

Original Article

Validity of Self-Report of Fractures: Results from a Prospective Study in Men and Women Across Europe

A. A. Ismail, T. W. O'Neill, W. Cockerill, J. D. Finn, J. B. Cannata, K. Hozzowski, O. Johnell, C. Matthis, H. Raspe, A. Raspe, J. Reeve, A. J. Silman and the EPOS Study Group*

ARC Epidemiology Unit, Manchester, UK

Abstract: In population-based studies of osteoporosis, ascertainment of fractures is typically based on self-report, with subsequent verification by medical records. The aim of this analysis was to assess the validity of self-report of incident nonspine fractures using a postal questionnaire. The degree of overreporting of fracture (false positives) was assessed by comparing self-reports of new fracture from respondents in the multicenter European Prospective Osteoporosis Study with data from other sources including radiographs and medical records.

In the analysis, 563 subjects reported nonspine fractures. Verification of the presence of fracture was possible in 510 subjects. Of these, fractures were not confirmed in 11% (false positives). The percentage of false positives was greater in men than in women (15% vs 9%, $p=0.04$), and less for fractures of the distal forearm and hip than for fractures at other sites. In a separate study, the degree of underreporting (false negatives) was assessed by follow-up of 251 individuals with confirmed fracture ascertained from the records of fracture clinics in three European centers (Lubeck, Oviedo, Warsaw). Questionnaire responses were received from 174 (69%) subjects. Of these, 12 (7%) did not recall sustaining a fracture (false negatives). The percentage of false negatives was lower for hip and distal forearm fractures with only 3 of 90 (3%) such fractures not recalled. Using the combined data from both studies, of those who reported a 'date' of fracture on the questionnaire, 91% of subjects were correct to within 1 month of the actual date of the fracture. A postal questionnaire is a relatively simple and accurate method for obtaining information about the occurrence of hip and distal forearm fractures, including their timing. Accuracy of ascertainment of fractures at other sites is less good and where possible self-reported fractures at such sites should be verified from other sources.

Correspondence and offprint requests to: Professor Alan Silman, ARC Epidemiology Unit, Stopford Building, University of Manchester, Oxford Road, Manchester, M13 9PT, UK.

*Members of the European Prospective Osteoporosis Study Group. *Project Coordinators:* J. Reeve and A.J. Silman, Cambridge and Manchester, UK. *Austria:* Graz Dr K. Weber; *Belgium:* Leuven Prof. J. Dequeker; *Croatia:* Zagreb Prof. I. Jajic; *Czech Republic:* Prague Prof. S. Havelka, Prof. J. Stephan *Slovakia:* Piestany Dr P. Masaryk; *France:* Montceau-Les-Mines Prof. P.D. Delmas, Dr F. Marchand; *Germany:* Berlin, Steglitz Prof. D. Felsenberg, Berlin, Potsdam Prof. D. Banzer, Berlin, Charité Prof. W. Reisinger, Bochum Prof. H. Schatz, Erfurt Dr G. Kragl, Heidelberg Dr C. Scheidt-Nave, Jena Dr K. Abendroth, Dr B. Felsch, Lubeck Prof. H. Raspe, C. Matthis; *Greece:* Athens Prof. G. Lyritis, Heraklion Prof. E. Dretakis; *Hungary:* Budapest Dr G. Poor; *Italy:* Siena Prof. C. Gennari; *Netherlands:* Amsterdam Dr P. Lips, Rotterdam Dr HAP Pols; *Norway:* Oslo Dr J. A. Falch; *Poland:* Szczecin Dr T. Miazgowski, Warsaw Dr K. Hozzowski, Prof. R. Lorenc; *Portugal:* Azores Dr J. Bruges Armas, Oporto Prof. A. Lopez Vaz; *Russia:* Moscow Prof. L. I. Benevolenskaya, Yaroslavl Dr O. Ershova; *Spain:* Madrid Dr A. Rapado, Seville Prof. R. Perez Cano, Prof. F. Galan Galan, Oviedo Prof. J. B. Cannata; *Sweden:* Malmö Dr O. Johnell; *Finland:* Kuopio Dr H. Kröger; *Turkey:* Istanbul Prof. G. Dilsen; *United Kingdom:* Aberdeen Dr D. M. Reid, Bath Dr A. K. Bhalla, Cambridge Dr C. Todd, Harrow Dr J. Reeve, Truro Dr A. D. Woolf.

