



Published in final edited form as:

Med Care. 2009 August ; 47(8): 866–874. doi:10.1097/MLR.0b013e3181a31d00.

Differential Symptom Reporting by Mode of Administration of the Assessment: Automated Voice Response System versus a Live Telephone Interview

Alla Sikorskii, PhD [Assistant Professor],

Michigan State University, Department of Statistics and Probability, College of Natural Science, A423 Wells Hall, East Lansing, MI 48824, Phone: 517-353-2963, E-mail: sikorska@msu.edu

Charles W. Given, PhD [Professor],

Michigan State University, College of Human Medicine, Department of Family Medicine, B108 Clinical Center, East Lansing, MI 48824, Phone: 517-353-0851 x420, E-mail: givenc@msu.edu

Barbara Given, PhD, RN, FAAN [University Distinguished Professor],

Michigan State University, College of Nursing, B515B West Fee Hall, East Lansing, MI 48824, Phone: 517-353-0306, E-mail: Barb.Given@hc.msu.edu

Sangchoon Jeon, MS [Associate Research Scientist], and

Yale University, School of Nursing, 100 Church Street South, PO Box 9740, New Haven, CT 06536-0704, Phone: 203-785-6280, E-mail: sangchoon.jeon@yale.edu

Mei You, MS [Information and Statistical Analyst]

Michigan State University, College of Nursing, East Lansing, MI 48824, Phone: 517-432-8352, E-mail: mei.you@hc.msu.edu

Abstract

Background—Automated Voice Response (AVR) systems have been used to collect patient-reported outcome data. Mode of administration of the assessment may affect patient reporting.

Objective—To evaluate if there is a differential reporting of symptoms by the mode of assessment: AVR versus a live telephone interview among cancer patients with solid tumors undergoing chemotherapy.

Research Design—Randomized clinical trial comparing a nurse assisted symptom management with an automated telephone symptom management. After completing intake telephone interview administered by a person, patients were randomized to either nurse arm or AVR arm to receive a 6-contact 8-week symptom management intervention. Patients in the nurse arm were called by specially trained nurses, and patients in the AVR arm were contacted via automated system to assess their symptoms and deliver symptom management strategies.

Subjects—Two hundred patients in nurse arm, and 186 patients in the AVR arm completed the first intervention contact.

Measures—Severities of 14 cancer-related symptoms were rated by patients at intake interview and at first intervention contact before the receipt of any interventions.

Results—When compared with patients contacted by a nurse, patients contacted by the AVR reported higher severity of nausea and vomiting, diarrhea, poor appetite, constipation, diarrhea, pain

and alopecia controlling for prior intake symptom assessment that was free of mode effect. **Symptom reporting varied by age with the oldest group of patients reporting higher severity to the nurse.**

Conclusion—Mode effect needs to be considered in designing trials for symptom management and in symptom monitoring in clinical practice.

Keywords

Symptoms; Automated Voice Response System; Mode of Administration; Cancer

Introduction

Interactive or Automated Telephone Voice Response (AVR) systems have been used to help clinicians provide care to patients experiencing chronic conditions including cancer.¹⁻⁵ These systems merge computer software to a pre-recorded voice and an automated telephone system to collect data from patients, monitor their health status, or deliver intervention strategies for the management of chronic conditions.⁶⁻¹¹ Patients are called at specified times and intervals, their responses are entered using the touchtone pad of a telephone, and recorded. Patient evaluations indicate that automated systems have high usability and acceptability, and are more accurate than in-person interviews in obtaining sensitive information.^{12,13} The automated voice response systems appear equal to clinical interviews in obtaining data to make psychiatric diagnoses,¹⁴ in promoting diabetic self-care management skills,^{6,15} and in delivering pain coping strategies.⁷ This research compares cancer patients' reports of symptoms using AVR systems with live telephone reports to nurses. Specifically this research seeks to determine if there is an effect of the mode of symptom assessment, that is, if patients report the severity of their symptoms differently to the automated system versus to a person. Given the same underlying value of severity, is there a difference in patient severity rating reported to a person (nurse) versus an AVR?

In randomized clinical trials, intake and follow-up assessments of patients usually involve the same mode of administration, for example paper-and pencil, telephone interviewing, electronic devices such as personal digital assistant (PDA),¹⁶ a touch-screen computer,¹⁷⁻¹⁸ or patient web-based self-reporting.¹⁹⁻²⁰ However, the technology for delivering interventions may differ by trial arm. For example, a nurse may deliver tailored cognitive behavioral strategies via the telephone, while an AVR system may deliver pre-programmed information and self-care strategies. Thus, the exposure is the same, the content similar, but mode of delivery differs. In attempting to specify the processes through which elements of the intervention seek to change the outcome variable during the contacts, it is critical to distinguish between real changes and changes that may be due to mode effects. Thus, comparisons based on different methods of assessment must be capable of separating real differences from those due to possible mode effects. The purpose of this research is to compare patients' assessments of symptom severity as reported during live telephone interviews compared with assessments using AVR techniques.

The computerized modes of data collection such as AVR, PDA, or web-based reporting have been compared to paper and pencil self-administration and to live telephone interviews. Comparisons of a paper and pencil administration to a PDA revealed mode effects for the Center for Epidemiological Studies Depression (CES-D) scale,²¹ but no differences in responses to other instruments.^{22,23} Paper and pencil self-administration has been shown to result in higher average depressive symptomatology scores compared to in-person interviews.²⁴ Among cancer patients who experience multiple symptoms related to cancer, its treatment and comorbid conditions, AVR systems have been used for symptom monitoring,^{1,17-19} but

the mode effects of AVR versus other formats of symptom data collection have not been studied.

In this research, cancer patients' reports of severity of 14 symptoms are investigated to answer the following research questions: 1) Is there a mode of administration effect on patients' reports of severity of 14 cancer-related symptoms: pain, fatigue, insomnia, peripheral neuropathy, dyspnea, cough, poor appetite, constipation, diarrhea, nausea/vomiting, dry mouth, alopecia (hair loss), difficulty remembering, and weakness, according to live telephone interview with a nurse versus AVR; 2) Are the mode effects in symptom reporting to a nurse versus AVR different by patient and disease characteristics such as age, sex, education or site of cancer?

