

# Calcium intake and bone mineral density as an example of non-linearity and threshold analysis

L. P. Breitling

Received: 21 March 2014 / Accepted: 8 December 2014 / Published online: 31 January 2015  
© International Osteoporosis Foundation and National Osteoporosis Foundation 2015

## Abstract

**Summary** Non-linearity is a likely phenomenon in bone metabolism, but is often ignored in pertinent epidemiological studies. Using NHANES III data on calcium intake and bone mineral density, the most important non-linear methods are introduced and discussed. The results should motivate researchers to consider non-linearity in this field more frequently.

**Introduction** Many relationships in bone metabolism and homeostasis are likely to follow non-linear patterns. Detailed dose-response analyses allowing for non-linear associations nonetheless remain scarce in this field.

**Methods** A detailed analysis of NHANES III data on dietary calcium intake and bone mineral density was used to demonstrate the application and some of the challenges of the most important dose-response methods, including LOESS, categorical analysis, fractional polynomials, restricted cubic splines, and segmented regression.

**Results** The spline estimate suggested increasing bone mineral density up to a calcium intake of about 1 g/day and a plateau thereafter. In segmented regression, the break-point marking

the beginning of the plateau was placed at an intake of 0.58 (95 % confidence interval, 0.33 to 0.82)g/day. Sensitivity analyses suggested a less curved dose-response in women.

**Conclusions** Knowing about the possibilities and limitations of non-linear dose-response approaches should encourage researchers to consider these methods more frequently in studies on bone health and disease. The example analysis suggested bone mineral density to reach a plateau slightly below current calcium intake recommendations, with fairly pronounced differences of the dose-response shape by sex and menopausal status.

**Keywords** Broken-line regression · Dose-response · Joinpoint regression · Piecewise regression · Spline regression

## Introduction

Regression models are one of the most important tools in modern epidemiological and biomedical data analysis [17]. They allow adjustment for multiple covariables and confounders, and diverse types of outcome and predictor variables can be handled in a usually straightforward fashion in widely available software packages (e.g., [13, 14]). The use of regression models allowing for more complex, namely non-linear associations, has become popular only somewhat more recently. Such models may obviously reflect natural relationships more closely than standard linear regression equations. The most often applied methods include fractional polynomials and spline regression, many subtypes of which nowadays are available in standard statistical software packages [10, 36]. Other methods, which try to fit specific non-linear shapes to dose-response data based on mechanistic assumptions, are rarely used in epidemiological settings, but firmly established in pharmacological studies [42]. Segmented regression, which also is less widely known, explicitly takes into

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00198-014-2996-7) contains supplementary material, which is available to authorized users.

---

L. P. Breitling (✉)  
German Cancer Research Center, Division C070 Clinical  
Epidemiology and Aging Research, PO Box 101949,  
69009 Heidelberg, Germany  
e-mail: lutz.breitling@gmail.com

L. P. Breitling  
Klinik für Geriatrie und Altersmedizin, Allgemeines Krankenhaus  
Celle, Siemensplatz 4, 29223 Celle, Germany

L. P. Breitling  
Klinik für Innere Medizin I, University Hospital Regensburg,  
Franz-Josef-Strauß-Allee 11, 93053 Regensburg, Germany

account that an association may change at discrete break-point values and locates such thresholds as part of the model-fitting process [23, 27].

Bone metabolism and homeostasis involve a complex interplay of nutritional intakes and hormonal feedback loops [33, 43]. Associations between the components of such a balancing regulatory system a priori must be expected to feature non-linear dose-response relationships. The existence of threshold phenomena also can reasonably be assumed. For instance, the availability of a nutrient involved may be the limiting factor of physiological processes, such as bone formation, in the lower range of intake. If its availability exceeds a possibly rather distinct threshold at which it is sufficient for an optimal level of bone formation, regulatory mechanisms of some kind obviously should counteract excessive osteoid mineralization or soft tissue calcification with potentially adverse effects on bone health [33].

Despite the preceding considerations, detailed dose-response analyses are not regularly presented in epidemiological studies on bone metabolism. A notable exception clearly demonstrating the importance of considering non-linearity in this context described a strongly curvilinear association between dietary calcium intake and risk of hip fracture employing spline regression [46]. Another study examined the dose-response relationship between serum vitamin D and parathyroid hormone using a segmented regression variant and other methods, revealing thresholds of optimal vitamin D levels that differ between Caucasians and African Americans [48]. The association between dietary calcium intake and bone mineral density in NHANES III furthermore has served as an example of non-linearity in the methodological literature [10]. In the present work, the latter data are revisited and reanalyzed with a focus on different approaches to non-linear dose-response modelling and threshold estimation. Brief non-technical introductions and comparisons are given, positing that more widespread consideration of these methods would benefit the research in the area of bone metabolism and related matters.

## Methods

### Data source and study sample

Data used in the present work originated from public-use files of the Third National Health and Nutrition Examination Survey (NHANES III) as provided by the Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD. NHANES III was a survey conducted from 1988 to 1994 in order to address questions of health and nutrition in national probability samples of the non-institutionalized population of the USA [30].

NHANES III participants were included in the current analysis if information on dietary calcium intake and bone mineral density of the femur neck region were available, as well as age, sex, race/ethnicity, education, and smoking status.

### Measurements and procedures

Key socio-demographic and health-related characteristics obtained from standardized interviews included age, sex, race/ethnicity (White, African American, Mexican American, or other), education (less than 12, 12, or more than 12 years of school completed), and self-reported current smoking. Subjects were classified as physically inactive, if they reported never walking a mile or more, as well as never engaging in jogging/running, cycling, swimming, aerobics, dancing, calisthenics, gardening, weight lifting, or any other physical exercises during the past month [32].

