A Randomized Clinical Trial of Outpatient Geriatric Evaluation and Management

Chad Boult, MD, MPH, MBA, Lisa B. Boult, MD, MPH,*† Lynne Morishita, MSN,* Bryan Dowd, PhD,‡ Robert L. Kane, MD,‡ and Cristina F. Urdangarin, MD, MPH§*

OBJECTIVES: To measure the effects of outpatient geriatric evaluation and management (GEM) on high-risk older persons' functional ability and use of health services.

DESIGN: Randomized clinical trial.

SETTING: Ambulatory clinic in a community hospital.

PARTICIPANTS: A population-based sample of community-dwelling Medicare beneficiaries age 70 and older who were at high risk for hospital admission in the future (N = 568).

INTERVENTION: Comprehensive assessment followed by interdisciplinary primary care.

MEASUREMENTS: Functional ability, restricted activity days, bed disability days, depressive symptoms, mortality, Medicare payments, and use of health services. Interviewers were blinded to participants' group status.

RESULTS: Intention-to-treat analysis showed that the experimental participants were significantly less likely than the controls to lose functional ability (adjusted odds ratio (aOR) = 0.67, 95% confidence interval (CI) = 0.47–0.99), to experience increased health-related restrictions in their daily activities (aOR = 0.60, 95% CI = 0.37–0.96), to have possible depression (aOR = 0.44, 95% CI = 0.20–0.94), or to use home healthcare services (aOR = 0.60, 95% CI = 0.37–0.92) during the 12 to 18 months after randomization. Mortality, use of most health services, and total Medicare payments did not differ significantly between the two groups. The intervention cost $1,350 per person.

CONCLUSION: Targeted outpatient GEM slows functional decline. **J Am Geriatr Soc 49:351–359, 2001.**

From the *Department of Family Practice and Community Health; †Division of History of Medicine, Department of Surgery; ‡Division of Health Services Research and Policy; and §Division of Epidemiology, University of Minnesota School of Public Health, Minneapolis, Minnesota.

Supported by grants from the National Institute on Aging and the Agency for Health Care Policy and Research (R01-AG/HS11047), the Health Care Financing Administration, and the American Academy of Family Physicians Foundation. A version of this report was presented at the annual meeting of the American Geriatrics Society, Philadelphia, Pennsylvania, May, 1999.

Address correspondence to Chad Boult, MD, MPH, MBA, Professor and Director, Aging and Geriatric Medicine Program, University of Minnesota Medical School, MMC 8152, 420 Delaware St. SE, Minneapolis, MN 55455.

Key words: geriatric evaluation and management; randomized trial; function; health services

Living is not the good, but living well.
The wise man therefore lives as long as he should, not as long as he can.
He will always think of life in terms of quality, not quantity.
—Seneca

As a cohort ages, the prevalence of disability and the cost of caring for it increase dramatically.[1,2] The retirement of the baby boom generation, which will begin in 2011, will bring to the U.S. population a pandemic of disability. If no changes in the prevalence of chronic disease occur during the coming decades, the number of functionally disabled older Americans will increase more than 300% to 7.2 million by 2049,[3] and the nation's long-term care costs will increase dramatically. Alternatively, health services projections suggest that if the prevalence of geriatric disability could be reduced by 1.5% per year, the Medicare Part A trust fund might remain solvent through 2070.[4] Actual reductions in disability exceeding the recently observed rate of 0.3% per year,[5] however, will require long-range campaigns involving biotechnological research, changes in health-related behaviors, and treatments that prevent chronic conditions from progressing to chronic disability.

Several such interventions have been developed for older people.[6,7] In comprehensive geriatric assessment (CGA), an interdisciplinary team of healthcare professionals assesses an older person's medical, functional, psychosocial, nutritional, and environmental needs; the team then creates a comprehensive plan of care that it communicates to the person's physician.[8] Perhaps because CGA has been provided to too many people who were too healthy to benefit (poor targeting) and because its recommendations have not been implemented completely,[9–11] the outcomes of many CGA programs have been disappointing.[12,13] More recently, a CGA program that was targeted to older people with specific conditions and reinforced with a campaign to maximize adherence to its recommendations suc-

0002-8614/01/$15.00

ceeded in reducing some of the deterioration of health and functional ability that often accompanies advancing age.[14] This program's effects on the use and costs of health care were not reported.

When an assessment team targets its services carefully and implements the plan of care itself, the expanded geriatric evaluation and management (GEM) process may be even more effective. Seventeen years ago, a randomized study showed that, compared with usual care, hospital-based GEM was associated with 42% fewer discharges to nursing homes and 51% fewer deaths during the year following the intervention.[15] Because inpatient care is prohibitively expensive, however, recent investigations of GEM have shifted to ambulatory settings.

All four previous randomized trials of outpatient GEM suggest that it may be effective, but most of the reported benefits are not statistically significant (see Table 1). Drawing inferences from these studies is difficult for several reasons: unrepresentative sampling frames, low statistical power, high rates of attrition (23–43%), unblinded collection of outcome data, and considerable diversity (in the populations studied, the interventions provided, and the outcomes measured).

Building on lessons from this body of research and from our earlier pilot experience,[16] we designed the present study to provide definitive information about the effectiveness and the costs of GEM in preventing disability among high-risk older outpatients.[17] We hypothesized that the experimental intervention would preserve function and lower total Medicare expenditures during the 18 months following randomization.

