PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	ORAL BISPHOSPHONATES ARE ASSOCIATED WITH INCREASED RISK OF ATYPICAL FEMORAL FRACTURES IN ELDERLY WOMEN
AUTHORS	Erviti, Juan; Alonso, Alvaro; Oliva, Belén; Gorricho, Javier; López, Antonio; Timoner, Julia; Huerta, Consuelo; Gil, Miguel; De Abajo, Francisco

VERSION 1 - REVIEW

REVIEWER	Donald Morrish MD 362 HMRC University of Alberta Edmonton, Alberta, Canada T6G2C8 Professor of Medicine, Division of Endocrinology and Metabolism No competing interests.
REVIEW RETURNED	06-Nov-2012

THE STUDY	The literature is adequately quoted
RESULTS & CONCLUSIONS	The study is well done within the limitations stated by the authors. However there is a rather small sample size compared to the largest study reported, by Park-Wyllie et al (2011) which had 716 atypical fractures. I am not sure this study therefore adds anything new to the literature.
GENERAL COMMENTS	The study is well done, but the sample size is rather small. The study duplicates other small studies and reaches the same conclusion. Since the database contains over 3.2 million patients, could more atypical fractures be found? This would make the study much stronger. (The Park-Wyllie study had 716 atypical fractures)

REVIEWER	Lydia Gedmintas Rheumatology clinical research fellow Brigham and Women's Hospital USA
REVIEW RETURNED	18-Nov-2012

THE STUDY	This is a very timely topic of research. At this point in the study of the association of bisphosphonates and atypical femur fracture, it is important to attempt to confirm radiographic confirmation of these fractures. If this is not possible with the dataset used, the authors should address this and consider changing the title of the study as the ASBMR has determined that atypical femur fracture definition requires radiographic confirmation. Therefore the title of this study should likely be "subtrochanteric and

	diaphyseal fractures" rather than "atypical femoral fractures" as the authors were not able to radiologically confirm the fractures. This is an important distinction as prior studies have shown significant misclassification of these fractures based on diagnostic coding alone. In regards to the study design, many prior papers have addressed ever/never use of bisphosphonates and associations with atypical femur fracture. Duration of bisphosphonate and its association with atypical femur fracture is currently the area of interest in the literature-- the authors therefore appropriately address duration in their study design. However, it is unclear why there were two definitions of treatment length; it would be helpful if this was explained further or potentially only use "cumulative duration of actual treatment" as this is likely the definition of most interest. It would be helpful for the authors to go into further detail of how bisphosphonate use was defined - for example, was there a minimum exposure, such as 30 days, required before labeling the patient as a bisphophonate user? Lastly, a cohort design is likely a stronger method to address duration of use, and this may be something that the authors could consider. In regards to the statistical methods, it was unclear what methods were used particularly when performing trend and this could be more clearly discussed. In addition, while all the covariates used in the multivariable model are important and associated with fracture, considering the number of covariates and the small sample size this could lead to overadjustment. Would benefit from editing prior to publication for ease of reading. Two recent papers, Meier RP et al, Arch Int Med, 2012 and Dell RM, JBMR, 2012 are important recent contributions to this field using radiographic confirmation of atypical femur fracture, and should likely be discussed in the setting of this study. It would be helpful if the authors addressed what niche their study fills in the current literature of atypical femur fracture.
RESULTS & CONCLUSIONS	While the main research question addressed, ie the association of ever use of bisphosphonates and atypical femur fractures, is clear from the conclusions, the conclusions made from the results with duration of use are less clear. The multivariate analysis with this small number of cases may not be the most accurate way to present the data, as it is unclear if this data is overadjusted. In addition, it is unclear if a strong conclusion can be made from this trend test using these small number of cases. It would be helpful in the conclusion to discuss some of the more recent studies suggested above that have addressed the issue of duration and discuss this paper in their context.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Donald Morrish MD

362 HMRC

University of Alberta

Edmonton, Alberta, Canada T6G2C8

Professor of Medicine, Division of Endocrinology and Metabolism No competing interests.

The literature is adequately quoted

The study is well done within the limitations stated by the authors. However there is a rather small sample size compared to the largest study reported, by Park-Wyllie et al (2011) which had 716 atypical fractures. I am not sure this study therefore adds anything new to the literature.

We agree the study published by Park-Wyllie et al (2011) has a much larger sample size. Apart from this study, there are just a few studies with a sample size around 40 cases. One peculiarity about our study is that it was carried out in a Mediterranean population, with a lower risk for bone fractures compared to Anglo-Saxon or Northern European countries (Kanis JA, et al. International variations in hip fracture probabilities: implications for risk assessment. JBMR 2002;17:1237-1244). It could be hypothesized that, because of the lower risk of fractures in the Spanish population, the association between bisphosphonates and subtrochanteric or diaphyseal fractures might not be evident. However that was not the case and our findings support this association even in a low-risk population.