Nutrient intakes were estimated from recall information pertaining to foods and beverages consumed during the previous 24-h time period [2]. For the present study, the total daily calcium, protein, and total energy intakes calculated based on the University of Minnesota Nutrition Coordinating Center nutrient composition database (which used a multi-version design and was updated as required to reflect changes in food composition throughout the survey period, see [ftp.cdc.gov/pub/Health\\_Statistics/NCHS/nhanes/nhanes3/2A/EXAMDR-acc.pdf](http://ftp.cdc.gov/pub/Health_Statistics/NCHS/nhanes/nhanes3/2A/EXAMDR-acc.pdf)) were used. Furthermore, the intake of vitamins or minerals in the past month was queried.

Body mass index was determined during visits to a mobile examination clinic. Areal bone mineral density (BMD) of the femoral neck area was measured using dual energy X-ray absorptiometry [24]. Furthermore, blood samples were taken and serum vitamin D levels examined by radioimmunoassay [49].

### Statistical analyses

The study sample first was described according to socio-demographic and medical characteristics analysed, using counts and percentages for categorical variables, mean (standard deviation) for normally distributed variables, and median (interquartile range) for quantitative variables whose distribution showed a pronounced deviation from normality. Normality was assessed by inspecting histograms of the respective variables.

### Local polynomial regression

The bivariate association between bone mineral density and calcium intake was explored by local polynomial regression fitting (LOESS). LOESS is a smoothing method that essentially summarizes the association between outcome and exposure by fitting a multitude of regression models to adjacent

subsets of the data [8]. For each  $x$ , the mean  $y$  then can be predicted from the regression equation estimated from the neighboring data points. The smoothing can be influenced by changing the width of the subsets used for each local regression. Furthermore, different importance (weights) can be assigned to data points included in a local subset according to their distance from the respective data point. The locally fitted regression models usually are linear (first-degree LOESS) or quadratic (second-degree LOESS).

#### *Linear regression analysis*

The covariable-adjusted association between calcium intake and bone mineral density was first examined by multiple linear regression of bone mineral density on calcium intake as a continuous untransformed predictor. Furthermore, models using quartiles or quintiles of calcium intake as predictors of BMD were examined.

These and all subsequent models were adjusted for age, sex, race/ethnicity, education, and smoking status, which a priori were considered potential major confounders of the association of interest. The adjustment for age was linear, and the categorical characteristics were coded using dummy variables. Additional characteristics were addressed in sensitivity analyses (see below).

#### *Fractional polynomial regression*

In fractional polynomial modelling, multiple models are fit using various transformations of the predictor for which a non-linear association with the outcome is assumed [34]. Using a standard approach, the transformations examined were  $1/x^2$ ,  $1/x$ , square root of  $x$ ,  $\log(x)$ ,  $x$ ,  $x^2$ , and  $x^3$ . Fractional polynomials were investigated up to the second degree, that is, all possible models including one or any combination of up to two of the transformations were examined. The final fractional polynomial model was selected using the RA2 algorithm [36], which keeps the predictor of interest only if an overall association exists and selects a second-degree polynomial only if it provides a significant improvement over the best-fitting first-degree polynomial. Non-linearity can be assessed using standard likelihood ratio tests.

#### *Restricted cubic spline regression*

In spline regression, the association between outcome and predictor essentially is allowed to differ in a piecewise fashion between a chosen set of knots [39]. Usually, the regression equation is formulated in such a way that no jumps occur in the relationship. In the case of restricted cubic splines, the association is piecewise defined by cubic functions, i.e., curvilinear associations can readily be fit. They are restricted in that the association is defined as linear beyond the outermost knots and no kinks occur in the relationship. In the present

analysis, increasingly flexible restricted cubic spline models with three to seven knots were examined, using the software's standard knot placement. The best model was selected according to the AIC value, which balances flexibility vs. overfitting [41]. In the case of a restricted cubic spline model with three knots, the  $p$  value of the spline term as shown in standard regression output allows a direct assessment of non-linearity [10].

#### *$E_{max}$ modelling*

As pointed out by a reviewer, specific non-linear dose-response shapes can sometimes be assumed due to considerations regarding the underlying biology. Non-linear regression methods may fit such shapes defined by explicit non-linear formulae. In the case of the so-called  $E_{max}$  model, a sigmoid dose-response curve is fit to the data [25]. The dose-response shape in this model consists of plateaus at low and high exposure values connected by a sigmoid increase or decrease in between. The steepness (slope factor  $N$ ) and location of the sigmoid part of the dose-response are estimated by the modelling procedure, which produces readily interpretable estimates, including the difference between the two plateau values (commonly considered the maximum effect attributable to the exposure,  $E_{max}$ ) and the exposure value  $ED_{50}$ , at which half the  $E_{max}$  is observed.

#### *Segmented regression*

In segmented regression, the association between predictor and outcome is allowed to change at discrete break-points [27]. In contrast to the knots in spline regression, the location of one or more such break-points is estimated as part of the model-fitting process by likelihood maximization. The resulting break-point estimates may depend on the starting values, fitting parameters and random aspects of the model estimation algorithm, which includes a so-called bootstrapping procedure [47]. Thus, a reasonable set of starting values and parameters were explored in the present work, and relevant models were fit repeatedly.

#### *Sensitivity analyses*

In sensitivity analyses, the effect of including age<sup>2</sup> in the adjustment set was explored. Restricted cubic spline and segmented regression models were examined separately for female and male participants. Furthermore, the importance of mineral and vitamin supplements was examined by restricting the analysis to subjects reporting no intake during the past month. Similarly, the impact of excluding subjects reporting no physical activity or those suffering from severe obesity (BMI 35+ kg/m<sup>2</sup>) was explored. Further models were fit stratified on vitamin D sufficiency (cutoff 20 ng/mL) status.

As effects of calcium on bone mineral density were limited to subjects with higher protein intake in some studies, models

were also fit after stratification on protein intake (expressed as % of total energy intake), additionally adjusting for total energy intake [9]. To explore the importance of age and menopause, spline and segmented regression models were also fit separately to males up to or above the median age of 44 years, and to pre- or postmenopausal females (where postmenopausal participants were identified by being 70+ years of age, or being non-pregnant and 40+ years of age with no period during the past 12 months, or having had hysterectomy or removal of both ovaries [45]). Further models were fit adjusting for or stratifying on body height.

