carcinoma with positive lymph nodes: Retrospective analysis of predictive variables regarding survival and failure patterns. *Am J Clin Oncol* 2004;27:140–148.
16. Kaplan E, Meier P. Non-parametric estimation from incomplete observation. *J Am Stat Assoc* 1958;53:457–481.
17. Mantle N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966;50:163–170.
18. Cox D. Regression models and life table. *J R Stat Soc* 1972;34: 197–220.
19. Whitney CW, Sause W, Bundy BN, *et al.* Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: A Gynecologic Oncology Group and Southwest Oncology Group study. *J Clin Oncol* 1999;17:1339–1348.
20. Sasaoka M, Fuwa N, Asano A, *et al.* Patterns of failure in carcinoma of the uterine cervix treated with definitive radiotherapy alone. *Am J Clin Oncol* 2001;24:586–590.
21. Kodaira T, Fuwa N, Nakanishi T, *et al.* Retrospective analysis of postoperative radiotherapy for node-negative cervical carcinoma with stage IB-IIB disease. *Jpn J Clin Oncol* 2002;32: 255–261.
22. Small W Jr., Winter K, Levenback C, *et al.* Extended-field irradiation and intracavitary brachytherapy combined with cisplatin chemotherapy for cervical cancer with positive para-aortic or high common iliac lymph nodes: Results of ARM 1 of RTOG 0116. *Int J Radiat Oncol Biol Phys* 2007;68:1081–1087.
23. Varia MA, Bundy BN, Deppe G, *et al.* Cervical carcinoma metastatic to para-aortic nodes: Extended field radiation therapy with concomitant 5-fluorouracil and cisplatin chemotherapy: A Gynecologic Oncology Group study. *Int J Radiat Oncol Biol Phys* 1998;42:1015–1023.
24. Grigsby PW, Heydon K, Mutch DG, *et al.* Long-term follow-up of RTOG 92-10: Cervical cancer with positive para-aortic lymph nodes. *Int J Radiat Oncol Biol Phys* 2001;51:982–987.
25. Malfetano JH, Keys H. Aggressive multimodality treatment for cervical cancer with paraaortic lymph node metastases. *Gynecol Oncol* 1991;42:44–47.
26. Podczaski E, Stryker JA, Kaminski P, *et al.* Extended-field radiation therapy for carcinoma of the cervix. *Cancer* 1990;66: 251–258.
27. Kodaira T, Karasawa K, Shimizu T, *et al.* Clinical efficacy of applying four-field portals to paraaortic irradiation in the treatment of cervical carcinoma. *Radiat Oncol Investig* 1999;7: 170–177.
28. Thomas GM. Concurrent chemotherapy and radiation for locally advanced cervical cancer: The new standard of care. *Semin Radiat Oncol* 2000;10:44–50.
29. Sasaki Y, Amano T, Morita M, *et al.* Phase I study and pharmacological analysis of cis-diammine(glycolato)platinum (254-S; NSC 375101D) administered by 5-day continuous intravenous infusion. *Cancer Res* 1991;51:1472–1477.
30. Fuwa N, Shikama N, Hayashi N, *et al.* Treatment results of alternating chemoradiotherapy for nasopharyngeal cancer using cisplatin and 5-fluorouracil—a phase II study. *Oral Oncol* 2007; 43:948–955.

第III相臨床試験

Phase III clinical trials

小谷凡子　勝俣範之

Key words：第III相臨床試験，無作為化比較試験，層別因子，ITT解析

はじめに

がんの治療成績を向上させるためには，より有効な新薬および新治療法の開発が必須である．新しい医薬品や治療法の有効性や安全性を評価し，確立した治療法とするためには，幾つかの関門を通らなければならない．医学的な治験や基礎研究・前臨床試験の結果に基づき，人を対象として新たな治療法の性能を評価する方法論が臨床試験である．この臨床試験は，大きく分けて第I相，第II相，第III相試験からなる．

本稿では第III相臨床試験について取り上げる．

1. 第III相試験とは

第II相試験にて安全性，有効性が確認され，治験が終了した薬剤(併用も含む)や，既に単剤での安全性，有効性が確立されている薬剤では，単剤または他剤との併用において対象とする疾患の既存の標準治療との無作為化比較試験(randomized controlled trial: RCT)が行われる．これを第III相試験，またはcomparative treatment efficacy(CTE) trialsと呼ぶ．新たに開発・考案された治療を受ける被験者集団を試験群，標準治療を受ける被験者集団を対照群と呼ぶ．

第III相試験の主な目的は，上記2つの集団の間で，生存期間(あるいは無病期間や無増悪期間)について治療法を比較することである．試験は現在の標準治療との直接比較にあたるため，対照群には現在の標準治療を用いるのが原則である．

第III相試験の結果が得られたあとで行うべき判断は通常，①現在の標準治療を引き続き標準治療とするか，②新治療を新たな標準治療とするかのどちらかであり，これ以外の判断を得ることは目的としない．また，第III相試験で得られた結果を実際の臨床現場に適用する際には，専門病院のみに限らず，日常がん診療を行っている一般の市中病院にも適応されることが好ましい．そのため，一般に第III相試験を実施するにあたっては可能なかぎり，これらの一般の市中病院を含む多施設共同として実施すべきである．

2. 方　　法

a. 無作為化

選択基準・除外基準に基づいて選定された，試験参加に同意した患者(被験者)は，無作為に試験群か対照群かに割付けられ，比較対照する群が設定される．この操作を'無作為化，無作為割付け，ないしランダム化'という．試験薬と対照薬との真の有効性・安全性の差を証明するためには，試験薬剤(試験薬と対照薬)以外の予後因子(患者背景因子の中で結果に影響する

Namiko Kotani, Noriyuki Katsumata: Breast Cancer and Medical Oncology Division, National Cancer Center Hospital　国立がんセンター中央病院 乳腺・腫瘍内科

0047-1852/09/¥40/頁/JCLS

もの)が両群で均等であること，すなわち比較可能性の確保が重要である．

b. 層別因子

サンプルサイズが十分大きくないかぎり，無作為化はそれ自体では治療群の比較可能性を保証するには全く不十分である．小，中規模の試験では，重要な患者背景因子に偶然に大きなアンバランスが生じることは珍しくなく，試験結果の解釈を困難にさせる．したがって，最も重要な因子について適度にバランスをとることが求められる．バランスをとるために無作為化の手法に組み込まれる患者背景因子を層別因子と呼ぶ．層別因子は，アウトカムに強く関連することが知られているものでなくてはならない．参加施設が少ないときは，施設によって日常診療が異なるため，施設を層別因子とするのが最良である．しかし，あまりにも多くの層別因子が含まれると無作為化の手法ではバランスがとれなくなる．通常のがんの臨床試験での症例数程度であれば，層別因子は最大3つとするのが適当である．

c. 無作為化の方法

がん領域では重要な予後因子が既知であることが多く，また病期など重要な予後因子が与える予後への影響の方が，新治療に見込まれる効果よりも大きいことがあらかじめ期待される状況も珍しくない．したがって，得られた結果の信憑性を極端に劣らせないこと，および精度上昇を目的として，これら明らかな予後因子に関して群間バランスをとる工夫(割付け調整)を付加した無作為化を一般的に用いる．これを実現する方法として層別無作為化法と動的割付け法がある．

1) 層別無作為化法

最も一般的なものはランダムブロック法である．この方法では，割付け調整したい予後因子(割付け調整因子)ごとに層を作り，その層ごとに別々に無作為化を行う．割付け調整因子が複数存在する場合には，その組み合わせごとに層を作る．更に各層においても各群への割付け数が等しい大きさから4から10のブロックを順に並べることによって，群間バランスを保ちやすくすることが一般的である．このランダムブロック法は事前にすべての割付けリストが決まっており，変更されることはないという意味において静的割付け法に分類される．

2) 動的割付け法

その時点までに既にどのような予後因子をもつ対照がどちらの群により多く割付けされたかを表す割付け履歴に依存して，新たな割付けを行う割付け法を動的割付け法と呼ぶ．この動的割付け法の中で最も一般的なものは最小化法である．この方法は新たな登録者のもつ割付け調整因子に関して，その時点における偏りの程度を数値化したものに基づき現在生じている偏りを減じる方向に，新たな割付けがされやすくすることで逐次的に群間のアンバランスを改善するものである．つまり，ある割付け調整因子をもつ対象者が既に一方の群に偏って多く割付けられていた場合，同じ割付け調整因子をもつ新たな登録者に関しては他方の群に割付けられやすくなる．

偏りの程度の尺度として，単純には新たな登録者に等しい予後因子をもつ既登録者の予後因子ごとの群別登録数を求め，それを総和したもの(その時点までに偏って多く登録された群の総和がより大きくなる)の群間差を用いればよい．この際，偏りが減る方向に100%の確率で割付けを行う方法(強制割付け)は，将来の割付けに対する予測可能性および統計学的な観点から好ましくなく，通常は偏りが減る方向に50%より大きく100%よりは小さい確率(例：66%，80%など)で割付けする方法(バイアスコイン法)を用いる．

この最小化法は，前述したランダムブロック法と比べて，割付け調整因子が複数存在する場合に，その組み合わせごとの群間バランスを保つとは限らない点において劣る一方で，その逐次的な仕組みから登録者数が少ない状況においても群間バランスが相対的に保たれやすい点において優れている．