METHODS

Design

We conducted an 18-month randomized clinical trial to compare the effects of outpatient GEM with those of "usual" health care on functional ability and on the use and cost of healthcare services among high-risk community-dwelling older people. Secondary endpoints included depressive symptoms and mortality. The University of Minnesota's Institutional Review Board approved the study.

Recruitment

We screened 23,801 community-dwelling Medicare beneficiaries age 70 and older in Ramsey County, Minnesota, and adjacent zip codes to identify those at risk for hospital admission and functional decline. Approximately semimonthly from May 1994 until July 1996, we mailed questionnaires to new geographic cohorts of fee-for-service Medicare beneficiaries (names provided by the Health Care Financing Administration (HCFA)). The cover letter asked beneficiaries to complete and return a four-page questionnaire containing the eight-item probability of repeated admission (P_{ra}) instrument that identifies older people that are more likely to use hospitals, nursing homes, home care, emergency rooms, and medications.[18–22] Each person's responses to the eight items, when inserted into a logistic formula, generate a P_{ra} value between zero and one.

Table 1. Previous Randomized Clinical Trials of Outpatient Geriatric Evaluation and Management (GEM)

	Design					Outcomes Associated with Receipt of GEM			
Source	Population	Site	Duration of Care	Team	Duration of Follow-Up	Function	Mortality	Utilization of Services	Total Payments
U.K.[45] Tulloch et al.	age 70+, outpatients, n = 194	Home	2 years	GP, RN, health visitor	2 years	same	same	↑ referrals* ↓ hosp. LOS*	not measured
U.S.[46,47] Rubin et al.	age 70+, indigent, minority, medical inpatients, n = 178	Hospital, clinic	1 year	IM, RN, SW, psychiatrist	1 year	less decline in IADLs	same	same	↓ Medicare payments
U.S.[43,48] Burns et al.	age 65+, high-risk, male, medical inpatients, n = 128	Clinic	2 years	IM, NP, SW, pharmacist, psychologist	2 years	↑ ADLs, IADLs	↓	↓ clinic visits	not measured
U.S.[49,50] Toseland et al.	age 55+, high-risk male outpatients, n = 160	Clinic	2 years	NP, SW, geriatrician	2 years	same	same	↑ visits* ↓ hosp. days ↑ NH days	no diff. in VA costs

*$P < .01$.

GP = general practitioner; IM = internist; RN = nurse; NP = nurse practitioner; SW = social worker; NH = nursing home; LOS = length of stay; VA = Department of Veterans Affairs; IADLs = instrumental activities of daily living; ADLs = activities of daily living.

Research assistants telephoned all high-risk respondents ($P_{ra} > 0.40$) to explain the study, determine eligibility, invite participation, and obtain informed consent. Exclusion criteria included residence in a nursing home, illness requiring frequent physician visits, communication barriers (related to cognitive impairment, deafness, or inability to speak English), anticipated travel more than 3 months per year, restrictive insurance, and refusal by the primary physician to consent to the person's participation. A detailed description of the recruitment procedures for this study has been published previously.[23]

Measures

Before randomization, research assistants administered a six-item cognitive screen[24] by telephone to evaluate participants' ability to provide meaningful interviews; they then conducted 15-minute telephone interviews with the participants or, if necessary, with their proxies. Each interview included measures of functional ability (the 45-item Sickness Impact Profile: Physical Functioning Dimension[25] (SIP:PFD)) and depressive symptoms (the 30-item Geriatric Depression Scale[26] (GDS)) and individual questions about general health,[27] use of nursing home and home health services during the previous 6 months, and bed disability days (BDDs) and restricted activity days (RADs) during the previous month.[28] Proxies were not asked questions from the GDS. Previously, we determined that administering the SIP:PFD and the GDS to frail older outpatients by telephone produces scores similar to those obtained through in-person interviews ($r = 0.96$ and 0.90, respectively, $P < .001$).[24] Upon completion of these baseline interviews, we randomized participants by computer algorithm within P_{ra}-stratified blocks of eight participants. Six, 12, and 18 months later, the research assistants, still blinded to group assignment, conducted identical follow-up interviews with each participant.

To measure the use and cost of health services, we obtained HCFA's records (Standard Analytical Files) of its payments for participants' Medicare-covered health care during the 12 months before and the 18 months after randomization. We did not attempt to measure the effects of GEM on total cost of care, which includes out-of-pocket expenses, payments by supplemental insurers, and the expenses and opportunity costs incurred by the participants' informal caregivers. We monitored mortality through follow-up interviews, death records of the Minnesota Department of Health, and the National Death Index (NDI).

Interventions

Immediately after randomization, we notified the primary care physicians of the control participants that their patients were at high risk for repeated hospitalization. This ethically motivated disclosure gave the control physicians and participants the opportunity to modify the ongoing "usual care" to address this risk. Thereafter, the control participants received whatever health care they and their physicians deemed appropriate.