All analyses were done using R v3.0.2 for Linux (R Foundation for Statistical Computing 2013). The add-on package “mfp” v1.4.9 was used for fractional polynomial analysis [3]. Restricted cubic spline regression employed the package “rms” v4.1-1 [19], whereas “segmented” v0.3-0.0 [27, 28] was used to fit segmented regression models.  $E_{\max}$  modelling was done using the package “DoseFinding” v0.9-12 [7].

## Results

### Descriptives

Table 1 shows the socio-demographic and medical characteristics of the 14,116 NHANES III participants for whom sufficient data for inclusion in the present study were available. The two sexes were equally represented in the analysis sample. Almost half of the population were classified as non-Hispanic white.

The dietary calcium intakes derived from nutritional recall information were right-skewed and ranged from 0.002 to 9.87 g/day, the latter being a single outlier with the next highest intakes being 6.69 and 6.62 g/day. Bone mineral densities at the femoral neck ranged from 0.23 to 1.84 g/cm<sup>2</sup> and followed a normal distribution rather well.

Missing values occurred for supplement intake ( $n=9$ ), physical activity ( $n=23$ ), BMI ( $n=19$ ), and vitamin D ( $n=532$ ).

### Local polynomial regression

The LOESS estimate for the bivariate association between calcium intake and bone mineral density is shown in Fig. 1a. The width of the local subsets here was defined as 40 % of the data points, and the default tricubic weighting scheme was used. The local polynomials fitted were linear functions.

Relative to the scatter of the measurements, the LOESS estimate seemed to suggest little systematic variation in BMD associated with calcium, though there seemed to be an overall positive association and a

**Table 1** Description of the 14116 NHANES III participants included in the present study

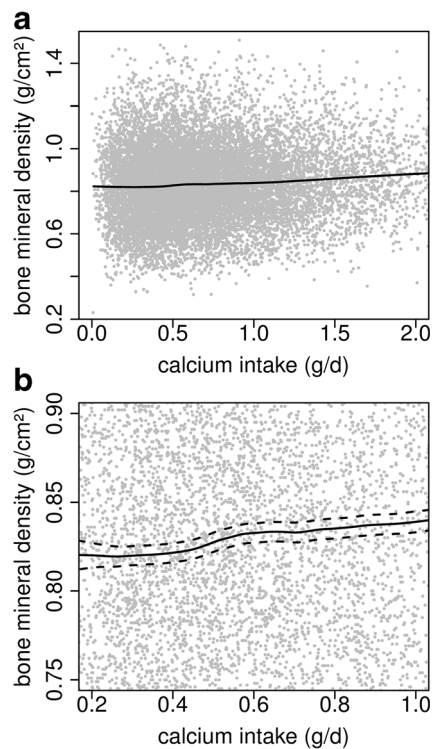
Socio-demographic and medical characteristics	
Age, years (median [IQR])	44 [32–64]
Sex ( $n$ [%])	
Female	7286 (52%)
Male	6830 (48%)
Race/ethnicity ( $n$ [%])	
Non-Hispanic white	5978 (42%)
Non-Hispanic black	3861 (27%)
Mexican American	3725 (26%)
Other	552 (4%)
Education ( $n$ [%])	
<12 years	5649 (40%)
12 years	4371 (31%)
>12 years	4096 (29%)
Current smoking ( $n$ [%])	
Yes	3702 (26%)
No	10,414 (74%)
Dietary daily calcium intake, g/day (median [IQR])	0.62 [0.39–0.96]
Femoral neck bone mineral density, g/cm <sup>2</sup> (mean [SD])	0.83 (0.17)
Mineral/vitamin supplements in past month ( $n$ [%])	5218 (37%)
No physical activity in past month ( $n$ [%])	2961 (21%)
Body mass index, kg/m <sup>2</sup> (median [IQR])	26.2 [23.2–30.0]
Serum vitamin D, ng/mL (median [IQR])	24.3 [17.7–32.2]

*IQR* interquartile range, *SD* standard deviation

distinct kink in the relationship, which became more apparent when examining the LOESS estimate at a higher resolution (Fig. 1b).

### Linear regression analysis

In multiple linear regression analysis, calcium intake was associated with bone mineral density with an increase in BMD by 0.0058 [95 % CI: 0.0014–0.0101]g/cm<sup>2</sup> per g/day calcium. Though the confidence intervals in both quartile and quintile models all excluded the null effect, they widely overlapped. In reference to the lowest quartile, the second, third, and fourth quartile of calcium intake (break-points of 0.39, 0.62, and 0.96 g/day) were associated with higher BMD with coefficients of 0.0091 [0.0028–0.0153], 0.0081 [0.0018–0.0144], and 0.0117 [0.0052–0.0182]g/cm<sup>2</sup> per g/day. In reference to the lowest quintile, the second, third, fourth, and fifth quintile (break-points of 0.34, 0.52, 0.74, and 1.07 g/day) were associated with increases in BMD with coefficients of 0.0075 [0.0005–0.0144], 0.0110 [0.0040–0.0179], 0.0103 [0.0032–0.0173], and 0.0138 [0.0065–0.0210]g/cm<sup>2</sup> per g/day.