がん領域では，割付け調整因子として一般的によく用いられる施設ごとの登録患者数が少ないことや既登録者数が比較的少ない状況で実施

されやすい中間解析を試験デザインに含むことが一般的であるため，最小化法の方が特に好まれて用いられる傾向がある．

マッチングと同様に既知の予後因子に関して割付け調整できたとしても，未知の予後因子に関しては必ずしも群間バランスを保つことができない．過度に割付け調整するほど割付けにおける無作為性が失われてしまい，未知の予後因子に関しては逆にアンバランスが生じやすくなってしまうことに注意が必要である．

d. 無作為化のタイミング

無作為割付けの最もよいタイミングは，比較する治療の開始に可能なかぎり近い時期である．もし無作為割付けと治療の開始が離れていたら，治療開始前に死亡したり，全身状態が悪化したり，他の治療による合併症が生じたりして割付けたとおりの治療を受けない患者が多くなる．こうした患者を解析から除いてしまうと，両群は比較可能ではなくなる．

e. マスキング

二重盲験とは，患者も臨床医もどの治療が割り付けられたかを知らないことを意味する．一方，単盲験は患者のみが知らないことを指し，プラセボ対照はある群の治療のすべてもしくは一部が非活性の部分からなる場合に，すべての群の患者が見分けのつかない治療を受けることである．これらの試験には費用がかかる．プラセボの製造，実薬とプラセボへのコード番号付きのラベル貼付，その記録の保持，コード番号のみを用いた薬剤の配送，薬剤師との連携，医学的に緊急な状況におけるキーオープンの仕組みなどは，準備に多くの時間を要し，費用もかさむ．盲験試験が必要な状況は様々あるが，実行に移す前には慎重に考慮すべきである．プラセボ対照や二重盲験は，重要なエンドポイントが主観的なものである場合に必要になる．

3. 仮説検定，αエラー，βエラー，サンプルサイズ設計

a. 仮説検定

仮説検定は帰無仮説を否定することで対立仮説を採択するという一連の統計的手順である．例えば優越性試験では，一般的に群間差がないという帰無仮説を否定することにより新治療の標準治療に対する優越性，同様に非劣性試験では，一般的に群間差Δ以上であるという帰無仮説を否定することによって新治療が標準治療と比べてΔ以上は劣らないことがそれぞれ証明される．仮説検定ではp値が求められる．p値は帰無仮説が正しい確率を表すのではなく，帰無仮説が正しい場合に観察されたデータより等しいか極端なものが得られる確率を表す．事前に設定した有意水準(α)を観察されたデータから求めたp値が下回った場合には，統計的有意と判断して帰無仮説を否定できる一方で，これをp値が上回り統計的有意と判断できなかった場合には，そのまま否定されなかった帰無仮説が正しいという結論が得られないことに注意が必要である．また，p値の大きさは登録患者数や追跡期間に依存するという性質があるため，p値が非常に小さいからといって群間差が非常に大きいという解釈はすべきでない．

b. αエラー

真には帰無仮説が正しいのにこれを否定して対立仮説を採択してしまう誤りをαエラーと呼ぶ．このαエラーを犯す確率は事前に設定した有意水準に等しくなる．αエラーの場合に害を被るのは真にはよくない治療を受けることになる患者，ひいては国民であるため，新薬治験の場合は規制当局，それ以外の場合には当該研究領域の研究者コミュニティ全体の監視下で厳しく守られるべきである．αエラーは消費者リスクととらえることもできる．一般に第III相試験で設定されるαエラーの大きさは5%ないしは2.5%である．

c. βエラー，サンプルサイズ設計

真には対立仮説が正しいのに，帰無仮説を否定できないことでこの対立仮説を採択できない誤りをβエラーと呼ぶ．βエラーの場合には，本当はよい治療を受けられないことで患者ひいては国民が害を被ることは考慮すべきであるが，標準治療が既にある場合にはαエラーの場合よりも害は大きくない．よい治療を開発した製薬企業，よい治療法を考案した研究者が損をす

ることになるため，生産者リスクととらえることができ，研究実施主体で決めることになる．状況にもよるが，一般に第III相試験で設定されるβエラーはαよりも多少ゆるくして10–20%である．また，真には対立仮説が正しい状況で，これを正しく採択できる確率は検出力(power)と呼ばれる．βエラーを犯す確率をβとすると検出力は$1-\beta$となる．

臨床試験では臨床的に見過ごすことのできない対立仮説を具体的に設定し，それを確実に検出するのに必要なサンプルサイズを計画段階で考慮しておくのが普通である．この手続きをサンプルサイズの決定と呼ぶ．サンプルサイズに関係する3つの変数は，有意水準，検出力，検出したい群間差である．サンプルサイズは具体的には，コンピュータプログラムや数値表などから算出する．

生存期間をプライマリ・エンドポイントとする第III相試験では，厳密にはイベント数によって検出力の大きさが定まる．つまり，登録患者数を多くするだけでなく，各患者に予定される追跡期間を長くした場合にも検出力は一般に大きくなる．試験計画段階で適切なサンプルサイズ設計法を用いて，当該試験に必要十分な登録患者数，登録期間，追跡期間などを決めることは非常に重要である．登録患者数が不足し検出力を適切な範囲に保てない場合には，何の結論も出せない結果が出る可能性があることから倫理的な問題も生じうるため，試験実施の可否も含めて試験計画を再検討すべきである．登録患者数，追跡期間，解析の実施時期などはプロトコールに事前に明記しなければならない．

4. エンドポイント

試験の主たる結論を下す目的で測定されるエンドポイントをプライマリ・エンドポイントと呼ぶ．探索的な目的で実施される第I相および第II相試験では，評価すべき有効性(真のエンドポイント)を間接的に反映するであろう簡便な代替エンドポイントをプライマリ・エンドポイントとして用いることが一般的であるのとは異なり，検証的な目的で実施される第III相試験では，真のエンドポイントをプライマリ・エンドポイントとするべきである．多くのがん領域における真のエンドポイントは延命と症状コントロールである．がん領域では治療による延命効果を評価するために全生存期間(overall survival)をプライマリ・エンドポイントとすることが一般的である．これは一般的に無作為化時点を起点として他死因を打ち切りとして除外することなくすべての死因による死亡(全死亡)をイベントとして定義し，このイベントが発生するまでの時間を測定したものである．

ただし，早期乳癌患者など特に予後のよい集団を対象として試験を実施する場合，全生存期間をプライマリ・エンドポイントとしてしまうと，試験期間内に十分なイベント数を観察するために現実的でないほどの登録患者数を要する．このような領域に関しては，全死亡に加えて再発もイベントとする無再発生存期間(relapse-free survival: RFS)，あるいは全死亡に加えて増悪もイベントとする無増悪生存期間(progression-free survival: PFS)などをプライマリ・エンドポイントとして用いることも当該領域におけるコンセンサスとなっている．ただし，再発や増悪などをイベントとする場合，それらが確認されるために必要な検査や来院スケジュール・間隔が粗であるほど，結果として観察される無再発生存期間や無増悪生存期間は長くなるというバイアスが生じる．したがってこれらをエンドポイントとして設定する前に，この種のバイアスの入り方自体に結果に影響を与えるほどの群間差が存在するのか否か，もしも存在するならばそれを可能なかぎり減らすための方策を十分に検討すべきである．

5. 優越性試験と非劣性試験

'標準治療に比べてよい治療'は大きく2種類存在する．一つは新治療が'毒性も強いが延命効果も高い'というものであり，もう一つは新治療が'延命効果は同程度であるが毒性が軽い'というものである．前者を評価する場合には優越性試験，後者を評価する場合には非劣性試験がそれぞれ用いられる．

優越性試験では，標準治療よりも毒性が強い新治療はその毒性に見合う分以上にプライマリ・エンドポイントで優れなければならない．優越性試験の目的は標準治療と新治療がプライマリ・エンドポイントにおいて等しいことを否定できるか否かを検証することにあり，これを統計的有意に否定できた場合に限って新治療の優越性を証明することになる．

一方，非劣性試験では毒性の軽減に見合う分を越えてプライマリ・エンドポイントで劣ることは許容されない．したがって'これ以上劣ることが否定できればよい治療'であるといえる差Δ(非劣性許容下限)を非劣性試験で否定したい差として事前に設定することになる．非劣性試験の目的は，新治療が標準治療に比べてプライマリ・エンドポイントにおいて非劣性許容下限Δを越えて劣ることが否定できるか否かを検証することにあり，これを統計的有意に否定できた場合に限って新治療がΔ以上は劣らないことを証明することになる．

6. 解　　析

a. ITT(intention-to-treat)解析

一般に，検証を目的とした第III相試験では，ITT解析に従って試験の主たる解析を行う．ITT解析とは'治療を意図した割付けに基づいた解析'という意味がある．脱落例や除外例などの不完全症例を解析から外さず，割付け時の群構成に基づいて解析する方策を指す．

b. 中間解析

予定症例数に達しない段階で，試験の継続の可否を検討するために行う解析を中間解析と呼ぶ．解析の結果，治療効果に明らかな差がある，毒性が極めて強いなどの問題が明らかになった場合は試験を中止する場合がある．中間解析の実施時期と回数はあらかじめプロトコールに定めておき，予定外の解析は行わない．中間解析を行うことは試験期間中に複数回繰り返して群間比較の検定を実施することに対応し，これによって生じる多重性の問題により，真には群間差が存在しない状況でも偶然に誤って有意な結果が得られてしまう危険性がより高まる．この多重性に対する配慮として，①事前に定めた少数回の中間解析を実施すること，②経時的な多重性の調整を行うことである．当該試験を実施する研究者が未成熟な中間解析の結果を知ることでその後の試験運営に甚大な影響が与えられる可能性がある．中間解析を実施するにあたり，研究者とは独立な組織(独立データモニタリング委員会)を構成し，この組織外には試験の途中結果を開示しないまま，試験の継続可否も含めて審査されることが一般的となっている．

c. 生存時間解析

がんにおける臨床試験の主なアウトカムは患者の生存時間である．患者は無作為に割付けられ，死亡まで追跡される．第III相試験では患者集積期間があり，その後更にデータの解析時点までの追跡期間がある．最終解析の時点で生存している患者の観察された時間は，全体の集積期間のどの時点で患者が試験に登録されたかによる．こうした患者の本当の生存時間は不明である．この種のデータでは被験者が打ち切りを受けたと表現する．生存時間を評価するにあたってはこの打ち切りの発生を適切に考慮できる特殊な統計解析手法が必要とされる．

理想的には，患者の情報は可能なかぎり最大限に活用すべきである．臨床試験における患者の生存時間を推定するために最もよく用いられる方法はカプラン・マイヤー推定量である．すべての観察時点について推定累積生存割合を計算したものである．

特定の時点での生存割合のほかにも，よく用いられる統計量として生存時間中央値(50％生存時間)がある．これは半数の患者が死亡すると推定される時間である．これはカプラン・マイヤー法で推定した生存曲線が最初に0.50以下になる時間である．

群間比較を行う際にはログランク検定と呼ばれる検定法が一般的である．この検定はある特定の時点での両群の生存割合や生存期間中央値を比較するものではなく，正しくは両群の生存曲線全体を比較するものである．

文　献

1) 福田治彦ほか(訳)：米国SWOGに学ぶがん臨床試験の実践，医学書院，2004.
2) 椿　弘計ほか：これからの臨床試験，朝倉書店，1999.
3) 吉村健一：新臨床腫瘍学(日本臨床腫瘍学会編)，p127-132，南江堂，2006.
4) Piantadosi S: Clinical Trials: A Methodologic Perspectives, John Wiley & Sons, New York, 1997.
5) Simon R: Design and analysis of clinical trials. In: Cancer, Principles & Practice of Oncology, 7th ed (ed by DeVita VT, et al), p478-489, Lippincott Williams & Wilkins, Philadelphia, 2005.

