THE PROGNOSTIC VALUE OF VERTEBRAL FRACTURES ON CHEST CT

CFM BUCKENS

The prognostic value of vertebral fractures on chest CT

PhD thesis, Utrecht University, The Netherlands
ISBN: 978-94-6299-057-9

Author:	CFM Buckens
Layout:	Tim Luijkx
Cover design and artwork:	Peter Query & Tim Luijkx
Print:	Ridderprint BV, Ridderkerk

The prognostic value of vertebral fractures on chest CT

De prognostische waarde van wervelfracturen op CT-thorax

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 24 april 2015
des middags te 4.15 uur

door

Constantinus Franciscus Maria Buckens

geboren op 8 januari 1985 te Nairobi, Kenia

Promotoren:

Prof. dr. Y. van der Graaf
Prof. dr. W.P.T.M. Mali

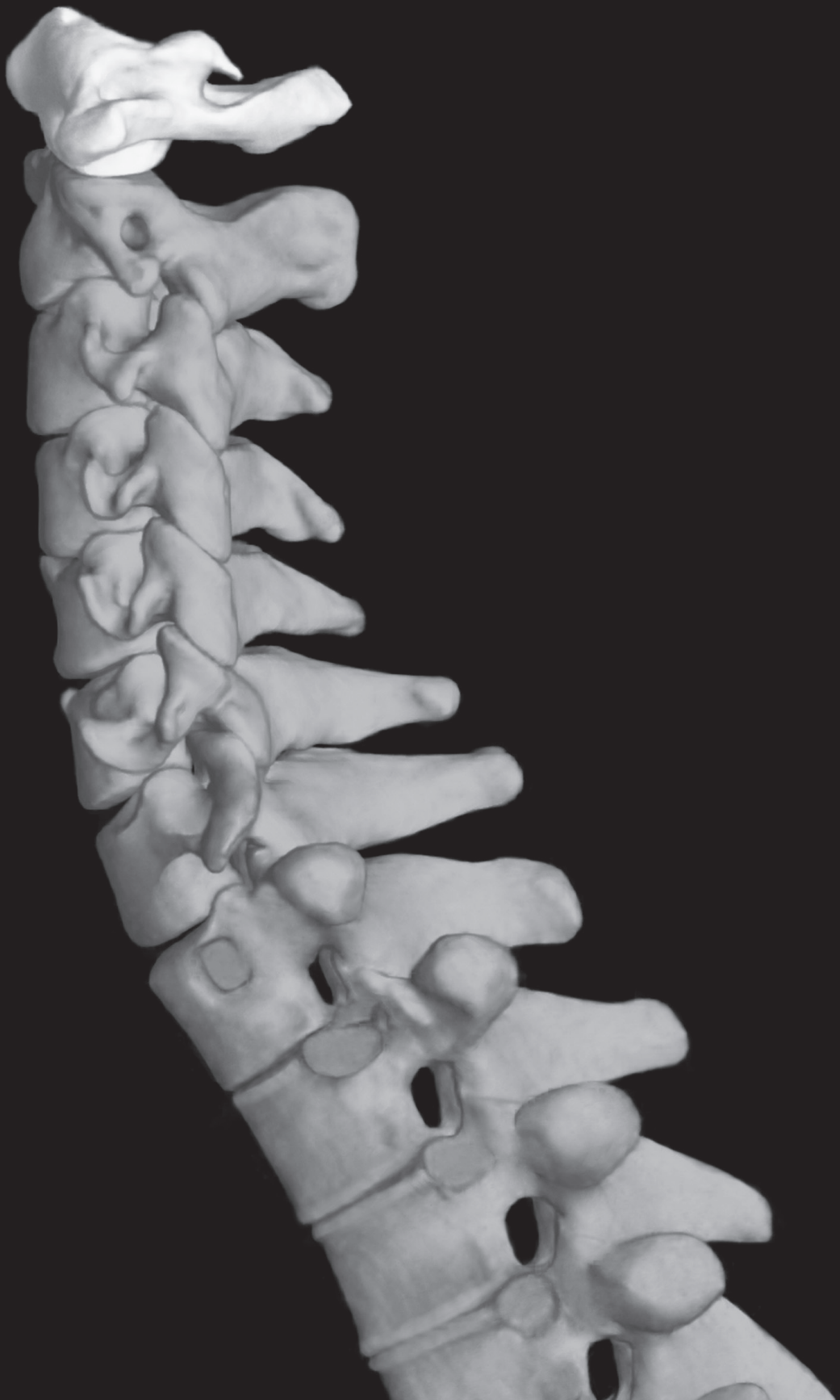
Copromotoren:

Dr. P.A. de Jong
Dr. H.M. Verkooijen

Financial support was generously provided by Röntgen Stichting Utrecht.

Look well to the spine for the cause of disease – Hippocrates

Contents



Chapter 1

Introduction

Introduction

Unrequested findings

Unrequested or incidental findings on diagnostic imaging in routine clinical practice are increasingly common due to improving imaging quality and increased utilization of imaging in routine care.[1,2] These unrequested findings range from the clinically relevant, with definite therapeutic and/or prognostic implications, and those which are (almost) definitely clinically irrelevant. Many unrequested findings cannot be placed unambiguously into one of these categories. A large subset of indeterminate findings consists of early manifestations of severe diseases which may or may not progress, such as small aortic aneurysms or cardiovascular calcifications. The predicament then facing radiologists is whether or not to report such indeterminate unrequested findings, and which recommendations to attach to them, if any. The danger of over-reporting is that it may lead to confusion and unnecessary, costly and invasive follow-up investigations or treatments.[3] Underreporting relevant findings however would cause relevant diagnoses to be missed and opportunities for early intervention to go unused.[2]

Prognostic relevance framework

Assessing unrequested findings of indeterminate clinical relevance begins with a careful definition of the outcome or disease being targeted and a precise delineation of the indeterminate finding and how it is assessed or graded on imaging. The latter must be proven to be reliable and reproducible. Only then can the finding be assessed for potential prognostic value by relating it to the outcome in question. If this is shown to be fruitful, a multivariate model containing all established and available variables alongside the finding in question can be constructed. Before this model can be considered for implementation it needs to undergo external validation in a separate population. If the performance of the model remains acceptable externally then an impact trial assessing whether early treatment of these indeterminate findings indeed leads to improvement in outcomes for the disease of interest. Then we know whether reporting these findings make sense.

Chest CT

Computed Tomography (CT) in particular embodies this conundrum. CT has undergone dramatic technical improvements in recent years. In response the utilization of CT has increased and its use continues to grow unremittingly, with an average annual rise of 9% in the Netherlands since 2001, increasing to 1.23 million investigations, roughly one for every

14 citizens per year.[4] Chest CT is one of the most commonly requested investigations for evaluating suspected thoracic pathology, including pathology of the lungs, heart, thoracic great vessels, mediastinum, thoracic spine and bones. This density of physiologically critical anatomy in the thorax results in a relatively high prevalence of unrequested findings.[5] In addition to the growing use in routine care, the growing momentum behind CT based screening for lung cancer[6] may lead to a further expansion in the number and demographic range of patients receiving chest CT.

PROVIDI study

To guide reporting of indeterminate unrequested findings evidence on the prognostic relevance of the finding in question is necessary. This is the goal of the PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) study.[7] The PROVIDI study is a large multicenter cohort study including all chest Computed Tomography (CT) acquired from patients aged ≤40 years in the course or routine care in eight participating centers in the Netherlands between 2002 and 2005. As such is reflects general radiological practice both in demography of the cohort but also in the heterogeneity of the acquisition protocols and referral indications. For these patients death and hospital admissions were tabulated using data from the Dutch national bureau of statistics, which centrally tracks causes of death and hospital admission and discharge diagnoses. The PROVIDI study employs a uniquely flexible but under-utilized study design, the case-cohort, which facilitates the cost-effective investigation of multiple endpoints.