**Fig. 1** Scatter plot of dietary calcium intake and bone mineral density in NHANES III, showing the corresponding LOESS estimate and approximate confidence bands (dotted lines; defined as  $\pm$  two standard errors around the estimate)

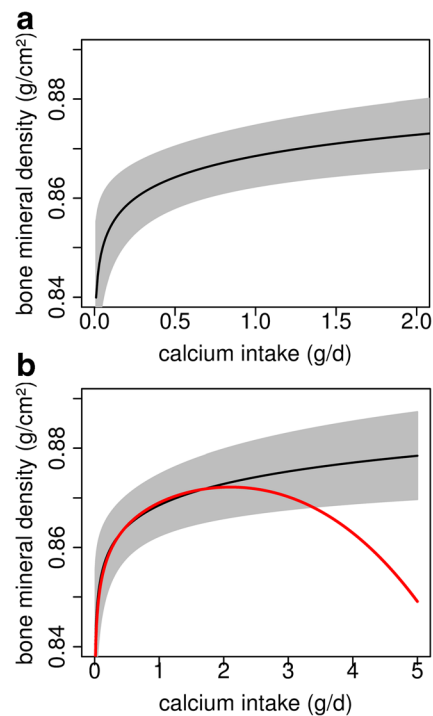
### Fractional polynomial analysis

In fractional polynomial modelling, the first-degree polynomial with log-transformed calcium intake as the predictor was selected as the final model ( $p$  for non-linearity: 0.0079). The association is graphically depicted in Fig. 2a. As in the following figures, the mean predicted bone mineral density is shown for covariable values of median age, male sex, non-Hispanic White, less than 12 years of education, and non-smoking. Reflecting the nature of the logarithmic exposure transformation, the shape of the dose-response curve was characterised by an extremely steep slope at very low calcium intake and a more and more shallow but monotonously positive association at higher exposure values.

The best-fitting second-degree fractional polynomial additionally included (calcium intake)<sup>3</sup> and did not differ visibly from the log-only model for intakes up to 2 g/day (Fig. 2b). Beyond this value, the dose-response turned into a negative association.

### Restricted cubic spline analysis

The restricted cubic spline model with three knots (placed at the 10th, 50th, and 90th percentile) was selected as the best model. It also was superior to the simple linear model ( $AIC_{\text{spline}}=-17027$  vs.  $AIC_{\text{linear}}=-17021$ ;  $p=0.0050$ ). The

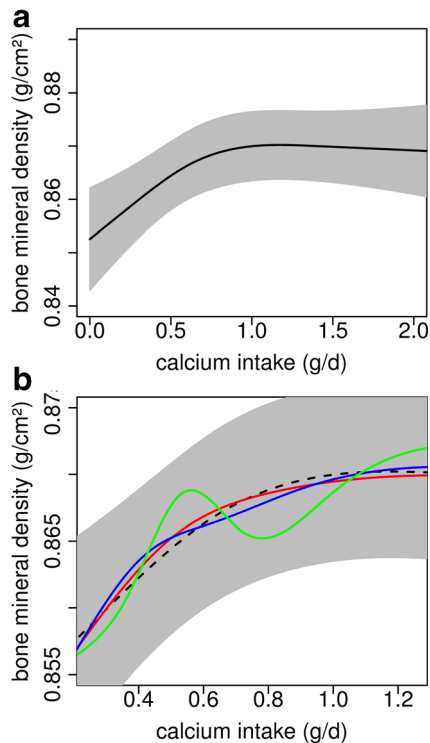


**Fig. 2** Fractional polynomial modelling of the association of calcium intake with bone mineral density. **a** Final model with 95 % confidence interval (shaded). **b** Like panel **a**, but with best-fitting second-degree polynomial overlaid in red

shape of the estimated dose-response relationship is shown in Fig. 3a. The spline estimate essentially consisted of a positive association at calcium intake values below 1.0 g/day and a plateau at higher values. The shape of the splines with four and five knots did not differ to any relevant degree from the best spline. The spline with six knots (5th, 23rd, 41st, 59th, 77th, 95th percentile), showed some more pronounced fluctuations, but even these stayed within the confidence interval of the best model (Fig. 3b).

### $E_{\text{max}}$ modelling

The sigmoid  $E_{\text{max}}$  model yielded the dose-response estimate presented in Fig. 4, suggesting a dietary calcium intake-associated maximum increase in bone mineral density (95 % confidence interval) of  $E_{\text{max}}=0.0159$  (0.0009–0.0308)g/cm<sup>2</sup>. With an estimated  $ED_{50}=0.36$  (0.12–0.60)g/day, the dose-response curve seemed to approach the upper plateau somewhat more steeply and at lower exposure values than in the spline regression models (Fig. 3). Note, however, that a simplified, so-called hyperbolic  $E_{\text{max}}$  model, which featured a positive association at low exposure values approaching a plateau at higher exposure values, yielded a minimally smaller AIC (–17,027 vs. –17,026; details not shown).

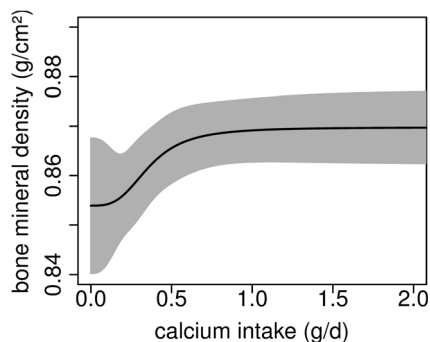


**Fig. 3** Restricted cubic spline modelling of the association of calcium intake with bone mineral density. **a**. Best model with 95 % confidence interval (shaded). **b** Like panel **a**, but with the results of models with four, five, and six knots overlaid in red, blue, and green, respectively

