Femorotibial Cartilage Thickness Change Distributions for Subjects without Signs, Symptoms, or Risk Factors of Knee Osteoarthritis

Robert J. Buck[1], Don Dreher[2,5], and Felix Eckstein[3,4]

Abstract

Objective: To describe the distribution of longitudinal femorotibial cartilage thickness annualized rate of change (ΔThCtAB) from quasi–population-based studies, and to construct a reference distribution for men and women without signs, symptoms, or risk factors of knee osteoarthritis (OA). *Methods:* Segmented baseline and 1-year follow-up MRI from 43 men and 69 women of the Osteoarthritis Initiative (OAI) asymptomatic control cohort without risk factors and also baseline and 2-year follow-up data from 77 asymptomatic women of the Pfizer A9001140 study were included. The mean, standard deviation (SD), and correlation of ΔThCtAB in medial and lateral femorotibial subregions were estimated; distributions were tested for normality and for differences between cohorts and gender. *Results:* Distributions of femorotibial ΔThCtAB rates were consistent between cohorts and were normally distributed, with rates <0.7%/y. Subregion ΔThCtAB SDs were correlated with mean baseline cartilage thickness (ratio = 3%-5%). However, ΔThCtAB SD did not increase with baseline thickness when estimated for different tertiles of any given subregion, indicating the relationship may rather be due to spatial location than to baseline thickness. *Conclusions:* Distributions of (subregional) longitudinal cartilage thickness rates of change appear to be normally distributed, not significantly different from zero, and similar for different cohorts of asymptomatic subjects. Given the spatial heterogeneity of subregional cartilage change observed in OA knees, the proposed reference distribution of subregional cartilage thickness change, ΔThCtAB may be used to describe and identify structural progression (i.e., cartilage loss) in individual OA knees with greater accuracy and sensitivity than conventional approaches, such as minimal detectable difference.

Keywords

knee cartilage thickness change, nonsymptomatic cohort, reference distribution

Introduction

Although osteoarthritis (OA) is considered a disease of the entire joint, loss of cartilage thickness is still considered a hallmark of structural progression.[1] To understand the OA-related changes in cartilage thickness and accurately differentiate between OA subjects with and without structural progression (e.g., significant cartilage loss), an in-depth understanding of the behavior of longitudinal cartilage change in an asymptomatic population is required.[2-6]

A series of studies have examined cartilage thickness measurements in several femorotibial subregions.[7-12] A recent study showed cartilage loss in OA patients were confined to few subregions in each patient and that these subregions vary in location between subjects.[4] Further, the pattern of subregional change has been associated with the local biomechanical environment; that is, the location of (subregional) cartilage change has been found to be associated

with the location of meniscus lesions.[13] Hence, observing progression based on subregional changes may be more sensitive than based on thickness in the total knee, compartments, or total plates and provides insight into characteristics of spatial aspects of progression. As these regions

[1]StatAnswers Consulting LLC, Minneapolis, MN, USA
[2]MerckSerono S.A., Geneva, Switzerland
[3]Institute of Anatomy & Musculoskeletal Research, Paracelsus Medical University (PMU), Salzburg, Austria
[4]Chondrometrics GmbH, Ainring, Germany
[5]Totzke & Dreher Scientific SA, Geneva, Switzerland

Corresponding Author:
Robert J. Buck, StatAnswers Consulting LLC, 408 North 1st Street, #706, Minneapolis, MN 55401
Email: robert.j.buck@gmail.com

exhibit differences in baseline cartilage thickness[3,14] and potentially in variances of cartilage thickness change, studying subregional reference distributions may allow one to establish a common metric of structural progression across regions.

Observed change in cartilage thickness in any subject can be attributed to 1 of 3 general sources: 1) OA-related change, 2) non-OA–related change, and 3) measurement error. Estimates of the distribution of change in asymptomatic subjects without signs, symptoms, or risk factors of knee OA provide understanding of the latter 2 sources of change and put observed changes in cartilage thickness in OA knees[4,7,10-13,15,16] into perspective. In particular, interest is in estimates of total "natural" subregional variability in cartilage thickness change observed between asymptomatic subjects rather than identifying or estimating the relative importance of specific sources of variability.

Recruiting a sample of asymptomatic reference subjects for every study is costly and not necessary for all research questions. While use of reference distributions cannot replace collecting a control group in studies where their primary purpose is to compare a study population with controls, a reference distribution may be useful in exploratory studies where asymptomatic cohorts were not collected, reporting and classifying OA subjects as structural progressors or nonprogressors.[4]

A method for identifying OA-related change in an individual knee using subregional reference distributions has been proposed recently.[4] However, this study was based on a sample ($n = 77$) of only female participants. A standard reference distribution based on a larger set of subjects/studies that includes men can elucidate potential sex differences in that distribution and provide more accurate, generalizable estimates and standards. Although means and standard deviations (SDs) of femorotibial cartilage thickness change have been reported for OA cohorts,[17,18] a detailed examination of the distributional properties of longitudinal change in subregional (femorotibial) cartilage thickness in asymptomatic subjects has not been published.