Methods

Sample

Approvals for the study were obtained from the Institutional Review Board (IRB) of the sponsoring university, and the IRBs of two comprehensive cancer centers, one community cancer oncology program, and six hospital-affiliated community oncology centers. Eligible patients 1) were 21 years of age or older, 2) had a diagnosis of a solid tumor cancer or non-Hodgkins lymphoma, 3) were undergoing a course of chemotherapy, 4) were able to speak and read English, and 5) had a touchtone telephone. Patients agreeing to participate signed an informed consent form, and had all socio-demographic information entered into a web-based tracking system. Prior to entering the study, patients were screened for symptom severity using an automated voice response version of the M. D. Anderson Symptom Inventory.²⁵ Patients who scored 2 or higher on a 0-10 scale on severity of any symptom at screening were entered into the study. All patients had an intake interview administered by trained interviewers live over the telephone. Following the intake interview, patients received a printed copy of the Symptom Management Guide (SMG) that contained specific strategies for the management of each symptom. Computer minimization procedure²⁶ was used for to randomize patients to receive either a cognitive-behavioral intervention delivered by oncology nurses, or information and self-care strategies delivered by the AVR.

In each of the two trial arms, patients had 6 telephone contacts over 8 weeks, with the first intervention contact occurring on average about two weeks after the intake interview. Each telephone contact began with the assessment of severity of symptoms on a 0-10 rating scale. In the nurse-directed arm, tailored cognitive-behavioral symptom management strategies supplemented with the reference to the SMG were delivered for symptoms above threshold of 4 in severity. In the AVR arm, a pre-recorded pleasant female voice queried patients regarding their severity for their symptoms. To rate symptom severity, patients pressed the appropriate numbers on their telephone keypads. For symptoms rated at 4 or higher, the AVR delivered information and self-care strategies: patients were directed to the sections of the SMG that informed them about strategies to manage the symptoms that were above threshold. Details regarding the trial can be found elsewhere.²⁷

Patients who completed the first intervention contact are included in this analysis. Data from their symptom severity assessment obtained prior to the delivery of any intervention strategies are used to test for mode effects. The assessments of severity conducted during intervention contacts 2-6 were done following the delivery of the different types of the interventions in nurse-directed and AVR arms at contact 1. In the symptom assessments at contacts 2-6, the mode effects were confounded with the effects of different interventions; therefore only contact 1 data are used in mode effect testing. Figure 1 summarizes the number of patients who entered and dropped out at each step, and the number analyzed.

Measures

Independent variables included age, sex, level of education and site and stage of cancer. Their values were obtained from the patients' medical records, entered into the tracking system, and confirmed during intake interview. Drawing upon our own and others work age was categorized as <45, 45 to 74, and 75 and older. In earlier analyses these age categories were predictive of differences in response between nurse and AVR administered interventions; the youngest group favoring the AVR and the oldest group favoring the nurses. Thus age categories needed to be considered in the analysis of mode effects. Second, differences in age may reflect variations in the aggressiveness of tumors, patient responses to treatment,^{28,29} and in treatment aggressiveness.³⁰ Finally age categories represent variations in cognitive neuro-processing and plasticity.^{31,32}

The level of education was collapsed into 2 categories: high school or less versus some college. Individuals with a high school level of education or less have been shown to be at risk of experiencing difficulties in comprehending medical information.³³

Site of cancer was summarized as lung versus non-lung. While patients with sites of cancer other than lung form a heterogeneous population, the differences in symptom reporting were found for lung versus other cancer sites in earlier work.²⁷

Severity of symptoms was scored by patients on a scale ranging from absence (0) to the worst severity possible (10) at intake interview and at each of the intervention contacts. Because the lists of symptoms were slightly different in intake interview and intervention contacts, we conducted the analyses on the 14 symptoms common to both assessments. Symptom severity reported at the first intervention contact was used as a dependent variable, and symptom severity reported during intake interview was used as a covariate as described below.

Data Analysis

Unadjusted arm comparisons of symptom severity scores at intake interview and first intervention contact were performed using t-tests. At intake into the trial, all patients were interviewed by trained interviewers over the telephone, thus no mode effect was present. The severity score of each symptom as reported in the intake interview was used to control for the underlying value of severity in mode effect analysis. For each symptom, a regression model that related severity of a symptom reported at the first intervention contact to severity reported at intake interview, trial arm, and the number of days since intake interview was fit. To answer the question about differential reporting, that is, if the underlying value of severity is the same, but patient reports to a person versus AVR differ, a measure of the underlying value of severity needs to be included in the model. In the absence of a true value of severity, several different estimates have been used,³⁴ which may be problematic to use with symptoms due to dimensionality issues.³⁵ In this study, a value free of mode effects was available from the intake interview, which was used along with the adjustment for the number of days between 2 symptom assessments. Further, trial arm by severity at intake interaction was added to the model. Models with and without the interaction terms were compared using a likelihood ratio test. Non-uniform mode effect, i.e. mode effect with varying direction across levels of the underlying symptom severity, would be present if the model was better with the interaction term. Comparison of models with severity and severity and trial arm produced a test for a mode effect uniform with respect to the underlying value of severity as reflected by severity measured at intake.

The influences of patient age, sex, level of education and site of cancer on symptom reporting were investigated by adding the appropriate variables as covariates and in interaction with trial arm to the regression models. To avoid a potential problem with colinearity, age, sex, education

and site of cancer were investigated separately in relation to mode effect. The adjusted means³⁶ of symptom severity by levels of covariates and trial arm were derived from the regression models, and differences by trial arm were tested. Effect sizes for between trial arm differences were calculated by subtracting the adjusted means and dividing the difference by the standard deviation at intake. To illustrate the direction of the differences, effect sizes are reported as positive if severity was greater in the AVR arm, and as negative when severity was greater in the nurse arm. Because the distribution of symptom severity was skewed to the left, the regression models were implemented as generalized linear models with gamma distributed errors.³⁷ Bonferroni adjustments were used to control for probability of type I error in tests for 14 symptoms with test results deemed significant if p-value was less than 0.0036. The selection of Bonferroni adjustment to control type I error was informed by the fact that Bonferroni method is not based on any additional assumptions, and that it has been shown to result in almost the same conclusions as obtained with cross-validation.³⁸ On the other hand, Bonferroni adjustment may be much too stringent.³⁹ To control for the false discovery rate, Benjamini-Hochberg procedure was implemented,⁴⁰ which is valid under the assumption of positive dependence of p-values.⁴¹ Since the list of symptoms was pre-specified in advance of testing, and because adjustments for multiplicity may result in an inflated type II error,³⁹ both adjusted and unadjusted results are presented. The conclusions derived using Bonferroni and Benjamini-Hochberg procedures were very similar, and both are presented in tables and results. The analyses were performed using SAS 9.1@.⁴²

Results

Table 1 contains the summary of socio-demographic and disease information by arm of the trial for those randomized and for those analyzed. No significant differences were found between the arms of the trial at intake according to patient or disease characteristics. There were no significant differences by trial arm among those who dropped out after the intake interview, or among those who skipped intervention contacts.