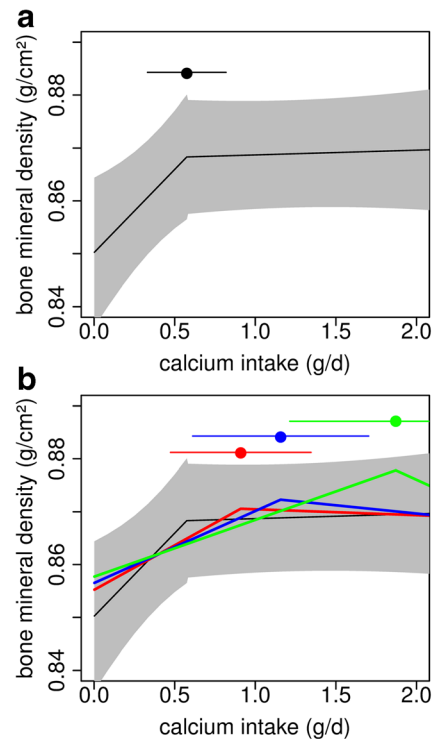
#### Segmented regression

The segmented regression estimate shown in Fig. 5a was obtained with a break-point location starting value of 0.7. The dose-response in this modelling approach was estimated to be a positive association turning into a plateau beyond a calcium intake of 0.58 [95 % CI, 0.33 to 0.82]g/day. The estimated slopes [95 % CI] below and above the break-point were 0.0314 [0.0069 to 0.0559], respectively, 0.0009 [−0.0050 to 0.0068]g/cm<sup>2</sup> per g/day.

On the basis of the spline in Fig. 3a, one might have chosen a starting point of 1.0. With that starting value, a break-point at 0.91 [95 % CI: 0.47 to 1.35] was obtained. Notably, starting



**Fig. 4**  $E_{\max}$  modelling of the association of calcium intake with bone mineral density. Main model with 95 % confidence interval (shaded)



**Fig. 5** Segmented regression modelling of the association of calcium intake with bone mineral density. **a** Main model with 95 % confidence interval (shaded), also showing the break-point estimate and its 95 % confidence interval. **b** Like panel **a**, but with the results of models converging on higher break-points overlaid; for details, see text

values from 0.4 to 0.9 consistently converged on break-points narrowly ranging from 0.57 to 0.59, whereas starting values of 1.1 and 1.2 or higher produced break-point estimates of 1.16 [95 % CI, 0.61 to 1.70] and around 1.87, respectively. As shown in Fig. 5b, the models with higher break-points featured increasingly implausible negative slopes beyond the estimated break-points. Models trying to estimate two break-points did not converge.

Based on the convergence patterns described above, taking into account the LOESS and especially the restricted cubic spline estimates, and furthermore considering biological plausibility, the model fits with break-points around 0.58 were considered the most relevant and plausible segmented regression estimates of the dose-response relationship. The results thus suggested that bone mineral density increased with dietary calcium intake up to 0.58 g/day, whereas intake in excess of this value was not associated with any further beneficial effects in terms of BMD.

#### Sensitivity analyses

The inclusion of age<sup>2</sup> in the adjustment set did not affect the shape of the spline model or the break-point estimates to any relevant degree. Though statistically significant, age<sup>2</sup> was not included in the main models for the sake of parsimony and

precision. In sex-stratified analyses (Supplemental Figure 1a), the spline estimate for the male participants ( $n=6830$ ) hardly differed from the main analysis. The break-point estimated in a corresponding segmented regression model likewise was essentially the same (0.62 [95 % CI, 0.33 to 0.92]g/day). The spline for the female participants ( $n=7286$ ) clearly appeared less curved (Supplemental Figure 1b). Segmented regression estimation placed a break-point at a calcium intake of 0.36 [0.06 to 0.67]g/day. The exclusion of 5218 subjects reporting vitamin or mineral intake during the past month did not affect the shape of the dose-response or break-point estimates to any relevant degree (Supplemental Figure 2). The models fit after excluding participants reporting no physical activity ( $n=2961$ ) or those being severely obese ( $n=1243$ ) also produced spline curves similar to the main models. The break-point estimates were somewhat more ambiguous, though the confidence intervals around these estimates were widely overlapping (Supplemental Figure 3). For participants with vitamin D deficiency ( $n=4554$ ), the spline estimate featured rather wide confidence intervals, but otherwise was rather similar to the previous model (Supplemental Figure 4a). A break-point at 0.61 g/day dietary calcium intake was estimated very consistently. Finally, the spline estimate for the 9030 subjects with sufficient serum vitamin D levels started off with a more shallow slope, though the overall dose-response shape was preserved (Supplemental Figure 4b). The segmented regression models converged either on a break-point of 0.44 or 0.91 g calcium per day.

As shown in Supplemental Figure 5, the impact of stratification on dietary protein intake and adjustment for total energy intake was limited, and the variation of the spline curves and break-point estimates showed no really clear pattern. Age-stratification had only a moderate effect on the dose-response and break-point estimates in men, but the stratification on menopausal status in women produced an almost linear dose-response estimate in the premenopausal subcohort, contrasting clear non-linearity in the postmenopausal participants (Supplemental Figure 6). Adjustment for or stratification on body height had no relevant impact on the dose-response relationships investigated (Supplemental Figure 7).

## Discussion

Non-linearity is a likely phenomenon in studies on bone metabolism and homeostasis. Going through the currently most important dose-response methods, the present work extended previous results regarding the non-linear association between calcium intake and bone mineral density in NHANES III, which were obtained in a methodological study [10]. The results especially of the less widely used segmented regression models suggested that femoral neck bone mineral

density (as a measure of bone health) levels off somewhat below current calcium intake recommendations [31]. Furthermore, sex-stratified models suggested that the association of calcium with BMD possibly differs more substantially between men and women than currently appreciated. Some important limitations of the various approaches to dose-response analysis also were encountered and can be exemplified by the present work.

### Smoothing the scatter plot

The LOESS approach [8] was used to summarize the bivariate association between calcium intake and bone mineral density as visualized in the scatter plot. As LOESS makes no assumption whatsoever about the shape of the dose-response relationship, it essentially is the most flexible of the methods employed in the present work. Such flexible smoothing may be particularly useful as a first step in dose-response analysis, though in the present example the non-linearity in the LOESS estimate was more or less lost in the scatter of the BMD measurements.