The objective of this study, therefore, was to characterize the distribution (mean and SD) of the annualized rate of change in thickness, ΔThCtAB, in subjects without symptoms, signs, or risk factors of knee OA for different femorotibial compartments, cartilage plates, and subregions. The presumption was that the distribution would best be described by a multivariate distribution with separate parameters for mean and SD for each subregion in men and women. To describe the distribution and to explore whether this full model may be simplified, we studied the following questions:

1. Is there any evidence that sex, age, or body mass index (BMI) are correlated with subregional ΔThCtAB in the Osteoarthritis Initiative (OAI) asymptomatic cohort?

2. Is it reasonable to assume that mean subregional ΔThCtAB is zero for all subregions in asymptomatic cohorts?

3. Are variances for ΔThCtAB similar for men and women?

4. Are distributions of ΔThCtAB for the 2 cohorts (see above) normally distributed, and are cohort results sufficiently similar to be viewed as random samples from the same population?

5. Are variances for ΔThCtAB similar for all subregions, and if not, can one describe a relationship between them?

6. How are rates of change in the different femorotibial subregions correlated?

7. What are the benefits and issues in constructing a reference distribution from the above samples?

Methods

One sample studied included baseline and 1-year follow-up MRI data from a single knee of subjects from the asymptomatic control cohort[14] of the OAI (public-use data sets 0.F.1 and 1.F.1). Of 122 participants, 112 (69 women, 43 men) had complete baseline and 1-year follow-up images suitable for cartilage thickness analysis.[14] Inclusion criteria were the following:

- No pain, aching, or stiffness in either knee in the past year.
- No radiographic findings of femorotibial OA of either knee using baseline bilateral fixed flexion radiographs.[19]
- No risk factors for onset of knee OA, including obesity, history of knee injury, knee surgery, family history of total knee replacement, Heberden nodes, or repetitive knee bending.[14] For obesity, weight was restricted to <77 or <81.6 kg for women <70 or ≥70 years, respectively, while men were restricted to 93 and 97.5 kg for the respective age groups.

The other sample studied, the Pfizer A9001140 study,[2,3,12,20] included baseline, 3-month, and 2-year follow-up images of a single knee from asymptomatic subjects who were age matched to a quasi–population-based sample of OA subjects (age = 41-75 years). Asymptomatic participants were free of knee pain, had BMI ≤28, and originally included subjects with a Kellgren-Lawrence grade (KLG) of 0 in anteroposterior (AP) radiographs[2,3,12,20] ($n = 97$). For this study, only subjects with a KLG of 0 in either AP or Lyon-Schuss (LS) radiographs were included ($n = 77$) to minimize the chance of including subjects with OA.

MRI Sequence and Image Analysis

Cartilage morphometric measurements were obtained from double oblique coronal water-excited spoil gradient-recalled (SPGRwe) images: OAI images had 1.5-mm slice thickness and were from 4 centers with 3.0-T Siemens Trio scanners (Munich, Germany).[21,22] Images from A9001140 had 1.0-mm slice thickness and were from seven 3.0-T scanners, 3 Siemens Trio and 4 General Electric (Fairfield, CT).[2,20] Manual segmentation of medial and lateral tibial (MT, LT) and medial and lateral weightbearing femoral (cMF, cLF) cartilage plates was performed by a team of expert readers with several years' experience in cartilage segmentation.[2,20-22]

Baseline cartilage thickness and longitudinal change in cartilage thickness were determined for medial and lateral femorotibial compartments (MFTC = MT + cMF; LFTC = LT + cLF), cartilage plates (MT, LT, cMF, cLF), and 16 femorotibial subregions (5 tibial: central [c], external [e], internal [i], anterior [a], posterior [p]; and 3 femoral: c, e, and i) in the medial and lateral compartments and for a central compartment summary (cMFTC = cMT + ccMF; cLFTC = cLT + ccLF).[9,10] Baseline and follow-up images were read in pairs, with readers blinded to acquisition order. Annualized rate of change in cartilage thickness (ΔThCtAB) was measured as the difference between follow-up and baseline divided by time between visits. Intrareader test-retest precision errors (CV%) in femorotibial subregions were previously found to range from 1.5% to 4.7%.[9]

Statistical Analysis

Estimates of mean and SD of ΔThCtAB, SD(ΔThCtAB), are provided for all regions. Rates of change are reported to compare study results of the OAI control cohort[23] with A9001140 study results[4,12] and to construct reference distributions for annual ΔThCtAB (mm/y). To directly compare estimates of SD(ΔThCtAB) between cohorts, they were multiplied by study length. Correlations of ΔThCtAB between subregions were computed. Paired *t* tests were used to test whether regional thickness rate of change was different from zero. Two-sample *t* tests were used to test whether cohorts had different mean values and *F* tests to test whether they had different variances. Linear regression was used to test whether differences in mean change existed between sex, age, or BMI. Adjustments for multiple comparisons were made using Bonferroni methods.

Distribution normality was assessed by examining normality plots and the Shapiro-Wilk test for the 16 subregions of OAI men and women separately. Multiple tests across regions were adjusted by using false discovery rate methods and α = 0.05.[24] All analysis was done using S-Plus (version 6.2, TIBCO Software, Palo Alto, CA).

