Prevalence of Knee Symptoms and Radiographic and Symptomatic Knee Osteoarthritis in African Americans and Caucasians: The Johnston County Osteoarthritis Project

JOANNE M. JORDAN, CHARLES G. HELMICK, JORDAN B. RENNER, GHEORGHE LUTA, ANCA D. DRAGOMIR, JANICE WOODARD, FANG FANG, TODD A. SCHWARTZ, LAUREN M. ABBATE, LEIGH F. CALLAHAN, WILLIAM D. KALSBEEK, and MARC C. HOCHBERG

ABSTRACT. Objective. To report contemporary estimates of the prevalence of knee-related osteoarthritis (OA) outcomes in African Americans and Caucasians aged ≥ 45 years.

Methods. Weighted prevalence estimates for knee symptoms, radiographic knee OA, symptomatic knee OA, and severe radiographic knee OA were calculated for age, ethnic, and sex subgroups, in 3018 participants (33% African Americans, 38% men) in the baseline examination (1991–97) of The Johnston County Osteoarthritis Project, a population-based study of OA in North Carolina. Radiographic knee OA was defined as Kellgren-Lawrence radiographic grade ≥ 2, severe radiographic knee OA as grades 3 and 4, and symptomatic knee OA as knee symptoms in a knee with radiographic OA.

Results. Knee symptoms were present in 43%, 28% had radiographic knee OA, 16% had symptomatic knee OA, and 8% had severe radiographic knee OA. Prevalence was higher in older individuals and women. African Americans had slightly higher prevalence of knee symptoms, radiographic knee OA, and symptomatic knee OA, but significantly higher prevalence of severe radiographic knee OA compared to Caucasians.

Conclusion. Policy should be directed to increasing education of the public and the medical community about the high prevalence of these conditions, especially in these subgroups, to decrease their impact and ultimately prevent them. (J Rheumatol 2007;34:172–80)

Key Indexing Terms: KNEE OSTEOARTHRITIS KNEE SYMPTOMS PREVALENCE ETHNICITY

Arthritis and other rheumatic conditions are a large, growing, and newly recognized public health problem in the United States. As it is variously defined, arthritis affected nearly 43 million adults in 2002, was the leading cause of disability in 1999 in the United States, severely affected health related quality of life, and in 1997 accounted for 750,000 hospitalizations, 56 million ambulatory care visits and 9340 deaths. Direct and indirect costs attributable to arthritis and other rheumatic conditions in 1997 totaled an estimated $86.2 billion, accounting for ~1% of gross domestic product. Since arthritis and rheumatic conditions are more common among older individuals, these effects are likely to increase as the population ages.

Arthritis and other rheumatic conditions comprise more than 100 different specific diseases and conditions, but few epidemiological data are available for these specific types, making it difficult to identify high risk groups and target interventions. Osteoarthritis (OA) is the most common specific condition, affecting a conservatively estimated 21 million people in the United States in 1990. Among the various types of OA, that of the knee and hip is considered to have the largest influence because of effects on ambulation, and accounted for most of the 478,000 total knee replacements and 234,000 total hip replacements for arthritis in 2004.
Despite this, the prevalence of these conditions overall and among demographic subgroups is not well characterized. The few prevalence estimates that exist for knee OA tend to be old, e.g., the first National Health and Nutrition Examination Survey (NHANES-1, 1971-75) focused on Caucasians only, e.g., Framingham Osteoarthritis Study and Mayo Clinic; focused on narrow age ranges in women only, e.g., southeast Michigan studies; or are limited by methodological shortcomings, e.g., NHANES surveys did not use weight-bearing knee radiographs. Prevalence estimates for demographic subgroups, especially those for ethnic subgroups, suffer from the same limitations.

Further, this is a field in which the definition of knee OA is evolving. While early studies used radiographic measures only, such as the Kellgren and Lawrence scale (K-L), as the field’s standard of measurement for OA, it has been shown that many persons with radiographic OA may have no symptoms and many of those even with severe symptoms may have normal radiographs. These observations have prompted a growing movement from both the clinical and public health communities to incorporate some measure of joint symptoms into a definition of clinical or symptomatic OA. Some suggest that a focus on joint pain itself is important, even though there is probably some misclassification of other conditions, e.g., injury or periarticular bursitis or tendonitis.