Notably, there is not one LOESS estimate for a given dataset, as the result depends on the various smoothing parameters, weighting schemes, and optional extensions such as corrections for outlying values [22]. In the present work, exploring a reasonable set of smoothing parameters did not affect the LOESS estimate to any relevant degree (not shown). Statistical procedures are available for testing LOESS estimates of different degrees against each other and an ordinary linear model. Multivariate extensions exist, but their results lack the intuitive interpretability of the standard LOESS curves [8, 21].

### Classical categorical analysis

The linear regression models using categorized calcium intake were consistent with a positive association with bone mineral density, but did not provide much more insight into the dose-response relationship. Grouping a quantitative exposure into quantiles, e.g. quartiles or quintiles, or according to predefined cutoffs is a very common approach in biomedical data analysis. Strengths of this approach include its practical simplicity, the ease of presenting the results, and the fact that no assumptions on the overall shape of the dose-response relationship are made. Non-linear patterns thus may well be captured by this approach.

Although models with a categorized predictor flexibly allow for non-linearity, they effectively fit a step function with discrete changes in the outcome at the category limits [39]. Mechanistically, this will hardly ever make sense. Whether or not non-linearity is discovered obviously may depend on the choice of category limits [16]. The non-linear parts of a dose-response relationship may disappear within a single broad

category, or they may get lost in the imprecision of estimates associated with more narrow categories.

In the absence of predefined cutoffs, researchers most often resort to grouping the study participants into exposure quartiles or quintiles, perceiving the equal group size and presumed comparability across studies as advantages. These are, to some extent, misperceptions. In the case of binary outcomes, such as occurrence of a hip fracture, it would be statistically preferable not to form equal-sized groups, but groups with the same number of outcome counts [39]. Furthermore, comparability across studies can only be assumed if the underlying dose-response relationship is linear across the entire exposure range, or if the categories have similar boundaries in the various studies. As discussed throughout, the linearity assumption cannot generally be made in studies on bone biology, and the exposure range for some important variables varies substantially in this field. Comparing category-based estimates between studies thus requires much more caution than often is appreciated.

#### Smooth dose-response modelling

The currently most widely used smooth non-linear modelling strategies, fractional polynomial, and spline regression, yielded somewhat different results. As the log-transformation selected as the best-fitting fractional polynomial model goes to negative infinity as  $x$  goes to zero, the dose-response curve obtained by this approach was extremely steep at its beginning. Restricted cubic spline modelling suggested a more shallow slope followed by a plateau beyond a calcium intake near 1 g/day.

There is no consensus regarding the superiority of either approach or their variants [15, 16, 39]. Neither method attempts to estimate mechanistically relevant parameters (in contrast to non-linear modelling based on differential equations in pharmacology, ecology, and so forth), and the overall shapes rather than exact coefficients of the dose-response curve may be considered the main result in this type of analysis.

As mentioned above, procedures are readily available to test for the significance of non-linearity or to judge whether an even more flexible model indeed fits the data better [10, 36]. One should consider the possibility that strictly relying on statistical testing might not be the most sensible approach in many situations in this context [16]. If a non-linear association is assumed based on prior knowledge, it might often make sense to present a dose-response curve estimated by fractional polynomial or spline methods regardless of statistical significance. With sufficient confidence in the applicability of a sigmoid relationship, for example, one could even employ an explicitly non-linear model, such as the  $E_{\max}$  model in the present study. Similarly, even if a spline model with three knots—as in the current analysis—is selected as the best

model based on the AIC, it may still make sense to explore more flexible splines that can reveal additional non-linearity of the dose-response relationship.

#### Threshold estimation by segmented regression

Among the methods used in the present work, segmented regression probably is the least well known. The results in the main analysis were consistent in particular with the spline estimate. The break-point, at which the positive association turned into a plateau, most consistently was placed at a calcium intake of about 0.6 g/day, somewhat lower than one might have guesstimated based on the smooth modelling procedure. Though estimating such a threshold by an objective procedure should be preferable to eyeballing curvilinear graphs, the example analysis also highlighted some major challenges of segmented regression.

Especially if the data are relatively noisy, the break-point estimated may depend on the starting value and parameters chosen, as multiple true or spurious break-points may be identified by the fitting algorithm [27, 28]. Recent versions of the software used here implement techniques trying to overcome this issue by a bootstrapping approach, which introduces a random component that may lead to different break-point estimates even when using identical starting values and modelling parameters (see package documentation; [47]). It then depends on subject matter knowledge and/or results of additional dose-response analyses, whether one chooses to discard a specific break-point estimate or not.

Altogether, segmented regression clearly is a useful and certainly underused addition to the epidemiological methods repertoire, which in most cases should complement rather than replace other dose-response approaches. Although this method generally can accommodate all kinds of outcomes and exposures, its use to date has been limited almost exclusively to time trend analyses (e.g., [4, 29, 37]).

Statistical tests have been proposed to check for the existence of a break-point [28], but these pertain to the comparison with an unsegmented linear rather than with a smooth non-linear model as the alternative. In some situations, there might be an exceptionally strong rationale for focusing on a segmented regression approach [12]. In a field like bone metabolism, discrete changes in a dose-response relationship often might be less expected, as complex feedback loops are involved that should be more likely to result in a smooth change in associations. They could nonetheless occur, for example, when an essential nutrient intake reaches a distinct sufficient level.

#### Dietary calcium intake and bone mineral density

In a previous study on the female participants of NHANES III, dietary calcium intake was not associated with bone mineral



density in the femoral neck region [5]. In the present analysis, calcium intake as a continuous variable was significantly associated with BMD also in a females-only model ( $p=0.0018$  in covariable-adjusted linear regression; details not shown). Bone mineral density in the study by Bass et al. [5] was dichotomized, which might explain the discrepancy of the results. Another NHANES III study described an association of calcium intake with (total hip) bone mineral density only in female participants with low serum vitamin D levels, but not in other women or in men of any vitamin D status [6]. The present sensitivity analyses were consistent with this to the extent that the initial slope of the dose-response curve was much steeper in vitamin D-deficient subjects, whereas the inconsistency with respect to an association among the male participants might be due to treating calcium intake as grouped into quartiles. An association of calcium intake (quartiles) with bone mineral density (dichotomized) also has been shown in another large study including both sexes [20], and an association of dietary calcium intake with bone health altogether is widely accepted, though evidence for some subgroups remains scarce or inconclusive [43].