Jpn J Clin Oncol 2009;39(7)443–448
doi:10.1093/jjco/hyp036

Randomized Controlled Study Comparing Two Doses of Intravenous Granisetron (1 and 3 mg) for Acute Chemotherapy-induced Nausea and Vomiting in Cancer Patients: A Non-inferiority Trial

Masahito Yonemura[1], Noriyuki Katsumata[2], Hironobu Hashimoto[1], Shoko Satake[1], Masayuki Kaneko[3], Yuka Kobayashi[1], Atsuo Takashima[2], Yasuhisa Kato[1], Masahiro Takeuchi[3], Yasuhiro Fujiwara[2], Hiroshi Yamamoto[1] and Taisuke Hojo[1]

[1]Department of Pharmacy, National Cancer Center Hospital, [2]Breast and Medical Oncology Division, National Cancer Center Hospital and [3]Division of Biostatistics, School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan

Received December 21, 2008; accepted March 18, 2009; published online April 24, 2009

Objective: The aim of this study was to assess the non-inferiority of 1 mg to 3 mg granisetron (GRN) injection for the treatment of acute chemotherapy-induced nausea and vomiting (CINV) and to evaluate the tolerability of GRN given at 1 mg in Japanese cancer patients.
Methods: Patients with cancer receiving highly emetogenic chemotherapy were enrolled in this single-blind randomized controlled study. Patients were randomly assigned to receive GRN at a single dose of 1 or 3 mg. The primary endpoint was the rate of complete protection from emetic events (no vomiting, no retching and no need for rescue medication) during the first 24 h following the initiation of chemotherapy.
Results: There were 89 patients in the 1 mg group and 90 patients in the 3 mg group. Complete protection was achieved in 70 patients (78.7%) in the 1 mg group and 73 (81.1%) patients in the 3 mg group. The one-sided test did not reveal non-inferiority of either dose of GRN to the other at a 5% significance level.
Conclusions: Our data failed to show the non-inferiority of 1 mg of GRN to 3 mg of GRN administered as a single dose. However, the rate of complete protection from nausea and vomiting was similar in the two groups. Given the recommended dosage in the guidelines and the economic need for reduction of medical care expenses in Japan, prophylactic administration of GRN at 1 mg may be an appropriate, alternative treatment for acute CINV in cancer patients.

Key words: granisetron – serotonin antagonist – antiemetic – vomiting – non-inferiority trial

INTRODUCTION

Vomiting is one of the most frequently encountered non-hematologic toxicities of cancer chemotherapy. Severe vomiting can lead to problems such as anorexia, dehydration, malnutrition and electrolyte abnormalities, which may lead to refusal of chemotherapy and poor compliance, as well as difficulty in continuing treatment (1,2). The incidence of chemotherapy-induced nausea and vomiting (CINV) depends on the type, dose and administration route of anticancer drugs. For instance, 60–90% of patients receiving carboplatin (a platinum anticancer drug) or doxorubicin (an anthracycline anticancer drug) (>60 mg/m^2) and 90% of patients receiving cisplatin (>50 mg/m^2) exhibit acute emesis (3). Association between the 5-HT$_3$ receptor and CINV was first reported in the late 1980s, and 5-HT$_3$ receptor antagonists began to be applied as antiemetics in the clinical setting from the 1990s. A meta-analysis showed that the risk of CINV associated with cisplatin treatment is reduced to a greater extent by 5-HT$_3$ antagonists than by conventional antiemetics such as dopamine receptor antagonists and antihistamines (4); thus, 5-HT$_3$ antagonists are now the drugs of first choice for the prevention of CINV.

For reprints and all correspondence: Masahito Yonemura, Department of Pharmacy, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. E-mail: myonemur@ncc.go.jp

In the USA and Europe, evidence-based antiemetic treatment guidelines have been established by the American Society of Clinical Oncology (ASCO) (5), National Comprehension Cancer Network (NCCN) (6) and European Society of Medical Oncology (ESMO) (7). These guidelines recommend administration of granisetron (GRN) at the dose of 1 mg or 10 μg/kg i.v. or 2 mg orally. In order to determine the optimal effective dose of GRN for the prevention or treatment of CINV in Japanese patients, Furue et al. (8) administered GRN at the dose of 20, 40 or 80 μg/kg once a day and reported that 40 μg/kg administered i.v. once a day was the most appropriate. Therefore, the approved dose of GRN in Japan is set at 40 μg/kg (3 mg i.v.), which differs substantially from the recommendation in the USA and Europe. However, we believe that the results of the aforementioned study could be related to the ambiguous criteria used for defining nausea. We hypothesized that 10 μg/kg of GRN would exhibit equivalent antiemetic efficacy to 40 μg/kg, the approved dose in Japan, and compared the efficacy and safety of 1 mg and 3 mg of GRN from the point of view of clinical rationality.

PATIENTS AND METHODS

This study was a single institutional, single-blind, randomized controlled study conducted to assess whether GRN used at the dose of 1 mg might be non-inferior to the drug used at the dose of 3 mg in regard to complete protection from emetic events. The participants were patients with cancers who were scheduled to undergo chemotherapy and were stratified into the high or moderate emetic risk groups for CINV according to the ASCO guidelines for antiemetic treatment (2006). The study was approved by the Institutional Review Board of the National Cancer Center Hospital. In accordance with a statement from the International Committee of Medical Journal Editors (ICMJE), the study was registered in the University Hospital Medical Information Network (UMIN000000304).

Chemotherapy Schedule

The chemotherapy was performed according to the following schedule. DC: docetaxel 75 mg/m^2 on day 1 and carboplatin AUC = 5 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; TC: paclitaxel 175 mg/m^2 on day 1 and carboplatin AUC = 5–6 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; weekly TC: paclitaxel 80 mg/m^2 on day 1 and carboplatin AUC = 2 on day 1 every week, dexamethasone 8 mg/body on day 1; AP: adriamycin 60 mg/m^2 on day 1 and cisplatin 50 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1 and 20 mg/body on days 2 and 3; CBDCA/CPT-11: carboplatin AUC = 5 on day 1 and irinotecan 150 mg/m^2 on days 1, 8 and 15 every 3 weeks, dexamethasone 24 mg/body on day 1; CDDP/CPT-11: cisplatin 60 mg/m^2 on day 1 and irinotecan 60 mg/m^2 on days 1, 8 and 15 every 4 weeks, dexamethasone 24 mg/body on day 1, 8 mg/body on days 2 and 3, 4 mg/body on day 4 and 2 mg/body on day 5; CDDP/GEM: cisplatin 70 mg/m^2 on day 1 and gemcitabine 1000 mg/m^2 on days 1, 8 and 15 every 4 weeks, dexamethasone 16 mg/body on day 1, 8 mg/body on day 2, 4 mg/body on day 3 and 2 mg/body on days 4 and 5; AC: adriamycin 60 mg/m^2 on day 1 and cyclophosphamide 600 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; CEF: cyclophosphamide 500 mg/m^2 on day 1, epirubicin 100 mg/m^2 on day 1 and fluorouracil 500 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1. The following is a single administration. Carboplatin: AUC = 6 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; adriamycin: 60 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; cisplatin: 80 mg/m^2 on day 1 every 4 weeks, dexamethasone 24 mg/body on day 1 and 8 mg/body on days 2 and 3.

Eligibility Criteria

Patients admitted between January and October 2006 meeting all the inclusion but not falling under any of the exclusion criteria were informed about the study and requested to sign a written consent form. Eligible patients were at least 20 years old and were under treatment with neoplastic agents associated with the high or moderate emetic risk of acute emesis, patients with PS (Eastern Cooperative Oncology Group performance status scale) 0–2 and those meeting each of the following laboratory findings, examined within 3 weeks prior to registration for the study (alanine aminotransferase 100 IU/l lower, creatinine 2.5 mg/dl lower, absolute neutrophil count 1000/μl upper). The ineligibility criteria were known hypersensitivity to 5-HT3 receptor antagonists, treatment with neoplastic agents of the high or moderate emetic risk group for CINV from days 2 to 7 and treatment with the radiation therapy from days 2 to 7. Serious complications, except malignancy (e.g. bowel obstruction, lung fibrosis, cerebrovascular accident, active gastric and duodenal ulcer), inability to precisely record the episodes in a diary were also the ineligibility criteria. The patients were randomly assigned to two treatment arms using the minimization method with correction, including for treatment with a cisplatin-based or non-cisplatin-based regimen and positive/negative history of prior use of the test drug. The randomization of the patients was performed at a participant registry center established at the Division of Biostatistics, School of Pharmaceutical Sciences, Kitasato University. In this study, only patients were blinded to the knowledge of whether they were receiving 1 or 3 mg of GRN. Researchers asked the patients directly or by telephone about whether they experienced any emetic events within 24–36 h following the start of administration of the chemotherapeutic agents that were classified into the high or moderate emetic risk group. Furthermore, patients were asked to record their symptoms for 6 days on a diary card,

and the cards were collected at each visit. Adverse events were evaluated based on the CTCAE v3.0 (JCOG/JSCO Japanese version) (9).