Vertebral fractures

This manuscript will focus upon vertebral fractures on chest CT. Vertebral fractures or deformities increase with age and in contrast to peripheral fractures they are often clinically silent, without any clear trauma, manifesting only as a slow and gradual loss of height and a stooping of the back. Nonetheless they have major prognostic implications about the quality of a person's bone and their risk of future fracture and disease.

Osteoporosis

First and foremost, vertebral fractures and deformities are markers for osteoporosis. The bone demineralization and deterioration that characterizes osteoporosis manifests at an early stage as height loss of the vertebral bodies. Osteoporosis is an increasingly common disease associated with ageing. The uniquely persistent and high levels of axial loading placed upon the spine in the course of daily life means that any deterioration of the

trabecular bone in the vertebral bodies quickly leads to height loss, quite possibly before the more densely mineralized cortical bone is detectably affected.[8] This is germane, as the current standard for diagnosing osteoporosis, Dual-energy X-ray absorptiometry (DXA),[9] measures the overall (areal) mineral density of bone, which includes both cortices and the trabecular medulla and does not take into account the presence of vertebral height loss and fractures.

Prior research on vertebral fracture outcome

It should then come as no surprise that the presence of vertebral fractures has previously been shown to predict future osteoporotic fractures in specialized study settings,[10–14] that vertebral fractures and osteoporosis have been linked to other disease clusters, including cardiovascular disease[15] and all-cause mortality,[16] or that vertebral fracture status can identify additional patients who are potential candidates for osteoporosis treatment[17]. The more traditional DXA may only become positive for osteoporosis at a later stage of the disease. The fact that the majority of osteoporotic fragility fractures occur in patients with bone mineral densities above the osteoporotic threshold on DXA[18] both bears this out and illustrates the scope for improvement in early detection.

Subtle vertebral fractures

Especially in the early stages of the osteoporotic process the height loss can be subtle, typically <25% height reduction, below the threshold usually applied to define frank fractures. Such vertebral deformities can be conceived of as the first manifestations of deteriorating bone quality. Detecting osteoporosis in an early stage is especially beneficial as all currently available treatment options can only halt the progress but cannot reverse it.

Early detection means early management

These treatment options include fall prevention, lifestyle guidance, calcium/vitamin D supplementation and (most importantly) anti-resorptive medications, such as bisphosphonates, denosumab, selective estrogen receptor modulators, hormonal therapy and other bone anabolic agents such as strontium ranelate and exogenous PTH-analogues.[19] As these treatments are effective at slowing disease progression, early diagnosis and prevention of osteoporosis are currently the best approach to reducing its mounting disease burden.[20,21] However, the absence of trauma and/or obvious signs of fracture mean that diagnosis and treatment are often delayed.[22–24]

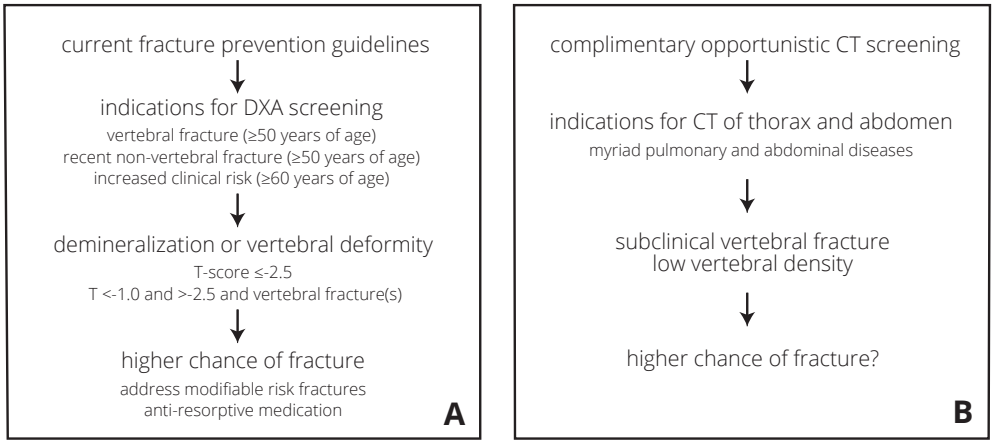
But still vertebral fractures are not routinely reported

Due to the clinically silent nature of vertebral fractures, these patients are often unaware of their vertebral fractures. Partly due to a lack of firm evidence supporting their prognostic value and subsequent useful treatment, clinicians requesting CT do not request vertebral fracture assessment and reading radiologists do not usually actively seek them, something that currently requires additional sagittal reformats to be specially made and read in most centers[13]. As a consequence current evidence shows that fewer than 15% of visible vertebral fractures are reported in routine practice[10–13] and the majority are clinically unknown.[14]

Complementing DXA

Clearly routine CT requested in the course of clinical care for other conditions or for lung cancer screening could serve to identify patients with a vertebral fracture and/or decreased vertebral density, which might benefit from preventative interventions. Both CT of the chest and the abdomen visualize (part of) the spine and can be used for opportunistic vertebral fracture assessment. CT investigations of the thorax and abdomen together account for roughly half of requested CTs and are important drivers for the continued growth of the modality.[4] Those patients receiving a CT of the chest can be expected to include a wide group of patients which only partly overlaps with the demographic typically considered for DXA screening. In contrast to conventional radiographs and DXA, CT can also provide information on the mineral content of vertebral medulla. As the mineral content decreases

Figure 1: current diagnostic flow for identifying patients at an increased risk of insufficiency fracture (A) vs. proposed parallel workflow utilizing opportunistic CT based screening (B)



and trabeculation becomes sparser, the attenuation or density of the trabecular portions of vertebrae decreases. The way in which opportunistic CT-based vertebral fracture assessment and bone attenuation measurement may provide an alternative avenue to DXA is illustrated in Figure 1.

This thesis

Aim

This thesis aims to investigate the reliability of vertebral fracture assessment on CT and their prognostic value for the prediction of osteoporotic fracture, cardiovascular disease and all-cause mortality, using CTs and outcomes data from both the PROVIDI and NELSON studies.

Outline

In chapter 2 we discuss a simulation study based upon a large clinical cohort study comparing different novel and classic approaches to case-cohort analysis. In chapter 3 we examine the literature for the prevalence of clinically relevant extra-cardiac findings on cardiac CT. We begin chapter 4 by establishing the reproducibility of semi-quantitative vertebral fracture assessment on routine clinical chest CT in chapter 4. In chapter 5 we examined whether subclinical vertebral fractures on sagittal reconstruction of routine chest CT are associated with future hip fractures, the most important osteoporotic fracture. In chapter 6 we perform and external validation study of the diagnostic accuracy of vertebral bone density on routine clinical CT for DXA defined osteoporosis, as well as the performance of vertebral fracture status for the same. We look beyond osteoporosis at the association of vertebral fracture status and all-cause mortality in a lung cancer screening setting in chapter 7, and at the complex relation between vertebral fractures (a proxy for osseous demineralization), cardiovascular calcifications and future cardiovascular events in chapter 8. Chapter 9 concludes this thesis with a discussion of the results, the overarching themes and a view to future research.

All work presented in this thesis was supported by a grant from the Netherlands Organization for Scientific Research-Medical Sciences (NWO-MW) grant 40-00812-98-07-005).