Participants in the GEM group progressed through a four-step enrollment and assessment process that included obtaining the primary physician's permission to participate, a home visit by a GEM social worker, and two visits to the GEM clinic—the first for a history and physical examination by a gerontological nurse practitioner (GNP) and the second for evaluation by a geriatrician and a nurse. Collectively, the team assessed medical conditions, psychosocial status, functional ability,[29,30] cognitive status,[31] nutritional risk,[32] use of alcohol,[33] social network,[34] gait and balance,[35] environmental safety, medications, advance directives, hearing, and vision. After the participant's second visit to the GEM clinic, the team met for about 15 minutes to set priorities for intervention, create a plan of care, and assign individual responsibility for specific follow-up actions.

Based on this plan, GEM participants received primary care from their GEM team (geriatrician, GNP, nurse, and social worker), including 24-hour-a-day on-call services, for several months. The team diagnosed and treated problems, adjusted medication regimens, provided counseling and health education, assisted with advance directives, and made referrals to other health professionals and community services as needed. Individual team members saw their patients approximately monthly in the clinic, after which the team met to appraise the patients' progress and to plan future interventions. Participants received free transportation, if necessary, to and from the GEM clinic.

Between visits, team members monitored and coordinated their patients' care plans through regular telephone calls. Either the participant's original primary care physician or the GEM team geriatrician, depending on the participant's preference, provided hospital care when necessary. GEM care continued until the participant's significant problems were resolved or until a stable treatment regimen that did not require the GEM team was established. On average, the GEM intervention lasted 6 months. The participant was then discharged to his or her original primary care physician, accompanied by a detailed discharge summary, recommendations for ongoing care, and (if completed) an advance directive. Additional operational details of the GEM intervention have been published previously.[36]

Analysis

We analyzed all data according to the "intention-to-treat" principle, including 46 experimental participants who dropped out of the study before being assessed or treated by the GEM team. We used *t* and chi-square statistics to compare the distributions of continuous and categorical variables, respectively, and logistic, linear, and Cox proportional hazards models to adjust for the possible confounding effects of differences between the experimental and control groups at baseline.

To evaluate functional outcomes, we measured the association between group assignment and the probability of experiencing a meaningful decline in status during follow-up after randomization. We computed SIP:PFD scores using the instrument's recommended item weights.[25] Based on previously established thresholds for clinically meaningful differences, we classified participants whose SIP:PFD scores increased by three or more points as having a deterioration in functional ability[37,38] and those whose GDS scores were 11 or higher as being possibly depressed.[26] We regarded an increase of one or more BDDs or RADs as a meaningful decrease in functional ability.[26,39]

We aggregated utilization and cost data so that each episode of care (e.g., hospital stay) was attributed entirely to the interval in which it began. We constructed multiple linear regression models of the relationship between expenditures and group assignment, adjusting for baseline SIP:PFD, GDS, general health, P_{ra}, and Medicare expenditures during the year before randomization. To account for the skewed distributions of Medicare payments in the two groups, we performed logarithmic transformations of Medicare's total expenditures for each participant.

RESULTS

Sixty-one percent of the screened population responded to the mailed survey. Compared with nonrespondents, respondents were younger (77.7 vs 80.2 years, $P < .001$), had fewer mean admissions to hospitals during the previous year (0.26 vs 0.34, $P < .001$), and were more likely to be male (29.4% vs 23.8%, $P < .001$) and white (97.0% vs 95.6%, $P < .001$). Thirty-four percent of the 1,806 eligible respondents gave informed consent to participate. Consenters did not differ significantly from refusers in race (96.5% vs 97.8% white), previous year's mean hospital admissions (0.86 vs 0.89) or Medicare payments ($9,266 vs $10,504), or mean probability of repeated admission (P_{ra}) to hospitals in the future (0.493 vs 0.498), but they were younger (79.2 vs 80.9 years, $P < .001$) and much more likely to be male (54.6% vs 38.0%, $P < .001$). (Anecdotally, women expressed greater reluctance to accept the possible change of physicians that their randomization might require.) The final study sample did not differ significantly from the general population of community-dwelling high-risk respondents in race (96.6% vs 96.2% white), previous year's mean hospital admissions (0.83 vs 0.87), or mean P_{ra} (0.494 vs 0.496), but it was younger (79.1 vs 80.5 years, $P < .01$), its Medicare payments during the previous year were lower ($8,443 vs $10,079, $P < .01$), and it included a higher percentage of men (56.9% vs 44.4%, $P < .01$).[23] Each participant had a previously established relationship with a primary care physician.

As shown in Table 2, the control and experimental groups were similar at baseline. Although not statistically significant, several small differences suggested a tendency toward worse health and functional ability among the control participants. When reanalyzed as continuous variables, three differences were statistically significant: mean SIP:PFD (14.4 vs 11.4, $P < .01$), GDS (4.9 vs 3.8, $P < .05$), and self-rated health (3.6 vs 3.4, $P < .05$). During the 18 months following the baseline, the study participants (96%) or their proxies (4%) completed 97% of the scheduled follow-up interviews.

