REVIEWS

Examining the Evidence: A Systematic Review of the Inclusion and Analysis of Older Adults in Randomized Controlled Trials

Donna M. Zulman, MD, MS^{1,2,3}, Jeremy B. Sussman, MD, MS^{1,2,3}, Xisui Chen⁴,
Christine T. Cigolle, MD, MPH^{2,5,6}, Caroline S. Blaum, MD, MS^{2,6}, and Rodney A. Hayward, MD^{1,2,3}

¹The Robert Wood Johnson Foundation Clinical Scholars Program, University of Michigan, Ann Arbor, MI, USA; ²Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA; ³Department of Veterans Affairs, Health Services Research and Development Center of Excellence, Ann Arbor, MI, USA; ⁴College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI, USA; ⁵Department of Family Medicine, University of Michigan, Ann Arbor, MI, USA; ⁶VA Ann Arbor Healthcare System, Geriatric Research Education and Clinical Center (GRECC), Ann Arbor, MI, USA.

BACKGROUND: Due to a shortage of studies focusing on older adults, clinicians and policy makers frequently rely on clinical trials of the general population to provide supportive evidence for treating complex, older patients.

OBJECTIVES: To examine the inclusion and analysis of complex, older adults in randomized controlled trials.

REVIEW METHODS: A PubMed search identified phase III or IV randomized controlled trials published in 2007 in JAMA, NEJM, Lancet, Circulation, and BMJ. Therapeutic interventions that assessed major morbidity or mortality in adults were included. For each study, age eligibility, average age of study population, primary and secondary outcomes, exclusion criteria, and the frequency, characteristics, and methodology of age-specific subgroup analyses were reviewed.

RESULTS: Of the 109 clinical trials reviewed in full, 22 (20.2%) excluded patients above a specified age. Almost half (45.6%) of the remaining trials excluded individuals using criteria that could disproportionately impact older adults. Only one in four trials (26.6%) examined outcomes that are considered highly relevant to older adults, such as health status or quality of life. Of the 42 (38.5%) trials that performed an age-specific subgroup analysis, fewer than half examined potential confounders of differential treatment effects by age, such as comorbidities or risk of primary outcome. Trials with age-specific subgroup analyses were more likely than those without to be multicenter trials (97.6% vs. 79.1%, $p < 0.01$) and funded by industry (83.3% vs. 62.7%, $p < 0.05$). Differential benefit by age was found in seven trials (16.7%).

CONCLUSION: Clinical trial evidence guiding treatment of complex, older adults could be improved by eliminating upper age limits for study inclusion, by reducing the use of eligibility criteria that disproportionately affect multimorbid older patients, by evaluating outcomes that are highly relevant to older individuals, and by encouraging adherence to recommended analytic methods for evaluating differential treatment effects by age.

Electronic supplementary material The online version of this article (doi:10.1007/s11606-010-1629-x) contains supplementary material, which is available to authorized users.

Received July 12, 2010

Revised November 29, 2010

Accepted December 28, 2010

Published online February 1, 2011

KEY WORDS: clinical trial methodology; exclusion criteria; subgroup analysis; comorbidities.

J Gen Intern Med (2011) 26:783–90

DOI: 10.1007/s11606-010-1629-x

© Society of General Internal Medicine 2011

BACKGROUND

The aging of the United States population has created a pressing need for evidence to guide clinical practice for older adults, especially those with multiple comorbidities or functional impairment. Although older individuals account for the major share of health care utilization and expenditures,^{1,2} a review of high-profile journals in 2004 found that only 5% of studies focused on older adults.³ Unfortunately, funding remains inadequate to sufficiently expand geriatrics research.⁴ Clinicians and policy makers who make treatment decisions about older patients therefore often rely on evidence from randomized controlled trials (RCTs) conducted primarily among middle-aged adults. While extrapolating these findings to relatively healthy older adults is appropriate in many circumstances, extrapolating these findings to individuals with complex multimorbidities or cognitive or functional impairment can be hazardous.^{5,6}

One of the challenges in treating older patients is their heterogeneity in multiple domains, including overall health and functional status, severity of individual disease states, need for assistance with daily living, and presence of geriatric syndromes such as impaired cognition or incontinence. These characteristics can affect both the efficacy and risks of a particular intervention, and also the ability of a patient to implement a treatment or successfully complete self-management tasks. As a result, clinical trials have historically excluded individuals over a certain age, or excluded individuals who are frail or have a high morbidity burden. The consequence of this, however, is that subjects enrolled in clinical trials, even those in the oldest cohort, often do not represent older patients in the general population.⁷ This phenomenon has been demonstrated in a variety of disease states, including cancer,^{8–10} acute coronary syndrome,¹¹ chronic kidney disease,¹² cardiovascular disease risk in diabetes,¹³ congestive heart failure,¹⁴ and substance abuse.¹⁵