Table 1. Summary of Demographic Characteristics of Osteoarthritis Initiative (OAI) Men and Women and A9001140 Women[12]

| | OAI control cohort | | | | A9001140 | |
| | Men (n = 43) | | Women (n = 69) | | Women (n = 86) | |
	Mean	SD	Mean	SD	Mean	SD
Age, y	57.0	9.6	53.8	6.0	55.9	8.8
Height, cm	174.7	6.9	163.5	6.5	165.4	6.9
Weight, kg	79.3	8.2	61.9	8.2	67.0	11.5
Body mass index	26.1	3.0	23.1	2.5	24.4	3.3

Results

Demographics for both cohorts are listed in **Table 1**. ΔThCtAB ranged from –0.013 mm/y in ccMF (OAI men) and pLT and cMT (A9001140 women) to 0.020 mm/y in cLT (OAI women) (**Table 2**). Percentage changes ranged from –0.73%/y in pLT to +1.19%/y in aMT (both in OAI women). Three *t* tests for ΔThCtAB ≠ 0 in women from 48 regions for OAI men and women yielded *P* values <0.05 (cLFTC: *P* = 0.02; ecMF: *P* = 0.03; aMT: *P* = 0.03), but none was significant after adjusting for multiple comparisons. All *t* tests for A9001140 women produced *P* values ≥0.15. Averaged over all cohorts, ΔThCtAB ranged from –0.008 to 0.013 mm/y across all regions; 95% confidence intervals (CIs) for subregions were less than ±0.015 mm/y, and the direction of regional ΔThCtAB was not consistent between cohorts (**Table 2**).

The *P* values for testing for inclusion of all covariates (sex, age, and BMI) were >0.12 and not viewed as significant. Tests for individual covariates were also not viewed as significant after adjusting for multiple comparisons (unadjusted *P* values >0.1 except age/sex for iLT = 0.02/0.03 and LT = 0.04/0.04).

Subregion SD(ΔThCtAB) varied from 0.055 mm/y (eLT) to 0.101 mm/y (cLT) in OAI men, from 0.046 mm/y (eMT) to 0.106 mm/y (eLT) in OAI women, and from 0.050 mm/y (ecMF) to 0.112 mm/y (cLT) in A9001140 women at 24 months. The SDs were similar at 3 months. Central subregions had the largest SDs and external tibial subregions the smallest SDs (**Table 3**). *F* tests for comparing subregion variance of ΔThCtAB for OAI women and A9001140 women were not significant (*P* > 0.05), except for ccMF (*P* = 0.005) and ecMF (*P* = 0.01); neither was significant after adjusting for multiple comparisons. Analogous *F* tests for comparing men and women in the OAI revealed 4 (of 16) subregions with *P* < 0.05 (eMT: *P* = 0.036; ccMF: *P* = 0.023; ecMF: *P* = 0.009; aLT: *P* = 0.01), but none was significant after adjusting for multiple tests. *F* tests for comparison of OAI men versus women showed *P* < 0.05 only

Table 2. Mean ΔThCtAB (mm/y) for Medial and Lateral Subregions of Osteoarthritis Initiative (OAI) Men and Women and A9001140 Women at 24-Month Visit Intervals[4]

| Region | OAI | | A9001140 | Overall | |
	Men	Women	Women	Women	Overall (95% CI)
MFTC	−0.010	0.009	−0.008	0.000	−0.002 (−0.014 to 0.009)
cMFTC	−0.011	0.010	−0.017	−0.004	−0.006 (−0.024 to 0.012)
MT	−0.004	0.006	−0.005	0.000	−0.001 (−0.007 to 0.005)
cMF	−0.005	0.004	−0.003	0.000	−0.001 (−0.009 to 0.008)
cMT	0.003	0.005	−0.013	−0.005	−0.003 (−0.015 to 0.009)
eMT	−0.003	0.009	−0.004	0.002	0.001 (−0.007 to 0.008)
iMT	−0.005	0.004	0.001	0.002	0.001 (−0.008 to 0.009)
aMT	−0.001	0.017	−0.008	0.004	0.003 (−0.006 to 0.012)
pMT	−0.012	−0.003	0.000	−0.001	−0.001 (−0.009 to 0.006)
ccMF	−0.013	0.005	−0.004	0.000	−0.003 (−0.015 to 0.009)
ecMF	−0.007	0.013	0.002	0.007	0.004 (−0.004 to 0.012)
icMF	0.003	−0.005	−0.006	−0.006	−0.004 (−0.014 to 0.007)
LFTC	−0.006	0.017	−0.003	0.006	0.004 (−0.008 to 0.015)
cLFTC	0.000	0.037	−0.001	0.017	0.013 (−0.006 to 0.033)
LT	−0.005	0.007	−0.007	0.000	−0.001 (−0.008 to 0.006)
cLF	−0.001	0.010	0.005	0.007	0.005 (−0.003 to 0.014)
cLT	0.000	0.020	−0.005	0.007	0.005 (−0.010 to 0.021)
eLT	−0.001	0.012	−0.004	0.003	0.002 (−0.006 to 0.010)
iLT	−0.007	0.006	−0.004	0.001	−0.001 (−0.011 to 0.009)
aLT	−0.009	0.004	−0.010	−0.003	−0.005 (−0.015 to 0.006)
pLT	−0.008	−0.002	−0.013	−0.008	−0.008 (−0.021 to 0.006)
ccLF	0.000	0.017	0.004	0.010	0.008 (−0.005 to 0.021)
ecLF	−0.002	0.009	0.005	0.007	0.005 (−0.005 to 0.014)
icLF	0.000	0.004	0.005	0.005	0.004 (−0.006 to 0.013)

for medial plates (MT: $P = 0.032$; cMF: $P = 0.002$) and compartments (MFTC: $P = 0.0009$; cMFTC: $P = 0.009$), with only MFTC and cMF being significant after adjusting for multiple comparisons.