Table 2. Characteristics of Experimental and Control Groups at Baseline

	Control Group (n = 274)	GEM Group (n = 294)	t/χ^2
Demographics			
Age, mean years (SD)	78.7 (5.8)	78.8 (5.3)	0.3
White race (%)	96.4	96.3	0.0
Male sex (%)	58.4	54.1	1.1
Functional ability			
SIP:PFD ≥30 (%)	14.1	10.0	2.2
Restricted activity days/mo., mean (SD)	1.2 (4.8)	0.6 (2.8)	1.9
Bed disability days/mo., mean (SD)	1.3 (4.2)	0.8 (2.9)	1.6
Independent residence (%)	98.9	99.3	0.3
Health			
Fair or poor self-rated health (%)	54.8	46.9	3.5
GDS ≥11 (%)	12.5	8.1	2.8
Failed cognitive screen (%)	2.6	0.7	1.2
History of diabetes mellitus (%)	26.8	25.3	0.2
coronary artery disease (%)	66.9	64.1	0.5
stroke (%)	21.9	17.7	1.6
emphysema (%)	8.6	8.1	0.0
arthritis (%)	62.4	58.2	1.1
hip fracture (%)	5.1	4.8	0.0
Pra, mean (SD)	0.5 (0.1)	0.5 (0.1)	0.9
Use of services in previous year			
Hospital admissions, mean (SD)	0.8 (1.0)	0.8 (1.0)	0.2
Physician visits, mean (SD)	4.4 (0.9)	4.4 (0.9)	0.3
Nursing home days, mean (SD)	0.8 (5.4)	1.1 (11.5)	0.4
Prescription medications, mean (SD)	4.8 (2.7)	4.4 (2.9)	1.9
Medicare expenditures, mean (SD)	$8,888 (12,942)	$8,027 (11,195)	

Notes: Higher SIP:PFD scores indicate greater disability.

None of the differences between groups is significant at $P < .05$.

GEM = geriatric evaluation and management; SD = standard deviation; GDS = geriatric depression scale; SIP:PFD = sickness impact profile:physical functioning dimension.

Table 3. Functional Ability During the 18 Months Following Randomization

		6 Months	*t*	12 Months	*t*	18 Months	*t*
SIP:PFD, mean (SD)	experimental	13.5 (15.2)		14.8 (15.5)		15.7 (17.9)	
	control	16.3 (14.4)	2.2*	18.4 (16.6)	2.6*	18.9 (17.5)	2.0*
Bed disab. days, mean (SD)	experimental	1.2 (4.6)		0.9 (3.7)		0.6 (2.8)	
	control	1.6 (5.3)	1.1	1.5 (5.2)	1.5	1.5 (5.2)	2.2*
Restricted activ. days, mean (SD)	experimental	1.8 (6.0)		2.1 (6.7)		2.1 (6.3)	
	control	2.7 (7.4)	1.5	3.4 (8.4)	2.0*	2.3 (6.6)	0.4

*$P < .05$.
Note: Higher SIP:PFD scores indicate greater disability.
SD = standard deviation; SIP:PFD = sickness impact profile:physical functioning dimension.

Functional Ability

As shown by their higher SIP:PFD, RAD, and BDD values in Table 3, the control group was significantly more disabled than the experimental group at the 6-, 12-, and 18-month follow-up interviews. The intergroup differences in SIP:PFD scores were statistically significant at all three interview times, while the differences in RADs and BDDs were significant at 12 and 18 months, respectively.

To investigate the effects of GEM on individuals' functional ability, we computed the change in each participant's SIP:PFD, RAD, and BDD scores between the baseline and each of the three follow-up interviews. As shown in Figure 1, significantly smaller percentages of GEM participants experienced meaningful deterioration in their SIP:PFD scores (i.e., increases of 3+ points) during each of the three intervals. Similarly, fewer GEM participants had increased RADs (14.2% vs 23.3%, $P < .01$) and BDDs (8.8% vs 12.3%, NS) during the year after randomization. After logistic adjustment for the possible confounding effects of participants' baseline functional and affective sta-