In recognition of the need for evidence pertinent to older adults, the Food and Drug Administration (FDA) and the Agency for Healthcare Research and Quality (AHRQ) have actively promoted clinical trial practices that yield more useful information for treating older individuals.^{16,17} Clinical researchers, in turn, are increasingly conducting analyses to examine whether differential treatment effects are present in older trial participants. These age-specific subgroup analyses occasionally lead researchers to conclude that a treatment is more or less efficacious in older adults than in their younger counterparts. In some circumstances the results of these analyses have rectified the under-use of treatment in older patients.¹⁸ To date, however, there has been no systematic evaluation of age-specific subgroup analyses and their contribution to our evidence base for treating older adults.

We sought to examine the inclusion and analysis of complex, older patients in clinical trials. We conducted a systematic review¹⁹ of RCTs in five high-profile journals in 2007 to assess the representation and analysis of older adults in major clinical research. Our objective was to better understand current practices and to provide recommendations for policy makers and researchers conducting clinical trials.

relevant articles published in 2007 in five journals (Lancet, JAMA, NEJM, Circulation, and BMJ). These journals were selected based on their high impact factors and their tendency to publish studies of disease processes that are managed by general internists (hence the inclusion of Circulation but not other high-profile subspecialty journals). A medical librarian validated the search strategy.

Inclusion and Exclusion Criteria

We included all phase III and phase IV RCTs of treatment strategies for which the primary outcome included major morbidity or mortality. We defined major morbidity as conditions that increase the risk of mortality or that result in substantive reduction in quality of life (e.g., hospitalization, a major procedure, chronic pain, depression or other serious mental health condition). Outcome data were reviewed for inclusion criteria by two authors (DZ, JS) who were blinded to article identifiers. Discrepancies were resolved by consensus with a third author (RH). Trials were excluded if they focused exclusively on children or young adults, or if the randomization unit was something other than individual patients (Fig. 1).

REVIEW METHODS

We conducted a PubMed search in 2008 using the terms "Clinical Trial," "RCT," "Phase III," and "Phase IV" to extract all

Data Abstraction

We constructed an abstraction document using an iterative revision process in which two reviewers (DZ, JS) compared

Figure 1. Study selection process.

abstracted information and clarified questions until consistency was achieved. At least two authors (DZ and JS or XC) abstracted relevant study methodology information for each of the primary articles. Abstracted information included the mean, median, and range of age in each study population, eligibility criteria that might disproportionately affect older individuals (i.e., exclusion based on inability to give informed consent, cognitive impairment, functional limitations, or decreased life expectancy), and information about study design (conditions of interest, intervention type, single country vs. international, single center vs. multicenter, and sponsoring agency). All primary and secondary outcomes were examined for measures of health status, physical function, and quality of life.

Discrepancies in abstracted information were resolved by consensus between two reviewers, or when two or more discrepancies occurred ($n=5$, 4.6%), by discussion with a third reviewer (JS, XC, or RH). Results from multiple trials presented in a single article were combined when appropriate (for example by reporting the average age from both trials). If a methodology article was cited ($n=40$, 36.7%), this article served as a secondary source when information was not available in the primary article. The ClinicalTrials.gov website was utilized to abstract information about eligibility criteria that was not otherwise available ($n=1$, 0.9%).

One author (DZ) abstracted additional information from the subset of articles that reported performing subgroup analyses. These abstractions were validated by a 15% audit by a separate reviewer (JS) ($\kappa=0.72-1.00$), and questions were resolved by discussion among three authors (DZ, JS, RH). We defined a subgroup analysis as a statistical analysis evaluating treatment differences for a specific endpoint among different levels of one baseline factor.²⁰ We recorded the number of subgroup analyses reported and examined whether established guidelines for subgroup analyses were followed,^{21,22} such as whether it was clear that each subgroup analysis was prespecified or post hoc, whether interaction tests were used to assess heterogeneity of treatment effect, and whether results were described as being exploratory or influenced by power or multiple comparisons. We also recorded whether subgroup analyses were conducted using the study's primary outcome, whether results were reported in a way that facilitates future performance of meta-analyses, and whether significant heterogeneity was reported. For studies that conducted an age-specific subgroup analysis, we assessed whether age was analyzed as a dichotomous, categorical, or continuous variable, whether age cutoffs (if present) were prespecified, and whether additional subgroup analyses were performed to evaluate potential confounders (comorbidities, health status, or overall risk of outcome).