Normality plots for A9001140 women and for OAI men and women did not indicate any deviation of the empirical distributions of ΔThCtAB from the normal distribution. The Shapiro-Wilk test for normality of ΔThCtAB had $P < 0.05$ in only 1 test (of 22 regions) from OAI men (ccLF: $P = 0.034$) and 1 test of regions from OAI women (icLF: $P = 0.0003$); only the test for icLF region in women was significant after adjusting for multiple tests. Because region ΔThCtAB appeared to be normally distributed, 95% CIs for SDs were constructed as a percentage of estimated SD(ΔThCtAB), which were $0.82s$ to $1.27s$ for OAI men, while for OAI women, it was $0.86s$ to $1.20s$ and for A9001140 women was $0.86s$ to $1.19s$, where s = SD(ΔThCtAB) for a given subregion and cohort.

Rate of change was moderately correlated ($r = 0.4\text{-}0.7$) between neighboring subregions in OAI men and women and some nonneighboring subregions in OAI men, but fairly uncorrelated in nonneighboring subregions in OAI women ($r < 0.3$). Estimates of correlation between nonneighboring

subregions were larger for men, particularly for the femoral subregions with all other subregions. While correlations were generally positive, they appeared to be modestly negative between aLT and pLT. See supplementary materials for further discussion of correlation of ΔThCtAB between subregions.

Estimates of subregional SD(ΔThCtAB) were significantly correlated with subregional mean baseline ThCtAB for OAI men ($r = 0.74$) and women ($r = 0.80$) ($P < 0.0001$ for both cohorts). The coefficient of variation, CV = SD(ΔThCtAB)/mean baseline ThCtAB, ranged from 2.8% to 4.7% in OAI men and was slightly higher for women (3.2%-5.5% for OAI and 3.5%-5.0% for A9001140). This statistical relationship indicated there may be some structural or intrinsic relationship between SD(ΔThCtAB) and baseline ThCtAB, for example, segmentation variability as a percentage of thickness. Therefore, this relationship was also studied within each subregion. The range of baseline thickness across OAI men and women within each subregion was generally ≥1 mm, with the largest ranges in the central subregions, thus, providing sufficient range of thicknesses to test whether SD(ΔThCtAB) was different for the upper and lower tertiles of ThCtAB distribution using a

Table 3. Standard Deviation of ΔThCtAB (mm/y) for Medial and Lateral Regions of Osteoarthritis Initiative (OAI) Men and Women (12-Month Visit Interval) and A9001140 Women[4] (3- and 24-Month Visit Intervals)

| | OAI | | A9001140 women | | Overall |
Region	Men	Women	24-mo	3-mo	Women
MFTC	0.101	0.066	0.080	NA	0.074
cMFTC	0.152	0.110	0.127	NA	0.119
MT	0.051	0.040	0.040	NA	0.040
cMF	0.069	0.047	0.062	NA	0.056
cMT	0.095	0.076	0.079	0.089	0.078
eMT	0.059	0.046	0.054	0.049	0.051
iMT	0.056	0.056	0.061	0.077	0.059
aMT	0.076	0.062	0.055	0.065	0.059
pMT	0.060	0.053	0.050	0.064	0.052
ccMF	0.091	0.070	0.095	0.083	0.084
ecMF	0.067	0.049	0.050	0.049	0.050
icMF	0.077	0.070	0.077	0.068	0.074
LFTC	0.092	0.075	0.079	NA	0.077
cLFTC	0.156	0.128	0.135	NA	0.132
LT	0.050	0.048	0.051	NA	0.050
cLF	0.061	0.056	0.065	NA	0.061
cLT	0.101	0.106	0.112	0.114	0.109
eLT	0.055	0.057	0.057	0.061	0.057
iLT	0.081	0.074	0.067	0.065	0.070
aLT	0.085	0.063	0.073	0.085	0.069
pLT	0.091	0.101	0.091	0.097	0.096
ccLF	0.093	0.092	0.092	0.095	0.092
ecLF	0.068	0.059	0.072	0.064	0.066
icLF	0.056	0.066	0.068	0.063	0.067