Keywords: Fracture; Manikin; Osteoporosis; Questionnaire; Validation

Introduction

In population-based studies of osteoporosis, ascertainment of fractures is typically based on self-report, with subsequent verification by medical records. However, self-report of fractures is subject to errors of recall and studies based on this method of ascertainment may result in misclassification of fracture status. Where present, information on fracture history may be obtained from record linkage systems, though, apart from some Scandinavian countries, such systems are not widely available.

Fractures which are not recalled (underreporting/false negatives) may be missed when ascertainment is based on self-report. If significant, this may result in an underestimation in incidence of fracture. Previous studies provide conflicting evidence as to the magnitude of underreporting in population studies. Nevitt et al. [1] looked at the detailed medical records of 283 women who denied experiencing a fracture in the previous year and confirmed that none had actually occurred. In contrast, other studies which have looked at recall over longer time periods suggest a potentially important degree of underreporting. In one Swedish study [2] 14% of women who had suffered a previous distal forearm fracture could not recall it (mean time since fracture 11 years) while, in another, 40% of all fractures were not recalled (approximate mean time since fracture 15 years) [3]. Thus the magnitude of underreporting of fractures in population studies remains uncertain.

Fractures which are reported but which did not occur (overreporting/false positives) may be detected by checking the subject's medical record, though in multicenter studies this is not always possible because of difficulties with availability or accessibility of this information. Several studies have looked at overreporting of fractures [1,4-7], with the percentage of false positives varying from 0 to 12%. However, these studies have been largely confined to women. There are, to our knowledge, no data concerning overreporting of fractures in men.

In subjects who report fractures, characterization of the site of fracture is usually undertaken by review of the radiographic or medical records. However, as outlined above, such data are not always available. One approach to this is to include in the questionnaire a body 'manikin' and invite subjects to mark on this the site of fracture. Manikins have been used extensively in other areas of research, including characterization of the site of pain in joint diseases [8,9], though not as yet in population-based studies of osteoporosis.

The European Prospective Osteoporosis Study (EPOS) is a population-based, multicenter study of incident fractures in men and women. Data concerning fractures were obtained by self-report using a postal questionnaire which included a manikin. The aim of this analysis was to determine the validity of self-report of nonspine fractures using this instrument.

Methods

Overall Design

Overreporting of Fractures. The degree of overreporting (false positives) was assessed by verification of self-reports of fracture from subjects participating in EPOS.

Underreporting of Fractures. In a separate study, subjects with confirmed fractures (who were not part of the EPOS cohort) were recruited from fracture clinics in three of the EPOS centers. These subjects were followed by postal questionnaire to determine the degree of underreporting of nonvertebral fractures (false negatives).

Assessment of Overreporting

Subjects and Baseline Examination. Subjects who took part in the analysis were participants in EPOS. This is a multicenter study of the occurrence and risk factors for incident fractures. A baseline survey of vertebral osteoporosis (European Vertebral Osteoporosis Study) was undertaken between 1989 and 1993 [10]. Men and women aged 50 years and over were recruited from population registers in 36 European centers and invited to attend for an interviewer-administered questionnaire and lateral spinal radiographs. Details of the population samples and study methods have been described elsewhere [10].

Follow-up and Postal Questionnaire. Subjects from 29 of the EPOS centers have been followed with a postal questionnaire, the purpose of this being to ascertain incident fractures. In two centers, subjects were followed by telephone interview, and in one center by repeat interview; these centers were not included in this analysis. In the majority of centers, from 1994, subjects have been contacted annually on at least two occasions, though in the 8 German centers they have been contacted biannually over a 3-year period.

The postal questionnaire included several questions: 'Since participating in the baseline survey, have you fractured a bone?' If the response was 'yes', subjects were then asked: 'What date was the fracture?' and 'Where did you attend for treatment of the fracture?' If a subject had experienced more than one fracture event, they were asked to record how many fracture events had taken place, though to include details of only the most recent event. In addition subjects were asked to mark on a manikin the site of fracture (Fig. 1).

The questionnaire was translated and back-translated into the various European languages in which its use was intended. The translation was altered for any ambiguities or disagreements. Any amendments were incorporated into all the language versions before the final version was produced to insure comparability of data.