Table 2 provides the results of trial arm comparisons of symptom severity at intake for those who skipped the intervention contacts, or dropped out of the study prior to the first intervention contact, and thus were not analyzed (first panel). Even though the percent of those not analyzed was higher in the AVR arm compared to the nurse-directed arm (15% versus 8.3%), their patient and disease characteristics (Table 1), and symptom severity (Table 2) were not different by trial arm.

Panels of Table 2 display the severity of symptoms for those not analyzed, and for those analyzed as reported by at the intake interview, and approximately two weeks later when patients in the AVR arm were called by the automated system, and patients in the nurse arm were queried by the nurse interveners. The number of days between the intake interview and first intervention contact did not differ by trial arm (nurse arm: mean of 15.05 (standard deviation 6.31), AVR arm: 13.60 days (standard deviation 5.53)). The comparison of intake symptom severity reported by those who completed the first intervention contact and those who did not revealed no differences between or within trial arms. For those analyzed, the severity of poor appetite was higher in the AVR arm ($p=.02$) at intake interview, but this difference was not significant with either Bonferroni or Benjamini-Hochberg adjustment for multiplicity. Severity of all other symptoms was balanced by trial arm at intake even without applying adjustments for multiple testing ($p\text{-values}>.05$, Table 2). However at first intervention contact prior to the delivery of any intervention strategies, patients in the AVR arm reported significantly higher severity of several symptoms: alopecia (hair loss), poor appetite, nausea/vomiting, constipation, and pain (significant with Benjamini-Hochberg adjustment). Note that the means of symptom severity at live intake interview and at first nurse-administered intervention contact are similar.

Further analyses were performed controlling for the symptom severity reported at intake interview to detect differential reporting of symptoms given the same underlying value of severity. The comparisons of models with and without trial arm by symptom severity at intake interaction revealed that models without the interaction term were better for all symptoms, thus non-uniform mode effect was absent. After the interaction term between trial arm and intake symptom severity was removed from the model, the uniform mode effect as reflected by the trial arm coefficient was tested. The coefficients of the models with intake severity and arm as explanatory variables for the outcome of symptom severity reported at the first intervention contact are presented in Table 3.

From these models, a significant uniform mode effect after either multiplicity adjustment was found for alopecia, poor appetite, nausea/vomiting, and constipation. For symptoms of pain, and diarrhea, mode effect was significant with Benjamini-Hochberg adjustment for multiplicity. For fatigue, dyspnea, and peripheral neuropathy the p-values for mode effect were less than 0.05, but not significant after either multiplicity adjustment. Remarkably, the direction of the mode effect was the same across symptoms, with higher levels of severity reported to the AVR compared to a nurse.

When site of cancer, sex or level of education were added to the model in interaction with trial arm, no significant mode effects that were differential by these characteristics were observed. The findings for age suggested that for several symptoms, the mode of administration effect was differential by age group. The adjusted means by trial arm and age categories are presented in Table 4.

Patients in the middle age group (45-74 years) reported higher levels of severity of alopecia, poor appetite, constipation, diarrhea, fatigue, nausea/vomiting, pain, and peripheral neuropathy to the AVR. In contrast, oldest age group (75+ years) reported higher levels of severity of dyspnea, fatigue, pain, peripheral neuropathy and insomnia to the nurse. Because of the sample sizes of 17 and 20 for patients 75 years of age or older in the nurse and AVR arm respectively, the magnitude of the effect size was used to assess the differences between trial arms. The effect size of .33 is recommended as a lower bound for clinically important difference.⁴³ In the oldest age group, the effect sizes for dyspnea, fatigue, pain, peripheral neuropathy and insomnia was greater than or equal to .35 favoring nurse. In the youngest age group (25-44 years), only 3 symptoms, alopecia, dyspnea, and pain, exhibited mode effects with effect sizes exceeding .33 in magnitude with higher levels of severity reported to the AVR. Alopecia was the only symptom for which the effect sizes in all three age groups were clinically significant favoring AVR arm. For pain, the magnitude of the effect size for the differences in all three groups was clinically significant with youngest and oldest age groups favoring the nurse, and the middle age group favoring the AVR.

Discussion

The results highlight the importance of accounting for mode effects in designing trials of interventions for symptom management. When different types of interventions are delivered using different modes, and symptom assessments are used to determine which symptoms should and which should not be targeted with interventions, mode effects need to be considered. For example, according to the National Comprehensive Cancer Network guidelines,⁴⁴ scores of 4 or higher in severity indicate the need for symptom management. However, if given the same underlying level of severity, patients report this severity differently to an AVR versus a nurse, then adjustments to the thresholds that trigger intervention delivery are needed. In addition, if mode effects are present early on, but diminish over time, the reported decrease in severity over time should be distinguished from the response to the interventions delivered. Automated telephone voice response systems are just one example of the computer technology

used to assess patient reported outcomes and deliver interventions to patients experiencing symptoms due to chronic conditions.

In this trial, randomization procedure allocated patients to trial arms that were equivalent at intake with regard to socio-demographic characteristics and reported symptom severity. However, at the time the intervention began, this equivalence was affected by the mode of administration of the symptom assessment with groups being no longer equivalent with respect to primary outcome of symptom severity burden.²⁷ This arm imbalance can be handled analytically by adjusting for the severity reported at first contact when evaluating severity at later contacts, however once different intervention strategies are delivered using different modes, the mode effect can not be separated from the effect of the intervention.