Table 4. Standard Deviation of ΔThCtAB (mm/y) for Lower, Middle, and Upper Third of Subjects Ranked by Baseline Thickness for Subregions of Osteoarthritis Initiative (OAI) Men and Women

| | OAI men | | | OAI women | | |
	Low	Middle	High	Low	Middle	High
cMT	0.057	0.032	0.049	0.035	0.033	0.044
eMT	0.030	0.027	0.035	0.023	0.026	0.021
iMT	0.024	0.033	0.025	0.026	0.033	0.024
aMT	0.035	0.041	0.036	0.028	0.033	0.036
pMT	0.031	0.029	0.032	0.028	0.023	0.028
ccMF	0.048	0.057	0.036	0.038	0.031	0.031
ecMF	0.038	0.030	0.036	0.017	0.030	0.025
icMF	0.039	0.047	0.033	0.030	0.038	0.037
cLT	0.049	0.051	0.048	0.050	0.065	0.040
eLT	0.026	0.023	0.034	0.030	0.024	0.033
iLT	0.050	0.035	0.034	0.030	0.036	0.042
aLT	0.049	0.026	0.048	0.026	0.041	0.022
pLT	0.050	0.040	0.046	0.042	0.045	0.061
ccLF	0.039	0.027	0.061	0.047	0.042	0.049
ecLF	0.026	0.029	0.041	0.029	0.030	0.028
icLF	0.024	0.028	0.027	0.034	0.026	0.037

2-sample t test. The SD(ΔThCtAB) of the upper tertile of ThCtAB was greater than the SD(ΔThCtAB) of the lower tertile for half the subregions in both men and women; however, only 2 of 8 central subregions (4 for men, 4 for women) had larger SD(ΔThCtAB) for the upper tertiles (**Table 4**). The average SD across all subregions was the same for the upper versus lower tertile in OAI men (0.040 mm) and was only slightly larger in the upper versus lower tertile of OAI women (0.033-0.036 mm).

When using a reference distribution for classification, bias, that is, deviations from the true distribution, can affect the probabilities of (in)correctly classifying individuals, for example, as progressors or nonprogressors. Therefore, 2 simulation studies were carried out to examine the effect of bias on false-positive and true-positive rates of classification. Both simulations assumed the true and reference distributions were standard normal and used the criteria $Z > 1.96$ to classify a subject as not consistent with the true distribution. To assess bias, the standardized response mean (SRM) = mean(ΔThCtAB)/SD(ΔThCtAB), and SDs were

varied by ±20% from the SRM of the proposed reference distribution. One simulation considered the case where the "true" distribution was the asymptomatic cohort (the NULL case), so the true and reference distributions should be equivalent; hence, the percentage of subjects with $Z > 1.96$ estimated the rate of false positives and was expected to be 2.5%. Another simulation, the ALTERNATE case, had the true distribution represent an OA cohort (SRM = 2.8 compared to SRM = 0 for proposed reference distribution). In this case, the expected percentage of subjects with $Z > 1.96$ is 80%.

In the NULL case, the percentage of subjects from a reference population with $Z > 1.96$ varied from 0.35% to 7.12% (expectation = 2.5%) and was most sensitive to bias in SD(ΔThCtAB) (**Table 5**). In the ALTERNATE case, the percentage of subjects with $Z > 1.96$ ranged from 70.4% to 90.4% (expectation = 80%), and the effect of bias in the mean and SD was comparable but in inverse proportion (**Table 6**).

Discussion

This is the first study to comprehensively report on the distribution of subregional femorotibial cartilage thickness longitudinal change in asymptomatic cohorts, including (and comparing) both sexes. While mean longitudinal change is similarly negligible across subregions and sex, SD for these distributions varies with subregions and may differ between men and women. The latter differences

Table 5. Percentage of Subjects Expected to Be Classified as Having ThCtAB (or ΔThCtAB) Larger Than the Healthy Distribution (i.e., Z > 1.96) When Subjects Belong to the Reference Distribution (Null Case)

% bias in σ	% bias in SRM						
	−20%	−10%	−5%	0%	5%	10%	20%
−20%	0.35	0.50	0.60	0.71	0.85	1.00	1.39
−10%	0.82	1.10	1.28	1.47	1.69	1.94	2.53
−5%	1.15	1.51	1.72	1.95	2.22	2.51	3.20
0%	1.54	1.97	2.22	2.50	2.81	3.14	3.92
5%	1.98	2.49	2.78	3.10	3.45	3.82	4.69
10%	2.48	3.06	3.38	3.74	4.12	4.54	5.48
20%	3.59	4.30	4.70	5.12	5.57	6.06	7.12

Note: However, the reference distribution is biased in either the mean, % bias in standardized response mean (SRM), or standard deviation, % bias in σ. Positive bias indicates that the true population mean or standard deviation is larger than assumed by reference distribution.

Table 6. Percentage of Subjects Expected to Be Classified as Having ThCtAB (or ΔThCtAB) Larger Than the Healthy Distribution (i.e., Z > 1.96) When Subjects Come from a Distribution with a Mean Thickness Value 2.8 Standard Deviations Greater Than the Mean of True Distribution

% bias in σ	% bias in SRM						
	−20%	−10%	−5%	0%	5%	10%	20%
−20%	78.9	82.3	83.9	85.4	86.7	88.0	90.4
−10%	76.2	79.5	81.0	82.5	83.9	85.2	87.6
−5%	75.0	78.2	79.8	81.2	82.6	83.9	86.4
0%	73.9	77.1	78.6	80.0	81.4	82.7	85.1
5%	72.9	76.0	77.5	78.9	80.2	81.5	83.9
10%	72.0	75.0	76.4	77.8	79.1	80.4	82.8
20%	70.4	73.2	74.5	75.8	77.1	78.4	80.7