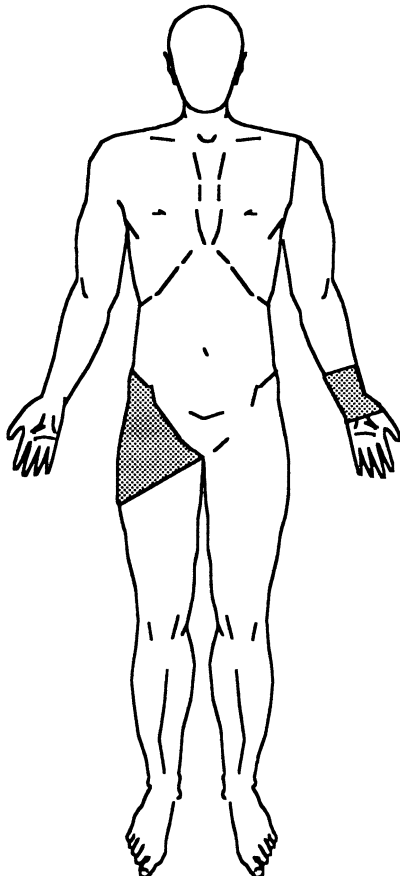


Fig. 1. Manikin. *Shaded areas* (not included on the questionnaire) were used to define hip and distal forearm fracture.

Ascertainment of Fracture. Each center was responsible for verification of reported fractures at their own center. Prior to the study, each center received a detailed instruction manual and a standard log book for confirming fractures.

A hierarchical three-stage procedure was used. Where possible confirmation of fracture was obtained by documented radiographic opinion or review of the original radiograph. Such evaluation was done locally – no centralized evaluation of radiographs was undertaken. Where radiographic data were lacking (because the films either were not available or had not been taken), the information about the reported injury was sought using documented contemporary clinical opinion. In a proportion of subjects it was not possible to verify fractures by either method. In these circumstances, where possible, subjects were interviewed by a local clinician on the basis that it may be possible to determine whether or not fracture had occurred, and also the type of fracture, by what had been told to them or by the methods which had been used for treatment. In some cases, however, it was not possible to obtain any further details about the injury.

In addition to verifying the occurrence of reported fractures, each center was asked to identify the site of

any reported injury and to assign an ICD code using the 9th edition of the International Classification of Diseases [11].

Accuracy of Manikin. A priori, we defined an area on the upper and lower limb of the manikin corresponding to an area within which we expected individuals would mark if they had sustained a distal forearm or hip fracture respectively (Fig. 1). We then looked at questionnaires from subjects (non-German centers) with radiographically confirmed fractures. Without knowledge of the type of fracture sustained, using a transparency, one of us (W.C.) recorded whether or not the mark was within or outside these defined areas. Based on these data we then determined the sensitivity and specificity of these areas in detecting hip and distal forearm fracture.

Assessment of Underreporting

To investigate the degree of possible underreporting of fractures we undertook a separate study at three centers: Lubeck (Germany), Oviedo (Spain) and Warsaw (Poland). Each of these centers recruited quota samples of men and women aged 50–79 years who had sustained a radiographically confirmed fracture (hip/distal forearm/other) in the previous 18 months, from local fracture clinics. These subjects were not part of the EPOS cohort. In total 251 subjects were sent a postal questionnaire similar to that used in EPOS. The questionnaire was modified slightly to include, instead of the first question (see above), the question ‘In the last 18 months have you fractured a bone?’

Analysis. Because of the different methods of ascertainment, analysis of data from the German and non-German centers was initially performed separately. However, because the results were broadly similar the data were combined and are the data presented in this paper. In the non-German centers we included subjects from the first follow-up only. We excluded subjects who reported incident vertebral fractures.

In all analyses, subjects were the unit of analysis. When subjects reported more than one fracture, we selected one for analysis according to the following method. Firstly we included those which could be verified by radiograph, then physician, and then interview. If fractures were verified by the same method, we included first hip, then distal forearm then ‘other’ fractures. If the fracture types were similar we selected the most recent fracture.

In subjects in whom verification of self-reports of fracture was possible, the percentage of false positives was calculated as the number of unconfirmed self-reports divided by the total number of confirmed and unconfirmed self-reports (multiplied by 100). We looked separately at false positives by gender, method of verification (radiograph/physician/subject interview) and by site of injury: hip (neck of femur), ICD code 820; distal forearm, ICD code 813; humerus, ICD code 812;

hand and feet, ICD codes 814–817, 825, 826; leg (other femur, patellae, tibia/fibula and ankle), ICD codes 821–824; other (skull, rib, pelvis, clavicle and scapula), ICD codes 800–804, 807, 808, 810, 811. We looked at possible geographical influences by comparing the proportion of false positives in centers from Southern Europe, Eastern Europe and the remaining centers in Western and Northern Europe.