Finally, we searched for any articles published between January of 2007 and June of 2010 that described age-specific subgroup analyses not included in the primary articles. For all trials registered with ClinicalTrials.org or the International Standard Randomised Controlled Trial Number Register, we reviewed abstracts of papers published after the primary article. Four reports of age-specific subgroup analyses were identified, only one of which included results not captured in our review of primary articles. These results are not discussed here as they are

beyond the scope of our review, but details are available upon request.

Data Analysis

We calculated the proportion of trials with an upper age limit for inclusion in the study, the proportion with eligibility criteria that might disproportionately exclude older adults, and the proportion that examined a measure of health status, physical function, or quality of life as an outcome. We also calculated the proportion of trials that reported conducting a subgroup analysis and an age-specific subgroup analysis. We used Pearson's chi-squared test to compare characteristics of trials that did and did not perform an age-specific subgroup analysis. All analyses were performed using Stata 11.0 (StataCorp, College Station, TX).

RESULTS

A total of 109 articles met our eligibility criteria and were reviewed in full (Fig. 1, Table 1). The articles covered a range of subspecialties, with cardiovascular (40.4%) and oncologic (20.2%) conditions being most common. The majority of trials (88.1%) enrolled both men and women. There were 74 (67.9%) trials that evaluated a drug, and 23 (21.1%) trials that

Table 1. Trial Characteristics (N=109)

	n (%)
Gender eligibility	
Male	3 (2.7)
Female	10 (9.2)
Both male and female	96 (88.1)
Conditions of interest	
Cardiovascular	44 (40.4)
Oncology	22 (20.2)
Hematology	14 (12.8)
Gastroenterology	13 (11.9)
Neurology	11 (10.1)
Infectious diseases	8 (7.3)
Musculoskeletal	7 (6.4)
Nephrology	6 (5.5)
Psychology	6 (5.5)
Pulmonary	5 (4.6)
Endocrine	3 (2.8)
Gynecology	3 (2.8)
Intensive care unit	1 (0.9)
Urology	1 (0.9)
Study intervention	
Drug	74 (67.9)
Device/procedure/surgery	23 (21.1)
Other	15 (13.8)
Study scope	
Single country	54 (49.5)
International	55 (50.5)
Number of centers involved	
Single center	15 (13.8)
Multicenter	94 (86.2)
Sponsoring agencies	
Public	53 (48.6)
Industry	77 (70.6)
Philanthropy	27 (24.8)
Other	3 (2.8)
Not reported	1 (0.9)

evaluated a device, procedure, or surgery. About half of the trials (50.5%) were international in scope, and most (86.2%) were conducted at multiple centers. Approximately two thirds of the trials (70.6%) were funded fully or in part by industry, and nearly half (48.6%) received public funding.

Representation of Older Adults and Relevant Outcomes in RCTs

The mean reported average age in all trials was 61 (SD±9.2) years. The age range (full or interquartile) of the study population was reported in only 44 (40%) articles. There were 22 (20.2%) trials that reported excluding individuals over a specific age (Table 2). An additional 50 (45.6%) trials did not have an upper age limit, but did have eligibility criteria that would be expected to disproportionately exclude older adults. The most common of these criteria were decreased life expectancy (n=24, 22.0%), physical disability or functional limitations (n=20, 18.4%), and inability to give informed consent (n=12, 11.0%). Examples of other eligibility criteria that were identified as potentially leading to disproportionate

Table 2. Representation of Older Adults and Age-Relevant Outcomes in Reviewed Trials (n=109)

	n (%) [*]
Average age reported in article, mean (SD) [†]	61 (9.2)
Upper age limit specified	22 (20.2)
Age cutoff for trial exclusion	
<70	4
70–75	8
76–80	7
>80	3
Exclusion criteria that may disproportionately affect older adults	
• Physical disability or functional limitations (e.g., ECOG, WHO, Karnofsky performance scores)	20 (18.4)
• Decreased life expectancy	24 (22.0)
• Inability to give informed consent	12 (11.0)
• Age-related cognitive impairment	6 (5.5)
• Other (see below)	21 (19.3)
Other exclusion criteria that may disproportionately affect older adults	
• Preexisting medical conditions judged to preclude safe participation in study	
• A clinically significant skeletal, cardiac, or endocrine disorder	
• Judgment by the clinician that participation was unwise on the basis of patient or treatment characteristics	
• Uncontrolled medical comorbidity	
• Any condition that might decrease chance of obtaining satisfactory data	
• Evidence of poor compliance with prescribed medication	
• Living in a residential or nursing home	
• Serious concurrent illness	
• Subjects who in the opinion of the investigator had clinical conditions that may have made evaluation of the safety and efficacy of this drug difficult	
Outcomes include one or more measures of health status, function, or quality of life	26 (23.9)