Note: If reference distribution is correct, this criterion corresponds to the ability (power) to detect 80% of the population. Results examine cases when reference distribution is biased in either the mean, % bias in standardized response mean (SRM), or standard deviation, % bias in σ. Positive bias indicates that the true population mean or standard deviation is larger than assumed by reference distribution.

indicate that it may be useful to separate reference distributions for men and women. Generally, correlation between neighboring subregions, particularly between central, external, and internal subregions, of the same plate in OAI women was moderate, while subregion ΔThCtAB in different plates had negligible correlation. Results indicate that distribution characteristics are robust across typical variations in study design, for example, recruiting strategies or small differences in inclusion/exclusion criteria; hence, a reference distribution for asymptomatic cohorts based on

these study results is proposed. It should also be noted that the estimated distribution is for observed rates of change in asymptomatic subjects, incorporating both real non-OA–related biological changes and the accompanying measurement error.

Developing standard reference distributions requires defining a population of interest and information that accurately reflects the population. Two studies, the OAI and Pfizer A9001140, were national multisite studies that included a fairly large number of asymptomatic subjects. Inclusion criteria were similar (except for sex), but recruitment methods and areas within the United States differed. Further, the image resolution (1.5-mm v. 1.0-mm slice thickness), scanners (1 single magnet type from 1 manufacturer at 4 centers v. 3 magnet types from 2 manufacturers at 7 centers), and observation periods (1 v. 2 years) differed between the studies. Yet, the results were fairly consistent between the cohorts, suggesting that the characterization of subregional rates of change in asymptomatic men or women is robust to common variations in study implementation.

The results indicate that subregional ΔThCtAB of asymptomatic cohorts is normally distributed. Evidence that the univariate distribution of ΔThCtAB for each subregion is approximately normal does not imply that the multivariate distribution for all subregions has a multivariate normal distribution, but it is a reasonable presumption. Under the assumption of multivariate normality, estimates of the mean and covariance matrix completely describe the distributions of ΔThCtAB.

The longitudinal changes in cartilage thickness for individual subregions in this study were generally not statistically different from zero. The magnitude of change was modest at <1.2% and not reproduced between subregions or cohorts, and the 95% CI of change based on all 3 cohorts indicates that the mean rate of change in asymptomatic subjects is <0.7%/y in any given subregion. Using zero for the mean of the reference distribution provides a simple, easily interpreted mean value for a reference distribution; yet, we cannot exclude that a small average loss (<0.7%) may exist in an asymptomatic population. However, the sensitivity analyses for bias indicate that small mean changes that may exist in the asymptomatic cohort would have minimal impact on classification outcomes.

The range of SD(ΔThCtAB) estimates generally fell within the bounds of a 95% CI when a single "average" variance for all subregions is assumed. This makes it difficult to reach conclusions about differences in subregion variability when considering studies separately. The consistency across cohorts, with central subregions having the largest SD and ecMF the smallest SD, indicates not all subregions have the same natural variability. Differences in natural variability may be caused by intrinsic differences between subregions; for example, the femur may be more difficult to segment than the tibia, or the variability may be

a function of some characteristic that varies with the individual, for example, cartilage thickness or surface area. A strong relationship between mean ThCtAB and SD(ΔThCtAB) was found across cohorts and regions, but the relationship between ThCtAB and SD(ΔThCtAB) is partially confounded with spatial location. If a causal relationship exists between SD(ΔThCtAB) and ThCtAB, it should also be seen within subregions. Evidence supporting this presumption was weak, however, with variability in the upper tertile actually lower in half of the cohort subregions and only a quarter of central subregions, which had the largest range of thickness values. Therefore, current evidence does not support a single model to describe variance across subregions, and thus, individual estimates for different subregions are proposed.

The use of annualized rate of change provides a standard framework for studies of different length. The SD in ΔThCtAB for different cohorts was found to be closely related to study length. This occurs when variation between subjects in true (not observed) rate of change is negligible compared to random variation in observed rate of change. This leads to SDs for annualized rates of change being smaller for longer studies compared to shorter studies. Hence, there is the need to adjust SD estimates for the reference distribution so that the estimates reflect the variability expected in 1-year studies. Adjustments of SD to reflect a 1-year study are made simply by multiplying the SD from the annualized rate of change by the study length.

When constructing reference distributions, potential biases need to be considered. Bias can occur when the reference distribution does not accurately represent the intended population or when measurements for a new subject deviate from the reference distribution due to differences in segmentation methods or other unwanted ways. Reference distribution bias is more likely to impact classification bias in absolute terms, for example, estimating proportions of a study population outside the reference distribution, than in relative terms, for example, comparing study cohorts or regional behavior, as bias will affect all groups equally in the latter. While reference distributions have a useful role in research studies, we emphasize that they should not take the place of collecting controls in studies aimed at comparing 2 (or more) cohorts.