References

1. Stone JH. Incidentalomas: clinical correlation and translational science required. *N Engl J Med.* 2006;354(26):2748–2749.
2. Lee CI, Forman, H.P. What we can and cannot see coming. *Radiology*. 2010;257(2):313–314.
3. Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'Byrne MM, Port JD et al. Incidental findings in imaging research: evaluating incidence, benefit, and burden. *Arch Intern Med*. 2010;170(17): 1525–1532.
4. Pruppers M, de Waard-Schalkx I, Bijwaard H. Developments in radiation exposure from medical imaging [Internet]. 2013. Available from: http://www.rivm.nl/dsresource?objectid=rivmp:209044&type=org&disposition=inline&ns_nc=1
5. Jacobs PCA, Mali WPTM, Grobbee DE, van der Graaf Y. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. *J Comput Assist Tomogr*. 2008;32(2):214–221.
6. National Lung Screening Trial Research Team et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
7. Gondrie MJ1, Mali WP, Buckens CF, Jacobs PC, Grobbee DE, van der Graaf Y. The PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) Study: rationale and design. *Eur J Epidemiol*. 2010;25(10):751–758.
8. Riggs BL, Melton LJ, Robb RA, Camp JJ, Atkinson EJ, McDaniel L et al. A population-based assessment of rates of bone loss at multiple skeletal sites: evidence for substantial trabecular bone loss in young adult women and men. *J Bone Miner Res*. 2008;23(2):205–214.
9. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. *World Health Organ Tech Rep Ser*. 1994;843:1-129.
10. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009 Jan;69(1):179-83.
11. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N et al. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009 Mar;69(3):555-9.
12. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008 Feb;63(2):160-4.
13. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008 Aug;18(8):1696-702.
14. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ. Unreported Vertebral Body Compression Fractures at Abdominal Multidetector CT. *Radiology*. 2013 Jul;268(1):120-6.
15. Persy V, D'Haese P. Vascular calcification and bone disease: the calcification paradox. *Trends Mol Med*. 2009 Sep;15(9):405-16

16. Johansson C, Black D, Johnell O, Odén A, Mellström D. Bone mineral density is a predictor of survival. *Calcif Tissue Int*. 1998 Sep;63(3):190-6.

17. Kanis JA, Barton IP, Johnell O. Risedronate decreases fracture risk in patients selected solely on the basis of prior vertebral fracture. *Osteoporos Int*. 2005 May;16(5):475-82.

18. Siris ES, Chen YT, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med*. 2004 May 24;164(10):1108-12.

19. Cranney A, Guyatt G, Griffith L, Wells G, Tugwell P, Rosen C et al. Meta-analyses of therapies for postmenopausal osteoporosis. IX: Summary of meta-analyses of therapies for postmenopausal osteoporosis. *Endocr Rev*. 2002 Aug;23(4):570-8.

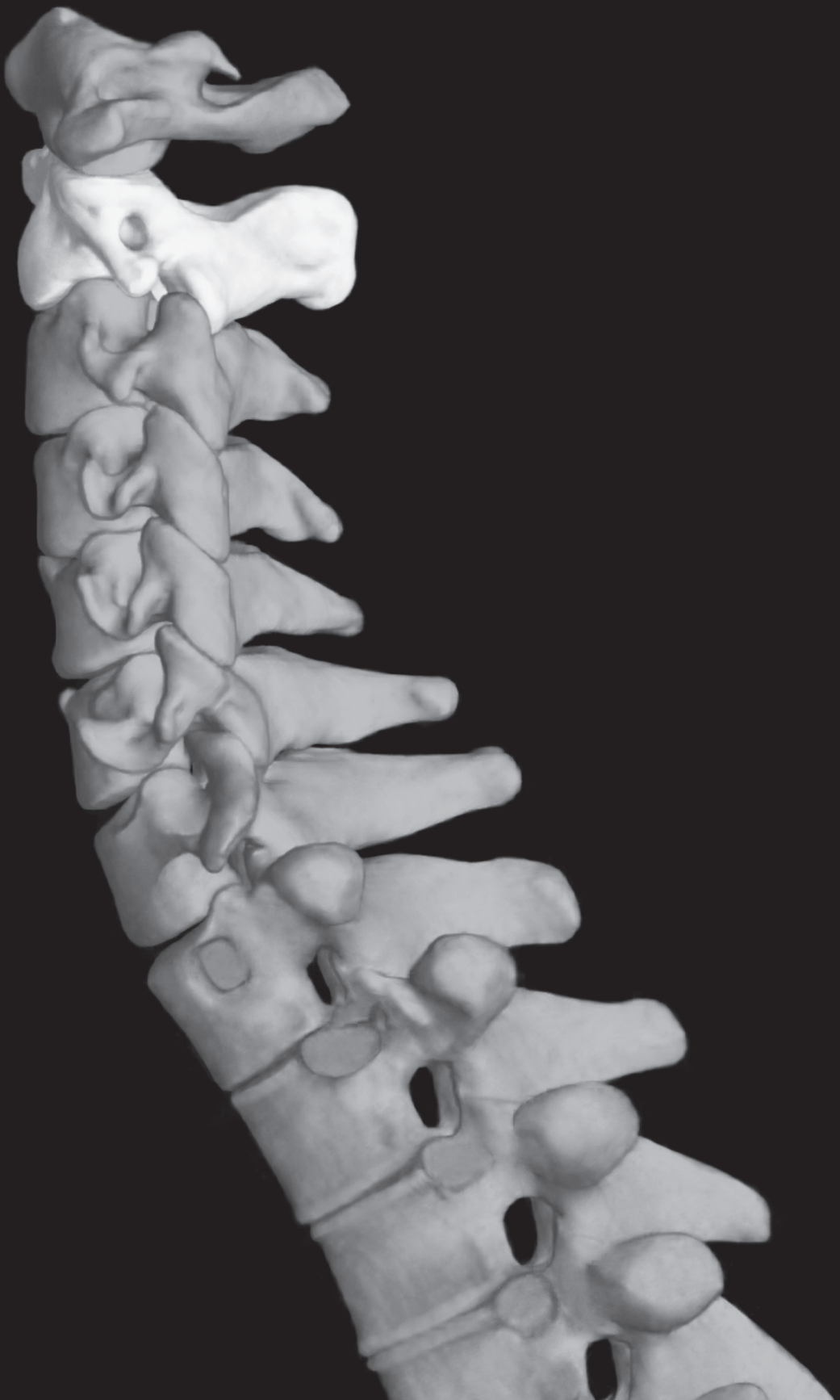
20. Lim LS, Hoeksema LJ, Sherin K; ACPM Prevention Practice Committee. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. *Am J Prev Med*. 2009 Apr;36(4):366-75.

21. Kern LM, Powe NR, Levine MA, Fitzpatrick AL, Harris TB, Robbins J et al.Association between screening for osteoporosis and the incidence of hip fracture. *Ann Intern Med*. 2005 Feb 1;142(3):173-81.

22. Cooper C, Melton LJ. Vertebral fractures. *BMJ*. 1992 Mar 28;304(6830):793-4.

23. Bessette L, Ste-Marie LG, Jean S, Davison KS, Beaulieu M, Baranci M et al. The care gap in diagnosis and treatment of women with a fragility fracture. *Osteoporos Int*. 2008 Jan;19(1):79-86.

24. Metge CJ, Leslie WD, Manness LJ, Yogendran M, Yuen CK, Kvern B et al. Postfracture care for older women: gaps between optimal care and actual care. *Can Fam Physician*. 2008 Sep;54(9):1270-6.



Chapter 2

Prognostic modeling in case-cohort data: comparison of the performance of different approaches

Buckens CF
Debray TP
Peelen LM
Moons KG
van der Graaf Y

Under preparation

Abstract

Prognostic modeling in case-cohort data: comparison of the performance of different approaches

Objective

The case-cohort design greatly improves the cost-effectiveness of observational follow-up studies involving expensive measurements and/or rare diseases, such as those encountered in clinical prognostic research. Because case-cohort studies require specialized approaches to analyze corresponding data, researchers are often daunted to implement the case-cohort design.

Methods

We set out to compare the relative performance of four approaches with varying degree of complexity for developing prediction models using case-cohort data. Hereto, we used a large clinical dataset to predict cardiovascular outcomes, and simulated different scenarios where four different approaches were compared: one completely ignoring the case-cohort status, one using classic weighting schemes, one involving multiple resampling and one employing multiple imputation techniques to restore the original cohort. The model performance of the resulting models from these approaches was compared in terms of model performance (discrimination and overall calibration).

Results

All four approaches, including the approach that ignored the case-cohort status of the data, showed very good and consistent discrimination. There were major deficiencies in the overall calibration when case-cohort status was ignored. Calibration was relatively consistent in the other approaches but somewhat more accurate when using sampling-based approaches (resampling or imputation).