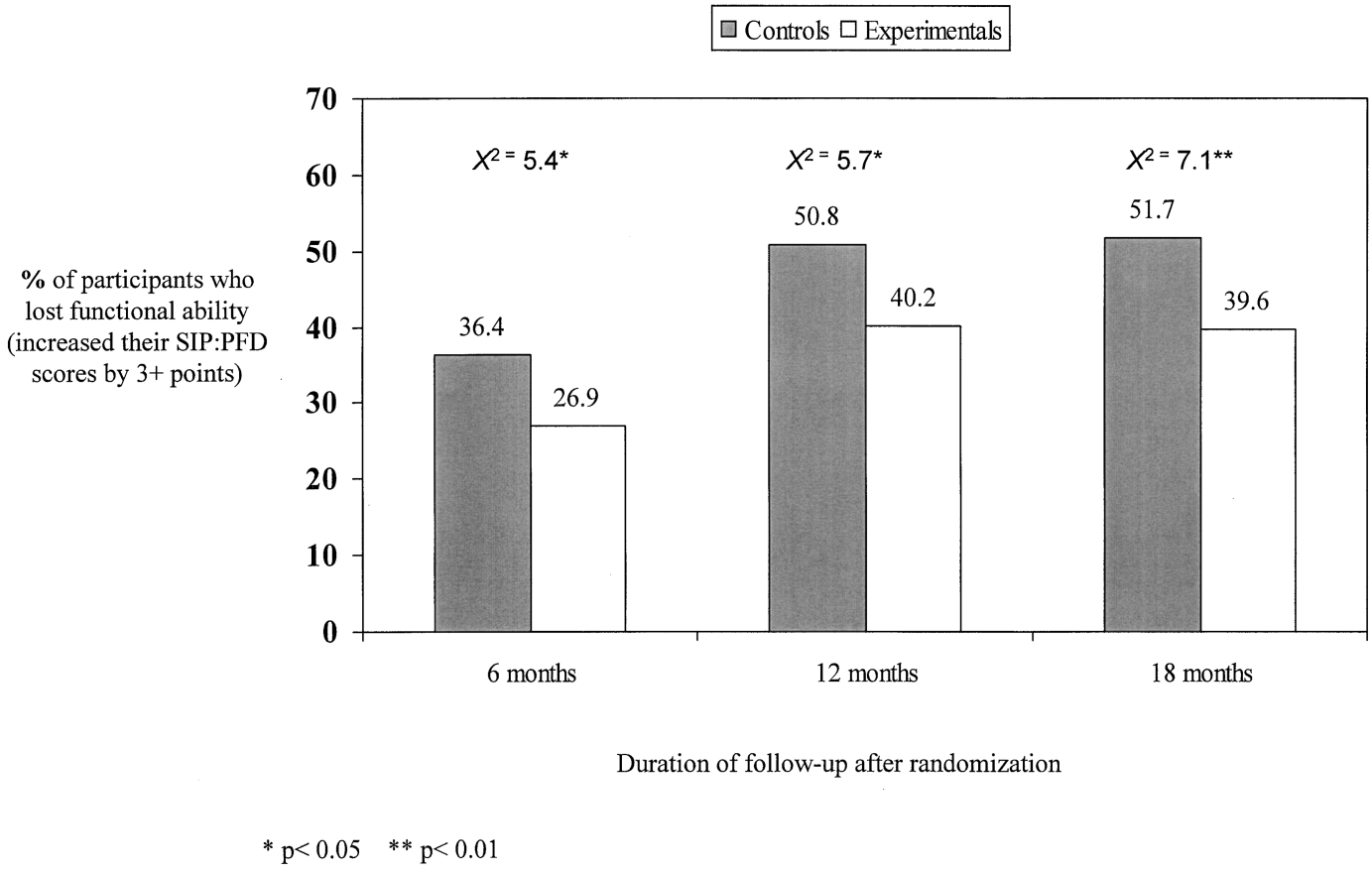

Figure 1. Loss of functional ability.

tus, GEM retained its negative association with meaningful SIP:PFD worsening at six, 12, and 18 months, with aORs of 0.62 (95% CI = 0.42–0.91), 0.73 (95% CI = 0.51–1.05), and 0.67 (95% CI = 0.47–0.99), respectively. Similarly, GEM remained significantly protective against increases in RAD during the year after randomization, aOR = 0.60 (95% CI = 0.37–0.96).

Healthcare Utilization

Logistic regression models that adjusted for baseline SIP:PFD, GDS, general health, P_{ra}, and Medicare expenditures (during the year before randomization) did not identify group assignment as a significant predictor of Medicare payments (any vs none) for any of the health services listed in Table 4. Logistic models of participants' self-reported use of home health care, adjusted for baseline use of these services, showed that GEM participants were less likely to use any home care during the 18-month follow-up period, with the difference reaching statistical significance 12 months after randomization (aOR = 0.60, 95% CI = 0.37–0.98). Self-reported use of nursing homes did not differ between the groups.

As shown in Table 4, Medicare's mean total payments during the 18 months after randomization were 3.8% greater for the recipients of usual care than for the recipients of GEM. Not surprisingly, Medicare spent more on the GEM recipients during the first 6 months of follow-up when GEM services were being provided. In contrast, it spent more on the controls from the seventh to the eighteenth month of the study.

A multiple linear regression model that adjusted for baseline SIP:PFD, GDS, general health, P_{ra}, and Medicare expenditures during the year before randomization showed that assignment to the GEM group was associated with an insignificant increase in total Medicare expenditures during the 18 months after randomization ($P = .65$), but the distribution of the model's error terms was not normal. Analogous modeling of log-transformed total Medicare expenditures revealed a statistically significant increase in payments only for GEM participants in the lowest quartile of total expenditures. There was no significant relationship between group assignment and total expenditures among the other 75% of participants. The Wilcoxon rank-sum test showed no significant difference between the experimental and control groups' total 18-month Medicare expenditures ($P = .93$). In addition to Medicare-covered services, each GEM participant underwent screening and received noncovered GEM services, costing $84 and $1,266, respectively, that were underwritten by research grants.

Depressive Symptoms

At all four measurement times, more controls than GEM participants had GDS scores suggesting depression (11+). These differences reached statistical significance 12 months (17.4% vs 9.3%, $P < .01$) and 18 months (18.3% vs 8.8%, $P < .01$) after randomization. After logistic adjustment for the possible confounding effects of participants' baseline GDS values, assignment to the GEM group retained its negative association with depression throughout the period of observation, reaching statistical significance 18 months after randomization (aOR = 0.43, 95% CI = 0.20–0.94).

Mortality

During the 18 months following randomization, 28 members of each group died. Kaplan-Meier survival analysis and Cox proportional hazards modeling (adjusting for baseline SIP:PFD, GDS, and general health) showed no significant difference between the groups' rates of mortality ($P = .88$).