^{*}Unless otherwise specified

[†]Seventy-four trials reported a mean age, 33 trials reported a median age, and 2 trials reported both. The average age reported here represents the mean of reported means and medians

exclusion of complex older adults are listed in Table 2. Approximately one in four trials (n=29, 26.6%) examined one or more outcomes that evaluated health status, physical function, or quality of life.

Frequency and Quality of Age-Specific Subgroup Analyses

Of the 69 trials reporting a subgroup analysis, roughly 60% (n=42) performed an age-specific subgroup analysis. Compared to trials with no age-specific subgroup analyses, trials that reported such an analysis were more likely to be evaluating a drug treatment (83.3% vs. 58.2%, p<0.01) and to be funded by industry (83.3% vs. 62.7%, p<0.05) and were significantly more likely to be large, multicenter, international trials (Table 3). Trials that examined age as a subgroup also tended to perform a large number of subgroup analyses, with 35 (83.3%) conducting more than 6 subgroup analyses (Table 4).

Among the 42 trials with age-specific subgroup analyses, age was analyzed as a continuous risk factor in only three studies (Table 4). More commonly, age was analyzed as either a dichotomous variable (n=22; 52.4%) or a categorical variable (n=14; 33.3%), with age groups clearly prespecified in only three instances. Most of these analyses utilized the appropriate interaction test to assess treatment effect heterogeneity by age (n=32, 91.4%) and examined the primary outcome for which the trial was powered (n=39, 92.9%). Few, however, described the analyses as exploratory (n=17, 40.5%) or underpowered (n=15, 35.7%). Trials that examined age as a subgroup did not regularly conduct additional analyses to explore potential confounders, such as comorbidities (n=19; 45.2%), health status (n=8; 19.1%), or overall risk of primary outcome (n=17; 40.5%).

Positive and Negative Subgroup Findings by Age

Among the 42 articles with age-specific subgroup analyses, 7 (16.7%) reported a statistically significant result, suggesting heterogeneity in treatment effect by age (Online Appendix Table). Of these seven trials, three performed additional subgroup analyses based on health status, three performed analyses based on comorbidities, and five performed analyses based on risk of the primary outcome. Only four of the articles described their subgroup analyses as exploratory, and only one article discussed the issue of multiple comparisons. Heterogeneity was examined using the appropriate interaction test in six of seven trials, and in one trial the analytic technique was not clear. Of the 35 trials that conducted an age-specific subgroup analysis and found no significant differential treatment effect by age, only 13 (37.1%) mentioned that the trial's subgroup analyses were exploratory, and only 12 (34.3%) discussed the issue of the analyses being underpowered.

DISCUSSION

Our investigation of phase III and IV RCTs published in high impact journals reveals several practices that may limit the

Table 3. Comparison of Trials that did and did not Conduct an Age-Specific Subgroup Analysis (SGA)

	Trials with an age-specific SGA (N=42)		Trials with no age-specific SGA (N=67)		p-value
	n	%	n	%	
Study population					
Age, mean (SD)	61.6 (8.0)		60.8 (9.9)		0.67
Trial characteristics					
Funding					
Industry	35	83.3	42	62.7	<0.05
Public	18	42.9	35	52.2	0.34
Intervention					
Drug	35	83.3	39	58.2	<0.01
Device/procedure/other	5	11.9	18	26.9	0.06
International trial	31	73.8	24	35.8	<0.001
Multicenter trial	41	97.6	53	79.1	<0.01 [†]
Study size, mean (SD)	4,347 (7,001)		1,337 (2,437)		<0.01

*Pearson's chi-squared test was used to compare the frequency of each characteristic among trials that did and did not perform age-specific SGAs