The primary endpoint was the rate of complete response (CR). The rate was complete protection from emetic events (vomiting, retching and need for rescue medications) during the first 24 h following the start of administration of the chemotherapeutic agents classified as high-emetic-risk agents. The secondary endpoints were: rate, the rate was complete protection from nausea and emetic events (no or mild nausea, no vomiting, no retching, no need for rescue medications or premature withdrawals), time-to-treatment success, number of emetic episodes, severity of nausea and severity of adverse events.

Definitions of the Efficacy Parameters

Emetic episodes were defined as vomiting or retching. A vomiting episode was considered to have ended when retching or vomiting had ceased for at least 1 min. One or more retching episodes within a 5 min period were defined as one emetic episode. Retching associated with vomiting within a minute interval was defined as one emetic episode within a 5 min period.

Episodes of nausea were recorded by the patients on diary cards, along with the severity of the episodes according to the following four-point scale: 0, none (no nausea); 1, mild (able to take meals as usual); 2, moderate (reduced intake of food) and 3, severe (unable to take either food or water).

Rescue antiemetic medications were defined as follows: the medication for emetic events following chemotherapy that had not previously been prescribed, or temporary medication according to the physical condition in particular patients. Temporary medications were included in rescue antiemetic medications. These medications were used when emetic events or nausea occurred, or the patients desired treatment for these symptoms. Any type or doses of antiemetic agents could be used. Detailed information regarding the use of rescue antiemetic medications, including the date of administration, was recorded when these agents were used. The time of the first rescue antiemetic medication was also recorded. In addition, patients were asked to record any drugs taken on their own judgment.

Statistics

The Italian Group for Antiemetic Research reported a rate of complete protection of 92.6% in patients receiving GRN plus dexamethasone for the prevention of emesis and nausea caused by cyclophosphamide, doxorubicin, epirubicin or carboplatin (10). On the basis of this finding, we assumed that the CR rates in both the test group (1 mg of GRN) and the control group (3 mg of GRN) would be 92.6%, with a non-inferiority margin (Δ) of 10%. The required sample size was calculated as 166 patients (83 per group) assuming α value = 0.05 and β value = 0.8. In addition, with an anticipated loss of approximately 14 excluded patients, the target sample size was set at 180 patients (90 per group).

The analysis was performed for all the randomized patients on an intent-to-treat basis. The patients' demographics (gender, age, type of antineoplastic agents, history of treatment with the target regimen and performance status) were collected to compare the distributions across the treatment groups. The data were analyzed by Fisher's exact test (gender, type of neoplastic agents and history of treatment with the target regimen), *t*-test (age) or the χ^2 test (performance status). For the rate of complete protection from emesis and the rate of complete protection from nausea, non-inferiority of the test group to the control group was tested by the Dunnett and Gent test with a 10% non-inferiority margin. The log-rank test was used to analyze the time-to-treatment success, which was defined as the time from the start of the high or moderate emetic risk chemotherapeutic agents to the first emetic episode and use of rescue medication. The frequency of vomiting episodes was compared using the χ^2 test. In regard to the safety variables, Grade 3 or more severe non-hematologic toxicities were evaluated. All tests were one-sided, with the statistical significance set at a P value of <0.05. The two-sided 95% confidence interval was estimated.

RESULTS

Baseline Demographics

A total of 182 patients were randomized to the GRN 1 mg group ($n = 90$) or 3 mg group ($n = 92$). Of these, one patient (Patient 106) in the 1 mg group and two patients (Patients 83 and 159) in the 3 mg group withdrew their consent after the randomization. Therefore, 89 patients in the 1 mg group and 90 patients in the 3 mg group were included in the full analysis set (Fig. 1).

Patient Characteristics

The characteristics of the patients in the two treatment groups were similar (Table 1). Elderly women were somewhat more likely to be included in the GRN 1 mg group; therefore, that group was slightly disadvantaged at the primary endpoint. The most commonly reported primary cancers in all the treatment groups were: breast cancer ($n = 94$), gynecologic cancer (cervical, endometrial and ovarian cancer) ($n = 64$), primary unknown cancer ($n = 16$), urothelial cancer ($n = 4$) and sarcoma ($n = 3$).

Efficacy Analysis

Primary Efficacy Endpoint: CR Rate

Table 2 shows the proportion of patients in whom complete protection from emetic events was achieved (no vomiting, no

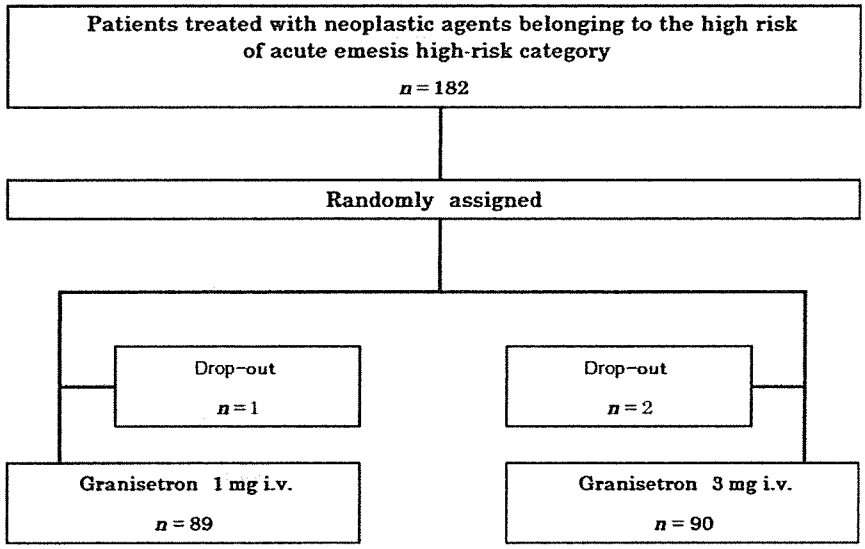

Figure 1. Patients' random assignment flow chart.

retching and no rescue medications) during the first 24 h following the start of chemotherapy, with the 95% confidence intervals. The one-sided, non-inferiority test at a 5% significance level with a non-inferiority margin of 10% failed to show the non-inferiority of the 1 mg (test dose) dose to the 3 mg (control dose) dose ($P = 0.103$).

Secondary Efficacy Endpoint: CR Rate for Nausea and Vomiting

Table 3 shows the proportion of patients in whom complete protection from nausea and emetic events was achieved (no vomiting, no retching, no rescue medications and no Grade 2 or more severe nausea) during the first 24 h following the initiation of chemotherapy, with the 95% confidence intervals. The one-sided, non-inferiority test at a 5% significance level with a non-inferiority margin of 10% failed to show the non-inferiority of the test group to the control group ($P = 0.108$).

Time to Start of Vomiting

Episodes of vomiting were observed in eight and six patients in the 1 and 3 mg groups, respectively. The log-rank test showed no statistically significant difference in the time to start of vomiting between the two groups ($P = 0.554$).

Frequency of Vomiting and Retching

Table 4 shows the distribution of the frequencies of vomiting and retching, respectively. The χ^2 test showed the absence of any statistically significant difference in the frequencies of vomiting and retching between the groups ($P = 0.666$ and 0.609, respectively).

Safety Analysis

In this study, only four patients exhibited Grade 3 or more severe non-hematologic toxicities, as follows: Grade 3 anorexia ($n = 1$) and Grade 3 dehydration ($n = 1$) in the 1 mg group and Grade 3 syncope ($n = 1$) and Grade 3 general malaise ($n = 1$) in the 3 mg group. The investigator did not consider any of these events to be related to GRN treatment.

DISCUSSION

In this study, we could not show the non-inferiority of GRN 1 mg to 3 mg; however, the difference in the rate of complete protection from emesis between the two GRN dose groups was only 2.4%. Thus, the failure to show the non-inferiority of the 1 mg dose might be mainly attributable to the lack of sufficient statistical power of the analysis arising from the small sample size. Prior to the start of our study, we expected that the number of patients receiving the target regimens of AC therapy and carboplatin-based chemotherapy would be larger than that of those receiving other regimens. On the basis of this expectation and the results of a study conducted overseas using the above regimen, we assumed that the CR rate with respect to emesis would be 92.6% when calculating the sample size. In this study, patients treated with rescue medications were not included as drop-outs, but as patients not showing CR; therefore, the actual CR rate might have been smaller than that estimated when calculating the sample size. Thus, the non-inferiority could not be proved statistically, even though the difference in the CR between the two dose groups was small.