Sensitivity was calculated as the percentage of subjects (from the three center study) with radiographically confirmed fractures which were recalled. The percentage of false negatives was calculated as 100%–sensitivity. We looked separately at false negatives by gender and site of fracture.

Using the combined data from the non-German centers in EPOS and the three centers participating in the assessment of underreporting, we assessed the accuracy of the timing of fracture by comparing the date of the reported fracture (from the postal questionnaire) with the actual date as determined by the investigator. All analyses were performed on the statistical package STATA [12].

Results

Data from 28 centers were used in the analysis [one center, Warsaw, was excluded because a relatively high proportion of the self-reported fractures (approximately 90%) were not verified]. In total 9627 subjects from the 20 non-German centers were sent the postal questionnaire, of whom 8157 (85%) replied. Of the 4454 subjects from the 8 German centers who were contacted, 3726 (84%) replied during the subsequent follow-up periods. The median follow-up time for the non-German centers was 2.0 years (range 0.6–4.1 years) and for the German centers 3.4 years (range 0.4–5.1 years). During follow-up, 563 subjects (177 men and 386 women) reported a new non spine fracture. Of these, 25 subjects reported two, and 2 subjects reported three fractures.

Of the 563 subjects reporting fracture, 53 (9%) subjects were excluded from this analysis because it was not possible to verify these reports using either radiograph, physician or interview. In total it was possible to verify fracture status in 510 (91%), by either radiograph ($n = 428$), physician/medical record ($n = 37$) or subject interview ($n = 45$).

Table 1. False positive self-reported fractures by gender and age

Self-reported fractures (verification available) $n = 510^a$	Both sexes		Men		Women	
	n	False positives (%)	n	False positives (%)	n	False positives (%)
All ages	510	55 (11%)	151	23 (15%)	359	32 (9%)
50–59 years	184	25 (14%)	70	13 (19%)	114	12 (11%)
60–69 years	171	17 (10%)	37	5 (14%)	134	12 (9%)
70+ years	155	13 (8%)	44	5 (11%)	111	8 (7%)

^a Fracture verification by radiograph, physician, or interview.

Overreporting of Fractures

Table 1 presents the percentage of false positives by age and gender. Overall, 55 (11%) of the 510 subjects reported a fracture which was not subsequently confirmed. There was a decline in the percentage of false positives with increasing age. Compared with those individuals in the 50–59 year age group, individuals aged 70 years and over had fewer false positives (men: 19% vs 11%, $p = 0.3$; women: 11% vs 7%, $p = 0.4$). Further, there were more false positives among men than women (15% vs 9%, $p = 0.04$). The male excess became less marked with age, though there was no statistical evidence of an interaction between the effects of age and gender on the occurrence of false positives. The percentage of false positives was higher in Southern Europe (19%), than in Northern and Western (11%) or Eastern Europe (4%) ($p < 0.05$).

Table 2 presents the percentage of false positives by method used for verification and site of injury. Of the 510 subjects who had reported a fracture, it was not possible to determine the site of injury in 13 subjects, and therefore the analysis was restricted to the 497 subjects in whom this information was available. For both sexes overreporting was slightly greater for reported fractures verified by medical record and interview alone, compared with those verified by radiograph (14% vs 10%, $p = 0.3$). When analyzed by site of injury, overreporting was lower for hip, distal forearm and leg fractures than for fractures at the humerus, hand and foot and ‘other’ sites. Only 7 (5%) of 157 self-reported hip and distal forearm fractures were not confirmed compared with 17 (20%) of 84 fractures in the ‘other’ fracture site category ($p < 0.001$).

Underreporting of Fractures

In this study, of the 251 subjects with definite radiographic fracture who were sent the questionnaire, 174 replied (69%). Of these, 41 (24%) had hip fracture, 49 (28%) distal forearm fracture and 84 (48%) fractures at other sites. Their mean age was 65.5 years (SD = 8.0) and 72% were female.