The study's findings were not materially altered when we alternatively restricted our analysis to only the highest-risk subsets of the participants (i.e., those with baseline $P_{ra} \geq 0.45$ or $P_{ra} \geq 0.50$) or to only nondropout experimental participants (i.e., those who visited the GEM clinic at least twice).

DISCUSSION

The results of this trial confirm the suggestions of previous smaller studies that outpatient GEM can reduce functional decline among high-risk, community-dwelling older persons. This model of outpatient GEM also appeared to have a salutary effect on depressive symptoms, but it did not have a significant effect on mortality or Medicare payments. This is the first trial to provide an accounting of outpatient GEM's effect on Medicare's net expenditures for health care, a $1,350 per person increase.

We previously reported that recipients of this outpatient GEM intervention were more satisfied with their health care than were recipients of usual care ($P < .001$),[40]

Table 4. Medicare Payments for Health Services During the 18 Months Following Randomization

	Control Group Mean (SD)	GEM Group Mean (SD)
Inpatient hospital care	$6,819 (13,599)	$6,319 (13,647)
Physicians' care	2,260 (2,997)	2,276 (2,549)
Care in outpatient facilities	705 (1,207)	801 (1,289)
Nursing home care	958 (3,075)	899 (3,219)
Home care	777 (2,654)	793 (3,290)
Durable medical equipment	157 (681)	114 (427)
Hospice care	110 (707)	153 (1,232)
Total Medicare payments	$11,786 (19,218)	$11,354 (18,753)

SD = standard deviation; GEM = geriatric evaluation and management.

that their primary care physicians were satisfied with GEM services,[16,40] and that their family caregivers felt less burdened than did those of similar patients receiving usual care ($P < .05$).[41]

The magnitude of these intergroup differences may have been constrained by the geographic location in which the study was conducted. All of the control participants had established relationships with primary physicians in a metropolitan area known for progressive medical care. In other communities where the usual primary care of older people is less ideal, GEM may produce cost savings and even greater preservation of function.

Recruiting a large cohort of frail older people who are willing to be randomized, treated medically, and interviewed repeatedly for 18 months is difficult. Some degree of selection bias is inevitable. Nevertheless, despite the 61.1% response rate and the 34.4% consent rate, the results of this study appear to be applicable to the whole frail elderly population—at least in the Saint Paul metropolitan area. The study sample accurately represented the high-risk target population in the local community,[23] and its size (N = 586) provided the statistical power necessary to detect meaningful differences in functional change. Data collection by HCFA, NDI, and our research assistants was blinded and nearly complete (96–98%). Conservative analytic strategies such as adherence to the intention-to-treat principle and adjustment for possible baseline differences between the groups minimize the chances of attrition bias and confounding.

The disclosure of the controls' high-risk status and their participation in the study may have motivated them and their primary physicians to improve their health care, possibly biasing the study's results in favor of the control group. Furthermore, primary physicians with patients in both groups may have acquired new geriatrics expertise through their interactions with the GEM team and applied it to the care of their patients in the control group. In contrast, the experimental participants' possible desire to please their interviewers could have biased their self-rated function and affect in favor of the intervention, especially during the first follow-up interview. Inexplicably, the random assignment of eligible, consenting participants produced an experimental group that was slightly better off than the control group in health, functional ability, and previous healthcare costs. Despite our statistical adjustments for these baseline differences, other unmeasured baseline differences could have biased the observed outcomes in either direction.

Some of the properties of the measures of function used in this study have not been evaluated thoroughly. Self-reported RADs and BDDs during the previous 12 months have been shown to correlate with illness, hospitalization, and other measures of functional limitation[42] and to respond to changes in health status.[39] Nevertheless, the ideal threshold values for determining clinically meaningful changes in RADs and BDDs during the previous month, the measure used in this study, have not yet been established. Knowledge of the SIP:PFD's responsiveness to changes in the functional ability of older people is also incomplete. Among younger rheumatoid arthritis patients, mean changes in SIP:PFD scores correlate with both patients' and clinicians' judgments about the direction of clinical change, and increases of three or more points in overall SIP scores reflect clinical worsening more accurately than do clinicians' ratings on the American Rheumatologic Association's functional scale.[38] Furthermore, secondary analysis of data from the present study showed that changes in SIP:PFD scores were associated with changes in other indicators of worsening function. The participants whose SIP:PFD scores increased by three or more points between the baseline and 18-month interviews showed greater increases in RADs, BDDs, and nursing home days than those whose SIP:PFD scores remained stable or improved ($P < .01$). The validity of the GDS as a measure of change in depressive symptoms over time has not been reported.

Another limitation is the restricted scope of the financial data we analyzed. Relying mostly on Medicare payments, we were unable to quantify the effects of outpatient GEM on health-related expenses covered by other sources, such as supplemental insurance, the Medicaid program, family resources, and community-sponsored programs.

To maximize the generalizability of this study's findings, we attempted to make this GEM program as representative of a realistic community resource as possible. The clinic functioned under the typical administrative and financial constraints of a community hospital and made efficient use of existing community resources. The clinical and clerical staff of the GEM clinic were mostly hired from the community at market rates for this project. Their work remained distinct from that of the project's research staff.