The first study on the effects of GRN has demonstrated the absence of a significant difference in the drug efficacy among groups treated with doses of 40 or 160 μg/kg in the UK (11–12). Meanwhile, the approved dose of GRN in Japan remains

Table 1. Patient characteristics ($n = 179$)

	1 mg ($n = 89$)	3 mg ($n = 90$)	P value
Gender [n (%)]			
Female	88 (98.9)	86 (95.6)	0.368†
Male	1 (1.1)	4 (4.4)	
Age			
Median	54	58	0.135‡
Range	27–79	23–80	
Type of neoplastic agents[a] [n (%)]			
CDDP included	7 (7.9)	6 (6.7)	0.782†
Others[b]	88 (98.9)	84 (93.3)	
First cycle [n (%)]			
Yes	36 (40.5)	37 (41.1)	1†
No	53 (59.6)	53 (58.9)	
PS [n (%)]			
0	86 (96.6)	87 (96.7)	0.99#
1	2 (2.3)	2 (2.2)	
2	1 (1.1)	1 (1.1)	

[a]The chemotherapy was performed according to the following schedule. DC: docetaxel 75 mg/m^2 on day 1 and carboplatin AUC = 5 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; TC: paclitaxel 175 mg/m^2 on day 1 and carboplatin AUC = 5–6 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; weekly TC: paclitaxel 80 mg/m^2 on day 1 and carboplatin AUC = 2 on day 1 every week, dexamethasone 8 mg/body on day 1; AP: adriamycin 60 mg/m^2 on day 1 and cisplatin 50 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1 and 20 mg/body on days 2 and 3; CBDCA/CPT-11: carboplatin AUC = 5 on day 1 and irinotecan 150 mg/m^2 on days 1, 8 and 15 every 3 weeks, dexamethasone 24 mg/body on day 1; CDDP/CPT-11: cisplatin 60 mg/m^2 on day 1 and irinotecan 60 mg/m^2 on days 1, 8 and 15 every 4 weeks, dexamethasone 24 mg/body on day 1, 8 mg/body on days 2 and 3, 4 mg/body on day 4 and 2 mg/body on day 5; CDDP/GEM: cisplatin 70 mg/m^2 on day 1 and gemcitabine 1000 mg/m^2 on days 1, 8 and 15 every 4 weeks, dexamethasone 16 mg/body on day 1, 8 mg/body on day 2, 4 mg/body on day 3 and 2 mg/body on days 4 and 5; AC: adriamycin 60 mg/m^2 on day 1 and cyclophosphamide 600 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; CEF: cyclophosphamide 500 mg/m^2 on day 1, epirubicin 100 mg/m^2 on day 1 and fluorouracil 500 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1. The following is a single administration. Carboplatin: AUC = 6 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; adriamycin: 60 mg/m^2 on day 1 every 3 weeks, dexamethasone 24 mg/body on day 1; cisplatin: 80 mg/m^2 on day 1 every 4 weeks, dexamethasone 24 mg/body on day 1 and 8 mg/body on days 2 and 3.
[b]Other regimen of the drug includes the carboplatin, cyclophosphamide, doxorubicin and epirubicin.
Data were analyzed by the †Fisher's exact test, the ‡t-test or the #χ^2 test.

Table 2. CR to the prophylactic therapy against chemotherapy-induced vomiting and to rescue the agents

	Sample	CR of vomiting [95% CI]	Non-CR
1 mg	89	70 (78.7%) [68.7–86.6]	19
3 mg	90	73 (81.1%) [71.5–88.6]	17

The one-sided non-inferiority test with a 5% significance level did not prove non-inferiority of the 1 mg group to the 3 mg group ($P = 0.103$).

Table 3. CR to the prophylactic therapy against chemotherapy-induced vomiting and nausea and to rescue the agents

	Sample	CR of nausea and vomiting [95% CI]	Non-CR
1 mg	89	69 (77.5%) [67.4–85.7]	20
3 mg	90	72 (80.0%) [70.3–87.7]	18

The one-sided non-inferiority test with a 5% significance level did not prove non-inferiority of the 1 mg group to the 3 mg group ($P = 0.108$). CR, complete response; CI, confidence interval.

Table 4. Frequency of vomiting during the 5-day period after chemotherapy

	Number of patients	Frequency of vomiting					
		0	1	3	6	7	
1 mg	89	81 (91.0%)	3	3	1	1	
3 mg	90	84 (93.3%)	3	1	2	0	
		Frequency of retching					
		0	1	2	3	4	5	10
1 mg	89	78 (87.6%)	3	4	1	1	2	0
3 mg	90	79 (87.7%)	5	4	1	0	0	1

There was no statistical difference between the groups ($P = 0.666$). Frequency of retching during the 5-day period after chemotherapy. There was no statistical difference between the groups ($P = 0.609$).

40 μg/kg, possibly based on the approved dose in the UK in 1991. This dose of 40 μg/kg has been approved in Japan based on the results of dose-finding study. In the analysis of this trial included not only the objective data of the frequency of emetic episodes, but also the frequency of nausea, which was a subjective variable applied, so that the results depended substantially on the investigators' judgment. Although many studies (13–15) have reported a relationship between the dose and the effectiveness of GRN, the dose of 10 μg/kg or 1 mg/body has been approved in the USA. Furthermore, GRN at the dose of 10 μg/kg or 1 mg/body is recommended in the guidelines of the ASCO, ESMO and NCCN. However, the approved dose in Japan remains unchanged. There is a growing global consensus that the doses of antiemetic agents should be minimized to achieve the desired efficacy. Hence, we conducted this study in the hope of achieving efficient use of antiemetic medications in Japan. Physicians in Japan use relatively higher doses of GRN, and the possible medical economic benefit that can be expected with avoidance of the excessive use of these medications has been estimated. For example, we calculated the consumption and purchase price of GRN 3 mg, which has been used at our hospital in 2007. The consumption was 18 455 ampoules each year. The price of each 3 and 1 mg ampoule for injection was 7177 and 3015 yen in 2007. Thus, if GRN 3 mg were switched to GRN 1 mg, the difference in the purchase price annually would be 76 809 710 yen.

Our results, based on only the objective parameter of complete protection from emetic episodes, showed that the CR rate was similar between the GRN 1 mg and the GRN 3 mg groups. Given the need for promoting efficient use of the limited medical resources and for stemming the rising medical costs in Japan, prophylactic administration of GRN at 1 mg may be the appropriate choice, not expected to be associated with any significant problems. GRN has already been established at a high position among the 5-HT_3 receptor antagonists. Nevertheless, one study indicated a possibly higher incidence of constipation in the 5-HT_3 receptor antagonist treatment group than in the metoclopramide treatment group (16); therefore, the minimum effective dose of the 5-HT_3 receptor antagonist should be recommended in the clinical setting. Recently, aprepitant has been newly developed as a neurokinin receptor antagonist, and combined administration of this agent with a 5-HT_3 receptor antagonist and dexamethasone has been recommended by the ASCO guideline for antiemetics treatment (2006). But as it is an unapproved drug in Japan, we cannot use the aprepitant. Therefore, we should discuss based on the ASCO guideline for antiemetics treatment (1999) now. If the expensive new drug (aprepitant) were approved, in our country, we would readily understand of rising medical costs with such a newly launched drug. As a result, we have to promote more efficient use of the drugs.

CONCLUSION

Our data failed to show the non-inferiority of GRN 1 mg to GRN 3 mg. However, considering the recommendation by the ASCO, ESMO and NCCN guidelines for the administration of GRN at the dose of 1 mg or 10 μg/kg and the economic need for reduction of medical care expenses in Japan, and also the lack of statistical power of the analysis in this study, prophylactic administration of GRN at 1 mg may be the appropriate choice for cancer patients receiving highly emetogenic chemotherapy in Japan.

Acknowledgements

We thank all the participating patients. We also thank Masahiro Koseki and Mariko Hochi of the Division of Biostatistics, School of Pharmaceutical Sciences, Kitasato University, for the data analysis, Masashi Ando, Chikako Shimizu, Tsutomu Kohno and Kan Yonemori of the Breast and Medical Oncology Division, National Cancer Center Hospital, for clinical management of the patients.

Funding

This research was supported by a grant from the Ministry of Health, Labour and Welfare, Japan.

Conflict of interest statement

None declared.

References

1. Mitchell EP. Gastrointestinal toxicity of chemotherapeutic agents. *Semin Oncol* 1992;19:566–79.
2. Richardson JL, Marks G, Levine A. The influence of symptoms of disease and side effects of treatment on compliance with cancer therapy. *J Clin Oncol* 1988;6:1746–52.
3. Hesketh PJ, Kris MG, Grunberg SM, Beck T, Hainsworth JD, Harker G, et al. Proposal for classifying the acute emetogenicity of cancer chemotherapy. *J Clin Oncol* 1997;15:103–9.
4. Jantunen IT, Kataja VV, Muhonen TT. An overview of randomised studies comparing 5-HT3 receptor antagonists to conventional anti-emetics in the prophylaxis of acute chemotherapy-induced vomiting. *Eur J Cancer* 1997;33:66–74.
5. Gralla RJ, Osoba D, Kris MG, Kirkbride P, Hesketh PJ, Chinnery LW, et al. Recommendations for the use of antiemetics: evidence-based, clinical practice guidelines. *J Clin Oncol* 1999;17:2971–94.
6. NCCN Clinical Practice Guidelines in Oncology: Antiemesis, Ver1. 2005, http://www.nccn.org/professionals/physician_gls/PDF/antiemesis-Japanese.pdf.
7. Herrstedt J, Aapro MS, Roila F, Kataja VV. ESMO Minimum Clinical Recommendations for prophylaxis of chemotherapy-induced nausea and vomiting (NV). *Ann Oncol* 2005;16(Suppl 1):i77–9.
8. Furue H, Oota K, Taguchi T, Niitani H. Clinical evaluation of granisetron against nausea and vomiting induced by anticancer drugs (I). *Rinsho Iyaku* 1990;6:49–61 (in Japanese).
9. Japanese translation of common terminology criteria for adverse events (CTCAE), and instructions and guideline. *Int J Clin Oncol* 2004;9(Suppl 3):1–82 (in Japanese).
10. The Italian Group for Antiemetic Research. Dexamethasone, granisetron, or both for the prevention of nausea and vomiting during chemotherapy for cancer. *N Engl J Med* 1995;332:1–5.
11. Soucop M. A comparison of two dose levels of granisetron in patients receiving high-dose cisplatin. *Eur J Cancer* 1990;26:S15–9.
12. Smith IE. A comparison of two dose levels of granisetron in patients receiving moderately emetogenic cytostatic chemotherapy. *Eur J cancer* 1990;26:S19–23.
13. Navari RM, Kaplan HG, Gralla RJ, Grunberg SM, Palmer R, Fitts D. Efficacy and safety of granisetron, a selective 5-hydroxytryptamine-3 receptor antagonist, in the prevention of nausea and vomiting induced by high-dose cisplatin. *J Clin Oncol* 1994;12:2204–10.
14. Riviere A. Dose finding study of granisetron in patients receiving high-dose cisplatin chemotherapy. *Br J Cancer* 1994;69:967–71.
15. Navari RM, Gandara DR, Hesketh PJ, et al. Comparative clinical trial of granisetron and ondansetron in the prophylaxis of cisplatin-induced emesis. *J Clin Oncol* 1995;13:1242–8.
16. Aapro MS, Thuerlimann B, Sessa C, De Pree C, Bernhard J, Maibach R. A randomized double-blind trial to compare the clinical efficacy of granisetron with metoclopramide, both combined with dexamethasone in the prophylaxis of chemotherapy-induced delayed emesis. *Ann Oncol* 2003;14:291–7.