Table 3 presents the false negatives by gender for hip, distal forearm and ‘other’ fractures. Overall, 12 (7%) did not recall sustaining a fracture (false negatives). In both

Table 2. Self-reported fractures: characteristics of false positives (FP) by site and level of verification

Site of injury	Radiograph				Radiograph/physician				Radiograph/physician/interview			
	<i>n</i>	Yes	No	FP	<i>n</i>	Yes	No	FP	<i>n</i>	Yes	No	FP
Hip (<i>n</i> = 29)	27	27	0	0	27	27	0	0	29	28	1	4%
Distal forearm (<i>n</i> = 128)	110	104	6	6%	120	114	6	5%	128	122	6	5%
Humerus (<i>n</i> = 50)	41	37	4	10%	46	39	7	15%	50	43	7	14%
Hand and foot (<i>n</i> = 117)	92	78	14	15%	100	84	16	16%	117	99	18	15%
Leg (<i>n</i> = 89)	81	78	3	4%	83	79	4	5%	89	85	4	5%
Other ^a (<i>n</i> = 84)	67	52	15	22%	78	62	16	21%	84	67	17	20%
All (<i>n</i> = 497)	418	376	42	10%	454	405	49	11%	497	444	53	11%

^a Includes skull, pelvis, clavicle, scapula and rib.

Table 3. False negatives by site and gender

Fracture type	Men		Women	
	No. of fractures	False negatives (%)	No. of fractures	False negatives (%)
Hip	10	0	31	1 (3%)
Distal forearm	8	1 (12%)	41	1 (2%)
Other	31	6 (19%)	53	3 (6%)
All	49	7 (14%)	125	5 (4%)

men and women, the percentage of false negatives was lower for hip and distal forearm fractures compared with 'other' fracture sites (3% vs 11%, $p = 0.05$). Overall, the proportion of false negatives was lower in women than men (4% vs 14%, $p = 0.02$).

Accuracy of the Manikin

Fifty-one of 60 radiographically confirmed distal forearm fractures had been marked within the shaded region of the upper limb (Fig. 1), resulting in a sensitivity of this site for detecting distal forearm fracture of 85%. Only 3 of 164 nondistal forearm fractures were marked at this site, resulting in a specificity of 98%. When the analysis was restricted to fractures of the upper limb, sensitivity remained unchanged, though specificity fell slightly to 95%. All 18 radiographically confirmed hip fractures were marked within the shaded region of the lower limb (Fig. 1), resulting in a sensitivity of this site for detecting hip fracture of 100%. Only 3 of 206 nonhip fractures were marked within the site, giving a specificity 99%. When the analysis was restricted to fractures of the lower limb, sensitivity remained unchanged and specificity fell slightly to 98%.

Accuracy of Timing of Fracture

Using combined data from both studies, the timing of fracture from the questionnaire was available in a total of

378 subjects with radiographically confirmed fractures. Amongst these, 344 (91%) reported their fracture date to within 1 month of the actual date (as recorded by the investigator) of fracture, 18 (5%) reported it to within 6 months, 7 (2%) to within 1 year, and 9 (2%) more than 1 year from the actual date. Similar proportions of subjects reported that their fracture occurred before the actual date and after the actual date.

Two hundred and thirty-nine (93%) of the 258 subjects who had experienced a fracture within the 12 months prior to receipt of the questionnaire recalled it to be within 1 month of the actual date, compared with 105 (88%) in those 120 subjects whose fractures had occurred over 12 months before receipt of the questionnaire.

Discussion

In this prospective study, using a postal questionnaire to ascertain fractures, 11% of subjects with self-reported fractures did not have their fracture confirmed by a review of radiographs, medical records or interview. Seven per cent of subjects with radiographically confirmed fractures were unable to recall their fractures. The overall accuracy of self-report was better in women than men, and for hip and distal forearm fracture than for fractures at other sites. In those who reported fractures, the 'manikin' was a relatively accurate method for detecting hip and distal forearm fractures.

There are a number of methodologic issues to be considered in interpreting these findings. In EPOS a three-level procedure was used to verify self-reported fracture: radiograph, medical notes and interview of subject. For practical reasons, verification was undertaken locally. Each center received detailed instructions about fracture verification; however, the extent to which this may have reduced possible between-center variation in accuracy of verification is unknown.

It was recognized at the onset of the study that it would not always be possible to access radiographs or medical notes. In some European countries, for example, individuals keep their radiographs after they have been performed and these may become lost or difficult to obtain for review. In these circumstances, where

possible, subjects were interviewed by a local clinician. The accuracy of this approach is likely to be less good than confirmation by radiograph or medical record. In practice, however, only a small proportion of subjects (8.6%) had fractures verified in this way. Furthermore a scheme incorporating this validation approach produced similar results to that when radiographs alone were reviewed.