[†]Small cell size: only one single-center trial performed an age-specific SGA

Table 4. Characteristics of Trials with Age-Specific Subgroup Analyses (N=42)

	n (%)
Significant differential age effect reported*	7 (16.7)
Age variable	
Continuous	3 (7.1)
Dichotomous	22 (52.4)
Categorical	14 (33.3)
Not stated	6 (14.3)
Age groups clearly prespecified	3 (7.1)
Other age-relevant subgroup analyses performed	
Comorbidity	19 (45.2)
Health status	8 (19.1)
Overall risk of primary outcome	17 (40.5)
Characteristics of subgroup analyses [†]	
Prespecified all/some subgroups	22 (52.4)
Analyzed subgroups using an interaction test	32 (91.4)
Described subgroup analyses as exploratory	17 (40.5)
Described subgroup analyses as underpowered	15 (35.7)
Report of subgroup analyses facilitates meta-analysis	27 (64.3)
Subgroup analysis uses primary outcome	39 (92.9)
Number of subgroups analyses performed	
1	1 (2.4)
2–5	6 (14.3)
6–10	23 (54.8)
11–23	12 (28.6)

*All of these trials reported that subgroup analyses revealed heterogeneity of treatment effect by age, meaning that the effectiveness of treatment varied by age. In four trials, a treatment appeared to benefit the overall population, but subgroup analyses revealed that older adults did not benefit in terms of: (a) the primary outcome overall survival, (b) the primary outcome progression-free survival, (c) the secondary outcome recurrence, or (d) in terms of net benefit as assessed in a post hoc secondary safety analysis. In two trials, overall results were negative, but subgroup analyses revealed that older adults either benefitted in terms of all-cause mortality, or had worse outcomes in terms of the secondary endpoints of death, myocardial infarction, or stroke. Finally, in one trial, the overall population benefitted from a treatment in the as-treated analysis (but not the intention-to-treat analysis), and younger patients appeared to benefit more from a treatment at early but not later time points. Additional details about these seven trials can be found in the supplementary online appendix

[†]Refers to quality of all reported subgroup analyses (not just those examining age)

relevance of study findings to older individuals. First, approximately 20% of clinical trials exclude older adults from participation based on age criteria alone, and nearly half of the remaining trials employ non-age-based eligibility criteria that could disproportionately exclude older adults with complex health status. Second, trials rarely assess highly age-relevant outcomes such as functional status and quality of life. Third, subgroup analyses that examine treatment effects by age do not routinely utilize rigorous methodology. These common practices should be addressed in order to expand and strengthen the evidence guiding policy and treatment decisions for complex, older patients.

Our review of trial age limits provides an update to a previous assessment of RCTs between 1994 and 2006 that found that 39% of trials excluded adults over the age of 65.²³ While it is encouraging that trials have reduced their use of explicit upper age limits, eligibility criteria that could disproportionately exclude older individuals, such as functional limitations, cognitive impairment, and clinician discretion, remain pervasive. These exclusion criteria likely contribute to the under-representation of complex, older (and often typical) patients in large clinical trials. Trials that exclude these patients may be less likely to evaluate outcomes and side effect profiles that are clinically meaningful to older adults. In fact, we found that only one in four trials examined at least one outcome that was a measure of health status, physical function, or quality of life. While budget or resource limitations, or aims to establish efficacy, may be cited as a reason to exclude individuals in poor health or those with functional limitations, studies of interventions that are likely to be extended to patients with these characteristics should attempt to include representative individuals. Likewise, if a condition is prevalent in older individuals, or if there is a physiological reason to suspect that an intervention's effects may differ by age, then it is especially important to recruit older adults.

To our knowledge, this study is the first detailed examination of age-specific subgroup analyses. When performed appropriately, subgroup analyses can provide important information about heterogeneity of treatment effect related to baseline risk or physiology.²¹ Previous evaluations of subgroup analyses, however, have suggested that they frequently utilize

flawed statistical techniques,^{20,24–30} are underpowered,^{29,31} or make false conclusions due to multiple comparisons or inattention to potential confounders.^{20,31,32} Our review suggests that many of these missteps continue to occur in age-specific subgroup analyses. For example, fewer than half ($n=30$, 43.5%) of the reviewed studies reported that subgroups were prespecified. In addition, age was usually analyzed as a dichotomous or categorical variable ($n=36$, 85.7%), a strategy that is sometimes used to simplify the reporting of results, but one that reduces the power available to detect a treatment effect.^{31,33}

We also found that researchers' confidence in the reliability of their age-specific subgroup analyses was often overstated, with few mentioning that study findings were exploratory or underpowered. In several instances, non-significant results were used to claim that a treatment effect was consistent across age groups, without acknowledging a lack of power to detect clinically significant differential age effects. At the same time, few of the studies claiming statistically significant heterogeneity by age fully explored potential confounding reasons for heterogeneity (such as more comorbidities or poorer health status, on average, in older subjects). While the implications of these methodological problems will vary based on the circumstances of each trial, such errors can promote

misinterpretation of evidence. For example, a clinician may conclude that a treatment is safe in complex older patients, not realizing that the trial was underpowered to detect differential effects by age or that the study did not examine potential confounders. Alternatively, a clinician may think that older adults receive less benefit when the actual mechanism is comorbidity severity or poor health status, resulting in under-treatment of robust elderly patients and over-treatment of sicker younger patients.