Accurate prevalence estimates are critical for understanding the spectrum of disease, including which individuals are most affected. They can also help to direct intervention efforts, in order to limit the progression of the disease and reduce its effect. For example, the U.S. Healthy People 2010 objective 2.6 focuses on eliminating Black/White disparities in the use of total knee arthroplasty, which could be due simply to differences in the underlying rates of disease, although several studies have suggested this is not likely to be the case in women. The purpose of our study was to improve characterization of knee OA by reporting prevalence estimates from the Johnston County Osteoarthritis Project for 4 knee OA-related measures in African American and Caucasian men and women — joint symptoms, radiographic OA, symptomatic OA, and severe radiographic OA — for the overall population and by age, sex, and ethnicity.

**MATERIALS AND METHODS**

At the time this study was designed in 1990, Johnston County, North Carolina, had a population of about 81,000 and a rural area of about 800 square miles. It had and continues to have a high prevalence of sociodemographic subgroups at high risk for poor health outcomes. Specifically, a majority of residents (66%) lived in completely rural areas, with the remainder in small towns. African American residents and residents 60 years of age or older constituted 20% and 17% of the population, respectively. Households with limited education and lower income were also common, with 35% of individuals over 25 years of age having less than a high school diploma and roughly 30% of jobs in the county in manufacturing, service, or farming.

The Johnston County Osteoarthritis Project is a population-based prospective cohort study of knee and hip OA in African Americans and Caucasians in this county. The study was approved by the Institutional Review Boards of the University of North Carolina Schools of Medicine and Public Health and the Centers for Disease Control and Prevention. All participants gave written informed consent at the time of recruitment. The baseline evaluation that forms the basis for this report was conducted between 1991 and 1997. Details of the sampling methods are described in the Appendix. Briefly, the probability-based sample was designed to be representative of the civilian, noninstitutionalized, African American or Caucasian population, aged 45 years and older, who were residents of one of the 6 towns of Johnston County because they contained the largest proportion of African American residents, according to the 1990 U.S. Census, and because each contained a town surrounded by a largely rural area. Since the project was designed as a long-term study of ethnic differences in OA occurrence and progression, the sampling design involved oversampling of African Americans and undersampling of Caucasian women age 65 years and older, the largest subgroup in many of the targeted areas. The baseline sampling occurred from May 1991 through December 1997 and involved 2 steps: stratified simple random sampling of streets as primary sampling units and stratified subsampling of Caucasian women age 65 years or older. Of 14,297 enumerated dwellings on 707 sampled streets, 4866 contained eligible households. Of the 5138 individuals from 3874 eligible households invited to participate, 992 eligible households were not selected in the second step of sampling because they had only Caucasian women age 65 years or older, 72% (3690 individuals from 3003 eligible households) completed a first home interview, and 83% of those (3068 individuals from 2501 households) subsequently completed the clinical examination and second home interview. African Americans, women, and unmarried individuals were more likely than Caucasians, men, and married individuals, respectively, to complete the first home interview, but Caucasians, individuals older than 65 years, married individuals, and those who had completed high school were more likely than their counterparts to complete the clinical visit, after first home interview participation. There were no differences in knee or hip symptoms or employment status between those completing the clinical visit after completing the first home interview, and those who did not complete the clinical visit (see Appendix). Sample weights were calculated in several steps, including calculation of raw weights as the reciprocal of the selection probability, calculation of nonresponse adjustment, and post-stratification adjustment.

Ethnic group categorization was self-reported. Knee symptoms were assessed separately for right and left sides of the body by the following question: “On most days, do you have pain, aching, or stiffness in your (right, left) knee?” Knee symptoms were defined for analysis as an affirmative response to the above question in at least one knee. All participants had radiographic examination of the knees with the anterior-posterior view with weight-bearing and foot map positioning. Knee radiographs were read without knowledge of participant clinical status, by a single bone and joint radiologist (JBR) using the K-L radiographic atlas for overall knee radiographic grades. This scale defines radiographic OA in 5 categories. Radiographs scored as grade 0 (normal) showed no radiographic features of OA; K-L grade 1 (questionable) included a minute radiographic osteophyte of doubtful pathologic significance. Radiographs showing an osteophyte but no joint space narrowing were assigned a K-L grade 2 (mild); moderate diminution of joint space was graded K-L 3 (moderate); and K-L grade 4 (severe) was defined by severe joint space narrowing with subchondral bone sclerosis. Interrater reliability assessed with another trained radiologist and intrarater reliability for the radiologist were high (weighted kappa for interrater reliability was 0.86; kappa for intrarater reliability was 0.89), as described. Radiographic knee OA was defined for analysis as K-L grade of at least 2 in at least one knee. Severe radiographic knee OA was defined as K-L grade 3 or 4 in at least one knee. Symptomatic knee OA was defined as the presence of knee symptoms in at least one knee with corresponding radiographic knee OA in that joint.
All analyses were performed using Sasa (version 9.1) and Sudaan (version 9.0). Weighted prevalence estimates for the knee outcomes and their corresponding 95% confidence intervals were derived for African American and Caucasian men and women in 4 age categories: 45–54, 55–64, 65–74, and 75 years and older.