The reference distributions proposed here are based on subjects aged 40 to 80 years; therefore, extrapolation to younger or older subjects should be done with caution. However, differences between older and younger subjects has been found to be relatively small, with most differences in mean ΔThCtAB less than ±20%.[25,26] Estimates of SDs in these studies[25,26] varied considerably, with deviations up to 40% to 50% from the proposed reference distribution SD. Sample sizes for these studies were small (<30 subjects) and were not necessarily population based. We have seen that biases of up to 20% have a relatively small impact on classification error rates; hence, classification outcomes may not be strongly biased if extrapolated to differently aged cohorts.

One concern in using the asymptomatic cohort as a reference distribution for assessing subjects with OA is that OA subjects may have a different "natural" variability, possibly from increased difficulty in segmenting cartilage or perhaps different lifestyles, for example, less active. Non-OA–related variability and OA-related change are confounded at an individual level. Test/retest studies have shown minimal differences (<6.6%) in variability between OA and asymptomatic subjects, but samples are limited, and thickness measurements only reflected measurement error.[4] SD(ΔThCtAB) at 3 months tended to be modestly higher (10%-30%) than test/retest variability, but the largest differences were in subregions with the highest expectation of seeing OA-related change.[4] Robustness analyses reported here indicate that if biases existed at these levels, classification error rates would be relatively unaffected. If interest is in looking at correlations or associations between subjects with and without OA-related change, some bias due to misclassification may be introduced, but this bias would be considerably less than that of other common practices, for example, arbitrarily dividing subjects into upper and lower tertiles and comparing these subcohorts.

A limitation of this study is that the proposed reference distributions are based on one segmentation and image analysis technique. Differences in segmentation and thickness calculation algorithms should, however, only affect reproducibility error. Therefore, assuming methods use similar region definitions, different methods are likely to be relatively unbiased. Also, the ΔThCtAB reference distribution is independent of methodology if set to 0 mm/y. Of more important concern are estimates of SD: a large component of observed variation in asymptomatic subjects is likely to be due to segmentation variability. As shown with robustness analyses reported here, increases of up to 20% in total variability would have a relatively small impact on classification probabilities. If bias due to segmentation and image analysis algorithms is of concern, they could be readily and fairly inexpensively handled by comparing results on a common set of subject images, for example, from OAI database images. Differences in results could be rescaled to reference distribution based on these results. While not perfect, this procedure should reduce the bias from segmentation to within bounds that are acceptable for most situations of interest.

A previous study[4] showed that only 40% of OA knees had cartilage thinning and 21% of subjects with thickening compared to the asymptomatic reference distribution, and the relative frequency of cartilage thinning and thickening was found to differ between KLG subcohorts.[4] The inclusion of all subjects regardless of magnitude and direction of change can impact the assessment of longitudinal change in

subjects, and hence, it may be beneficial to examine progression and nonprogression cohorts independently. Using a standard reference distribution provides a framework for constructing progression classifications based on objective consistent criteria.

In conclusion, this study shows that different cohorts and men and women have similar distributions for subregional ΔThCtAB; they are normally distributed with negligible mean ΔThCtAB (<0.7%), and variation in ΔThCtAB ranges between 3% and 5% of baseline ThCtAB depending on cartilage subregion. Reference distributions based on these results may be used to standardize reporting and identify individual longitudinal change outside the distribution of asymptomatic subjects. Classification of subjects as structural "progressors" or "nonprogressors" provides opportunities for alternative statistical procedures, for example, logistic regression, and the proposed distribution should permit more accurate identification of structural progression than conventional approaches that are based on measurement error or other metrics.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Felix Eckstein is CEO of Chondrometrics GmbH, a company providing MRI analysis services. He provides consulting services to MerckSerono, Novartis, and Sanofi Aventis. He received research support from Pfizer for analysis of the A9001140 study data and from Centocor for analysis of the OAI asymptomatic reference data. Robert Buck is sole proprietor of StatAnswers Consulting LLC. He received support from MerckSerono and worked for Pfizer during the A9001140 study.

References

1. Peterfy C, Woodworth T, Altman R. Workshop for consensus on osteoarthritis imaging: MRI of the knee. Osteoarthritis Cartilage. 2006;14(Suppl 1):44-5.

2. Eckstein F, Buck RJ, Burstein D, Charles HC, Crim J, Hudelmaier M, *et al*. Precision of 3.0 Tesla quantitative magnetic resonance imaging of cartilage morphology in a multicentre clinical trial. Ann Rheum Dis. 2008;67(12):1683-8.

3. Hellio Le Graverand MP, Buck RJ, Wyman BT, Vignon E, Mazzuca SA, Brandt KD, *et al*. Subregional femorotibial cartilage morphology in women: comparison between healthy controls and participants with different grades of radiographic knee osteoarthritis. Osteoarthritis Cartilage. 2009;17:1177-85.

4. Buck RJ, Wyman BT, Le Graverand MP, Hudelmaier M, Wirth W, Eckstein F. Osteoarthritis may not be a one-way-road of cartilage loss: comparison of spatial patterns of cartilage change between osteoarthritic and healthy knees. Osteoarthritis Cartilage. 2010;18(3):329-35.

5. Hanna F, Ebeling PR, Wang Y, O'Sullivan R, Davis S, Wluka AE, *et al*. Factors influencing longitudinal change in knee cartilage volume measured from magnetic resonance imaging in healthy men. Ann Rheum Dis. 2005;64(7):1038-42.