Because of rather small sample size, studies on atypical fractures bear a high risk of bias. Thereby it is important to replicate them in different populations and databases. BIFAP database is a non-profit research project operated by the Spanish Medicines Agency, a public agency belonging to the Spanish Department of Health. The Spanish Medicines Agency guarantees the quality of the database. The fact that our study presents outcomes consistent with the findings in other similar studies supports the association between the use of bisphosphonates and the incidence of subtrochanteric and diaphyseal fracture risk.

The study is well done, but the sample size is rather small. The study duplicates other small studies and reaches the same conclusion. Since the database contains over 3.2 million patients, could more atypical fractures be found? This would make the study much stronger. (The Park-Wyllie study had 716 atypical fractures)

Please find below the main differences between the Park-Wyllie study and ours that may explain the different number of cases found in each study:

Park-Wyllie study includes cases registered during 6 years (April 1st 2003 to March 31st 2009) whereas in our study this is a 4-year period instead (January 1st 2005 to December 31st 2008). The average follow-up period was 7 years in Park-Wyllie study and 2.75±1.94 years in our study.

Park-Wyllie study includes a cohort of some 800,000 women aged >68 years using bisphosphonates. Our study had a total population of 3.2 million patients but our cohort of women aged >65 years included approximately 280,000 individuals. As well as being a larger cohort, Park-Wyllie included

women at higher risk of fracture, eg previous fractures in 70% of patients in Park-Wyllie and 20% in our study (see Baseline Characteristics of Cases and Controls in each study).

Finally, Spanish populations have a lower risk of fracture compared with the Anglo-Saxon populations after adjusting for number of risk factors.

Reviewer: Lydia Gedmintas

Rheumatology clinical research fellow

Brigham and Women's Hospital

USA

This is a very timely topic of research.

At this point in the study of the association of bisphosphonates and atypical femur fracture, it is important to attempt to confirm radiographic confirmation of these fractures. If this is not possible with the dataset used, the authors should address this and consider changing the title of the study as the ASBMR has determined that atypical femur fracture definition requires radiographic confirmation. Therefore the title of this study should likely be "subtrochanteric and diaphyseal fractures" rather than "atypical femoral fractures" as the authors were not able to radiologically confirm the fractures. This is an important distinction as prior studies have shown significant misclassification of these fractures based on diagnostic coding alone.

We absolutely agree with this comment and the title will be changed according to the reviewer's suggestion.

In regards to the study design, many prior papers have addressed ever/never use of bisphosphonates and associations with atypical femur fracture. Duration of bisphosphonate and its association with atypical femur fracture is currently the area of interest in the literature-- the authors therefore appropriately address duration in their study design. However, it is unclear why there were two definitions of treatment length; it would be helpful if this was explained further or potentially only use "cumulative duration of actual treatment" as this is likely the definition of most interest.

We assessed duration of bisphosphonate use in two different ways. First, as the cumulative use of bisphosphonates since the first prescription and, second, as the time since the first prescription. The idea behind this choice was to capture two different aspects of bisphosphonate use: the total use and the total time that the bone has been exposed to bisphosphonates. As we explain in the manuscript discussion, pathophysiological reasons support both measures. Therefore, we prefer to report both.

Nonetheless, the methods section clarifies the definitions of cumulative exposure, which now reads as follows:

> In order to assess the effects of treatment length on the outcomes four different subgroups were considered based on cumulative duration of actual treatment, namely 30 days or less; >30 days to ≤1 year; >1 to ≤3 years and over 3 years. The effects of time of bisphosphonate exposure on atypical hip fracture risk were also analyzed. Exposure was measured as the time (in days) since the first prescription.

These changes were also included in table 2.

It would be helpful for the authors to go into further detail of how bisphosphonate use was defined - for example, was there a minimum exposure, such as 30 days, required before labeling the patient as a bisphophonate user?

A patient was categorized as bisphosphonate user if she had any prescription for bisphosphonates (no minimum duration was required), though a majority of users had prescriptions for more than 30 days (see table 2 in the manuscript). Oral bisphosphonate presentations in Spain provide medication for 28 days. We have clarified this issue in the methods section of the manuscript:

> Individuals were classified as ever vs never users. Ever users were those with at least one prescription, with no minimum duration.

Lastly, a cohort design is likely a stronger method to address duration of use, and this may be something that the authors could consider.