Conclusions

The rare use of the case cohort design is lamentable because it possesses several important advantages, and relatively simple approaches can be applied to overcome its inherent imbalance in the original proportion of cases.

Introduction

The case cohort design

The case-cohort design is an underutilized study design that reduces the often prohibitive cost of classic cohort design. The classic cohort study is a powerful and popular study design for studying prognostic factors with time-to-event data, but can be expensive when large numbers of participants are needed to collect enough cases for studying diseases with a low incidence, or when expensive covariates need to be determined in a large sample.

Since the number of cases (cohort members suffering an event of interest) is almost universally the group limiting the precision of the results, different extensions of the cohort design exist that reduce the need of collecting expensive data in the non-cases. These novel designs generally require an existing cohort study where traditional covariates (such as age, gender, etc.) were measured during inclusion in the study (phase I covariates) and where some "expensive" covariates (phase II covariates) such as blood or imaging test results can retrospectively be assessed at the end of the study, e.g. using stored tissue samples or imaging data acquired at study inclusion.

Upon implementation of the case-cohort design, a completely random sample from the entire cohort at baseline, also termed the subcohort[1] is taken. The subcohort is usually a small fraction of the full cohort, typically below 10%.[2] The case-cohort dataset is then constructed by complementing the subcohort with all cases that appeared in the entire cohort during follow-up but were not selected for inclusion in the subcohort. The case-cohort dataset can be viewed as a sample of the entire cohort where many non-cases are missing completely at random.

The case-cohort design greatly reduces the amount of Phase II covariates that need to be determined. For the standard full cohort design all covariates of interest (both phase I and phase II) would have be determined in all cases and non-cases. In the case-cohort design, phase II covariates are only determined for all cases and a random sample of non-cases.

Preconditions and advantages of the case-cohort design

Implementation of the case-cohort design requires that expensive phase II covariates can be retrospectively determined. Clinical studies involving expensive analyses on retained blood samples or time consuming analyses of stored imaging data, where the outcome status can be determined through simple follow-up, are good examples of clinical research

settings where the case-cohort design can be applied. The prevalence of the outcome of interest also determines the suitability of the case-cohort design, with rare outcomes being particularly attractive to study using the case-cohort.

Although it is not an absolute precondition, rarer outcomes magnify the benefit of the case-cohort design. Where an outcome is rare a larger cohort is necessary to achieve adequate precision; driving up monetary costs and/or follow-up durations required. The resulting cohorts are particularly imbalanced, with few cases and huge numbers of non-cases which individually add little precision but relentlessly drive up costs.

The case-cohort design creates scope for unrivaled versatility in cohort studies, as the same subcohort can be repurposed to study multiple outcomes and there is no risk of introducing bias through imperfect selection of controls, as applies to case-control studies. Once a subcohort has been selected it can be used for addressing multiple research questions. If a new outcome of interest is identified for the entire cohort, only the phase II covariates for the newly identified cases need to be additionally determined to complete the case-cohort dataset.

Adjustment during the analysis of case-cohort datasets is required

As said, the case-cohort dataset is formed by combining the subcohort with additional cases from the entire cohort. Because the subcohort is a random sample from the original cohort, corresponding estimates will not be biased (although their precision will be reduced). The case-cohort dataset improves the precision of estimates from the subcohort, but requires special care during statistical analyses because it no longer comprises a random sample. In particular, there is a disruption of the original proportion of cases and non-cases in the case-cohort dataset (cases are overrepresented).

To ensure valid estimates and unbiased standard errors (SE), adjustment during the analysis is necessary for the three types of subjects included in the case-cohort dataset: cases which happen to fall within the subcohort, non-cases in the subcohort and the bulk of the cases outside the subcohort that are identified during follow-up. The classic approach involves weighting each type of subject differently in the statistical model to correct for the overrepresentation of cases. Since the publication of the original case-cohort design[1] improvements to the weighting scheme of the non-cases in the subcohort and the cases outside the subcohort have been widely adopted[3,4] without radically overhauling the weighting approach.

Prognostic research is uniquely suited the case-cohort design

Multivariate prognostic research presents a special set of challenges and opportunities to the case-cohort design. The case-cohort design is uniquely suited to investigating the rare outcomes and the expensive to determine covariates that characterize clinical prognostic research. Being able to repurpose cohorts for studying multiple outcomes of interest with minimal effort confers additional benefit and scope for cost-saving. Since prognostic research is principally concerned with the performance of a model in its entirety rather than the particular point estimates of its associations, overall model performance is the principle concern. The standard built in ready-to-use validation and performance assessment functions in most commonly used statistical environments are not adapted to the uniquely, intentionally imbalanced case-cohort dataset. Furthermore, statistical packages for dealing with case-cohort data typically focus on producing accurate estimates of relative risk (such as hazard ratios), and do not estimate the baseline survival function of the original cohort (which is needed to obtain absolute outcome probabilities over time). This creates a barrier towards implementing the case-cohort approach, perhaps partly explaining it rarity in the literature, despite its substantial advantages.

Alternatives to classic weighting approaches to case-cohort analysis

A simple and pragmatic way to avoid having to adjust standard ready-to-use prognostic modelling functions would be to restore the original cohort using sampling techniques. A simple approach is to sample the non-cases from the subcohort with replacement to recreate the proportion of cases to non-cases found in the original cohort (up-sampling). This approach is then repeated to generate several plausible full cohort data sets: 'multiple resampling'.

Breslow et al[5] have suggested optimizing the sampling weights using the phase I covariates for the non-cases outside the subcohort, which would have been entirely discarded using classic approaches. Paik and Tsai[6] and Marti and Chavance[7] subsequently proposed extending this utilization of the previously discarded phase I covariates by viewing the entire cohort dataset as a special case of incomplete data. Because phase II covariates are missing at random (MAR) by case status, traditional imputation techniques can be used to generate plausible completed data sets. Corresponding imputation models should include a status indicator to distinguish between cases and non-cases, and may account for additional phase-I and follow-up covariates (e.g. time until censoring or event). A major advantage of this approach is that missing data in phase I covariates can also be accommodated. The growing sophistication, acceptance and availability of advanced imputation tools for missing data[8,9] have helped to drive this avenue of inquiry.

Empirical comparison

Using a large clinical cohort with time-to-event data we set out to give a comparison of the overall model performance of models generated using the classic weighting approach, an approach using multiple resampling of non-cases to recreate the full cohort and a multiple imputation approach that treats the partially missing phase II variables in the entire cohort as a special case of missing data. We additionally included an approach that ignored the case-cohort status as an illustration of the scope and nature of the model performance problems that ensue.