6. Wijayaratne SP, Teichtahl AJ, Wluka AE, Hanna F, Bell R, Davis SR, *et al*. The determinants of change in patella cartilage volume: a cohort study of healthy middle-aged women. Rheumatology (Oxford). 2008;47(9):1426-9.

7. Pelletier JP, Raynauld JP, Berthiaume MJ, Abram F, Choquette D, Haraoui B, *et al*. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. Arthritis Res Ther. 2007;9(4):R74.

8. Koo S, Gold GE, Andriacchi TP. Considerations in measuring cartilage thickness using MRI: factors influencing reproducibility and accuracy. Osteoarthritis Cartilage. 2005;13(9):782-9.

9. Wirth W, Eckstein F. A technique for regional analysis of femorotibial cartilage thickness based on quantitative magnetic resonance imaging. IEEE Trans Med Imaging. 2008;27(6):737-44.

10. Wirth W, Hellio Le Graverand MP, Wyman BT, Maschek S, Hudelmaier M, Hitzl W, *et al*. Regional analysis of femorotibial cartilage loss in a subsample from the Osteoarthritis Initiative progression subcohort. Osteoarthritis Cartilage. 2009;17(3):291-7.

11. Wirth W, Benichou O, Kwoh CK, Guermazi A, Hunter D, Putz R, *et al*. Spatial patterns of cartilage loss in the medial femoral condyle in osteoarthritic knees: data from the Osteoarthritis Initiative. Magn Reson Med. 2010;63(3):574-81.

12. Le Graverand MP, Buck RJ, Wyman BT, Vignon E, Mazzuca SA, Brandt KD, *et al*. Change in regional cartilage morphology and joint space width in osteoarthritis participants versus healthy controls: a multicentre study using 3.0 Tesla MRI and Lyon-Schuss radiography. Ann Rheum Dis. 2010;69(1):155-62.

13. Chang A, Moisio K, Chmiel JS, Eckstein F, Guermazi A, Almagor O, *et al*. Subregional effects of meniscal tears on

cartilage loss over 2 years in knee osteoarthritis. Ann Rheum Dis. 2011;70(1):74-9.

14. Eckstein F, Yang M, Guermazi A, Roemer FW, Hudelmaier M, Picha K, *et al*. Reference values and Z-scores for subregional femorotibial cartilage thickness: results from a large population-based sample (Framingham) and comparison with the non-exposed Osteoarthritis Initiative reference cohort. Osteoarthritis Cartilage. 2010;18(10):1275-83.

15. Eckstein F, Wirth W, Hudelmaier M, Stein V, Lengfelder V, Cahue S, *et al*. Patterns of femorotibial cartilage loss in knees with neutral, varus, and valgus alignment. Arthritis Rheum. 2008;59(11):1563-70.

16. Wirth W, Larroque S, Davies RY, Nevitt M, Gimona A, Baribaud F, *et al*. Comparison of 1-year vs 2-year change in regional cartilage thickness in osteoarthritis results from 346 participants from the Osteoarthritis Initiative. Osteoarthritis Cartilage. 2011;19(1):74-83.

17. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed. 2006;19(7):822-54.

18. Eckstein F, Guermazi A, Roemer FW. Quantitative MR imaging of cartilage and trabecular bone in osteoarthritis. Radiol Clin North Am. 2009;47(4):655-73.

19. Peterfy C, Li J, Zaim S, Duryea J, Lynch J, Miaux Y, *et al*. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. Skeletal Radiol. 2003;32(3):128-32.

20. Eckstein F, Buck RJ, Wyman BT, Kotyk JJ, Le Graverand MP, Remmers AE, *et al*. Quantitative imaging of cartilage morphology at 3.0 Tesla in the presence of gadopentate dimeglumine (Gd-DTPA). Magn Reson Med. 2007;58(2):402-6.

21. Eckstein F, Hudelmaier M, Wirth W, Kiefer B, Jackson R, Yu J, *et al*. Double echo steady state magnetic resonance imaging of knee articular cartilage at 3 Tesla: a pilot study for the Osteoarthritis Initiative. Ann Rheum Dis. 2006; 65(4):433-41.

22. Eckstein F, Maschek S, Wirth W, Hudelmaier M, Hitzl W, Wyman B, *et al*. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. Ann Rheum Dis. 2009;68(5):674-9.

23. Eckstein F, Nevitt M, Gimona A, Picha K, Lee JH, Davies RY, *et al*. Rates of change and sensitivity to change in cartilage morphology in healthy knees and in knees with mild, moderate, and end stage radiographic osteoarthritis. Arthritis Care Res (Hoboken). Epub 2010 Oct 18.

24. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7): 811-8.

25. Hudelmaier M, Glaser C, Hohe J, Englmeier KH, Reiser M, Putz R, *et al*. Age-related changes in the morphology and deformational behavior of knee joint cartilage. Arthritis Rheum. 2001;44(11):2556-61.

26. Burgkart R, Glaser C, Hinterwimmer S, Hudelmaier M, Englmeier KH, Reiser M, *et al*. Feasibility of T and Z scores from magnetic resonance imaging data for quantification of cartilage loss in osteoarthritis. Arthritis Rheum. 2003; 48(10):2829-35.