RESULTS
The target population of residents age 45 years and older in the 6 townships of the Johnston County Osteoarthritis Project consisted of 57.4% women and 18.5% African Americans (Table 1). A total of 46 participants (1.5%) with radiographic evidence of inflammatory arthritis in the knee (n = 17) and/or hip (n = 35), and 4 with missing knee symptoms and radiographic data were excluded, leaving 3018 for further analysis of OA outcomes. Forty-three percent reported knee symptoms, 28% had radiographic knee OA, 16% had symptomatic knee OA, and 8% had severe radiographic knee OA; the prevalence of these 4 outcomes was consistently and significantly higher for older age groups, women, and African Americans (Table 2). In stratified analyses, prevalence of these 4 knee outcomes was consistently and often significantly higher in older age groups than younger age groups for both sexes and both ethnic groups, with the exception of Caucasian men 75 and older (Table 3). Women had prevalence estimates that were consistently and often significantly higher than men for both ethnic groups and all age groups (Table 3), especially for the outcomes of knee symptoms and symptomatic knee OA among those age 75 and older (Figure 1). As Figure 1 shows, knee symptoms were reported more often than symptomatic knee OA occurred, for all sex-ethnic groups across all age groups. African Americans had consistently but often only slightly higher prevalence of these 4 knee outcomes than Caucasians for all age groups and both sexes, except for the outcome of severe radiographic knee OA, where rates were much higher among African Americans (Table 3).

DISCUSSION
In Johnston County residents aged 45 years and older, the 4 knee OA-related measures occurred at different frequencies,
with knee symptoms occurring most commonly, followed by radiographic OA, symptomatic OA (the combination of the previous 2), and severe radiographic OA. Ours is the first study to describe detailed prevalence estimates of these outcomes by age, sex, and ethnicity, and confirms previous findings that the prevalence of these conditions is associated with greater age and female sex but, in addition, shows that these associations occur in both African Americans and Caucasians.

Although differences in study populations, radiographic acquisition, and scoring techniques and methodology across studies limit strict comparisons between our results and previous reports (Table 4), some cautious inferences can be drawn. First, knee symptoms were common in this population and generally higher than frequencies observed in other studies. Andersen and colleagues reported that 18–24% of individuals aged 40 to 53 years old, 29% of African Americans and 19% of Caucasians had current knee pain, defined as "any joint pain in knees" in our study. The higher frequencies in our study could be due to a slightly broader symptom definition ("pain, aching, or stiffness" vs "significant pain" or "any joint pain in knees"), a longer time horizon ("most days" vs "6 weeks" or "past month"), geographic differences in disease rates, willingness to report symptoms, distribution of risk factors (for example, heavy occupational and everyday physical activity in a rural environment), or possibly other factors.

Second, radiographic knee OA occurred 3 to 7 times more frequently in our population than in the 1971-74 NHANES, but those radiographs were obtained in the non-weight-bearing position and may have additionally been under-read, both of which likely resulted in an underestimate of the true prevalence. The population-based Framingham Osteoarthritis Study, conducted in 1983-85, used the same weight-bearing anterior-posterior radiographic technique that we used in our study and estimated that 27% of Caucasians aged 63–70 years and 34% of those aged 70–79 years had radiographic knee OA, compared with 36% of Caucasians aged 65–74 in our study. Further subdividing our Caucasian population by age groups identical to those used in the Framingham analysis similarly demonstrated higher prevalence of radiographic knee OA in the Johnston County sample in both men and women (data not shown). Our slightly higher estimates compared to NHANES-1 and Framingham may be the result of the geographically different study populations (the continental United States and New England, respectively, vs the South), whose risk factors for OA may vary, or may reflect the increasing trends in risk factors like overweight and obesity.