The analyses were carried out on a symptom by symptom basis using generalized linear models that controlled for each symptom's severity that was reported a week earlier free of mode effects. Other methodological approaches to detection of mode effects that are based on classical test theory or item response theory, and differential item functioning analyses have been used.^{21,24,34} These approaches rely on dimensionality structure, e.g. established factor structure of the construct being measured. In the application to cancer-related symptoms, the dimensionality issue remains open to question.³⁵ Among breast cancer patients who are not undergoing active treatment, Stanton et al.⁴⁵ have shown the stability of factor-analytic results across multiple samples, but stability of subscales remains an issue and an open question among cancer patients undergoing treatment. One proposed explanation for such instability is that the results of factor analyses depend on correlations among symptoms, and those in turn may depend on specific sites of cancer, treatments and timing of administration of the symptom assessment relative to treatment and disease course. Notably, the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-30)⁴⁶ has six single items that assess symptoms reported by cancer patients that do not form subscales. The approach for the detection of mode effects implemented in this study does not rely on factor structure. Instead, we used a prior assessment of symptoms that was conducted over the telephone by the interviewers with all patients to control for the underlying value of severity. The observed stability of means of symptom severity scores between two assessment times was reflective of the fact that cancer-related symptoms persist over time,⁴⁷ and may have multiple etiologies linked to chemotherapeutic agents, comorbid conditions, or affective states.⁴⁸

A potential explanation of why patients reported higher levels of severity to the AVR compared to a person could lie in the neutrality of the AVR. With no interaction and no comments from the interviewer, patients may be more candid and truthful about the severity of their symptoms. This explanation would be consistent with the fact that patients report more severe symptoms to others than to oncologists or nurses.⁴⁹ The breakdown of the mode effects by age revealed that the age group of 45 to 74 years made the largest contribution to the overall mode effect. Except for 3 symptoms (alopecia, dyspnea and pain), no mode effects were found in the youngest group, while patients who were 75 years of age or older, reported higher severity of several symptoms to the nurse compared to the AVR. These symptoms are mostly physical in nature (dyspnea, pain, fatigue, peripheral neuropathy, and insomnia), and the effect sizes for the mode effect in the oldest age group were moderate to large suggesting clinically important differences. The literature on aging and plasticity is not clear. However, the oldest group of patients clearly was more willing to report severity to the nurse, just as the middle age group was more comfortable reporting greater severity to the AVR. The oldest group of patients may, based on some evidence, possess less adaptability⁵⁰ and, therefore, may value interactions with nurses who inquire about their symptoms. Further, shifts that require patients to use AVR technology may be easier for younger patients where they encounter these systems in their daily lives.

While the mechanisms behind these age related mode effects remain speculative, we believe that age is an important characteristic in assessing possible mode effects and needs to be considered carefully in future assessment of differences in delivery as interventions are moved to computer and web based modes of delivery. It should be noted that the sample size for the oldest age group was relatively small, and further research is needed to confirm this finding. Also, further research is needed to assess if mode effects persist over time, which was not possible to do in this study as from contact 2 on, the mode effects were confounded with the effects of different interventions administered in AVR and in nurse arms.

Limitations of this research include slight differences in symptom lists evaluated at intake interview and first intervention contact. In intake interview, nausea and vomiting were asked as 2 items but combined into one for the purposes of comparison with the intervention contact where both symptoms were inquired about in one question. There were approximately two weeks between intake interview and the first intervention contact, during which the severity of symptoms may have changed. However because of randomization, it is reasonable to assume that the changes due to time and course of cancer treatment happened equally in the two trial arms.

With technology playing an important role in assessing patient reported outcomes and symptoms in particular, it is important that comparisons of various modes of data collection are performed so that true differences in patient outcomes can be established.

Acknowledgments

This research supported by Grant R01 CA79280 from the National Cancer Institute and in affiliation with the Walther Cancer Institute, Indianapolis Indiana.

References

1. Kornblith AB, Dowel JM, Herndon JE, et al. Telephone monitoring of distress in patients aged 65 years or older with advanced cancer: A cancer and leukemia group study. *Cancer* 2006;107(11):2706–2714. [PubMed: 17078057]
2. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. *Nurs Outlook* 2003;51(6):277–283. [PubMed: 14688763]
3. Friedman RH, Kazi LE, Jette A, et al. A telecommunications system for monitoring and counseling patients with hypertension: impact on medication adherence and blood pressure control. *Am J Hypertens* 1996;9:285–292. [PubMed: 8722429]
4. Friedman RH. Automated telephone conversations to assess health behavior and deliver behavioral interventions. *J Med Syst* 1998;22:95–102. [PubMed: 9571516]
5. Mahoney D, Tennstedt S, Friedman R, Heeren T. An automated telephone system for monitoring the functional status of community-residing elders. *Gerontologist* 1999;39:229–234. [PubMed: 10224719]
6. Piette JD. Interactive voice response systems in the diagnosis and management of chronic disease. *American Journal of Managed Care* 2000;6(7):817–827. [PubMed: 11067378]
7. Naylor MR, Keefe FJ, Brigidi B, Naud S, et al. Therapeutic Interactive Voice Response for chronic pain reduction and relapse prevention. *Pain* 2008 Feb;134(3):335–345. [PubMed: 18178011]
8. Baer L, Jacobs DG, Cukor P, et al. Automated telephone screening survey of depression. *JAMA* 1995;273:1943–1944. [PubMed: 7783305]
9. Farzanfar R, Frishkopf S, Friedman R, Ludena K. Evaluating an automated mental health care system: making meaning of human-computer interaction. *Computers in Human Behavior* 2004:1–16.
10. Mundt JC, Moore HK, Greist JC. A novel interactive voice response (IVR) system for dementia screening, education, and referral: one-year summary. *Alzheimer Dis Assoc Disord* 2005;19(3):143–147. [PubMed: 16118531]