We were unable, using the three-stage strategy, to confirm (or refute) the presence of a fracture in 9% of subjects. It is possible that the accuracy of the nonverified fracture reports may differ from those in whom verification was possible. Although this is a relatively small proportion, some caution is required in extrapolating the results to this group. Similar considerations apply in extrapolation of the results to the 15% of subjects who did not reply.

In the study we asked about recall of fractures over the relatively short period of 12–18 months. In addition the subjects who were contacted had already participated in a survey of osteoporosis and this may have influenced the accuracy of their responses to the questionnaire. Caution is therefore required in extrapolating our results, either beyond the time period of the study or to cross-sectional studies where subjects have not previously participated.

There have been several previous studies looking at overreporting of fractures. Two of these suggest better recall than found in our study. In a study of 504 women, aged 35–94 years, in Rochester, Minnesota, the percentage of false positives for hip and distal forearm fractures (ever) were 0.3% and 0.7% respectively [6]. The method of ascertainment of fracture – by interview in this study – may, however, explain the better recall compared with our study, which used a questionnaire. In a sample from the Nurses Health study [4], all 30 women (age 30–55 years) who reported, by mailed questionnaire, having sustained a distal forearm fracture in the previous year had a medical record of this fracture. Their age and occupation may, however, have influenced recall. Other studies suggest similar results to our own, though the populations and time period of recall differ. Thus, in a younger group (45–56 years), and looking at recall over a 10-year period, 12% of self-reported fractures could not be confirmed [7]. A prospective study of fractures in elderly women aged 65 years and over [1] using regular (4 monthly) postal surveillance found that the percentage of false positives for fracture in the previous year was 11%. The similarity of these results to our own suggests, perhaps, that such detailed surveillance may not be required; however, the mean age of subjects in that study was greater so we cannot be confident of this. One advantage of more frequent surveillance might be to reduce losses to follow-up.

Our data confirm previous findings indicating substantial variation in the percentage of false positives by fracture site [1]. In addition there were fewer false positives for the major osteoporotic fractures (hip and distal forearm) than for fractures at most other sites.

To our knowledge there are no data concerning overreporting of fractures in men. Bush [5] found good agreement in a small group of men and women with fracture though did not report separately on these. In our study there was a greater percentage of false positives in men, though this may in part be because men sustain a greater proportion of fractures (distal extremity and ‘other’; data not presented) which are themselves associated with poorer recall.

In our study there was evidence of an age effect, with a greater proportion of false positives in those less than 60 years compared with those 70 years and over. As with men, this may in part be due to a greater proportion of ‘poor recall’ fractures in the younger age groups. There was in addition some evidence of a geographical variation in accuracy of reporting, with false positives being greater in Southern Europe. This did not appear to be related to differences in age or the proportion of ‘poor recall’ fractures. It is possible that other factors such as the level of education, social or cultural factors (on which we have no information) may be important.

Previous reports concerning the underreporting of self-report of fracture vary widely. Jonsson et al. [3] reported that 40% of fractures sustained in Malmö (median delay 15 years) and 22% in Sjöbo (median delay 14 years) were not recalled. In a separate Swedish study [2], 5 of 36 women (14%) failed to recall a distal forearm fracture (median delay 10 years). In contrast, Nevitt et al. [1] found no evidence of underreporting of fracture during a 1-year period in a population sample of 283 women. The numbers of fractures which one would expect, however, in this group is small and the study was perhaps limited in its ability to address the issue of underreporting.

Our results suggest that overall 7% of fractures were not recalled, though as with the false positives, there was variation by gender and site of fracture. The percentage of false negatives was greater in men, and for nonhip/distal forearm fracture than at other fracture sites. It is possible that this may be due to the interventions associated with these fractures, i.e. surgery and immobilization in plaster, which is less likely to have been forgotten. It is also possible in the case of hip and distal forearm fractures that subjects may have been informed that these may be due to osteoporosis and this may have influenced recall.