Methods

SMART Cohort

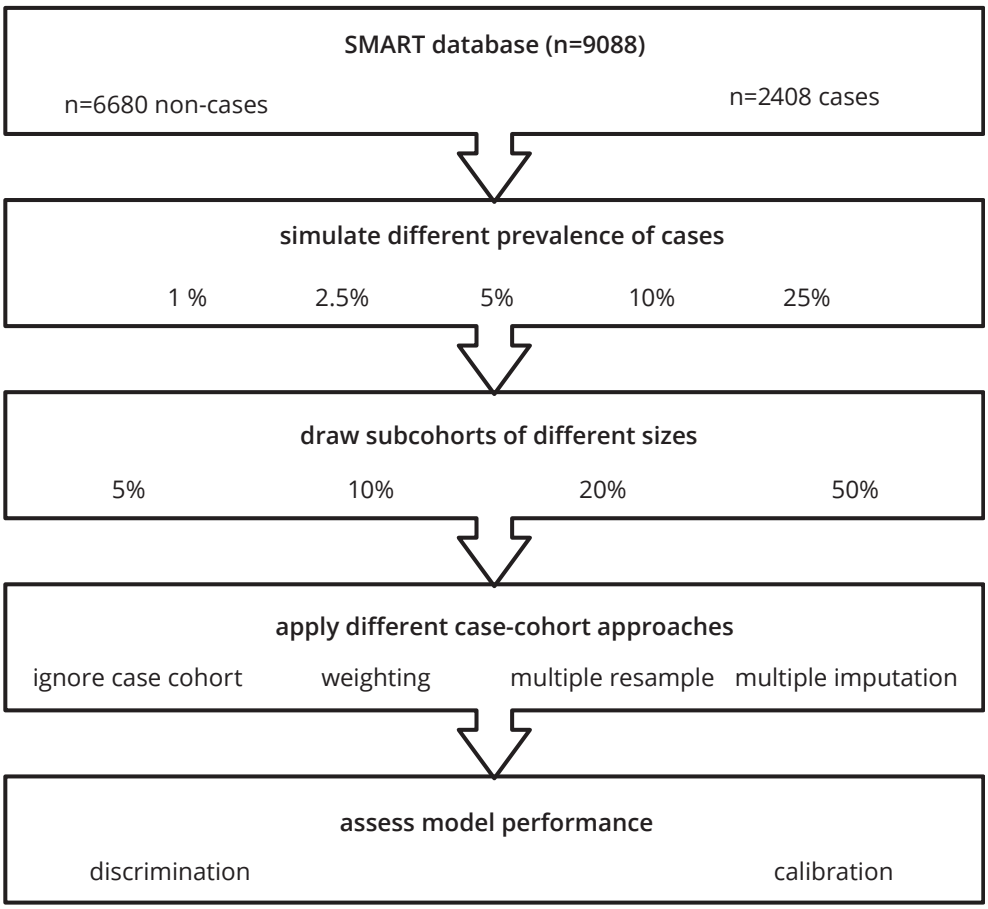
Patients originated from the SMART (Secondary Manifestations of ARTerial disease) study, an ongoing prospective cohort study at the University Medical Centre Utrecht, The Netherlands. A detailed description of the study has been published previously.[10]

Briefly, patients completed a vascular screening questionnaire at inclusion covering medical history, smoking status and medication use. Office blood pressure and fasting venous blood and urine samples were taken. To assess outcome status, patients completed a biannually distributed questionnaire on hospitalizations and outpatient clinic visits. The outcome of interest for this study was any major cardiovascular event; including cardiovascular death, ischemic or hemorrhagic stroke, or myocardial infarction. From the SMART dataset we selected those covariates that had previously been identified as being prognostic for a recurrent vascular event[11]: age, gender, diabetes status, smoking status, systolic blood pressure, total cholesterol, High-density lipoprotein (HDL) cholesterol, C-reactive protein (CRP) Glomerular Filtration Rate (GFR; ml/min/1.73m^2) and a having a history of either cerebrovascular disease, coronary artery disease and/or abdominal aortic aneurysm. Although data were available for all covariates in the original SMART cohort, measurements on cholesterol level are rather expensive and a case-cohort design may thus be undertaken to improve efficiency. Using these definitions, from a cohort of n=9088, a total of 2408 events were observed. The median follow-up per subject was 1598 days.

Simulation scenarios

Working with these covariates from the complete SMART cohort we simulated new cohorts with varying prevalence of events using stratified resampling without replacement (stratified upon case status). In order to assess the effect of disease prevalence on model

Figure 1. Flowchart



performance in the different approaches we simulated full cohorts with 1%, 2.5%, 5%, 10% and 25% prevalence of the outcome. From these simulated full cohorts we then randomly drew subcohorts of different sizes (5%, 10%, 20% and 50% of the full cohort size). We subsequently complemented the subcohorts with the remaining cases from the simulated full cohort to form the case-cohort dataset (Figure 1).

Analysis approaches

We employed four approaches to fit a Cox-proportional hazards model containing the same covariates in each case-cohort analysis; firstly we completely ignored the case-cohort status of the simulated case-cohort data and analyzed the simulated case-cohort sample as a complete cohort, without any correction or weighting. This extreme approach serves to

highlight the problems that occur with prognostic research in case-cohort, and represents a worst case scenario against which to compare the other methods. For each approach, we extracted the estimated regression coefficients (log hazard ratios) and 10-years baseline survival.

Secondly, we applied classic weighting as described by Prentice et al1. This relatively simple weighting scheme has been shown to be especially reliable and robust[12]. In the Prentice method the subcohort members and the cases beyond the subcohort at failure contribute to the baseline hazard function. This approach was included to represent the weighting approaches.

Thirdly we implemented a sampling approach whereby we copied the original cases and randomly sampled non-cases from the subcohort with replacement to reconstruct the original cohort. This approach is then repeated several times to balance out random errors and preserve an adequate degree of uncertainty (multiple resampling). The multiple resampling approach is attractive in that it allows all the analysis techniques and software tools developed for full cohort analysis to be directly applied to case-cohort setting without modification and allows less experienced researchers apply existing methods directly.

In the final approach we treated the partially missing phase II covariates for the non-cases outside the subcohort as a special case of missing data and used existing multiple imputation techniques in the full cohort to impute these data7. In particular, we used outcome status, phase I covariates (age, gender, smoking status, diabetes and history of cardiovascular disease) and follow-up time as predictors for imputing the missing phase II covariates (cholesterol level, blood pressure, kidney function (MDRD) and CRP level). In this manner, information on the phase I covariates is preserved for non-cases outside the subcohort, and serves to inform imputation of corresponding phase II covariates.

Aside from these four case-cohort approaches we also fit the model in each simulation scenario to the full cohort from which the case-cohort data was drawn, in order to compare the case-cohort performance measures with the idealized full cohort situation where all data is available. The simulation was repeated 100 times for each scenario. The analyses were performed using the R statistical software[13] (version 3.02, The R foundation for Statistical Computing), with the mice[14] (version 2.18, Multivariate Imputation by Chained Equations) and survival[15] (version 2.37-4) packages.

Performance measures

We evaluated the performance of the described approaches by applying each resulting model from the case-cohort to the original full cohort for that simulation scenario. We

quantified model discrimination (the consistency with which the model can correctly rank eventual cases higher than non-cases) using the concordance statistic.[10] We also quantified model calibration for 10-years survival by calculating the discrepancy between the number of predicted cases (on the basis of the derived models from each approach) and the number of observed cases in that simulation scenario. We used Kaplan-Meier curves for calculating the number of observed cases to account for censoring of events.

Other measures of calibration were not deemed relevant here because the case-cohort design only affects the outcome occurrence, and the random selection procedure avoids introducing bias in predictor-outcome associations. For all performance statistics, the variability was assessed by taking the 2.5% and 97.5% quantiles of the results obtained during the simulation.

Results

Discrimination

Table 1 shows the performance statistics for the model fitted using the four case-cohort approaches, as well as the performance of the model fitted on the full cohort. In all scenarios we found that all the implemented approaches, including the approach that ignored the case-cohort nature of the simulated data, had discrimination in a similar range to that observed in the full cohort. The largest reduction of the mean c-indices was observed in the approach that ignored the case-cohort status (table 1). Smaller sampling fractions from the subcohort and larger prevalences of cases slightly increased this deviance. The discrepancy was largest in the situation with a 25% prevalence of cases and a 5% sub cohort sampling fraction. In that scenario a mean 4.5% decrease over all simulation runs was observed (mean c-statistic of 0.6659 (95% quantiles 0.6589-0.6741) vs. 0.6973 for the full cohort). By comparison, the classic weighting approach showed a mean c-statistic that was at most 1.2% lower (0.7177 (0.7086-0.7266) vs.0.7265) in the scenario with a prevalence of 10% and low sampling fraction of 5% (Table 1). It was 1.0% lower using the multiple re-sample approach (Table 1a) and 0.6% with the multiple imputation approach (Table 1a). In all other scenarios the reduction in c-statistic was smaller.