Support Care Cancer (2009) 17:1483–1491
DOI 10.1007/s00520-009-0613-7

ORIGINAL ARTICLE

Feasibility and validity of the Patient Neurotoxicity Questionnaire during taxane chemotherapy in a phase III randomized trial in patients with breast cancer: N-SAS BC 02

Kojiro Shimozuma · Yasuo Ohashi · Ayano Takeuchi · Toshihiko Aranishi · Satoshi Morita · Katsumasa Kuroi · Shozo Ohsumi · Haruhiko Makino · Hirohumi Mukai · Noriyuki Katsumata · Yoshihide Sunada · Toru Watanabe · Frederick H. Hausheer

Received: 1 October 2008 / Accepted: 8 March 2009 / Published online: 28 March 2009

Abstract
Goals The aim of the study was to determine the feasibility and validity of a newly developed patient-based instrument—the Patient Neurotoxicity Questionnaire (PNQ)—for grading chemotherapy-induced peripheral neuropathy (CIPN).

Patients and methods We prospectively collected data from 300 female patients who were treated with taxane chemotherapy for primary breast cancer as part of a national multicenter phase III randomized trial (N-SAS BC 02). We evaluated patient compliance with the PNQ and several

K. Shimozuma (✉)
Department of Biomedical Sciences,
College of Life Sciences, Ritsumeikan University,
1-1-1 Noji Higashi,
Kusatsu, Shiga 525-8577, Japan
e-mail: k-shimoz@sk.ritsumei.ac.jp

Y. Ohashi · A. Takeuchi · T. Aranishi
Department of Biostatistics, School of Public Health,
The University of Tokyo, 7-3-1 Hongo,
Bunkyo-ku, Tokyo 113-0033, Japan

Y. Ohashi
e-mail: ohashi@epistat.m.u-tokyo.ac.jp

S. Morita
Department of Biostatistics and Epidemiology,
Yokohama City University Medical Center,
4-57 Urafune-cho, Minami-ku,
Yokohama, Kanagawa 232-0024, Japan
e-mail: smorita@urahp.yokohama-cu.ac.jp

K. Kuroi
Division of Clinical Trials Research,
Department of Surgery,
Tokyo Metropolitan Komagome Hospital, 3-18-22 Hokomagome,
Bunkyo-ku, Tokyo 113-8677, Japan
e-mail: kurochan@dd.iij4u.or.jp

S. Ohsumi
Department of Breast Oncology, National Hospital Organization
Shikoku Cancer Center, 160 Ko, Minami-umemoto-machi,
Matsuyama, Ehime 791-0245, Japan
e-mail: sosumi@shikoku-cc.go.jp

H. Makino
Division of Breast Oncology, Niigata City General Hospital,
463-7 Shumoku, Chuo-ku,
Niigata 950-1197, Japan
e-mail: makino@hosp.niigata.niigata.jp

H. Mukai
Chemotherapy Division, National Cancer Center Hospital East,
6-5-1 Kashiwanoha,
Kashiwa, Chiba 277-8577, Japan
e-mail: hrmukai@east.ncc.go.jp

N. Katsumata
Breast and Medical Oncology Division,
National Cancer Center Hospital, 5-1-1, Tsukiji,
Chuo-ku, Tokyo 104-0045, Japan
e-mail: nkatsuma@ncc.go.jp

Y. Sunada
Department of Neurology, Kawasaki Medical School,
577 Matsushima,
Kurashiki, Okayama 701-0192, Japan
e-mail: ysunada@med.kawasaki-m.ac.jp

T. Watanabe
Hamamatsu Oncology Center, 3-6-13 Chuo Naka-ku,
Hamamatsu, Shizuoka 430-0929, Japan
e-mail: twatanab@oncoloplan.com

F. H. Hausheer
BioNumerik Pharmaceuticals Inc.,
8122 Datapoint Drive, Suite 1250,
San Antonio, TX 78229, USA
e-mail: fred.hausheer@bnpi.com

validation parameters, including concordance between CIPN grades noted by physicians (National Cancer Institute Common Toxicity Criteria) and patients (PNQ), and the concurrent validity and responsiveness of the PNQ versus the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (FACT/GOG-Ntx) utilizing data at pre-treatment and before three, five, and seven treatment cycles.

Main results The questionnaire completion rate was >90% at all assessments. Evaluation by physicians always resulted in lower neuropathy assessment scores compared with those reported directly by patients (weighted kappa coefficients, 0.02–0.06). Both PNQ sensory and motor scores were significantly correlated with the FACT/GOG-Ntx (r=0.66 and 0.51, respectively). In the repeated measures analysis of variance model, PNQ grades increased considerably as treatment continued, indicating progressively worsening CIPN over time.

Conclusions The PNQ has an applicable degree of feasibility and validity, useful for the diagnosis of CIPN as well as for clinical treatment decision-making, where the development of CIPN is a potential treatment-limiting consideration. Physicians underreport and underestimate the severity of CIPN symptoms compared with patients, thereby supporting the importance of assessing patient-reported outcomes using the PNQ.

Keywords Neurotoxicity · Patient Neurotoxicity Questionnaire (PNQ) Validation · Patient-reported outcomes · Peripheral neuropathy

Introduction

Chemotherapy-induced peripheral neuropathy (CIPN) has been associated with several commonly used oncology agents including taxanes, vinca alkaloids, platinum analogs, thalidomide, and bortezomib [12, 14, 16]. It is characterized by peripheral neurosensory disturbances and, less commonly, by neuromotor disturbances [16, 29]. Neurosensory disturbances may include paresthesias and dysesthesias, such as burning, numbness, tingling, and shooting pains, which often occur in a "stocking-glove" distribution [16]. Moreover, chemotherapy-related neurotoxicity manifested as neuropathy may adversely affect patients' quality of life and activities of daily living [4]. Deterioration of symptoms can lead to treatment delays, the need for dose modifications, or discontinuation of treatment [25]. Despite the fact that CIPN is a common and serious clinical problem, diagnosing and assessing toxic symptoms related to CIPN is complex.

Quantitative assessments of CIPN, such as nerve conduction velocity, vibration perception threshold, and electromyography, have been attempted in the clinical setting. However, these methods can be invasive and uncomfortable for patients, generally lack diagnostic value, and are costly in terms of both time and resources. To assist in the diagnosis and grading of CIPN, various physician-based scales have been developed including National Cancer Institute Common Toxicity Criteria (NCI-CTC), Physician Neurotoxity Examination Form [20], and Ajani criteria [2]. Unfortunately, it can be difficult to accurately interpret the different parameters used to describe peripheral neurotoxicity with these scales, for instance, "mild or moderate objective sensory loss" or "moderate paresthesias." Similarly, the difference between grade 2 and grade 3 neuropathy from these scales can be ambiguous, despite this determination having crucial clinical implications on NCI-CTC. Accordingly, these scales can be interpreted differently among observers, leading to poor inter-scale agreement [23, 24]. Studies have consistently demonstrated that physicians tend to underestimate and underreport the severity and frequency of physical symptoms compared with patients [28]. In particular, physicians tend to underreport subjective symptoms such as fatigue [3, 10] and numbness [23] as opposed to symptoms that can be observed directly. In addition, the impact that physical symptoms exert on patients' quality of life (QOL) is often underestimated by physicians [28]. Physicians report patients to have fewer symptoms than patients do for most domains of health-related QOL (HRQOL) [21]. These findings suggest that supplementing physicians' assessments with patients' perspectives on the toxic symptoms they experience during chemotherapy may provide valuable clinically relevant information. In recent years, the importance of specific instruments to assess patient-reported outcomes has been recognized if accurate data on patient-reported adverse events (AEs) during cancer therapy are to be collected [18, 26].

Recently, the Patient Neurotoxicity Questionnaire (PNQ) was developed in order to assess the incidence and severity of CIPN as reported directly by patients. The PNQ consists of two items (sensory and motor) and was initially developed for use in registration trials of potential neuroprotective agents in order to measure CIPN as a clinically significant endpoint [12]. The PNQ appears to be relatively convenient and specific in its ability to assess CIPN than other patient-reported CIPN scales such as the Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity (FACT/GOG-Ntx) [6, 13] and the EORTC quality of life questionnaire (EORTC QLQ)-CIPN20 [22]. For instance, the FACT/GOG-Ntx includes several questions relating to hearing loss, fatigue, and astereognosis that are not

diagnostic of CIPN. Moreover, the PNQ may be able to assess the presence or absence of significant functional impairment, as reported directly by the patient, and may provide useful information of whether the patient is experiencing interference with activities of daily living due to peripheral neuropathy [12]. We expect that this possibility of the PNQ is crucial for medical decision-making by physicians (e.g., dose modification, treatment delay, or cessation), as well as in clinical trials in order to assess valid neurosensory endpoints. Therefore, the aims of the current study were to evaluate the feasibility and validity of the newly developed PNQ for grading severity of CIPN and to compare patient-reported outcomes with those from a physician-based approach.