We agree with the reviewer that cohorts are in general a better study design than case-control studies, because they are less likely to suffer from recall bias, differential misclassification of exposure, and selection bias due to inappropriate selection of controls. However, our case-control study does not have those problems since cases and controls were nested in a well-defined dynamic cohort, removing the risk of selection bias, and information on treatments and comorbidities was collected before the outcome occurred, reducing the threat of differential misclassification of the exposure. In fact, conducting a case-control study instead of a cohort study is a more efficient use of resources since the same results are obtained with a much smaller sample size. Duration of use was defined in the same way that it would have been defined in a cohort study, that is, the time using bisphosphonate before having a hip fracture before a woman had the event, or the equivalent time for a woman without a fracture. We have added a short paragraph to the discussion highlighting this issue:

> Finally, our study had a case-control design and not a cohort design, which is supposed to be a stronger method. However, our cases and controls were selected from a well-defined cohort, reducing the possibility of selection bias, and information on treatment use and comorbidities was recorded before hip fractures occurred, making differential misclassification of the exposure less likely.

In regards to the statistical methods, it was unclear what methods were used particularly when performing trend and this could be more clearly discussed. In addition, while all the covariates used in the multivariable model are important and associated with fracture, considering the number of covariates and the small sample size this could lead to overadjustment.

Linear trends were calculated assigning the median value to each category, and including that value as a continuous variable in all models. This approach takes into account the distribution of the variable within each category and prevents outliers having a large impact in our estimates. This has been clarified in the methods section:

> Treatment duration was assessed as well and results were tested to identify a trend. Tests for trend were performed assigning the median to each category of ordinal variables, and including that value as a continuous variable in the models.

Regarding the multivariable adjustment, we decided to adjust for all possible confounders in spite of the limited sample size to avoid residual confounding in the estimates as much as possible. Notably, results from model 1 (minimally adjusted) and model 2 (multivariable) were qualitatively similar, showing increased risk of atypical fractures with increased duration use. Therefore, 'overadjustment' is unlikely to be responsible for our study results. Nonetheless, we recognize this is a limitation, leading to imprecise estimates with wide confidence intervals, and the revised version states it clearly. We have included the following sentence in the discussion:

> One of the main limitations of this study is the small number of cases, which made it unfeasible to perform subgroup or individual drugs analyses, and led to wide confidence intervals in the estimates of association.

Would benefit from editing prior to publication for ease of reading.

OK.

Two recent papers, Meier RP et al, Arch Int Med, 2012 and Dell RM, JBMR, 2012 are important recent contributions to this field using radiographic confirmation of atypical femur fracture, and should likely be discussed in the setting of this study. It would be helpful if the authors addressed what niche their study fills in the current literature of atypical femur fracture.

OK. Thank you for the contribution. The main niche of this study is the effects of bisphosphonates in a low fracture risk population. Please see the tracked new version of the manuscript

While the main research question addressed, ie the association of ever use of bisphosphonates and atypical femur fractures, is clear from the conclusions, the conclusions made from the results with duration of use are less clear. The multivariate analysis with this small number of cases may not be the most accurate way to present the data, as it is unclear if this data is overadjusted. In addition, it is unclear if a strong conclusion can be made from this trend test using these small number of cases.

We agree with the reviewer that the reduced number of cases limits the ability to make strong inferences for these results. This uncertainty is reflected by the wide confidence intervals in our estimates of association for the duration of use analyses. However, there are two reasons why we think these results are valuable. First, the trend analysis, which is more powerful to detect linear associations than the categorical analysis, showed highly significant associations between duration of use and risk of atypical fracture. Second, results from model 1 (minimally adjusted) and model 2 (multivariable) were qualitatively quite similar, i.e. both showed higher risk of fracture with increased duration of use. Differences in the magnitude of the association between models 1 and 2 are most likely due to the limited number of cases and, therefore, should be evaluated with caution. We believe that presenting multivariable adjusted models is necessary to demonstrate that confounding is not the main reason responsible for this association. To highlight the uncertainty of our results, though, we have including the following sentences in the discussion:

> One of the main limitations of this study is the small number of cases, which made it unfeasible to perform subgroup or individual drugs analyses, and led to wide confidence intervals in the estimates of association.

It would be helpful in the conclusion to discuss some of the more recent studies suggested above that have addressed the issue of duration and discuss this paper in their context.

OK. Please see the tracked new version of the manuscript.

VERSION 2 – REVIEW

REVIEWER	Donald W Morrish, MD, FRCPC Professor of Medicine Division of Endocrinology and Metabolism University of Alberta, Edmonton Alberta, Canada No competing interests
REVIEW RETURNED	20-Dec-2012

GENERAL COMMENTS	I appreciate the authors review to my comments. They make a reasonable case for publishing a smaller case series than that of Park-Wylie et al, ie that the Spanish population may have some different incidences of fracture. I think they should incorporate perhaps a shortened version of their response to my comments in the discussion as this has some valuable points.

VERSION 2 – AUTHOR RESPONSE

A track edited manuscript dated 2012-12-21 has been uploaded at "point 6", File upload