More recently, and in a timeframe concomitant with our

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Knee Symptoms</th>
<th>Radiographic Knee OA</th>
<th>Symptomatic Knee OA</th>
<th>Severe Radiographic Knee OA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age Group</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Male</td>
<td>All</td>
<td>36.7</td>
<td>34.4, 39.0</td>
</tr>
<tr>
<td></td>
<td>45–54</td>
<td>29.0</td>
<td>25.0, 33.2</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>55–64</td>
<td>37.6</td>
<td>33.7, 41.6</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>65–74</td>
<td>47.0</td>
<td>42.5, 51.4</td>
<td>36.1</td>
</tr>
<tr>
<td></td>
<td>75+</td>
<td>38.0</td>
<td>30.5, 46.0</td>
<td>36.0</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>All</td>
<td>46.8</td>
<td>44.4, 49.2</td>
</tr>
<tr>
<td></td>
<td>45–54</td>
<td>37.2</td>
<td>33.6, 41.0</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>55–64</td>
<td>44.6</td>
<td>41.2, 48.2</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>65–74</td>
<td>49.2</td>
<td>44.5, 53.9</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>75+</td>
<td>64.3</td>
<td>58.7, 69.6</td>
<td>53.5</td>
</tr>
<tr>
<td>African American</td>
<td>Male</td>
<td>All</td>
<td>40.9</td>
<td>37.0, 44.9</td>
</tr>
<tr>
<td></td>
<td>45–54</td>
<td>31.3</td>
<td>25.4, 37.9</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>55–64</td>
<td>46.0</td>
<td>38.7, 53.5</td>
<td>31.0</td>
</tr>
<tr>
<td></td>
<td>65–74</td>
<td>50.6</td>
<td>43.8, 57.4</td>
<td>37.4</td>
</tr>
<tr>
<td></td>
<td>75+</td>
<td>38.8</td>
<td>29.5, 49.0</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>All</td>
<td>51.0</td>
<td>48.2, 53.8</td>
</tr>
<tr>
<td></td>
<td>45–54</td>
<td>44.7</td>
<td>40.2, 49.3</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td>55–64</td>
<td>49.5</td>
<td>44.7, 54.2</td>
<td>27.7</td>
</tr>
<tr>
<td></td>
<td>65–74</td>
<td>53.3</td>
<td>48.3, 58.1</td>
<td>44.5</td>
</tr>
<tr>
<td></td>
<td>75+</td>
<td>63.3</td>
<td>54.4, 71.4</td>
<td>60.3</td>
</tr>
</tbody>
</table>

* Weighted to the 1990 target population.
study, the 1991-94 portion of NHANES-3 showed much higher frequencies of radiographic knee OA (although again measured with non-weight-bearing radiographs) than in NHANES-I and consistent with, or only slightly lower, than in our study. Hirsch, et al reported radiographic knee OA among those 60 years of age and older to be present in 31.1% of non-Hispanic White men, 38.5% of non-Hispanic White women, 42.8% of non-Hispanic Black men, and 60.6% of non-Hispanic Black women. This compares with our study, in which radiographic knee OA among those 65 and older was noted in 36% of Caucasian men, 34–54% of Caucasian women, 37–51% of African American men, and 45–60% of African American women.

Two other studies have specifically examined ethnic differences in radiographic knee OA. In an analysis of NHANES-1 data, Anderson and Felson found that African American women were twice as likely to exhibit radiographic knee OA as Caucasian women, even after adjustment for age, body mass index, skinfold thickness, income, education, marital status, uric acid level, and smoking, while there was no significant difference between African American and Caucasian men. In southeast Michigan, women aged 40 to 53 years old, 23.2% of Black women, and 8.5% of White women had radiographic knee OA, compared with 26% and 15% respectively, in women aged 45–54 years in our study. These data, in conjunction with our results, suggest that African Americans have a higher prevalence of radiographic knee OA than Caucasians, although the ethnic differences may be more pronounced in women than in men.

Third, our estimates of symptomatic knee OA, perhaps the best outcome to use for policy purposes since it includes both symptoms and radiographic evidence of OA, were nearly twice as high for Caucasians of similar age as those reported in the Framingham Osteoarthritis Study (6–8% of men and

8–13% of women aged 60–79\textsuperscript{14}. The higher frequencies in our study are most likely due to the higher frequency of each of the individual components of this outcome — symptoms and radiographic OA. We suspect differences in the prevalence of overweight and obesity over the time course in which these 2 studies were conducted may be a significant factor explaining these observations, especially since the prevalence of obesity has increased by 10% and 7% in men and women, respectively, aged 60–74 years, between the time periods 1976–80 and 1999–2000\textsuperscript{36}.