Materials and methods

Development of the Patient Neurotoxicity Questionnaire

The PNQ is a simple self-administered instrument that was designed and developed with reference to the neurosensory and neuromotor components of the NCI-CTC (Version 2) by BioNumerik Pharmaceuticals, with input from the US Food and Drug Administration and physicians and nurses familiar with CIPN [12]. The developers of the PNQ spent many years interviewing and carefully eliciting information regarding key symptoms and activities of daily living that are diagnostic of chemotherapy-induced neuropathy. These interactions were made in cancer patient support groups as well as clinical evaluations of patients who were undergoing active treatment or who had been previously treated with neurotoxic chemotherapeutic agents. The input from patients was invaluable, and the most common and diagnostically unambiguous symptoms and activities of daily living were incorporated into the instrument. In addition, the composition and formulation of the PNQ was extensively discussed with the US FDA, which approved the use of the PNQ as the primary endpoint analysis method in pivotal oncology trials involving neurotoxic chemotherapy.

The PNQ comprises two items to identify the incidence and severity of sensory and motor disturbances (Fig. 1). The subjective responses to each item are graded from A (no neuropathy) to E (severe neuropathy) by the patient. There is specific demarcation between grades C and D corresponding to the absence (grade ≤C) and presence of symptoms (grade ≥D) that interfere with activities of daily living. Patients with grade ≥D symptoms are asked to identify which activities are affected as a result of therapy. The PNQ was originally written in English and a Japanese translation was used in the present study. The Japanese version of the PNQ was developed using a rigorous forward and backward translation process [5] with independent review by several oncologists, neurologists, and linguistic experts fluent in both languages.

Study design

The first 300 patients enrolled (November 2001 to May 2003) in a Japanese randomized, multi-institutional phase III trial of adjuvant taxane chemotherapy in patients with operable breast cancer (N-SAS BC 02) were included in the PNQ validation study. CIPN and HRQOL were prospectively assessed in these patients.

Initially, patients with tumors that were positive for both the estrogen and progesterone receptors were ineligible. However, the study protocol of N-SAS BC 02 was amended to permit the enrollment of hormone-positive disease from June 2003 following data from the NSABP-B28 trial showing that the survival benefits associated with the taxanes are not only observed in hormone-negative disease [19]. Main inclusion criteria for the trial were: surgery for breast cancer; stage I–IIIA node-positive disease; age <70 years; and ECOG performance status 0–1. Patients were randomized to one of four treatment arms: (1) four cycles of doxorubicin 60 mg/m^2 (or epirubicin 75 mg/m^2) plus cyclophosphamide 600 mg/m^2 every 3 weeks for anthracycline–cyclophosphamide combination therapy followed by four cycles of paclitaxel 175 mg/m^2 every 3 weeks, (2) cyclophosphamide combination therapy followed by docetaxel 75 mg/m^2 every 3 weeks, (3) eight cycles of single-agent paclitaxel 175 mg/m^2 every 3 weeks, and (4) eight cycles of single-agent docetaxel 75 mg/m^2 every 3 weeks.

Assessments

The primary objective of the N-SAS BC 02 trial was to evaluate overall survival after eight cycles of taxane monotherapy compared with four cycles of an anthracycline–cyclophosphamide combination followed by four cycles of taxane therapy. Here we examine the protocol-defined secondary study endpoints for CIPN and HRQOL in order to assess the feasibility and validity of the PNQ.

CIPN was assessed using two patient-based instruments (PNQ [12]; FACT/GOG-Ntx [6]) and one physician-based instrument (neurosensory and neuromotor components of the unvalidated NCI-CTC [Version 2]). The FACT/GOG-Ntx is a 38-item questionnaire comprising two components: a general measure of quality of life (FACT-G) including physical, emotional, functional, and social/family well-being, and an 11-item neurotoxicity (Ntx) subscale [6]. The Japanese version of the FACT-G has been validated by Fumimoto and colleagues [11]; the

Patient Neurotoxicity Questionnaire (PNQ)®

Taxanes, Cisplatin and Carboplatin

Item 1.

☐ A	☐ B	☐ C	☐ D*	☐ E*
I have no numbness, pain or tingling in my hands or feet.	I have mild tingling, pain or numbness in my hands or feet. This does not interfere with my activities of daily living.	I have moderate tingling, pain or numbness in my hands or feet. This does not interfere with my activities of daily living.	I have moderate to severe tingling, pain or numbness in my hands or feet. This interferes with my activities of daily living.	I have severe tingling, pain or numbness in my hands or feet. It completely prevents me from doing most activities of daily living.

Item 2.

☐ A	☐ B	☐ C	☐ D*	☐ E*
I have no weakness in my arms or legs	I have a mild weakness in my arms or legs. This does not interfere with my activities of daily living.	I have moderate weakness in my arms or legs. This does not interfere of my activities of daily living.	I have moderate to severe weakness in my arms or legs. This interferes with my activities of daily living.	I have severe weakness in my arms or legs. It completely prevents me from doing most activities of daily living.

* Please indicate by placing an X in the box or writing in the space provided which activity or activities have been interfered with as a result of therapy.

My ability to:

☐ Button clothes	☐ Open doors	☐ Fasten buckles	☐ Write	☐ Sew
☐ Use a knife	☐ Put in or remove contact lenses	☐ Sleep	☐ Walk	☐ Work
☐ Use a fork	☐ Dial or use telephone	☐ Climb stairs	☐ Put on jewelry	☐ Tie shoes
☐ Use a spoon	☐ Operation of remote control	☐ Type on a keyboard	☐ Knit	☐ Drive
	☐ Other eating utensils, etc	☐ Perform activities of importance to me, specify:		
		_____		

Fig. 1 The Patient Neurotoxicity Questionnaire

Ntx subscale was developed by formal collaboration between the authors of this study and Ms Sonya Eremenco and David F. Cella from the Center on Outcomes, Research, and Education (CORE). The NCI-CTC scale ranges from 0 to 4 (0 = no symptoms, 4 = maximum symptoms). CIPN and HRQOL assessments were made at pre-treatment (baseline) and before cycles 3, 5, and 7, and were then repeated at 7 and 12 months after starting adjuvant chemotherapy. As a note, CIPN was assessed by using the NCI-CTC at the same assessment points with other instruments to make a comparison possible in this study, although the NCI-CTC is ordinarily used to assess the worst level of CIPN in a certain period of time.

Data collection

Patient-based questionnaires were distributed to the patients by the clinical research coordinators or physicians. The patients completed the patient questionnaires independently and sent them directly to the study data center without any input or discussion with their physicians or nurses.

Statistical considerations

A total of 1,060 patients were enrolled in the study before March, 2006. It was estimated that 300 patients would be an adequate sample size to accurately assess CIPN and HRQOL following review of the validation studies for

several major cancer-specific HRQOL instruments such as the EORTC QOL-Core 30 (N=305) [1], FACT-G (N=545) [7], and the QOL Questionnaire for patients treated with Anti-Cancer Drugs (QOL-ACD) (N=212) [15].

The following properties of the instruments were evaluated using data at baseline and before cycles 3, 5, and 7: compliance of instruments in order to confirm feasibility; concordance between CIPN grades noted by physicians (NCI-CTC) and patients (PNQ); concurrent validity; and responsiveness. The data for all assessment points for the full sample were examined for the evaluation other than that of responsiveness. Compliance was defined as the proportion of patients with evaluable questionnaires from the total of 300 questionnaires completed by the patients or physicians at each assessment point, and was assessed for the two patient-based instruments compared with the one physician-based scale.

For reference, convergent validity and discriminant validity were assessed by the degree of correlation between instruments [9]. The correlation represented the level of association, e.g., if physicians reported every symptom one grade lower than patients, there would be a high correlation despite the difference in response. Hence, in this study, the concordance was examined by comparing the absolute score distribution for each grade of severity between the patient-based PNQ and the physician-based NCI-CTC using the weighted kappa coefficient [8]. The categories used for interpreting kappa values were those previously proposed by Landis and Koch [17]: kappa <0.00 was poor, 0.00–0.20 was slight, 0.21–0.40 was fair, 0.41–0.60 was moderate, 0.61–0.80 was substantial, and 0.81–1.00 was almost perfect.

Concurrent validity between the PNQ and patient-reported HRQOL scales, the Ntx subscale scores, and the FACT-G total score was evaluated by Spearman's correlation coefficients [27]. We also attempted to examine the association between the PNQ and the NCI-CTC.

Responsiveness is the ability of an instrument to detect positive or negative changes in symptom grading scores [9]. To examine clinical responsiveness according to treatment, the PNQ scores were evaluated using linear time-trend tests with contrast coefficient "−3, −1, 1, 3" via a repeated measures analysis of variance model. We expected that the PNQ scores would be higher in later treatment cycles as CIPN is related to cumulative chemotherapy doses [25]. Additionally, the number of patients receiving chemotherapy with a taxane would approximately double after the patients in the anthracycline–cyclophosphamide combination arms had completed the four cycles of the non-neurotoxic phase. Because the score ranges of the PNQ and the NCI-CTC are different from that of the Ntx subscale, we converted the scores of the three scales into a 10-scale score [from 0 (minimum severity) to 10 (maximum severity)] to make a comparison possible. The average scores for all scales were evaluated using linear time trend-tests and the Cohen's D as an effect size.

Results

Patient and tumor characteristics

Patient and tumor characteristics at baseline are shown in Table 1 for the 300 patients included in this assessment. Breast-conserving surgery was performed in 57.7% of eligible patients. Over 70% of patients had hormone-receptor negative tumors because patients with tumors positive for both estrogen and progesterone receptors had been excluded in the first phase of the N-SAS BC 02 study.