The NIH-recommended dietary allowances for calcium are 1.0 g/day for young adults, and 1.2 g/day for women and men aged 51+ and 71+ years, respectively [31]. The present study was not designed to derive novel dietary recommendations, and the relevance and generalizability of findings like the rather pronounced heterogeneity by sex and menopausal status in the present work need to be evaluated in future studies and additional populations. Metabolic studies on calcium homeostasis seem to be in line with a lower calcium intake of about 0.75 g/day being sufficient in both sexes [21], and age-dependent threshold behaviour of calcium balances has been suggested [26]. In a large study on Scandinavian women, the risk of hip fracture diminished up to a dietary calcium intake of about 0.8 g/day (guesstimate based on eyeballing the published spline curve), with no additional protective effect beyond this value [46].

### Limitations

The present study focussed on bone mineral density at the femoral neck region. Some association patterns may differ for other femoral sites, but the determinants generally strongly overlap [5]. In addition, BMD at different femoral sites show similar associations with fracture risk, further justifying the restriction to one site in the present work [40]. Only areal measurements of bone mineral density were available for the present study, and future studies should preferably employ volumetric assessments less prone to distortions by bone volume and body size [1]. Other limitations are the cross-sectional design and the assessment of calcium intake by 24-h recall, though the reliability of the data source used has been demonstrated time and again [2].

The present analysis with its focus on dose-response aspects tried to maximize generalizability by excluding as few subjects as possible a priori. This increases the possibility of residual confounding, but the resulting large sample size in combination with multivariate regression modelling addressing major confounding variables remains a strength of the study. Generalizability of the results could have been further improved by applying methods fully taking into account the complex survey design of NHANES III, which included oversampling of persons over 60 years old, African and Mexican Americans [18, 30]. Such design-based methods currently have been implemented only for some of the dose-response models described in the present work, and given its methodological and comparative focus, standard procedures without survey design adjustment were used throughout. Confounding patterns that would introduce the non-linearity consistently shown in the various models are difficult to envision, and the general dose-response shape was rather robust in sensitivity analyses that addressed additional potential confounders or effect modifiers, such as mineral supplements [38], physical activity [11], body mass index [44], or protein and energy intake [9].

Even though the dose-response analyses presented above currently might be the most detailed ones published on calcium intake and bone mineral density, only a small selection of pertinent methods could be demonstrated. For more technical overviews, the reader is referred to Greenland [16] and Steenland [39]. Many variations and additional non-linear methods exist, including other procedures to identify breakpoints in segmented relationships (see Muggeo [27] for a pointer to the literature; Wright [48] for a recent example of the application of grid-search segmented regression and Helmert contrast methods; Salanti [35] for a non-parametric approach based on isotonic regression). Non-linear regression models in the narrow sense received only a cursory treatment in the present study in the form of the  $E_{\max}$  model. An in-depth introduction can be found in the drug development literature [42]. Finally, only a continuous example outcome was examined here, but most of these methods have been adapted to all relevant types of outcomes, in particular binary and survival data.

### Conclusion

Non-linearity of dose-response relationships is often neglected in research pertaining to bone metabolism and homeostasis. The example of dietary calcium intake and bone mineral density shows how interesting additional aspects of an association can be revealed using nowadays well-established and accessible non-linear dose-response methods. The results point to further research and replication needs, e.g., regarding

sex differences in the dose-response curves and exact threshold locations. More importantly, the applied demonstration of the currently most important dose-response methods should enable and stipulate researchers to consider non-linearity in studies of bone health and disease more frequently. This could contribute to refining and advancing our understanding of many phenomena related to this field.

**Acknowledgments** The NCHS is responsible only for the initial data, which this work was based upon. However, solely the author is responsible for any analyses, interpretations, and conclusions described in the present work.