Calibration

As expected, there were major deficiencies in the overall calibration using the approach that ignored the case-cohort status, with the mean proportion of predicted vs. observed event rates showing large discrepancies, especially in simulation situations where the subcohort fractioning size was smaller (Table 1). Of course, ideally the observed number

Table 1. Mean concordance statistic. 95% (0.025 and 0.975 quantile). Predicted vs.observed (quantiles) for full cohort and alternative approaches

prevalance of disease	subcohort fraction	c-statistic	pred vs. observ	method used							
				ignore		weighted		multiple resample		multiple imputation	
				c-statistic	pred vs. observ	c-statistic	pred vs. observ	c-statistic	pred vs. observ	c-statistic	pred. vs. observ.
1%	5%	0.7753	0.9991	0.7703 (0.7638-0.7744)	0.7196 (0.6953-0.7375)	0.7673 (0.7563-0.7740)	0.9877 (0.9780-0.9935)	0.7673 (0.7576-0.7739)	1.002 (0.9987-1.0052)	0.7704 (0.7589-0.7749)	1.0030 (1.0014-1.0048)
	10%			0.7726 (0.7678-0.7754)	0.8365 (0.8218-0.8477)	0.7710 (0.7658-0.7746)	0.9898 (0.9863-0.9930)	0.7710 (0.7659-0.7750)	1.0030 (1.0007-1.0052)	0.7727 (0.7689-0.7752)	1.0030 (1.0021-1.0043)
	20%			0.7743 (0.7717-0.7756)	0.9175 (0.9108-0.9234)	0.7736 (0.7706-0.7754)	0.9909 (0.9878-0.9929)	0.7736 (0.7707-0.7755)	1.0040 (1.0026-1.0051)	0.7744 (0.7725-0.7758)	1.0040 (1.0028-1.0043)
	50%			0.7750 (0.7739-0.7760)	0.9765 (0.9744-0.9783)	0.7749 (0.7738-0.7759)	0.9913 (0.9902-0.9927)	0.7749 (0.7734-0.7759)	1.0040 (1.0034-1.0048)	0.7752 (0.7744-0.7758)	1.0040 (1.0034-1.0042)
2.5%	5%	0.7501	0.9963	0.7422 (0.7336-0.7486)	0.5321 (0.5042-0.5585)	0.7427 (0.7316-0.7509)	0.9795 (0.9671-0.9883)	0.7434 (0.7326-0.7515)	1.003 (0.9947-1.010)	0.7479 (0.7411-0.7525)	1.0050 (1.0027-1.0075)
	10%			0.7471 (0.7417-0.7523)	0.6898 (0.6751-0.7061)	0.7475 (0.7408-0.7525)	0.9826 (0.9749-0.9890)	0.7479 (0.7411-0.7528)	1.005 (0.9999-1.0102)	0.7498 (0.7452-0.7530)	1.0060 (1.0033-1.0076)
	20%			0.7493 (0.7460-0.7525)	0.8236 (0.8148-0.8347)	0.7493 (0.7455-0.7523)	0.9841 (0.9791-0.9875)	0.7495 (0.7458-0.7524)	1.0060 (1.0026-1.0086)	0.7509 (0.7468-0.7535)	1.0060 (1.0045-1.0072)
	50%			0.7502 (0.7481-0.7529)	0.9437 (0.9397-0.9471)	0.7502 (0.7481-0.7530)	0.9848 (0.9817-0.9865)	0.7504 (0.7482-0.7528)	1.0060 (1.0047-1.0078)	0.7513 (0.7495-0.7535)	1.0060 (1.0056-1.0069)
5%	5%	0.7255	0.9952	0.7130 (0.7036-0.7198)	0.3515 (0.3267-0.3720)	0.7190 (0.7081-0.7265)	0.9765 (0.9607-0.9871)	0.7195 (0.7087-0.7267)	1.007 (0.9936-1.0174)	0.7220 (0.7128-0.7277)	1.0100 (1.0062-1.0146)
	10%			0.7127 (0.7138-0.7250)	0.5274 (0.5084-0.5443)	0.7233 (0.7177-0.7280)	0.9788 (0.9719-0.9844)	0.7236 (0.7176-0.7281)	1.010 (1.0021-1.0171)	0.7244 (0.7161-0.7282)	1.0110 (1.0076-1.0132)
	20%			0.7227 (0.7196-0.7263)	0.706 (0.6918-0.7175)	0.7242 (0.7207-0.7275)	0.9811 (0.9771-0.9847)	0.7243 (0.7210-0.7274)	1.011 (1.0049-1.0153)	0.7252 (0.7217-0.7280)	1.0110 (1.0091-1.0126)
	50%			0.7250 (0.7231-0.7275)	0.8995 (0.8940-0.9037)	0.7253 (0.7235-0.7277)	0.9817 (0.9795-0.9837)	0.7255 (0.7236-0.7277)	1.0110 (1.0084-1.0136)	0.7260 (0.7242-0.7278)	1.0110 (1.0102-1.0122)

pred vs. observ= predicted versus observed

Table 1. Continued

prevalance of disease	subcohort fraction	c-statistic	pred vs. observ	method used							
				ignore		weighted		multiple resample		multiple imputation	
				c-statistic	pred vs. observ	c-statistic	pred vs. observ	c-statistic	pred vs. observ	c-statistic	pred. vs. observ.
10%	5%	0,7265	0.9965	0.709 (0.7012-0.7147)	0.2685 (0.2453-0.2915)	0.7177 (0.7086-0.7266)	0.9586 (0.9360-0.9757)	0.7198 (0.7123-0.7267)	1.012 (0.9874-1.031)	0.7233 (0.7189-0.7265)	1.0140 (1.0070-1.0195)
	10%			0.7132 (0.7144-0.7229)	0.4243 (0.4047-0.4410)	0.7221 (0.7148-0.7266)	0.9639 (0.9503-0.9746)	0.7230 (0.7162-0.7272)	1.0140 (0.9969-1.0259)	0.7248 (0.7215-0.7269)	1.0150 (1.0088-1.0196)
	20%			0.7237 (0.7210-0.7260)	0.6122 (0.6001-0.6230)	0.7244 (0.7211-0.7277)	0.9666 (0.9599-0.9727)	0.7247 (0.7211-0.7276)	1.0160 (1.0058-1.0246)	0.7263 (0.7246-0.7275)	1.017 (1.0140-1.0202)
	50%			0.7261 (0.7249-0.7270)	0.8536 (0.8478-0.8592)	0.7260 (0.7244-0.7275)	0.9678 (0.9638-0.9708)	0.7261 (0.7247-0.7271)	1.0170 (1.0122-1.0216)	0.7267 (0.7261-0.7273)	1.0170 (1.0159-1.0189)
25%	5%	0,6973	1005	0.6659 (0.6589- 0.6741)	0.1676 (0.1492-0.1839)	0.6899 (0.6781-0.6966)	0.9469 (0.9273-0.9649)	0.6924 (0.6842-0.6978)	1.004 (0.9942-1.0480)	0.6947 (0.6910-0.6975)	1.0040 (0.9943-1.0145)
	10%			0.6777 (0.6779-0.6861)	0.2747 (0.2614-0.2870)	0.6941 (0.6890-0.6979)	0.9493 (0.9362-0.9592)	0.6953 (0.6906-0.6984)	1.0090 (0.9856-1.0283)	0.6957 (0.6927-0.6975)	1.0060 (0.9983-1.0125)
	20%			0.6914 (0.6886-0.6937)	0.4469 (0.4296-0.4622)	0.6958 (0.6919-0.6986)	0.9514 (0.9431-0.9580)	0.6964 (0.6941-0.6984)	1.0100 (0.9912-1.0261)	0.6966 (0.6953-0.6976)	1.0090 (1.0043-1.0136)
	50%			0.6963 (0.6951-0.6975)	0.7575 (0.7474-0.7673)	0.6956 (0.6956-0.6983)	0.9512 (0.9473-0.9545)	0.6970 (0.6957-0.6982)	1.0110 (1.0013-1.0194)	0.6973 (0.6964-0.6980)	1.0120 (1.0088-1.0158)

pred vs. observ= predicted versus observed

of cases is exactly equal to the number of expected cases and there is no discrepancy. The estimated event rates were slightly more congruous with the observed rates in the scenarios where the subcohort sampling fraction was largest; those scenarios where the traditional full cohort situation was most closely approximated.