Fourth, severe radiographic knee OA occurred 5 to 8 times more frequently in our population than in the NHANES-1 population, again likely due to the same reasons cited above. In contrast, our estimates were slightly lower than those reported in the Framingham Osteoarthritis Study, in which 12–18% of those aged 63–79 years had severe radiographic knee OA\textsuperscript{14}, compared with 8–11% of Caucasians aged 65–75 in our study. Compared to the Framingham investigators, we report higher frequencies for radiographic OA but lower frequencies for severe radiographic OA, but the observed differences between the 2 cohorts may be related to relatively small numbers of individuals in these more severely affected groups, leading to potentially less stable estimates in both studies.

This study builds on our earlier report, based on an interim sample of 1432 individuals\textsuperscript{32}, and exposes some ethnic differences that were not apparent in the smaller sample, such as the now significant differences in radiographic knee OA, defined as a K-L grade of 2 or higher\textsuperscript{32}, by ethnicity. In this larger sample, African Americans appeared to have equal or slightly more frequent knee symptoms, radiographic knee OA, and symptomatic knee OA than Caucasians in all age groups, but twice as frequent severe radiographic knee OA. The last is a critically important observation, since the severity of radiographic knee OA is a strong determinant of pain, disability, and need for total joint replacement\textsuperscript{38}. It is well-established that African Americans are much less likely than Caucasians to seek and obtain joint replacement therapy for knee OA\textsuperscript{39,40}. Our data emphasize that the ethnic disparity in the use of this intervention is not because African Americans are less likely to have severe radiographic knee OA. Our data also highlight potentially significant effects on quality of life in this group and unmet need for joint replacement, which has been shown to be a highly effective and cost-effective intervention for knee pain and disability\textsuperscript{41,42}.

Our study has several limitations. Like the few other population-based studies on knee OA, it occurred in a limited geographic region that may not be representative of the U.S. as a whole. However, over 70% of our study sample was overweight or obese, a strong risk factor for knee OA, aligning closely with current rates in the nation\textsuperscript{43}, suggesting that the sample may indeed be generalizable beyond the strict con-

---

**Table 4. Selected prevalence studies with knee symptoms and radiographic knee osteoarthritis (OA) data, United States.**

<table>
<thead>
<tr>
<th>Study</th>
<th>Time Period</th>
<th>N</th>
<th>Age Range, yrs</th>
<th>Women, %</th>
<th>African American, %</th>
<th>Knee Symptoms</th>
<th>Radiographic Knee OA</th>
<th>Symptomatic Knee OA</th>
<th>Severe Radiographic Knee OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHANES-1\textsuperscript{*13}</td>
<td>1971–75</td>
<td>5193; 315 with radiographic knee OA</td>
<td>35-74</td>
<td>53.2</td>
<td>NR</td>
<td>“ever had significant pain in knees on most days for at least 1 month”</td>
<td>Non-weight bearing anterior-posterior, KL ≥ 2</td>
<td>Knee pain and KL ≥ 2</td>
<td>KL ≥ 3</td>
</tr>
<tr>
<td>Framingham Osteoarthritis Study\textsuperscript{14}</td>
<td>1983–85</td>
<td>1805; 1424 with knee radiographs</td>
<td>63–94</td>
<td>58.5</td>
<td>0</td>
<td>“ever had pain in or around a knee of most days for at least a month”</td>
<td>Weight-bearing anterior-posterior, KL ≥ 2</td>
<td>Symptoms + radiographic knee OA in same knee</td>
<td>KL ≥ 3</td>
</tr>
<tr>
<td>NHANES-3\textsuperscript{33,37}</td>
<td>1988–94</td>
<td>6596 with knee pain data</td>
<td>60–90+</td>
<td>NR</td>
<td>NR</td>
<td>“significant knee pain on most days over the preceding 6 weeks”</td>
<td>Non-weight bearing anterior-posterior, KL ≥ 2</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>MBHS/ SWAN\textsuperscript{16,34}</td>
<td>1992/1996</td>
<td>1053</td>
<td>27–52/42–53</td>
<td>100</td>
<td>39</td>
<td>“any joint pain in knees in the past month”</td>
<td>Weight-bearing anterior-posterior, KL ≥ 2</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