11. Moore HK, Hughes SW, Mundt JC, et al. A pilot study of an electronic, adolescent version of the quick inventory of depressive symptomatology. *J Clin Psychiatry* 2007;68(9):1436–1440. [PubMed: 17915985]
12. Balas EA, Jaffrey F, Kuperman G, et al. Electronic communication with patients. *JAMA* 1997;278:152–159. [PubMed: 9214532]
13. Searles J, Perrine M, Mundt J, Helzer J. Self-report of drinking using touch-tone telephone: Extending the limits of reliable daily contact. *J Stud Alcohol* 1995;56:375–382. [PubMed: 7674671]
14. Kobak KA, Taylor LV, Dottle SL, et al. A computer-administered telephone interview to identify mental disorders. *JAMA* 1997;278:905–910. [PubMed: 9302242]
15. Piette J, Weinberger M, McPhee S. The effect of automated calls with telephone nurse follow-up on patient-centered outcomes of diabetes care (a randomized controlled trial). *Medical Care* 2000;38:218–223. [PubMed: 10659695]
16. Farzanfar R, Finkelstein J, Friedman RH. Testing the usability of two automated home-based patient-management systems. *Journal of Medical Systems* 2004;28(2):143–153. [PubMed: 15195845]
17. Allenby A, Matthews J, Beresford J, McLachlan SA. The application of computer touch-screen technology in screening for psychosocial distress in an ambulatory oncology setting. *European Journal of Cancer Care* 2002;11:245–253. [PubMed: 12492461]
18. Wilkie DJ, Huang HY, Berry DL, Schwartz A, et al. Cancer symptom control: feasibility of a tailored, interactive computerized program for patients. *Family and Community Health* 2001;24(3):48–62. [PubMed: 11563944]
19. Wilkie DJ, Judge MK, Berry DL, Dell J, et al. Usability of a computerized PAINReportIt in the general public with pain and people with cancer pain. *Journal of Pain and Symptom Management* 2003;25(3):213–224. [PubMed: 12614956]
20. Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity of symptoms during chemotherapy. *J Clin Oncol* 2005;23(15):3552–3561. [PubMed: 15908666]
21. Swartz RJ, de Moor K, Cook KF, et al. Mode effects in the center for epidemiologic studies depression (CES-D) scale: personal digital assistant vs. paper and pencil administration. *Qual Life Res* 2007;16(5):803–813. [PubMed: 17295102]
22. Weiler K, Christ AM, Woodworth CG, et al. Quality of patient-reported outcome data captured using paper and interactive voice response diaries in a allergic rhinitis study: is electronic data capture really better? *Ann Allergy Asthma Immunol* 2004;92(3):335–339. [PubMed: 15049397]
23. Shaw WS, Verma SK. Data equivalency of an interactive voice response system for home assessment of back pain and function. *Pain Res Manag* 2007;12(1):23–30. [PubMed: 17372631]
24. Chan KS, Orlando M, Ghosh-Dastidar B, et al. The interview mode effect on the Center for Epidemiologic Studies Depression (CES-D) scale: an item response theory analysis. *Medical Care* 2004;42:281–289. [PubMed: 15076828]
25. Cleeland C, Mendoza T, Wang X, et al. Assessing symptom distress in cancer patients. *Cancer* 2000;89:1634–1646. [PubMed: 11013380]
26. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacological Therapy* 1974;15:443–453.
27. Sikorskii A, Given C, Given B, Jeon S, Decker V, Decker D. Symptom management for cancer patients: A trial comparing two multimodal interventions. *Journal of Pain and Symptom Management* 2007;34(3):253–264. [PubMed: 17618080]
28. Lyman, G. Essentials in clinical decision analysis: A new way to think about cancer and aging. In: Balducci, L.; Lyman, G.; Ershler, W.; Extermann, M., editors. *Comprehensive Geriatric Oncology*. New York: Taylor and Francis; 2004. p. 11-25.
29. Anisimov, V. Age as a risk factor in multistage carcinogenesis. In: Balducci, L.; Lyman, G.; Ershler, W.; Extermann, M., editors. *Comprehensive Geriatric Oncology*. New York: Taylor and Francis; 2004. p. 75-101.
30. Balducci, L.; Extermann, M. Introduction. In: Muss, H.; Hunter, C.; Johnson, K., editors. *Cancer in the Elderly*. New York: Taylor and Francis; 2006. p. 1-6.
31. Bherer L, Kramer A, Peterson M, et al. Training effects on dual-task performance: are there age-related differences in plasticity of attentional control? *Psychology and Aging* 2005;20(4):695–709. [PubMed: 16420143]

32. Jones S, Nyberg L, Sandblom J, et al. Cognitive and neural plasticity in aging: General and task-specific limitations. *Neuroscience & Biobehavioral Reviews* 2006;30(6):864–871.
33. Williams M, Baker D, Parker R, Nurss J. Relationship of functional health literacy to patients' knowledge of their chronic disease: A study of patients with hypertension and diabetes. *Archives of Internal Medicine* 1998;158(2):166–172. [PubMed: 9448555]
34. Crane P, Gibbons L, Jolley L, van Belle G. Differential item analysis with ordinal logistic regression techniques. *Medical Care* 2006;44(11 Suppl 3):S115–S123. [PubMed: 17060818]
35. Fayers P, Hand D. Factor analysis, causal indicators and quality of life. *Quality of Life Research* 1997;6:139–150. [PubMed: 9161114]
36. Searle SR, Speed Fm, Miliken GA. Population marginal means in the linear model: an alternative to least squares means. *The American Statistician* 1980;34(4):216–221.
37. McCullagh, P.; Nelder, J. *Generalized Linear Models*. London: Chapman and Hall; 1989.
38. Groenvold M, Bjorner J, Klee M, Kreiner S. Test for item bias in a quality of life questionnaire. *J Clin Epidemiol* 1995;48:805–816. [PubMed: 7769411]
39. Rothman, KJ.; Greenland, S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven; 1998.
40. Thissen D, Steinberg L, Kuang D. Quick and easy implementation for the Benjamini-Hochberg procedure for controlling the false positives in multiple comparisons. *Journal of Educational and Behavioral Statistics* 2002;27(1):77–83.
41. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 2001;29(4):1165–1188.
42. SAS software, Version 9 of the SAS System for Windows. Copyright ©2002-2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
43. Sloan JA, Cella D, Hays RD. Clinical significance of patient-reported questionnaire data: Another step toward consensus. *J Clin Epidemiol* 2005;58:1217–1219. [PubMed: 16291464]
44. National Comprehensive Cancer Network. Clinical practice guidelines in oncology: Cancer-related fatigue version 1. 2006. Retrieved June 28, 2008 from: http://www.nccn.org/professionals/physician_gls/PDF/fatigue.pdf
45. Stanton A, Bernaards C, Ganz P. The BCPT Symptom Scales: A measure of physical symptoms for women diagnosed with or at risk for breast cancer. *JNCI* 2005;97(6):448–456. [PubMed: 15770009]
46. Aaronson NK, Ahmedzai S, Bergnam B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;5:365–76. [PubMed: 8433390]
47. Given CW, Given B, Azzouz F, Kozachik S, Stommel M. Predictors of pain and fatigue in the year following diagnosis among elderly cancer patients. *Journal of Pain and Symptom Management* 2001;21(6):456–466. [PubMed: 11397603]
48. Komaroff AL. Symptoms: In the head or in the brain? *Ann Intern Med* 2001;134(Part 1):783–785. [PubMed: 11329237]
49. Vogelzang NJ, Breitbart W, Cella D, et al. Patient, caregiver, and oncologist perceptions of cancer-related fatigue: results of a tripart assessment survey. *The Fatigue Coalition. Semin Hematol* 1997;34 (Suppl 2):4–12. [PubMed: 9253778]
50. Jones S, Nyberg L, Sandblom J, et al. Cognitive and neural plasticity in aging: general and task specific limitations. *Neuroscience & Biobehavioral Reviews* 2006;30:864–871. [PubMed: 16904746]