The weighting approach showed a moderate discrepancy between predicted and observed events, again proportional to the event prevalence in the simulation scenario and inversely proportional to the subcohort sampling fraction (Table 1). The largest mean discrepancy in the weighting approach was observed in the situation with the smallest subcohort size (5% of the whole cohort) and the largest prevalence of the outcome (25% of cohort). In this scenario the weighting approach showed a mean predicted vs. observed of 95% (93-96), an underestimation of 5%.

The multiple resampling approach showed very good and consistent overall calibration, remaining much closer to the ideal of 100%. There was a slight tendency towards overestimation of the number of events when the subcohort sampling fraction became larger (Table 1). The largest mean discrepancy was observed in the scenario of the largest 50% subcohort size with a 10% prevalence, but was limited to a 1.7% overestimation of event rates (predicted vs. observed 1.0170 (1.0122-1.0216)).

The multiple imputation approach had the best overall calibration across all the scenarios. It also showed a slight tendency towards overestimation of event rates in scenarios where the subcohort was larger. The largest mean discrepancy over all the runs were observed in the scenarios where the subcohort size was set at 20% and 50% and the prevalence at 10%. These were 1.017 (1.0140-1.0202) and 1.0170 (1.0159-1.0189) respectively, both overestimating event rates by 1.7% (Table 1).

Discussion

Clinical research is often expensive, and besides unbiasedness cost-effectiveness is a driving factor behind many study design choices. The case-cohort design fulfills both requirements by beginning with a full cohort but retrospectively measuring expensive information only in a random subset of the non-cases. The case-cohort design can be used to fit models that can estimate absolute risks (important in prediction modeling research) as well as relative risks (important in prognostic factor research). Since it is independent of outcome status, the subcohort can be re-purposed to study multiple endpoints and its selection is free from risk of selection bias. The collection of phase II covariates can commence before the end of follow-up period (since subcohort membership is not contingent on case status).

Unfortunately, the case-cohort design is currently not much used,[2] perhaps in part due to the daunting complexity of having to adapt standard statistical tools to the case-cohort design.

In this study, we demonstrated that different adjustment approaches can be used when developing prediction models using the case-cohort design. The choice of adjustment approach only marginally affects model performance, and relatively simple approaches can be used to develop prediction models that yield reliable absolute risks.

In particular, we showed that the loss of original outcome occurrence inherent to the case-cohort design can be effectively and reliably restored by employing classic weighting, multiple resampling of non-cases (to recreate the original cohort) or using multiple imputation techniques to impute the unmeasured, expensive phase II covariates.

In the simulation we conducted to compare different approaches towards case-cohort analysis using clinical data from the SMART cohort we found little difference in the performance of the resulting models in terms discrimination. In terms of calibration we found large discrepancies when the case-cohort nature of the data was ignored, suggesting that this is the performance domain where different approaches may show the largest differences when using the case-cohort design. We found that all three of the adopted approaches showed good calibration compared to a full cohort approach. The multiple re-sample and the multiple imputation approach showed marginally better mean overall calibration than the classic weighting approach. These differences were the most pronounced in the simulation scenarios when the subcohort sampling fraction was small.

Whilst all three approaches yield valid and consistent discrimination and calibration, the sampling-based approaches (multiple resampling and multiple imputation) allow standard prognostic modelling functions to be applied to the case-cohort dataset directly without modification. For many researchers modifying these complex functions is difficult and constitutes a barrier to implementing the case-cohort design.[2] This barrier further inflates when multiple datasets from different studies with varying designs are available for model development and/or validation, i.e. when performing an Individual Participant Data (IPD) meta-analysis.[16] The weighting approach then needs to be applied to a selective set of (case-cohort) studies, whilst allowing the estimation of baseline risk and/or predictor effects over all included studies and properly accounting for the presence of between-study heterogeneity. Furthermore, it is possible that phase-II covariates available from one or more case-cohort studies may not have been measured in other studies (e.g. due to cost constraints). Imputation methods can be applied not only to restore the included case-

Table 2. advantages and disadvantages of the compared approaches

approach	advantages	disadvantages
ignoring case-cohort status	· least complicated approach	· introduces bias in absolute estimates of risk, thereby substantially degrading the model's calibration
weighting schemes	· very efficient · resolves issues regarding model performance	· very complicated approach · mixes up models for data 'restoration' and estimation of predictor effects, making it difficult to perform non-standard analyses · cannot directly be combined with existing packages for external validation; and therefore requires advanced statistical expertise · not implemented for all analysis models, mostly used for Cox regression · ignores information on phase-I covariates from non-cases outside subcohort
multiple resampling	· simple to perform · separates data restoration from data analysis; thereby facilitating implementation of alternative statistical models · resolves issues regarding model performance	· can be inefficient in larger datasets · ignores information on phase-I covariates from non-cases outside subcohort
multiple imputation	· relatively simple to perform · makes optimal use of all available evidence · resolves issues regarding model performance	· can be inefficient in larger datasets · computationally more exhaustive

cohort dataset(s), but also to complete those datasets with missing phase-II covariates. Finally, sampling-based approaches allow a departure from adopting Cox proportional hazard models when developing a novel prediction model. In particular, they facilitate the implementation of parametric survival models, which have been recommended in prediction modeling research to allow the estimation of absolute risk over time.

Although not demonstrated here, sampling-based approaches also have advantages when externally validating an existing prediction model using case-cohort data. In particular, sampling-based approaches facilitate the calculation of appropriate model calibration statistics without having to adjust corresponding model coefficients using complex weighting schemes. For this reason, we recommend sampling-based approaches when

developing or validating prediction models using case-cohort data. In table 2 we outline the main advantages and disadvantages of each approach.

This article has several important limitations. We have limited our analysis to one clinical cohort, the SMART cohort. Although this large cohort is in many ways typical of clinical research cohorts, it may not be generalizable to other clinical research where cohorts are radically differently composed. We have no reason to believe that our findings will not hold true in similar research settings but have no way to be sure. We have limited our investigation to the overall model performance, measured pragmatically using the concordance statistic

In conclusion, the case cohort design possesses several important advantages, and several relatively simple approaches shown here can be applied to overcome its inherent imbalance in the original proportion of cases.

Acknowledgments

We gratefully acknowledge the contribution of the SMART research nurses, Rutger van Petersen (data manager) and Harry Pijl (vascular manager).