Compliance

A total of 295 (98%) patient-reported questionnaires were evaluable at baseline and 295 (98%), 279 (93%), and 270

Table 1 Patient and tumor characteristics at baseline

Parameter	Value (N=300)
Mean age (SD), years	51.7 (8.9)
Type of surgery, *n* (%)	
Breast conservation	173 (57.7)
Mastectomy	127 (42.3)
No. of positive lymph nodes, *n* (%)	
1–3	165 (55.0)
4–9	80 (26.7)
≥10	55 (18.3)
Radiation therapy to the breast, *n* (%)	
Yes	157 (52.3)
No	143 (47.7)
Estrogen receptor status, *n* (%)	
Negative	224 (74.7)
Positive	76 (25.3)
Progesterone receptor status, *n* (%)	
Negative	40 (13.3)
Positive	258 (86.0)
Unknown	2 (0.7)
HER2 status, *n* (%)	
Negative	82 (27.3)
Positive	
1+	50 (16.7)
2+	18 (6.0)
3+	56 (18.7)
Unknown	94 (31.3)

HER2 human epidermal growth factor receptor type 2, *SD* standard deviation

Table 2 Distribution of the scores for the PNQ and the NCI-CTC

NCI-CTC National Cancer Institute Common Toxicity Criteria, *PNQ* Patient Neurotoxicity Questionnaire

[a] PNQ scale ranges from A (no neuropathy) to E (severe neuropathy)

Kappa=0.16	NCI-CTC Sensory					Kappa=0.02	NCI-CTC Motor				
	0	1	2	3	4		0	1	2	3	4
PNQ Sensory[a]						PNQ Motor[a]					
A	489	38	0	0	0	A	492	5	1	0	0
B	432	252	4	0	0	B	701	37	2	0	0
C	113	171	5	0	0	C	231	17	3	0	0
D	44	66	11	3	1	D	62	10	5	1	0
E	9	1	0	0	0	E	9	0	0	0	0

(90%) were evaluable before cycles 3, 5, and 7, respectively. The main reasons for not completing assessments included patients' failure to report due to treatment AEs, disease recurrence, or change in primary physician. Overall, there were 293 (98%) completed and evaluable physician-reported NCI-CTC questionnaires at baseline and 287 (98%), 281 (94%), and 269 (90%) were evaluable before cycles 3, 5, and 7, respectively.

Concordance

The PNQ scores reported were distributed over the full range (A to E), whereas most of the NCI-CTC scores were distributed between 0 and 1. Especially, ten patients reported their symptoms were of maximum severity (E) for sensory disturbance, whereas nine physicians and one physician evaluated those patients had no symptoms (0) and slight symptoms (1), respectively. Similarly, nine patients reported their symptoms were of maximum severity (E) for motor disturbance, whereas all physicians evaluated those patients had no symptoms (0). The weighted kappa coefficient was 0.16 for sensory disturbance and 0.02 for motor disturbance (Table 2).

Concurrent validity

Table 3 presents the correlation matrix for each subscale score by Spearman's correlation coefficient. Both PNQ sensory and motor scores were strongly correlated with the Ntx subscale scores (r=0.66 and r=0.51, respectively). An even higher correlation (r=0.70) was observed between the PNQ sensory scores and the Ntx subscale scores if the five items not diagnostic of taxane-induced CIPN (joint pain/muscle cramps, trouble hearing, ringing in ears, and trouble feeling the shape of small objects) were excluded from the Ntx subscale. Although the Ntx subscale scores were significantly associated with the FACT-G total scores (r=0.43), only a weak relationship was observed between the PNQ sensory and motor scores and the FACT-G total scores (r=0.29 and r=0.39, respectively). The PNQ sensory scores were significantly correlated with the NCI-CTC sensory scores (r=0.44), but the PNQ motor scores were not associated with the NCI-CTC motor scores (r=0.16).

Responsiveness to treatment cycles

Table 4 shows the average scores for all subscales at baseline and before cycles 3, 5, and 7. Each subscale scores increased over time (Fig. 2). The PNQ sensory and motor scores significantly increased as the number of treatment cycles increased (P<0.0001), thereby indicating progressively worsening CIPN over time (Fig. 2). In addition, the PNQ sensory and motor scores were higher than the Ntx subscale scores at all assessment points. Although the NCI-CTC sensory scores demonstrated similar changes over time, physicians reported lower grades of CIPN compared with patient-reported CIPN scores during the entire treatment period (Fig. 2).

Discussion

This study confirmed the clinical feasibility and validity of the PNQ in a subgroup of women receiving taxane

Table 3 Correlation matrix for each subscale score by Spearman's correlation coefficient for overall treatment cycles

Scale	PNQ motor	FACT/GOG Ntx subscale	NCI-CTC sensory	NCI-CTC motor	FACT-G
PNQ sensory	0.48	0.66	0.44	0.19	0.29
PNQ motor	–	0.51	0.22	0.16	0.39
FACT/GOG Ntx subscale	–	–	0.45	0.23	0.43
NCI-CTC sensory	–	–	–	0.28	0.09
NCI-CTC motor	–	–	–	–	0.11

FACT-G Functional Assessment of Cancer Therapy-General scale, *FACT/GOG Ntx* Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity scale, *NCI-CTC* National Cancer Institute Common Toxicity Criteria, *PNQ* Patient Neurotoxicity Questionnaire

Table 4 Changes of mean scores for the FACT/GOG-Ntx, the PNQ, and the NCI-CTC over time

Scale	Baseline		3rd cycle		5th cycle		7th cycle		p value[a]	Cohen's D[b]
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
PNQ sensory	1.69	2.02	2.29	2.28	2.85	2.52	3.40	2.32	<0.0001	0.79
PNQ motor	2.07	1.99	2.14	1.94	2.57	2.33	2.90	2.31	<0.0001	0.38
FACT/GOG-Ntx	0.79	0.88	1.40	1.43	1.84	1.69	2.29	1.76	<0.0001	1.08
NCI-CTC sensory	0.03	0.25	0.62	1.18	0.86	1.36	1.48	1.45	<0.0001	1.39
NCI-CTC motor	0.04	0.48	0.10	0.58	0.16	0.68	0.26	0.87	<0.0001	0.31

Range of mean score is from 0 to 10, with a higher score indicating greater severity of CIPN

CIPN chemotherapy-induced peripheral neuropathy, *FACT/GOG Ntx* Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity Scale, *NCI-CTC* National Cancer Institute Common Toxicity Criteria, *PNQ* Patient Neurotoxicity Questionnaire, *SD* Standard Deviation

[a] Test for linear time-trend

[b] The higher score indicates the greater effect size

chemotherapy for breast cancer participating in a large, prospective, multi-institutional randomized phase III study. The PNQ was shown to be a sensitive and responsive instrument in the diagnosis and grading of CIPN, with a greater sensitivity than the FACT/GOG-Ntx and NCI-CTC scales. Compliance for the utilization of both patient-reported and physician-reported instruments was consistently high during the study, with over 90% of patients completing the PNQ questionnaire at all assessment points. This high level of compliance supports the feasibility of the PNQ as an effective method for patients to report their CIPN-related symptoms in clinical practice.

As we anticipated, physicians demonstrated a tendency to underestimate and underreport the severity of CIPN symptoms compared with patients' self-reporting their CIPN symptoms via the PNQ. Consequently, the PNQ showed that CIPN was associated with a greater impact on activities of daily living than was observed via the physician-based NCI-CTC. Indeed, a low degree of concordance in grading CIPN was observed between physicians and patients, especially for the NCI-CTC motor evaluations. This discriminant validity between the PNQ and the NCI-CTC indicates that the use of patient-reported outcomes is highly valuable in the clinical setting.

One interpretation for the observed discrepancy in the severity of CIPN scores between patients and physicians is that patients tend to be more aware of their CIPN symptoms in relation to how they impact on activities of daily living [4]. Conversely, physicians might generally judge the absolute detectable level of sensory abnormality or muscle weakness to be of greater importance than symptom levels. Alternatively, the observed large patient–physician discrepancy in CIPN symptom scores may be explained, in part, by the difference of internal standards for the severity of CIPN symptoms between patients and physicians. That is, while most breast cancer specialists might have their own standards based on their sufficient clinical experience of seeing many patients with very severe CIPN symptoms, most patients might never have experienced very severe ones. However, a question about the notable discrepancy, i.e, patients rated their symptoms as E (maximum severity), whereas physicians evaluated those patients had no or slight symptoms, still remains. A possible reason is that the wording of "completely" in category E of the PNQ might have been interpreted differently among patients, resulting in a large discrepancy in CIPN symptom scores between patients and physicians. We will need to confirm the reason for underestimation of the prevalence in future research and further refine the PNQ.

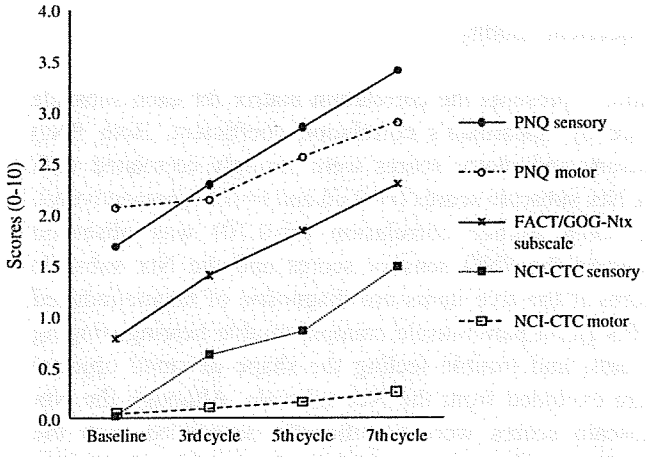

Fig. 2 Clinical responsiveness of the PNQ. Range of mean scores is from 0 (minimum severity) to 10 (maximum severity) for the PNQ, FACT/GOG-Ntx, and NCI-CTC subscales. *FACT/GOG Ntx* Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity scale, *NCI-CTC* National Cancer Institute Common Toxicity Criteria, *PNQ* Patient Neurotoxicity Questionnaire