Recent changes in FDA and funding agency guidelines reflect a growing awareness of the need for better evidence to guide treatment decisions for older individuals. At the same time, there is great interest in comparative effectiveness research, which aims to identify effective treatment strategies to inform health policy and more personalized medicine. This research is unlikely to improve quality of care in a cost-effective manner, however, if our most complex, expensive patients are not adequately represented. Researchers are now encouraged to report clinical trial data by age and to identify differences in safety or effectiveness associated with age.^{16,17} These guidelines could be strengthened further by requesting that trials report how well their study sample represents the population of interest, and by encouraging clinical justification for upper age limits and eligibility criteria that might dispropo-

Representation of older adults in clinical trials

- Trials studying conditions relevant to older adults should not restrict participation based on age.
- Eligibility criteria that may disproportionately exclude older adults should be minimized, and should be justified when used. Examples of such criteria include functional limitations, cognitive impairment, and comorbidities (unless they specifically interfere with study participation or are contraindications to receiving treatment).
- Trials studying interventions that are applicable to older adults should consider examining outcomes that are highly relevant to this patient population, including quality of life, health status, and physical function.
- Surveillance for trial-related adverse effects should be expanded to include conditions of special concern to older and frail individuals, such as polypharmacy, incontinence, and fall risk.
- Study recruitment and enrollment procedures should decrease participation burden whenever possible so as to enable involvement of individuals with high levels of frailty and morbidity.
- Characteristics of older trial participants should be described in a transparent fashion (including mean age and distribution, health status, and presence of comorbidities) to facilitate the performance of meta-analyses and to inform clinicians whether trial results are relevant for their individual patients.

Performance of age-specific subgroup analyses

- Age-specific subgroup analyses should be conducted when there is a prespecified hypothesis that a differential treatment effect may be present in older adults.
- Trials intending to definitively examine age-specific treatment heterogeneity should be designed with adequate power to detect this effect.
- In most cases, age should be analyzed as a continuous variable in order to maximize power. The exception to this is when there is a prespecified threshold above which a differential age effect is hypothesized to be present.
- Subgroup analyses should utilize an interaction test to evaluate heterogeneity by age.
- When age is one of multiple baseline characteristics evaluated in subgroup analyses, results should be adjusted to account for multiple comparisons and reduce the likelihood of false positive results.
- If a differential treatment effect is found for older adults, additional exploratory analyses should be performed to determine whether this effect can be attributed to other characteristics, such as health status, comorbidities, or overall risk of the outcome.
- Given that studies will often be underpowered to find age-related heterogeneity, results should be reported such that meta-analyses can be easily performed in the future (i.e., subgroup size, hazard ratio, and confidence interval should be reported).
- All conclusions derived from age-specific subgroup analyses should be described as exploratory, underpowered, and subject to false positive results due to multiple comparisons, when appropriate.

Text Box. Recommendations for Representation and Analysis of Older Adults in Clinical Trials

portionately affect older individuals. In addition, these guidelines could serve as a reminder that any age-specific subgroup analysis should be adequately powered and specified a priori in order to be considered a reliable finding of the study.

There are also a number of steps that research investigators can take in order to increase the relevance of study results for older patients in the general population (Text Box). For example, to the degree that it is clinically and economically feasible, studies should include multimorbid individuals of all ages reflective of the general population. Whenever possible, trial participants should also be monitored for outcomes that have greater relevance to older individuals, including quality of life, health status, and physical function, as well as side effects of special concern to these older patients, such as the exacerbation of geriatric conditions like incontinence, dementia, falls, and adverse polypharmacy effects.