* National Health and Nutrition Examination Survey. ** Michigan Bone Health Study/Study of Women’s Health Across the Nation, Michigan site. NR: not reported in this reference; KL: Kellgren-Lawrence scale\textsuperscript{17}. 
The Journal of Rheumatology 2007; 34:1

fines of the timeframe and areas sampled. Although we focused only on those age 45 years and older, these are the ages when OA begins to be detected more commonly. Our analysis did not include patellofemoral joint radiographs, which would likely increase the prevalence of radiographic outcomes and perhaps increase the concordance between radiographic outcomes and symptoms. Finally, strict comparisons between our results and previous studies are not possible due to the differences in radiographic techniques and readers, time periods, and geographic regions.

However, our study has several significant strengths. It occurred relatively recently in a well-defined population with a large sample, enriched with groups at high risk for OA outcomes, and a high rate of participation. Two ethnic groups were recruited from the same geographic location, thereby decreasing the inevitable systematic bias that occurs by comparing different ethnic groups recruited from different geographic regions. Both ethnic groups underwent identical examination using the same techniques, with very high reproducibility of the radiographic reading procedure. Importantly, the frequencies of knee or hip symptoms were similar in participants and nonparticipants, further emphasizing the importance of the high prevalence of knee OA outcomes we observed, since the sample was not biased by differential participation of symptomatic individuals eager to participate in an arthritis study. If anything, some of our estimates may even be low, since African Americans and those with low educational attainment were slightly less likely to complete the clinic examination after completing the first home interview. Finally, participants were well-characterized for OA using radiographs and symptoms, allowing 4 outcomes to be examined.

Our results have demonstrated that these 4 outcomes represent a common problem for all persons ages 45 and older, for both sexes, and for African Americans as well as Caucasians. Although these estimates can strictly apply only to the target population in which the study was conducted, it appears certain that the frequencies of these outcomes have substantially increased over the last 20 to 30 years. In the future, as our population ages and the obesity epidemic goes unchecked, the prevalence of knee OA and accompanying symptoms and disability can only be expected to increase for all ages, sexes, and ethnic groups. Reducing this impact will require educating the public about the modifiable risk factors for knee OA occurrence (overweight/obesity and joint injury) and progression (weight loss, physical activity, etc.), finding new modifiable risk factors, and developing effective interventions to treat, slow progression and ultimately prevent OA.

APPENDIX. Details of sampling strategy

The probability-based sample of the Johnston County Osteoarthritis Project was designed to be representative of the civilian, noninstitutionalized, African American or Caucasian population, aged 45 years and older who were permanent residents of one of the 6 townships of Banner, Beulah, Boon Hill, Clayton, Selma, and Smithfield for at least one year, and physically and mentally capable of completing the study’s protocol. The sampling design involved oversampling of African Americans and undersampling of Caucasian women age 65 years and older to provide adequate sample sizes for domains of interest, e.g., ethnic groups. The selection of these 6 townships out of the possible 17 from Johnston County, North Carolina, as the target population was determined by the fact that they were the townships with the largest proportion of African American residents and each contained a town surrounded by a largely rural area.

The sampling design for the project involved 2 steps. The first step was a stratified simple random sampling of streets as primary sampling units. Each street was the full length of a named thoroughfare, from start to end, e.g., Alford St. The stratification of the streets and the sampling rates differed from township to township depending on its relative size. In the 3 smallest townships, namely Beulah, Boon Hill, and Selma, the streets were stratified as urban streets and rural streets, and all the streets from each stratum were selected. In the 2 middle-sized townships, namely Banner and Clayton, the streets were also stratified as urban streets and rural streets, and a simple random sample of streets was selected from each stratum. In Smithfield, the largest township, the stratification was more extensive and consisted of 6 strata. First, the streets were classified as African American, mixed ethnicity, and Caucasian streets based upon their racial/ethnic composition. Segregation by race was almost complete with most streets either 100% African American or 100% Caucasian. The Caucasian streets were further subclassified as urban streets and rural streets. Because there were many more Caucasians than African Americans in the study area, and because Caucasians had more socioeconomic diversity than did the African Americans living in town (i.e., most of the African Americans were of low socioeconomic status), the Caucasian urban streets were further subclassified into 3 categories based on socioeconomic status (high, medium, and low). A stratified simple random sample of Smithfield’s streets was selected subject to the constraint that all African American streets, all mixed ethnicity streets, and all Caucasian rural streets were selected. Overall, 707 streets were selected out of a total of 882 streets.