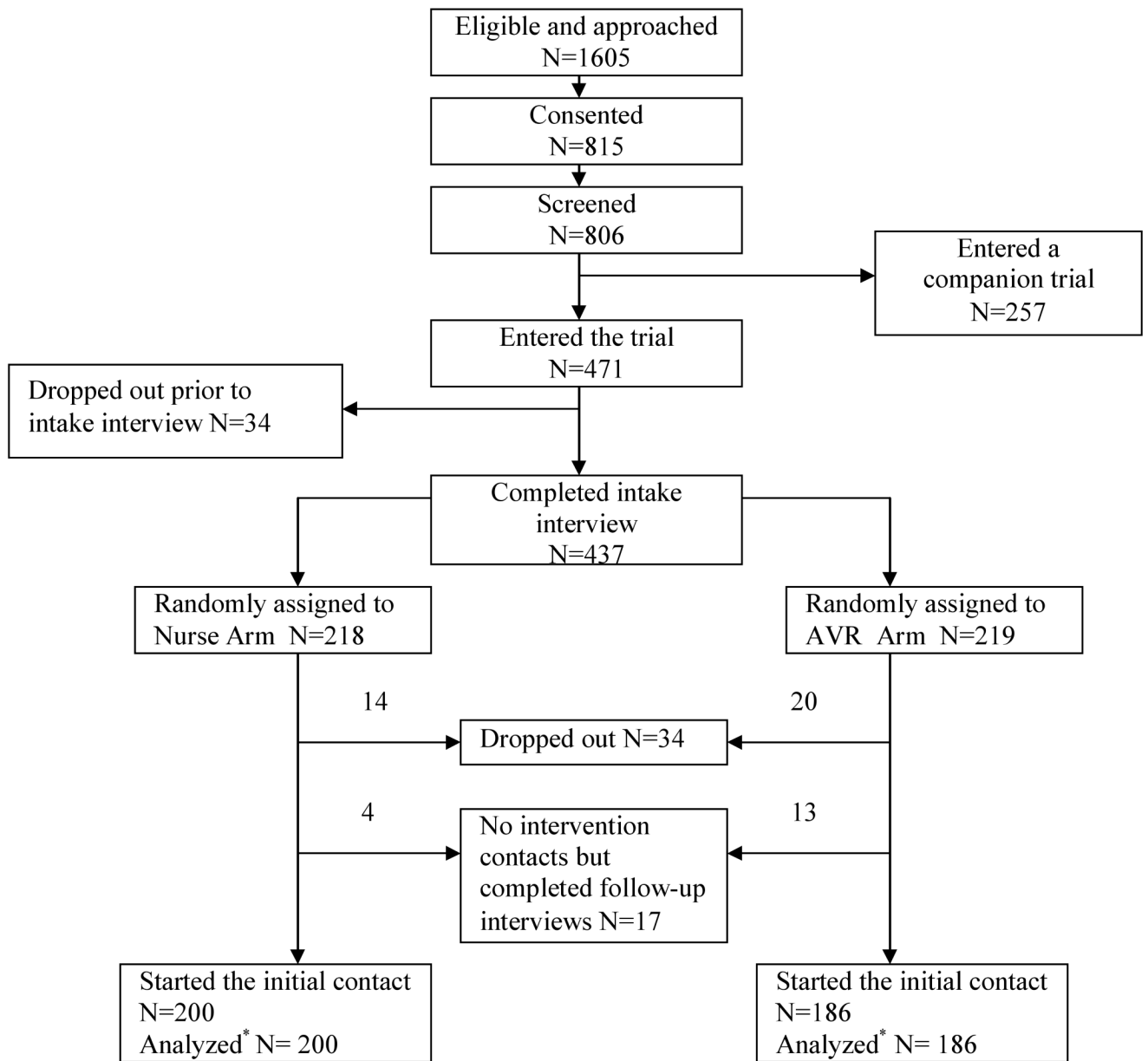


Figure 1.
Flow chart of the trial.

Table 1
 Characteristics of the sample by trial arm: those randomized and those analyzed.

Variable	Category	Randomized (N=437)		Analyzed (N=386)	
		Nurse Arm (N=218)	AVR Arm (N=219)	Nurse Arm (N=200)	AVR Arm (N=186)
Patient Age	25 ~ 44	N(%) 33 (15.14)	N(%) 31 (14.16)	N(%) 30 (15.00)	N(%) 27 (14.52)
	45 ~ 54	57 (26.15)	63 (28.77)	55 (27.50)	54 (29.03)
	55 ~ 64	72 (33.03)	74 (33.79)	64 (32.00)	62 (33.33)
	65 ~ 74	38 (17.43)	27 (12.33)	34 (17.00)	23 (12.37)
	75 +	18 (8.26)	24 (10.96)	17 (8.50)	20 (10.75)
Education	High school or below	71 (32.57)	67 (30.59)	64 (32.00)	57 (30.65)
	Some college or technical training	72 (33.03)	66 (30.14)	66 (33.00)	53 (28.49)
	College	41 (18.81)	42 (19.18)	40 (20.00)	40 (21.51)
	Graduate professional degree	34 (15.60)	44 (20.09)	30 (15.00)	36 (19.35)
Cancer Site	Colon	30 (13.76)	32 (14.61)	28 (14.00)	25 (13.44)
	Breast	90 (41.28)	87 (39.73)	83 (41.50)	77 (41.40)
	Lung	37 (16.97)	34 (15.53)	32 (16.00)	26 (13.98)
	Other	61 (27.98)	66 (30.14)	57 (28.50)	58 (31.18)
Cancer Stage	Early	63 (29.30)	80 (36.70)	59 (29.95)	69 (37.50)
	Late	152 (70.70)	138 (63.30)	138 (70.05)	116 (62.70)
Cancer Metastasis	Yes	128 (58.72)	112 (51.14)	117 (58.50)	92 (49.46)