The model of outpatient GEM tested here differed in two major respects from other models that have been studied. It used a reliable, validated screening instrument to identify community-dwelling older people who were likely to benefit from GEM, and it provided care through small expert teams consisting of fellowship-trained, board-certified geriatricians, as well as nurses and social workers with strong gerontological backgrounds. Its effectiveness may have been hampered, however, by the limited continuity of the care it provided. If so, still better outcomes may be achieved in the future by GEM programs that provide care for all their patients who require admission to hospitals and nursing homes, that retain responsibility for particularly complex patients indefinitely, and that offer periodic reevaluations for patients after discharge from GEM. Previous studies have suggested that the financial and functional benefits of outpatient GEM continue to accrue for up to 2 years.[43,44]

Embedding an outpatient GEM program within a coordinated, comprehensive system of health care would afford opportunities to maximize the continuity—and the cost-effectiveness—of GEM care. Future GEM programs would also benefit from more careful screening to identify the appropriate recipients of GEM, perhaps combining the efficiency of a short, mailed, self-administered questionnaire with the clinical judgment of the primary physician or GEM staff. By treating fewer false positives and missing fewer false negatives the cost-benefit characteristics of outpatient GEM would almost certainly improve further.

The benefits of outpatient GEM demonstrated in this

study—improvements in functional ability, satisfaction, and caregiver burden—are valuable, but they are not free. Like most healthcare interventions, outpatient GEM costs money, an average of about $1,350 per person for the screening, the tests, and the professional services.[36] The future availability of outpatient GEM will depend on the desires of consumers or third party payers to invest in these tertiary preventive services. If the results reported here are confirmed (or improved) in follow-up effectiveness studies, outpatient GEM could become an important force in forestalling disability among the retirees of the coming decades.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the scientific guidance provided by Aaron Folsom, MD, MPH, Phyllis Pirie, PhD, and Laurence Rubenstein, MD, MPH; the dedicated efforts of the staffs of the GEM program and the School of Public Health's Data Collection and Support Services; the technical assistance provided by Mr. David McCaffrey (data analysis); the administrative support of Mr. Scott Batulis (HealthEast Bethesda Rehabilitation Hospital); the cooperation of the primary care physicians of Ramsey County, Minnesota; and the personal contributions of the 586 randomized participants.