It is important to consider the consequences of the errors suggested by these data on both estimates of incidence and relationship with possible risk factors. The effect of underreporting on estimates of fracture incidence will vary according to the site of fracture; however, for hip and distal forearm fracture – the major osteoporotic fractures – our data suggest incidence will be underestimated by only a small degree (3%). Assuming that underreporting is unaffected by the occurrence of any putative risk factor under investigation, the effect of such misclassification would be to reduce slightly or attenuate the risk ratios. Similarly the influence of false positives when relying on postal questionnaire alone to ascertain fractures should be

assessed. In EPOS, for hip and distal forearm fractures, this proportion is relatively small (5%) and the overall effect of this would be to overestimate the true incidence of these fractures by approximately 2%, which again would only slightly attenuate the risk ratios for putative risk factors. By contrast both these errors would be more marked for non-hip distal forearm fractures, with comparable influences on both incidence and risk estimates.

In prospective studies, analysis of incidence rates involves calculation of the 'time to fracture' and inaccuracy in the date of reported fracture may therefore lead to bias in estimates of incidence. We found that 91% of subjects recalled their date of fracture to within 1 month of the actual date. There was no evidence of 'telescoping' – a phenomenon in epidemiologic studies whereby individuals report exposures from the past into a more recent time period [13]. Further, there were no differences in this study in accuracy of recall between those who fractured within the previous year and those in preceding years, though the average follow-up period of time was relatively short (median = 2.1 years).

We were unable to verify fracture type in 53 of the 563 self-reports of fracture (9% of cases). In these subjects information concerning site of fracture is limited to the information from the manikin. Using a predefined region we were able to identify both hip and distal forearm fractures with high sensitivity and specificity. This suggests that in the absence of other sources of confirmation, the manikin may be used to identify individuals with hip and distal forearm fractures with a relatively high degree of accuracy.

A postal questionnaire is a relatively simple and accurate method for obtaining information about the occurrence of hip and distal forearm fractures, including their timing. Accuracy of ascertainment of fractures at other sites is less good and where possible self-reported fractures at such sites should be verified from other sources.

Acknowledgements. The study was financially supported by a European Concerted Action Grant under Biomed-1 (BMH1CT920182). The central coordination was also supported by

the UK Arthritis Research Campaign, the Medical Research Council, and the European Foundation for Osteoporosis and Bone Disease. Individual centers acknowledge the receipt of locally acquired support for their data collection. A.A.I. is a Wellcome Trust Clinical Research Fellow.

References

1. Nevitt MC, Cummings SR, Browner WS, et al. The accuracy of self-report of fractures in elderly women: evidence from a prospective study. *Am J Epidemiol* 1992;135:490–9.
2. Akesson K, Gardsell P, Sernbo I, et al. Earlier wrist fracture: a confounding factor in distal forearm bone screening. *Osteoporos Int* 1992;2:201–4.
3. Jonsson B, Gardsell P, Johnell O, Redlund-Johnell I, Sernbo I. Remembering fractures: fracture registration and proband recall in southern Sweden. *J Epidemiol Community Health* 1994;48:489–90.
4. Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986;123:894–900.
5. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health* 1989;79:1554–6.
6. Beard CM, Melton LJ III, Cedel SL, et al. Ascertainment of risk factors for osteoporosis: comparison of interview data with medical record review. *J Bone Miner Res* 1990;5:691–9.
7. Honkanen K, Honkanen R, Heikkinen L, et al. The validity of self-reports of fractures. In: Ring EFJ, Elvins DM, Bhalla AK, editors. *Current research in osteoporosis and bone mineral measurement*, vol. 4. London: British Institute of Radiology, 1997:6–7.
8. Stucki G, Stucki S, Bruhlmann P, et al. Comparison of the validity and reliability of self-reported articular indices. *Br J Rheumatol* 1995;34:760–6.
9. Donald F, Esdaile JM, Kimoff JR, et al. Musculoskeletal complaints and fibromyalgia in patients attending a respiratory sleep disorders clinic. *J Rheumatol* 1996;23:1612–6.
10. O'Neill TW, Felsenberg D, Varlow J, et al. The prevalence of vertebral deformity in European men and women: the European Vertebral Osteoporosis Study. *J Bone Miner Res* 1996;11:1010–7.
11. World Health Organization. *International statistical classification of diseases and related health problems*. 9th ed. Geneva: World Health Organization, 1978.
12. Stata Corporation. *Stata reference manual*, release 3.1. 6th ed. College Station, TX.
13. Armstrong BK, White E, Saracci D, editors. *Principles of exposure measurement in epidemiology*. Oxford: Oxford University Press, 1994.

*Received for publication 22 March 1999
Accepted in revised form 20 August 1999*