Variable	Category	Randomized (N=437)		Analyzed (N=386)	
		Nurse Arm (N=218)	AVR Arm (N=219)	Nurse Arm (N=200)	AVR Arm (N=186)
	No	N(%) 90 (41.28)	N(%) 107 (48.86)	N(%) 83 (41.50)	N(%) 94 (50.54)
Comorbid Conditions	0~2	139 (63.76)	139 (64.47)	128(64.00)	115(61.83)
	3+	79 (36.24)	80 (36.53)	72(36.00)	71(38.17)
Sex	Male	57 (26.15)	53 (24.20)	52 (26.00)	40 (21.51)
	Female	161 (73.85)	166 (75.80)	148 (74.00)	146 (78.49)

Table 2
Severity of symptoms at intake interview and the first contact of intervention contact by trial arm.

Symptom	Intake Interview, those not analyzed		Intake Interview, those analyzed		First Intervention contact, those analyzed				
	Nurse Arm (N=18) Mean (Std)	AVR Arm (N=33) Mean (Std)	P-value	Nurse Arm (N=200) Mean (Std)	AVR Arm (N=186) Mean (Std)	P-value	Nurse Arm (N=200) Mean (Std)	AVR Arm (N=186) Mean (Std)	P-value
Pain	1.78 (3.08)	1.64 (2.62)	0.86	2.07 (2.80)	2.15 (2.82)	.78	1.66 (2.69)	2.50 (2.88)	<.01*
Fatigue	4.61 (2.97)	4.82 (2.91)	0.81	4.36 (2.71)	4.54 (2.78)	.53	3.66 (2.68)	4.28 (2.63)	.02
Peripheral Neuropathy	1.67 (2.40)	2.94 (3.15)	0.14	1.44 (2.47)	1.70 (2.60)	.32	1.14 (2.13)	1.52 (2.24)	.09
Cough	1.11 (1.88)	1.48 (2.31)	0.56	1.14 (2.14)	1.16 (2.20)	.91	1.20 (2.16)	1.18 (2.06)	.94
Dyspnea	1.78 (2.67)	1.58 (2.46)	0.79	1.53 (2.41)	1.72 (2.58)	.45	1.00 (1.99)	1.51 (2.30)	.02
Insomnia	3.50 (3.43)	3.06 (3.61)	0.67	3.53 (3.17)	3.91 (3.27)	.25	2.45 (2.93)	2.69 (2.84)	.41
Dry Mouth	3.56 (3.38)	2.79 (3.21)	0.43	2.21 (2.73)	2.68 (3.11)	.12	1.86 (2.50)	1.90 (2.56)	.89
Alopecia	1.44 (2.94)	2.30 (3.84)	0.41	1.82 (3.29)	1.90 (3.22)	.79	1.78 (2.93)	3.27 (3.84)	<.01**
Difficulty Remembering	2.22 (2.94)	1.12 (1.88)	0.16	1.48 (2.32)	1.55 (2.37)	.78	1.71 (2.29)	1.52 (2.21)	.42
Poor Appetite	2.33 (3.12)	2.91 (2.87)	0.51	2.14 (2.89)	2.87 (3.18)	.02	1.88 (2.73)	2.83 (2.81)	<.01**
Nausea / Vomiting	2.22 (2.92)	2.18 (3.18)	0.96	1.81 (2.72)	2.24 (2.88)	.14	0.79 (2.05)	1.58 (2.47)	<.01**
Diarrhea	2.67 (3.09)	1.73 (3.33)	0.33	1.29 (2.35)	1.46 (2.66)	.51	0.60 (1.78)	0.93 (1.81)	.07
Constipation	2.56 (3.48)	1.52 (2.94)	0.26	1.79 (2.76)	1.89 (2.92)	.73	1.14 (2.23)	2.16 (2.81)	<.01**
Weakness	3.72 (3.46)	3.03 (3.01)	0.46	2.33 (2.90)	2.82 (3.04)	.10	2.16 (2.69)	2.68 (2.85)	.07

* Statistically significant according to Benjamini-Hochberg procedure

** Statistically significant according to Bonferroni procedure (all comparisons that are significant with Bonferroni procedure are also significant with Benjamini-Hochberg procedure)

Table 3 Results of regression analyses relating severity at the first intervention contact to trialarm and severity at intake interview adjusting for the number of days between intake interview and first intervention contact.

	Pain		Fatigue		Peripheral Neuropathy	
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.14	78.0 <0.0001	0.10	94.9 <0.0001	0.18	179.3 <0.0001
Trial Arm	0.25	8.1 0.0045*	0.12	4.8 0.0290	0.15	5.6 0.0182
Cough						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.19	135.4 <0.0001	0.17	140.5 <0.0001	0.12	102.2 <0.0001
Trial Arm	0.05	0.7 0.4085	0.15	5.2 0.0223	0.02	0.04 0.8395
Dry Mouth						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	80.8 <0.0001	0.08	34.5 <0.0001	0.17	125.8 <0.0001
Trial Arm	-0.09	1.3 0.25	0.42	23.2 <0.0001**	-0.08	1.3 0.2594
Poor Appetite						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	98.3 <0.0001	0.14	106.2 <0.0001	0.11	72.2 <0.0001
Trial Arm	0.22	9.1 0.0027**	0.33	21.2 <0.0001**	0.19	8.1 0.0046*
Constipation						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value		
Intake severity	0.12	89.1 <0.0001	0.13	109.8 <0.0001		
Trial Arm	0.35	22.0 <0.0001**	0.11	2.3 0.1327		
Nausea/Vomiting						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	80.8 <0.0001	0.08	34.5 <0.0001	0.17	125.8 <0.0001
Trial Arm	-0.09	1.3 0.25	0.42	23.2 <0.0001**	-0.08	1.3 0.2594
Diarrhea						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	80.8 <0.0001	0.08	34.5 <0.0001	0.17	125.8 <0.0001
Trial Arm	-0.09	1.3 0.25	0.42	23.2 <0.0001**	-0.08	1.3 0.2594
Difficulty Remembering						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	80.8 <0.0001	0.08	34.5 <0.0001	0.17	125.8 <0.0001
Trial Arm	-0.09	1.3 0.25	0.42	23.2 <0.0001**	-0.08	1.3 0.2594
Alopecia						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.12	80.8 <0.0001	0.08	34.5 <0.0001	0.17	125.8 <0.0001
Trial Arm	-0.09	1.3 0.25	0.42	23.2 <0.0001**	-0.08	1.3 0.2594
Dyspnea						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.19	135.4 <0.0001	0.17	140.5 <0.0001	0.12	102.2 <0.0001
Trial Arm	0.05	0.7 0.4085	0.15	5.2 0.0223	0.02	0.04 0.8395
Insomnia						
	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value	Beta	Wald χ^2 P-value
Intake severity	0.19	135.4 <0.0001	0.17	140.5 <0.0001	0.12	102.2 <0.0001
Trial Arm	0.05	0.7 0.4085	0.15	5.2 0.0223	0.02	0.04 0.8395