Since the number of Caucasian women age 65 years and older living on the selected streets in all 6 townships was much higher than the 1990 census data suggested, a decision was made to stop recruiting them after their number had exceeded predetermined “quotas” in order to avoid a sample with excessive numbers of participants in this demographic subgroup. In this second step, the selection of Caucasian women age 65 years or older was based on a subset of the original sample from each stratum. About 745 Caucasian women age 65 years and older were further selected out of a total of 1725 already selected in the first step. Although the stratified selection was not based on random subsampling, the available data do not provide evidence for difference in age between those selected and those not selected (p = 0.719 from the Wilcoxon test).

Households were enumerated for each street in the sample, and information regarding the age, sex, ethnic group, and marital status of their members was collected. Age and ethnicity information was used to assess eligibility. Households were visited on no less than 3 occasions (weekdays and weekends, day and evening) before they were considered “not able to contact.” Proxy information from neighbors was sought about age-eligibility of residents of such households. Eligible individuals who initially declined to participate in the study were given an opportunity to convert to participation. For about 90% of the eligible households who still declined participation, information concerning education, employment status, and the presence of knee and hip symptoms was obtained from at least one member of the household.

Of 14,297 identified dwellings, 4866 were eligible; 1583 were unable to be contacted; 1038 were vacant; 5939 lacked persons of eligible age for the study; 685 included only persons physically or mentally unable to participate; 153 had a language problem; and the remaining 33 households formed a category of miscellaneous (i.e., temporarily away, deceased, no eligible respondent home, demolished/merged/not a housing unit, vacation/second home). Of those 5138 people in eligible households who were invited to participate in the study, 72% (3690 people) completed first home interviews and 83%
(3068 people) of those who completed the interview were subsequently examined at the local clinic.

Based on the sampling design, the calculation of the sample weights involved 3 steps. In the first step, raw weights were calculated as the reciprocal of the selection probability of streets as determined by the sampling design. In the second step, nonresponse adjustments were performed by multiplying the raw weights with the inverse of the predicted probabilities of response obtained from response propensity models. Since participation involved 2 steps, namely, the completion of the first home interview and subsequent participation in the clinical examination at the local clinic, 2 separate logistic regression models were fitted corresponding to each of the 2 steps. The first logistic regression model included ethnic group, sex, age group, and marital status, the variables available for all the participants and the refusals. African Americans, women, and unmarried people were more likely to complete the first home interview than were Caucasians, men, and married people, respectively. The second logistic regression model (for participation in the clinical examination, given completion of the first home interview) included ethnic group, sex, age group, marital status, and education. Caucasians, people over 65 years old, married people, and people who had finished high school were more likely than their counterparts to participate in the clinical examination after they completed the first interview. Knee symptoms, hip symptoms, and employment status were not predictive of participation in the clinical examination given their initial participation. The estimate of the probability of response was the product of the 2 step-specific estimated probabilities of response, namely, the probability of completing the first home interview and the probability of coming to the clinic, given the previous completion of the first home interview. The missing probabilities from the propensity models were imputed by the weighted response rate of the category of the strongest predictor (marital status was the first and race was the second strongest predictor in the propensity model for the home interview) in the corresponding propensity model. The strengths of the predictors were determined by the order of p values of the predictor in the propensity model. The weighted response rate was computed by total estimated sampling weights of all the eligible participants in the same category of the strongest predictor. If the first strongest predictor was missing, then we used the category of the second instead, and so on. The last step consisted of post-stratification adjustment using adjustment strata defined by township, ethnic group, sex, and age group. A total of 72 strata (6 × 2 × 2 × 3) were collapsed into 61 cells if sample counts were < 20 in a stratum. The collapsing was done by the order of predictive importance (the least to the most) for prevalence of knee and hip OA by logistic regression models. The strength of predictors were race < sex < township < age group by summing the rank of each predictor for the 2 outcomes. The 61 cells were formed by first collapsing by race, then by sex, and finally by townships. Data from the 1990 Census (the closest census date to the sampling date for the study) was used to calibrate the nonresponse adjusted weights to generate external better estimates of the relative sizes of each cell.

REFERENCES