In addition, we propose several guidelines for research investigators that we believe will improve the quality of age-specific subgroup analyses (Text Box). These guidelines expand on previous subgroup analysis recommendations by CONSORT and others,^{21,22,34,35} while incorporating additional principles that are especially relevant to age-specific analyses. For example, we recommend that investigators who hypothesize that an age-related heterogeneous treatment response may occur attempt to adequately power their study to detect this effect. If the study is underpowered to find this effect, nonsignificant subgroup analysis results should not be interpreted as evidence that treatment effects are consistent across age groups. A useful strategy is to optimize power by analyzing age as a continuous variable³³ (unless another functional form is hypothesized), and then to report results for different age categories for the sake of clarity. We also recommend that any differential treatment effect found in older adults be explored to determine whether this finding may be attributed to potential confounders, such as health status or comorbidities. Subgroup analyses that examine heterogeneity by overall risk (as determined by a multivariable risk prediction tool) may be especially powerful at assessing risks and benefits for older individuals with multiple comorbidities and will generally be far more effective than age alone at identifying how treatment effects differ for such individuals.^{35,36}

Findings from this systematic review may be limited by several factors. First, we selected journals for review based on their impact factor, their relevance for a general medicine audience, and their likelihood of publishing trials investigating major morbidity or mortality as an outcome. Circulation was included because it is a high-impact journal that frequently publishes RCTs that inform treatment standards in primary care (i.e., prevention and treatment of coronary artery disease, hypertension, heart failure, and stroke). To ensure that our results were not significantly biased by the inclusion of one subspecialty journal, we repeated our main analyses after excluding studies published in *Circulation*. Our primary findings did not change substantially for any of our major findings. In general, however, our results should not be generalized to trials published in most subspecialty journals.

There are also potential limitations to our proposed recommendations. Including complex, older individuals in clinical trials is not without challenges. There are likely to be costs associated with expanding trial eligibility, and with

monitoring for age-relevant outcomes and side effects. In addition, certain comorbidities, functional or cognitive limitations, and social circumstances may be contraindications to treatment in some instances. Some of these conditions may also represent significant barriers to trial participation and increase drop-out or non-adherence, especially when a trial requires multiple clinic visits or substantial changes to a person's daily routine. In certain instances, these barriers may be overcome by reducing trial participation burden, but we also recognize that it will not always be practical or feasible to expand trial eligibility or track all relevant outcomes and side effects.

Nevertheless, failing to adapt study design and analysis practices to meet the needs of our increasingly complex aging population will perpetuate the current shortage of information guiding clinical care for these patients. Our findings suggest opportunities for the FDA, funding agencies, and research investigators to improve the current evidence by increasing the representation of these individuals in clinical studies, and by adhering to appropriate analytic techniques when examining age-related heterogeneity. These measures will enhance our understanding of treatment effects in older patients and will improve the quality of evidence that informs guideline development at the policy level and treatment decisions for individual patients.

Contributors: The authors would like to thank Gupreet K. Rana, Clinical Education Librarian at the University of Michigan Taubman Medical Library, for her assistance with the literature search for this manuscript.

Funders: Donna Zulman and Jeremy Sussman are supported by the Robert Wood Johnson Foundation Clinical Scholars Program and an associated VA Advanced Fellowship. This work was also supported by the Department of Veterans Affairs Health Services Research & Development Service (QUERI DIB), and utilized the Measurement Core of the Michigan Diabetes Research and Training Center funded by DK020572 from the NIDDK. Dr. Cigolle is supported by the NIH-NCRR K12 Mentored Clinical Scholars Program at the University of Michigan and the VA Ann Arbor Healthcare System's Geriatric Research Education and Clinical Center (GRECC).

Prior Presentations: None.

Conflict of Interest: None disclosed.

Corresponding Author: Donna M. Zulman, MD, MS; The Robert Wood Johnson Foundation Clinical Scholars Program, University of Michigan, 6312 Medical Science Building I, 1150 W. Medical Center Drive, Ann Arbor, MI 48109-5604, USA (e-mail: dzulman@umich.edu).

REFERENCES

1. Hartman M, Catlin A, Lassman D, Cylus J, Heffler S. US health spending by age, selected years through 2004. *Health Aff (Millwood)*. 2008;27(1):w1-w12.
2. Machlin S. Trends in health care expenditures for the elderly age 65 and over: 2006 versus 1996. *Statistical Brief #256*. Rockville, MD. 2009.
3. McMurdo MET, Witham MD, Gillespie ND. Including older people in clinical research—benefits shown in trials in younger people may not apply to older people. *Br Med J*. 2005;331(7524):1036-7.
4. Besdine R, Boulton C, Brangman S, et al. Caring for older Americans: the future of geriatric medicine. *J Am Geriatr Soc*. 2005;53(6 Suppl):S245-56.