* Statistically significant according to Benjamini-Hochberg procedure

** Statistically significant according to Bonferroni procedure (all comparisons that are significant with Bonferroni procedure are also significant with Benjamini-Hochberg procedure)

Table 4

Least square means of symptom severity at the first intervention contact by age and trial arm adjusted for symptom severity at intake interview and the number of days between intake interview and first intervention contact.

Symptom	Age	Nurse Arm	AVR Arm	P-value for Arm Comparison	Between Arm Effect Size
		Mean (St Error)	Mean (St Error)		
Alopecia	25 - 44	2.19 (0.72)	3.46 (0.68)	0.13	0.35 [†]
	45 - 74	1.75 (0.30)	3.23 (0.29)	<0.01**	0.45 [†]
	75 +	0.28 (0.51)	1.70 (0.55)	<0.01	0.67 [†]
Poor Appetite	25 - 44	1.49 (0.39)	1.85 (0.37)	0.47	0.11
	45 - 74	1.74 (0.23)	2.68 (0.21)	<0.01**	0.31
	75 +	2.40 (0.43)	1.67 (0.46)	0.30	-0.24
Constipation	25 - 44	1.09 (0.35)	1.58 (0.34)	0.27	0.20
	45 - 74	1.04 (0.19)	2.02 (0.17)	<0.01**	0.34 [†]
	75 +	0.76 (0.38)	1.29 (0.40)	0.27	0.18
Cough	25 - 44	0.75 (0.20)	0.69 (0.19)	0.81	-0.03
	45 - 74	0.98 (0.12)	1.20 (0.11)	0.16	0.10
	75 +	0.96 (0.22)	0.57 (0.24)	0.28	-0.17
Diarrhea	25 - 44	0.80 (0.16)	0.37 (0.16)	0.09	-0.17
	45 - 74	0.50 (0.10)	0.94 (0.10)	<0.01**	0.18
	75 +	0.31 (0.26)	0.85 (0.28)	0.09	0.19
Dry Mouth	25 - 44	1.97 (0.38)	1.72 (0.36)	0.65	-0.10
	45 - 74	1.81 (0.16)	1.57 (0.15)	0.29	-0.08
	75 +	1.63 (0.40)	1.51 (0.44)	0.84	-0.05
Dyspnea	25 - 44	0.55 (0.32)	1.55 (0.30)	<0.01**	0.42 [†]
	45 - 74	0.86 (0.12)	1.20 (0.11)	0.03	0.13
	75 +	1.82 (0.25)	0.78 (0.28)	0.03	-0.49 [†]
Fatigue	25 - 44	3.44 (0.49)	3.77 (0.47)	0.62	0.12
	45 - 74	3.37 (0.24)	4.22 (0.23)	<0.01*	0.32
	75 +	4.72 (0.55)	3.53 (0.59)	0.19	-0.39 [†]
Nausea/Vomiting	25 - 44	0.99 (0.31)	1.37 (0.30)	0.33	0.14
	45 - 74	0.65 (0.14)	1.42 (0.13)	<0.01**	0.27

Symptom	Age	Nurse Arm		AVR Arm		P-value for Arm Comparison	Between Arm Effect Size
		Mean (St Error)	Mean (St Error)	Mean (St Error)	Mean (St Error)		
Pain	75 +	0.56 (0.28)	0.83 (0.30)	0.83 (0.30)	0.48	0.13	
	25 - 44	2.29 (0.38)	1.28 (0.30)	1.28 (0.30)	0.08	-0.38 [†]	
	45 - 74	1.18 (0.27)	2.54 (0.21)	2.54 (0.21)	<0.01**	0.48 [†]	
Peripheral Neuropathy	75 +	2.62 (0.42)	1.24 (0.39)	1.24 (0.39)	0.06	-0.55 [†]	
	25 - 44	1.17 (0.26)	1.20 (0.25)	1.20 (0.25)	0.94	0.01	
	45 - 74	0.78 (0.12)	1.23 (0.11)	1.23 (0.11)	<0.01**	0.17	
Difficulty Remembering	75 +	1.81 (0.32)	1.33 (0.35)	1.33 (0.35)	0.36	-0.35 [†]	
	25 - 44	1.89 (0.29)	1.32 (0.27)	1.32 (0.27)	0.20	-0.22	
	45 - 74	1.47 (0.13)	1.35 (0.12)	1.35 (0.12)	0.51	-0.05	
Insomnia	75 +	0.89 (0.29)	1.00 (0.31)	1.00 (0.31)	0.80	0.05	
	25 - 44	2.74 (0.56)	2.94 (0.52)	2.94 (0.52)	0.78	0.06	
	45 - 74	2.05 (0.21)	2.36 (0.20)	2.36 (0.20)	0.27	0.10	
Weakness	75 +	3.90 (0.41)	1.52 (0.46)	1.52 (0.46)	<0.01*	-0.81 [†]	
	25 - 44	1.81 (0.43)	2.14 (0.40)	2.14 (0.40)	0.55	0.13	
	45 - 74	2.04 (0.21)	2.39 (0.19)	2.39 (0.19)	0.20	0.12	
	75 +	1.75 (0.49)	2.14 (0.53)	2.14 (0.53)	0.57	0.12	

* Statistically significant according to Benjamini-Hochberg procedure

** Statistically significant according to Bonferroni procedure (all comparisons that are significant with Bonferroni procedure are also significant with Benjamini-Hochberg procedure)

[†] Effect sizes exceeding .33 deemed clinically significant