REFERENCES

1. Gruenberg L, Tompkins C, Porell F. The health status and utilization patterns of the elderly: Implications for setting Medicare payments to HMOs. Adv Health Econ Health Serv Res 1989;10:41–73.
2. Riley G, Lubitz J, Prihoda R et al. Changes in distribution of Medicare expenditures among aged enrollees, 1969–82. Health Care Financing Rev 1986;7:53–63.
3. Boult C, Altmann M, Gilbertson D et al. Decreasing disability in the 21st century: The future effects of controlling six fatal and non-fatal chronic conditions. Am J Public Health 1996;86:1388–1393.
4. Singer BH, Manton KG. The effects of health changes on projections of health service needs for the elderly population of the United States. Proc Natl Acad Sci USA 1998;95:15618–15622.
5. Manton KG, Corder L, Stallard E. Chronic disability trends in elderly United States populations: 1982–1994. Proc Natl Acad Sci USA 1997;94:2593–2598.
6. Boult C, Boult LB, Pacala JT. Systems of care for older populations of the future. J Am Geriatr Soc 1998;46:499–505.
7. Calkins E, Boult C, Wagner EH et al. New Ways to Care for Older People: Building Systems Based on Evidence. New York: Springer, 1998.
8. Solomon DH. Geriatric assessment: Methods for clinical decision-making. JAMA 1988;259:2450–2452.
9. Cefalu CA, Kaslow LD, Mims B et al. Follow-up of comprehensive geriatric assessment in a family medicine residency clinic. JABFP 1995;8:263–269.
10. Epstein AM, Hall JA, Fretwell M et al. Consultative geriatric assessment of ambulatory patients. JAMA 1990;263:538–544.
11. Shah PN, Maly RC, Frank JC. Managing geriatric syndromes: What geriatric assessment teams recommend, what primary care physicians implement, what patients adhere to. J Am Geriatr Soc 1997;45:413–419.
12. Rubenstein LZ, Stuck AE, Siu AL et al. Impacts of geriatric evaluation and management programs on defined outcomes: Overview of the evidence. J Am Geriatr Soc 1991;39S:8S–16S.
13. Stuck AE, Siu AL, Wieland GD et al. Comprehensive geriatric assessment: A meta-analysis of controlled trials. Lancet 1993;342:1032–1036.
14. Reuben DB, Frank JC, Hirsch SH et al. A randomized trial of outpatient comprehensive geriatric assessment coupled with an intervention to increase adherence to recommendations. J Am Geriatr Soc 1999;47:269–276.
15. Rubenstein LZ, Josephson KR, Wieland D et al. Effectiveness of a geriatric evaluation unit: A randomized controlled trial. N Engl J Med 1984;311:1664–1670.
16. Boult C, Boult L, Murphy C et al. A controlled trial of outpatient geriatric evaluation and management. J Am Geriatr Soc 1994;42:465–470.
17. Hedrick SC, Barrand N, Deyo R et al. Working group recommendations: Measuring outcomes of care in geriatric evaluation and management units. J Am Geriatr Soc 1991;39S:48S–52S.
18. Boult C, Dowd B, McCaffrey D et al. Screening elders for risk of hospital admission. J Am Geriatr Soc 1993;41:811–817.
19. Boult L, Boult C, Pirie P et al. Test-retest reliability of a questionnaire that identifies elders at risk for hospital admission. J Am Geriatr Soc 1994;42:707–711.
20. Coleman EA, Wagner EH, Grothaus LC et al. Predicting hospitalization and functional decline in older health plan enrollees: Are administrative data as accurate as self-report? J Am Geriatr Soc 1998;46:419–425.
21. Pacala JT, Boult C, Reed RL et al. Predictive validity of the P_{ra} instrument among older recipients of managed care. J Am Geriatr Soc 1997;45:614–617.
22. Pacala JT, Boult C, Boult L. Predictive validity of a questionnaire that identifies elders at risk for hospital admission. J Am Geriatr Soc 1995;43:374–377.
23. Boult C, Boult L, Morishita L et al. Soliciting defined populations to recruit samples of high-risk older adults. J Gerontol A Biol Sci Med Sci 1998;53A:M379–M384.
24. Morishita L, Boult C, Ebbitt B et al. Concurrent validity of administering the geriatric depression scale and the physical functioning dimension of the sickness impact profile by telephone. J Am Geriatr Soc 1995;43:680–683.
25. Bergner M, Bobbitt RA, Carter WB et al. The sickness impact profile: Development and final revision of a health status measure. Med Care 1981;19:787–805.
26. Yesavage JA, Brink TL. Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res 1983;17:37–49.
27. Kovar MG, Fitti JE, Chyba MM. The longitudinal study of aging. Vital Health Stat 1992;Series 1(28):1–248.
28. Kovar MG, Poe G. The design (1973–84) and procedures (1975–83) of the National Health Interview Survey. Vital Health Stat 1987;Series 1(21):1–115.
29. Duke University. Multidimensional Functional Assessment: The OARS Methodology. Durham, NC: Duke University, 1978.
30. Katz S, Ford AB, Moskowitz RW et al. Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychological function. JAMA 1963;185:914–919.
31. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198.
32. White JV, Dwyer JT, Posner BM et al. Nutrition screening initiative: Development and implementation of the public awareness checklist and screening tools. J Am Diet Assoc 1992;92:163–167.
33. Mayfield D, McLeod G, Hall P. The CAGE questionnaire: Validation of a new alcoholism screening instrument. Am J Psychiatry 1974;131:1121–1123.
34. Lubben JE. Assessing social networks among elderly populations. Fam Community Health 1988;11:42–52.
35. Podsiadlo D, Richardson S. The timed "up and go": A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142–148.
36. Boult C, Boult L, Morishita L et al. Outpatient geriatric evaluation and management (GEM). J Am Geriatr Soc 1998;46:296–302.
37. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chron Dis 1986;39:897–906.
38. Deyo RA, Inui TS. Toward clinical applications of health status measures: Sensitivity of scales to clinically important changes. Health Serv Res 1984;19:275–289.
39. Wagner EH, Lacroix AZ, Grothaus LC et al. Responsiveness of health status measures to change among older adults. J Am Geriatr Soc 1993;41:241–248.
40. Morishita L, Boult C, Boult L et al. Satisfaction with outpatient geriatric evaluation and management (GEM). Gerontologist 1998;38:303–308.
41. Weuve JL, Boult C, Morishita L. The effect of outpatient geriatric evaluation and management on caregiver burden. Gerontologist 2000;40:429–436.
42. Scholes D, Lacroix AZ, Wagner EH et al. Tracking progress toward national health objectives in the elderly: What do restricted activity days signify? Am J Public Health 1991;81:485–488.
43. Burns R, Nichols LO, Martindale-Adams J et al. Interdisciplinary geriatric primary care evaluation and management: Two-year outcomes. J Am Geriatr Soc 2000;48:8–13.
44. Toseland RW, O'Donnel JC, Engelhardt JB. Outpatient geriatric evaluation and management: Is there an investment effect? Gerontologist 1997;37:324–332.
45. Tulloch AJ, Moore V. A randomized controlled trial of geriatric screening and surveillance in general practice. J R Coll Gen Pract 1979;29:733–742.
46. Rubin CD, Sizemore MT, Loftis PA et al. A randomized, controlled trial of outpatient geriatric evaluation and management in a large public hospital. J Am Geriatr Soc 1993;41:1023–1028.
47. Rubin CD, Sizemore MT, Loftis PA et al. The effect of geriatric evaluation

and management on Medicare reimbursement in a large public hospital: A randomized clinical trial. J Am Geriatr Soc 1992;40:989–995.
48. Burns R, Nichols LO, Graney MJ et al. Impact of continued geriatric outpatient management on health outcomes of older veterans. Arch Intern Med 1995;155:1313–1318.
49. Engelhardt JB, Toseland RW, O'Donnell JC et al. The effectiveness and efficiency of outpatient geriatric evaluation and management. J Am Geriatr Soc 1996;44:847–856.
50. Toseland RW, O'Donnell JC, Engelhardt JB et al. Outpatient geriatric evaluation and management: Results of randomized trial. Med Care 1996;34:624–640.