later in early-phase clinical trials of experimental therapies.

There were no major differences in clinical outcome between participants and nonparticipants (Figure 1). Clinical response to F1,AQ:C the initial therapy was analyzed in all 190 participants and 96 nonparticipants. The RR was 30.5% for the participants and 21.9% for the nonparticipants ($P = .121$). The median follow-up time at our hospital was not significantly different between the participants (317 days) and the nonparticipants (292 days). The median survival time (MST) was 367 days for the participants and

Table 2. Characteristics and First Treatment Regimen of Nonparticipants

Characteristic	FU		CPT-11 Plus CDDP		S-1		P*
	No.	%	No.	%	No.	%	
No.	31		8		57		
Sex							
Male	22	34	5	8	37	58	.819
Female	9	28	3	9	20	63	
Age, years							
< 60	10	24	3	7	28	68	.297
≥ 60	21	38	5	9	29	53	
Clinical stage							
IV	20	29	6	9	44	63	.438
Recurrent	11	42	2	8	13	50	
PS							
0	15	29	6	12	30	59	.641
1	16	36	2	5	26	59	
2	0	0	0	0	1	100	
Histology							
Well differentiated	10	29	4	12	20	59	.814
Poorly differentiated	21	34	4	7	36	59	
Undifferentiated	0	0	0	0	1	100	
Peritoneal dissemination							
Yes	16	31	1	2	34	67	.043
No	15	33	7	16	23	52	
Physicians							
A	5	26	1	5	13	68	.363
B	4	40	3	30	3	30	
C	4	31	0	0	9	69	
D	8	30	2	7	17	63	
E	8	40	2	10	10	50	
F	2	29	0	0	5	71	

Abbreviations: FU, fluorouracil; CPT-11, irinotecan; CDDP, cisplatin; PS, performance status.
* Pearson's χ^2 test.

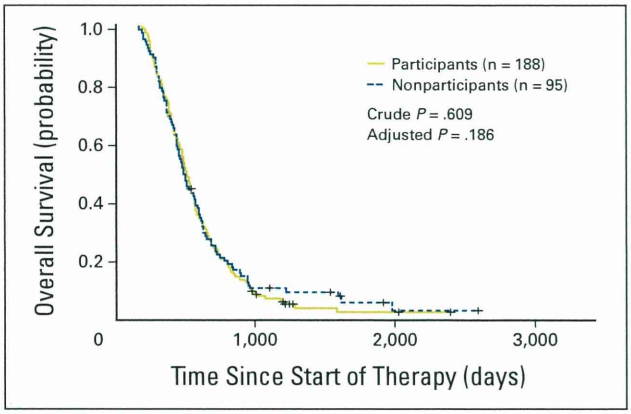

Figure 1. Overall survival of nonparticipants in randomized trials compared with that of participants. No significant difference were observed.

347 days for the nonparticipants. There were no significant difference in OS between the participants and the nonparticipants (Figure 1), and the HR was 1.07 (participants *v* nonparticipants; 95% CI, 0.83 to 1.38). With the Cox proportional hazards model adjusted for sex, age, tumor histology, clinical stage, PS, peritoneal dissemination, and treatment regimen, the HR of participants versus nonparticipants was 1.21 (95% CI, 0.91 to 1.60; $P = .19$). Furthermore, the RR and OS were not significantly different between the participants and the nonparticipants for each regimen. The RR in participants versus nonparticipants was 9.5% versus 6.5% for FU ($P = .646$), 54.0% versus 62.5% for CPT-11 plus CDDP ($P = .648$), and 28.1% versus 24.6% for *S*-1 ($P = .657$). MST in participants versus nonparticipants was 358 days versus 335 days for FU, 435 days versus 458 days for CPT-11 plus CDDP, and 338 days versus 345 days for *S*-1. The HR values for OS were 0.91 (95% CI, 0.57 to 1.44; $P = .679$) for FU, 0.99 (95% CI, 0.38 to 2.56; $P = .981$) for CPT-11 plus CDDP, and 1.22 (95% CI, 0.81 to 1.83; $P = .333$) for *S*-1 (Appendix Figures A1-A3, online only).

We analyzed the interaction between participation and regimen. The *P* value for the interaction term was greater than the α level of 0.2; it was 0.75 for participants and CPT-11 plus CDDP, and 0.28 for participants and *S*-1 (Appendix Table A1, online only).