**Conflicts of interest** Lutz P. Breitling declares that he has no conflict of interest.

**Funding information** No funding for this work was received.

## References

- Ahmad O, Ramamurthi K, Wilson KE, Engelke K, Prince RL, Taylor RH (2010) Volumetric DXA (VXA): A new method to extract 3d information from multiple in vivo DXA images. *J Bone Min Res* 25(12):2744–51
- Alaimo K, McDowell MA, Briefel RR et al (1994) Dietary intake of vitamins, minerals, and fiber of persons ages 2 months and over in the United States: Third National Health and Nutrition Examination Survey, phase 1, 1988–91. *Adv Data* 258:1–28
- Ambler G (2010) (modified by Benner A). mfp: multivariable fractional polynomials; <http://CRAN.R-project.org/package=mfp>
- Autier P, Koechlin A, Smans M et al (2012) Mammography screening and breast cancer mortality in Sweden. *J Natl Cancer Inst* 104:1080–93
- Bass M, Ford MA, Brown B et al (2006) Variables for the prediction of femoral bone mineral status in American women. *South Med J* 99(2):115–22
- Bischoff-Ferrari HA, Kiel DP, Dawson-Hughes B et al (2009) Dietary calcium and serum 25-hydroxyvitamin D status in relation to BMD among US adults. *J Bone Mineral Res* 24(5):935–42
- Bornkamp B, Pinheiro J, Bretz F (2014) DoseFinding: planning and analyzing dose finding experiments; <http://CRAN.R-project.org/package=DoseFinding>
- Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 83(403):596–610
- Dawson-Hughes B, Harris SS (2002) Calcium intake influences the association of protein intake with rates of bone loss in elderly men and women. *Am J Clin Nutr* 75:773–9
- Desquilbet L, Mariotti F (2010) Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med* 29:1037–57
- Dionyssiotis Y, Paspati I, Trovas G et al (2010) Association of physical exercise and calcium intake with bone mass measured by quantitative ultrasound. *BMC Women's Health* 10:12
- Feuer EJ, Kessler LG, Baker SG et al (1991) The impact of break-through clinical trials on survival in population based tumor registries. *J Clin Epidemiol* 44:141–53
- Field A, Miles J (2010) *Discovering statistics using SAS*. London
- Fox J, Weisberg HS (2011) *An R companion to applied regression*, 2nd ed. Thousand Oaks
- Govindarajulu US, Malloy EJ, Ganguli B (2009) The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. *Int J Biostat* 5(1), article 2
- Greenland S (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 6:356–65
- Greenland S (2008) Introduction to regression models. In: Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd ed. Philadelphia. Chapter 20, 381–417
- Hahs-Vaughn DL (2005) A primer for using and understanding weights with national datasets. *J Exp Educ* 73(3):221–48
- Harrell FE (2001) *Regression modeling strategies: with applications to linear models, logistic regression and survival analysis*. Springer, New York
- Hong H, Kim EK, Lee JS (2013) Effects of calcium intake, milk and dairy product intake, and blood vitamin D level on osteoporosis risk in Korean adults: analysis of the 2008 and 2009 Korea National Health and Nutrition Examination Survey. *Nutr Res Pract* 7(5):409–17
- Hunt CD, Johnson LK (2007) Calcium requirements: new estimations for men and women by cross-sectional statistical analyses of calcium balance data from metabolic studies. *Am J Clin Nutr* 86:1054–63
- Jacoby WG (2000) Loess: a nonparametric, graphical tool for depicting relationships between variables. *Elect Stud* 19(4):577–613
- Kim H, Fay MP, Feuer EJ et al (2000) Permutation tests for jointpoint regression with applications to cancer rates. *Stat Med* 19:335–51
- Looker AC, Wahner HW, Dunn WL et al (1998) Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* 8:468–89
- MacDougall J (2006) Analysis of dose-response studies—Emax model. In: Ting N (ed) *Dose finding in drug development*. Springer, New York
- Matkovic V, Heaney RP (1992) Calcium balance during human growth: evidence for threshold behavior. *Am J Clin Nutr* 55:992–6
- Muggeo VM (2003) Estimating regression models with unknown break-points. *Stat Med* 22:3055–71
- Muggeo VMR (2008) Segmented: an R package to fit regression models with broken-line relationships. *R News* 8:20–5
- Myer PA, Mannalithara A, Singh G et al (2012) Proximal and distal colorectal cancer resection rates in the United States since widespread screening by colonoscopy. *Gastroenterology* 143:1227–36
- National Center for Health Statistics (1994) Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94. *Vital Health Stat* 1(32)
- National Institutes of Health (2013) Calcium dietary supplement fact sheet. <http://ods.od.nih.gov/factsheets/Calcium-HealthProfessional/> (accessed 2014-02-10)
- Nelson KM, Reiber G, Boyko EJ (2002) Diet and exercise among adults with type 2 diabetes: findings from the Third National Health and Nutrition Examination Survey (NHANES III). *Diabetes Care* 25:1722–8
- Peacock M (2010) Calcium metabolism in health and disease. *Clin J Am Soc Nephrol* 5:S23–30
- Royston P, Ambler G, Sauerbrei W (1999) The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 28:964–74
- Salanti G, Ulm K (2005) A non-parametric framework for estimating threshold limit values. *BMC Med Res Methodol* 5:36
- Sauerbrei W, Meier-Hirmer C, Benner A et al (2006) Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. *Comput Stat Data Anal* 50:3464–85
- Shafique K, Oliphant R, Morrison DS (2012) The impact of socio-economic circumstances on overall and grade-specific prostate cancer incidence: a population-based study. *Br J Cancer* 107:575–82

38. Shea B, Wells G, Cranney A et al (2004) Calcium supplementation on bone loss in postmenopausal women. *Cochrane Database Syst Rev* 1, CD004526
39. Steenland K, Deddens JA (2004) A practical guide to dose-response analyses and risk assessment in occupational epidemiology. *Epidemiology* 15:63–70
40. Stone KL, Seeley DG, Lui LY et al (2003) BMD at multiple sites and risk of fracture of multiple types: long-term results from the Study of Osteoporotic Fractures. *J Bone Miner Res* 18:1947–54
41. Strasak AM, Rapp K, Brant LJ et al (2008) Association of  $\gamma$ -glutamyltransferase and risk of cancer incidence in men: a prospective study. *Cancer Res* 68(10):3970–7
42. Ting N (2006) Dose finding in drug development. Springer, New York
43. Uusi-Rasi K, Kärkkäinen MUM, Lamberg-Allardt CJE (2013) Calcium intake in health maintenance—a systematic review. *Food Nutr Res*
44. Varenna M, Binelli L, Casari S et al (2007) Effects of dietary calcium intake on body weight and prevalence of osteoporosis in early postmenopausal women. *Am J Clin Nutr* 86:639–44
45. Wang MC, Dixon LB (2006) Socioeconomic influences on bone health in postmenopausal women: findings from NHANES III, 1988–1994. *Osteoporos Int* 17:91–8
46. Warensjö E, Byberg L, Melhus H, Gedeberg R, Mallmin H, Wolk A, Michaëlsson K (2011) Dietary calcium intake and risk of fracture and osteoporosis: prospective longitudinal cohort study. *BMJ* 342:d1473
47. Wood SN (2001) Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting. *Biometrics* 57:240–44
48. Wright NC, Chen L, Niu J et al (2012) Defining physiologically “normal” vitamin D in African Americans. *Osteoporos Int* 23:2283–91
49. Yetley EA, Pfeiffer CM, Schleicher RL et al (2010) NHANES monitoring of serum 25-hydroxvitamin D: a roundtable summary. *J Nutr* 140(11):2030S–45S