5. The ACCORD Study Group. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med*. 2010.
6. Thiemann DR, Coresh J, Schulman SP, Gerstenblith G, Oetgen WJ, Powe NR. Lack of benefit for intravenous thrombolysis in patients with myocardial infarction who are older than 75 years. *Circulation*. 2000;101(19):2239–46.
7. Scott IA, Guyatt GH. Cautionary tales in the interpretation of clinical studies involving older persons. *Arch Intern Med*. Apr 12;170(7):587–595.
8. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med*. 1999;341(27):2061–7.
9. Lewis JH, Kilgore ML, Goldman DP, et al. Participation of patients 65 years of age or older in cancer clinical trials. *J Clin Oncol*. 2003;21(7):1383–9.
10. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291(22):2720–6.
11. Lee PY, Alexander KP, Hammill BG, Pasquali SK, Peterson ED. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA*. 2001;286(6):708–13.
12. O'Hare AM, Kaufman JS, Covinsky KE, Landefeld CS, McFarland LV, Larson EB. Current guidelines for using angiotensin-converting enzyme inhibitors and angiotensin II-receptor antagonists in chronic kidney disease: is the evidence base relevant to older adults? *Ann Intern Med*. 2009;150(10):717–24.
13. Cigolle CT, Blaum CS, Halter JB. Diabetes and cardiovascular disease prevention in older adults. *Clin Geriatr Med*. 2009;25(4):607–41. vii–viii.
14. Masoudi FA, Havranek EP, Wolfe P, et al. Most hospitalized older persons do not meet the enrollment criteria for clinical trials in heart failure. *Am Heart J*. 2003;146(2):250–7.
15. Humphreys K, Weingardt KR, Horst D, Joshi AA, Finney JW. Prevalence and predictors of research participant eligibility criteria in alcohol treatment outcome studies, 1970–98. *Addiction*. 2005;100(9):1249–57.
16. Agency for Healthcare Research and Quality. AHRQ Policy on the Inclusion of Priority Populations in Research: agency for healthcare research and quality, US Department of Health and Human Services. 2003.
17. US Government Accountability Office. Prescription drugs: FDA guidance and regulations related to data on elderly persons in clinical drug trials (GAO-07-47R). Washington, D.C. 2007.
18. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360(9326):7–22.
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–9. W264.
20. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357(21):2189–94.
21. Rothwell PM. Treating Individuals 2—subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176–86.
22. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663–94.
23. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals—a systematic sampling review. *JAMA*. 2007;297(11):1233–40.
24. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21(19):2917–30.
25. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317(7):426–32.
26. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355(9209):1064–9.
27. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266(1):93–8.
28. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. *Am Heart J*. 2000;139(6):952–61.
29. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004;57(3):229–36.
30. Hernandez AV, Boersma E, Murray GD, Habbema JDF, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J*. 2006;151(2):257–64.
31. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol*. 2006;6:18.
32. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357(9263):1191–4.
33. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127–41.
34. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 340:c117.
35. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010;11:85.
36. Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. *Lancet*. 2005;365:256–65.

REFERENCES FOR APPENDIX TABLE

- Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, Staroslawska E, Sosman J, McDermott D, Bodrogi I, Kovacevic Z, Lesovoy V, Schmidt-Wolf IG, Barbarash O, Gokmen E, O'Toole T, Lustgarten S, Moore L. Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med*. 2007;356(22):2271–81.
- Konstam MA, Gheorghiade M, Burnett JC Jr, Grinfeld L, Maggioni AP, Swedberg K, Udelson JE, Zannad F, Cook T, Ouyang J, Zimmer C, Orlandi C. Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study With Tolvaptan (EVEREST) Investigators. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST Outcome Trial. *JAMA*. 2007;297(12):1319–31.
- Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, Perez EA, Shenkier T, Cella D, Davidson NE. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med*. 2007;357(26):2666–76.
- Quasar Collaborative Group, Gray R, Barnwell J, McConkey C, Hills RK, Williams NS, Kerr DJ. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet*. 2007;370(9604):2020–9.
- Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA, Birkmeyer NJ, Hilibrand AS, Herkowitz H, Cammisa FP, Albert TJ, Emery SE, Lenke LG, Abdu WA, Longley M, Errico TJ, Hu SS. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med*. 2007;356(22):2257–70.
- Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann FJ, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM. TRITON-TIMI 38 investigators. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med*. 2007;357(20):2001–15.
- Zacharski LR, Chow BK, Howes PS, Shamayeva G, Baron JA, Dalman RL, Malenka DJ, Ozaki CK, Lavori PW. Reduction of iron stores and cardiovascular outcomes in patients with peripheral arterial disease: a randomized controlled trial. *JAMA*. 2007;297(6):603–10.