Discussion

We previously analyzed the characteristics and outcomes of patients who had been referred and were eligible for, but declined to participate in, two RCTs for naive, advanced, non–small-cell lung cancer and compared them with those of participants.[12] Trial 1 was a comparison of four similar combinations of injection therapies (cisplatin-irinotecan, carboplatin-paclitaxel, cisplatin-gemcitabine and cisplatin-vinorelbine), and Trial 2 compared two sequences of injection and oral therapies (four courses of carboplatin-paclitaxel followed by gefitinib or gefitinib until disease progression, followed by carboplatin-paclitaxel). We found that the rate of declining to participate in a trial in which similar injection therapies were compared was lower than that in a trial in which injection and oral therapies were compared (16% *v* 37%). We also reported that there was no evidence to suggest any difference, except for that of the attending physician, in the characteristics and clinical outcomes between participants and nonparticipants.

In the present study, we compared three different regimens, two of which were given by injection and the other as an oral agent. The rate of declining in the present study was 34%, which was as high as that of Trial 2 in our previous study. It is easy to understand that more difficulty is experienced in accepting the randomization of different types of therapy.[8,15] The therapy arms of the present study used different methods of administration; moreover, the estimated toxicities and the need for hospitalization were quite different among the various arms. We thus confirmed our previous finding that trial design influences trial accrual.

Nearly 60% of those who declined entry into the trial selected *S*-1 monotherapy, which may reflect the patients' desire for convenience and a higher quality of life. Younger patients, in particular, preferred this oral agent. We speculate that they may attach greater importance to avoiding hospitalization than to uncertain efficacy. This difference between age groups was a new finding of the present study.

As noted in our previous report, the rate of declining also appeared to be greatly affected by the attending physician. No record was available of which person actually took the initiative and offered the trial at each consultation; however, even when a resident or trainee offered the trial, the attending physician would have taken the responsibility for the consultation. No relationship was found between the length of experience of the physician as a gastrointestinal oncologist and the rate of declining. Each attending physician attempted to present the three regimens equally without showing favor toward any particular regimen; this suggests that individual consultations were not the source of bias. Physicians' clinical communications have been noted as affecting patients' decision making regarding participation in clinical trials.[16] Improved communications and more frequent interventions by clinical research coordinators and other medical staff members for all eligible patients might improve the accrual rate.[17-19] This study did not clarify whether differences in communication skills between physicians led to differences in rates of declining; further investigations of this effect are warranted.

On the other hand, inadequate data are available on the actual outcomes for RCT nonparticipants compared with those of participants.[7-11] Although several reports and a review[7] have suggested the existence of a "trial effect" in which participants enjoy more favorable outcomes, other studies, especially those that attempted to exclude confounding factors, have refuted this finding.[8-11] Our study revealed that the outcomes for participants were no better than those of nonparticipants. Furthermore, our results showing that interactions between participants or nonparticipants and the treatment regimen were not significant (Table 3) may suggest that the conclusion of this RCT could be generalized. The HR for OS between participants and nonparticipants was very close to 1 (0.91; 95% CI, 0.57 to 1.44) in the FU arm, which was the control arm in the trial, and numerically favorable for nonparticipants in the CPT-11 plus CDDP arm and the *S*-1 arm (CPT-11/CDDP: 0.99; 95% CI, 0.38 to 2.56; *S*-1: 1.22; 95% CI, 0.81 to 1.83), which were the testing arms in the trial. This suggests the possibility of a self-selection bias.

Our study has several limitations. First, we selected the participants and nonparticipants retrospectively among those who underwent chemotherapy for advanced gastric cancer during the period in which we conducted the RCT. The fact that all the patients accepted treatment of some sort is, in itself, a selection process, and information on patients who declined all active treatments at our institution remains elusive. There may have been some patients who did not want to continue active treatment and who instead opted for supportive care only, or other patients who declined to participate in the RCT and went to other hospitals. We did not review this population, and if there were any such patients, this may have affected the survival analysis.

Second, the present study was conducted at a single academic institution, and there was an insufficient number of patients. As a result, the numbers of patients in the various subsets were quite small, and it is difficult to rule out significant differences in some of these because of a lack of statistical power. Our investigation should therefore be interpreted as exploratory and hypothesis generating. Our results require further validation at other institutions, preferably on a multi-institutional basis, because the situation may well be different at other institutions.

Third, no data were available regarding the reasons for participation or nonparticipation. Such information would be useful for analyzing factors that affect consent or refusal to participate and would help in improving the accrual rate. However, so that patients are not coerced into participating in the study, reasons for their participation or refusal need to be collected by independent investigators.

In conclusion, we confirmed that the rate of declining to participate in RCTs was influenced by the design of the trial and by the referring physician. The age of the patients also had an effect on the rate of declining, suggesting that some patients may attach a greater importance to not having their normal schedule disrupted than to expectations of efficacy. There was no evidence of any difference in the RRs and survival times between participants and nonparticipants, and the interaction between participants or nonparticipants and the treatment regimen was not significant.

Accepted for publication on September 28, 2010.