Collaborators

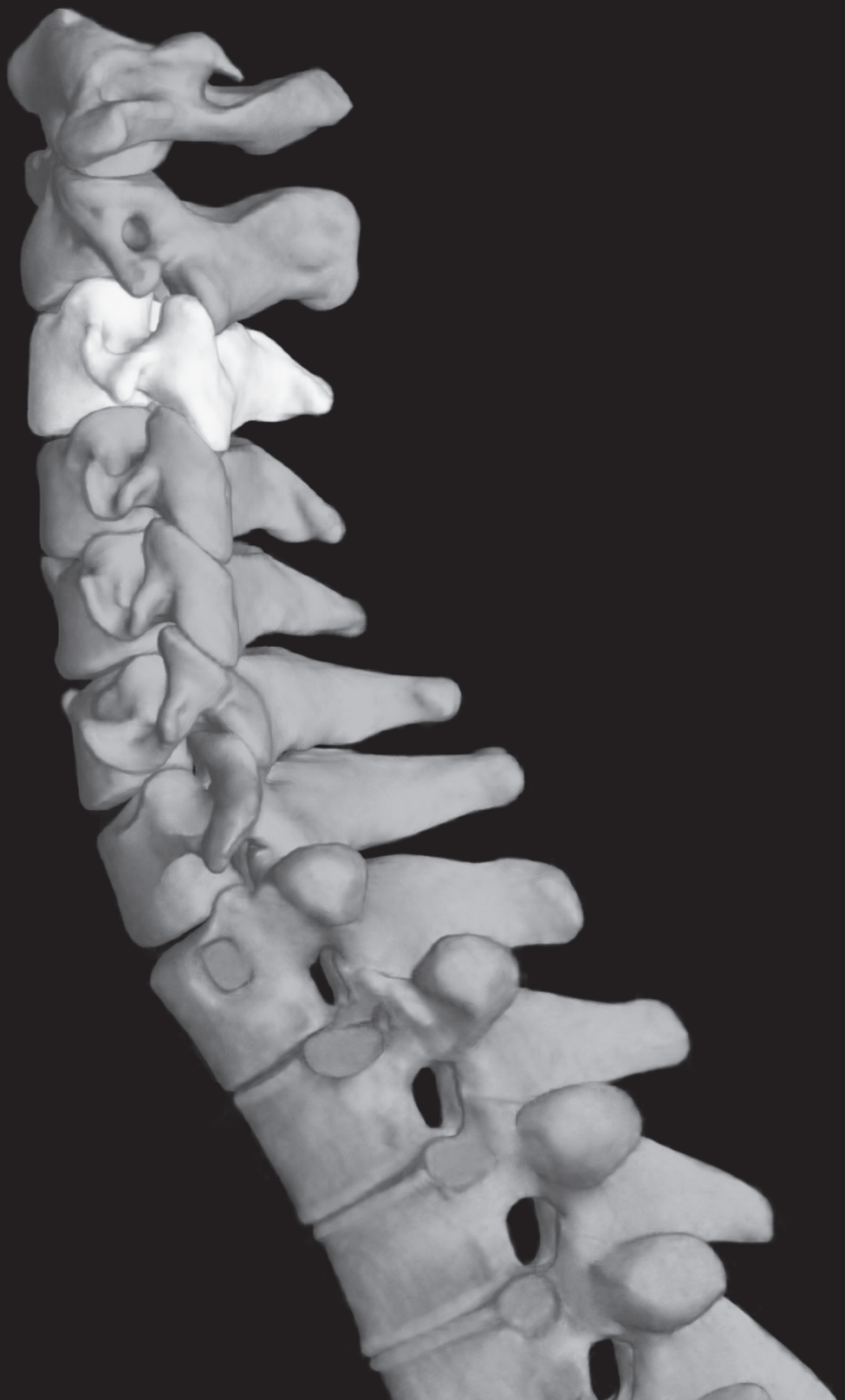
Members of the Secondary Manifestations of Arterial Disease Study Group are as follows: A Algra, MD, YvdG, MD, DE Grobbee, MD and GEHM Rutten, MD, Julius Centre for Health Sciences and Primary Care; FLJ Visseren, MD, Department of Vascular Medicine; FL Moll, MD, Department of Vascular Surgery; LJ Kappelle, MD, Department of Neurology; WPTM Mali, MD, Department of Radiology; PA Doevendans, MD, Department of Cardiology; University Medical Centre, Utrecht, The Netherlands.

Funding

The SMART study was supported by a grant of the Board of Directors of the University Medical Centre Utrecht, who had no role in the conduct of the analyses or drafting of the report. All statistical analyses were done by the investigators. CFB was further funded by an unrestricted program grant from the Netherlands Organization for Scientific Research-Medical Sciences (NWO-MW); grant 40-00812-98-07-005. This funder also had no role in the conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

References

1. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika.* 1986; 73: 1–11.
2. Sharp SJ, Poulaliou M, Thompson SG, White IR, Wood AM. A review of published analyses of case-cohort studies and recommendations for future reporting. PloS One 2014; 9: e101176.
3. Self S, Prentice R. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann Stat.*1988; 16: 64–81.
4. Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics.* 1994; 50: 1064–1072.
5. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the whole cohort in the analysis of case-cohort data. *Am J Epidemiol.* 2009; 169: 1398–1405.
6. Paik M. On using the Cox proportional hazards model with missing covariates. *Biometrika.* 1997; 84: 579–593.
7. Marti H, Chavance M. Multiple imputation analysis of case-cohort studies. *Stat Med.* 2011; 30: 1595–1607.
8. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res .*2007; 16: 219–242.
9. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* 2006; 59: 1087–1091.
10. Simons PC, Algra A, van de Laak MF, Grobbee DE, van der Graaf Y. Second manifestations of ARTerial disease (SMART) study: rationale and design. *Eur J Epidemiol.* 1999; 15: 773–781.
11. Dorresteijn JAN, Visseren FLJ, Wassink AMJ, Gondrie MJA, Steyerberg EW, Ridker PM et al. Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score. *Heart Br Card Soc.* 2013; 99: 866–872.
12. Onland-Moret NC, van der A DL, van der Schouw YT, Buschers W, Elias SG, van Gils CH et al. Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol.* 2007; 60: 350–355.
13. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2013http://www.R-project.org/.
14. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011; 45: 1–67.
15. Therneau T. A Package for Survival Analysis in S. 2014http://CRAN.R-project.org/package=survival.
16. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med.* 2013; 32: 3158–3180.W



Chapter 3

Unrequested findings on cardiac computed tomography: looking beyond the heart

Buckens CF
Verkooijen HM
Gondrie MJ
Jairam P
Mali WP
van der Graaf Y

PLoS One. 2012;7(4):e32184

Abstract

Unrequested findings on cardiac computed tomography: looking beyond the heart

Objectives

To determine the prevalence of clinically relevant unrequested extra-cardiac imaging findings on cardiac Computed Tomography (CT) and explanatory factors thereof.

Methods

A systematic review of studies drawn from online electronic databases followed by meta-analysis with meta-regression was performed. The prevalence of clinically relevant unrequested findings and potentially explanatory variables were extracted (proportion of smokers, mean age of patients, use of full FOV, proportion of men, years since publication).

Results

Nineteen radiological studies comprising 12922 patients met the inclusion criteria. The pooled prevalence of clinically relevant unrequested findings was 13% (95% confidence interval 9–18, range: 3–39%). The large differences in prevalence observed were not explained by the predefined (potentially explanatory) variables.

Conclusions

Clinically relevant extra-cardiac findings are common in patients undergoing routine cardiac CT, and their prevalence differs substantially between studies. These differences may be due to unreported factors such as different definitions of clinical relevance and differences between populations. We present suggestions for basic reporting which may improve the interpretability and comparability of future research.

Introduction

Improvements in the quality of cardiac Computed Tomography (CT) are driving its increasingly widespread use in an expanding patient-group.[1] These same improvements and the increased number of cardiac CT scans are also resulting in the increasing detection of unrequested findings. These unrequested ('ancillary' or 'incidental') findings are more frequently visible on advanced high-resolution scans but fall beyond the reasonable remit of the initial indication for imaging and thus beyond what has been explicitly requested by referring clinicians.

Whilst they apply to all diagnostic imaging modalities, unrequested findings are particularly germane to cardiac CT due to the density of organ systems in the chest and the practice of exclusively evaluating the cardiac/coronary structures. Furthermore, typical patients referred for cardiac CT may also be relatively prone to co-morbidities, due to the confluence of wide-ranging (cardiovascular) risk factors, such a smoking, hypertension, diabetes and obstructive pulmonary disease.[2]

Concerns over the growth of healthcare consumption and radiation exposure are driving calls for the efficient use CT.[3-6] Preventing unnecessary follow-up stemming from irrelevant un-requested findings and systematically reporting on prognostically relevant imaging information could contribute to this. Unfortunately, there is little clarity about which (classes of) findings hold relevance and which do not, although this is beginning to be addressed.[7]

This uncertainty poses a challenge to radiologists and referring physicians alike, with responses ranging from calculated disregard to evaluation of all imaging data available and aggressive follow-up of unrequested findings.[8,9] Often the only rationale provided is expert opinion or prevailing tradition. The fact that these unrequested findings can be detected without additional radiation exposure is pitted against the indeterminate significance of many unrequested findings and the risk and cost of provoking unnecessary follow-up.

Here we review those publications examining the prevalence of incidental findings amongst patients referred for routine cardiac CT scans and assess the effect of candidate explanatory factors abstracted from these articles through a systematic search, review and meta-analysis with meta-regression.

Figure 1. Flowchart illustrating literature search and selection procedure

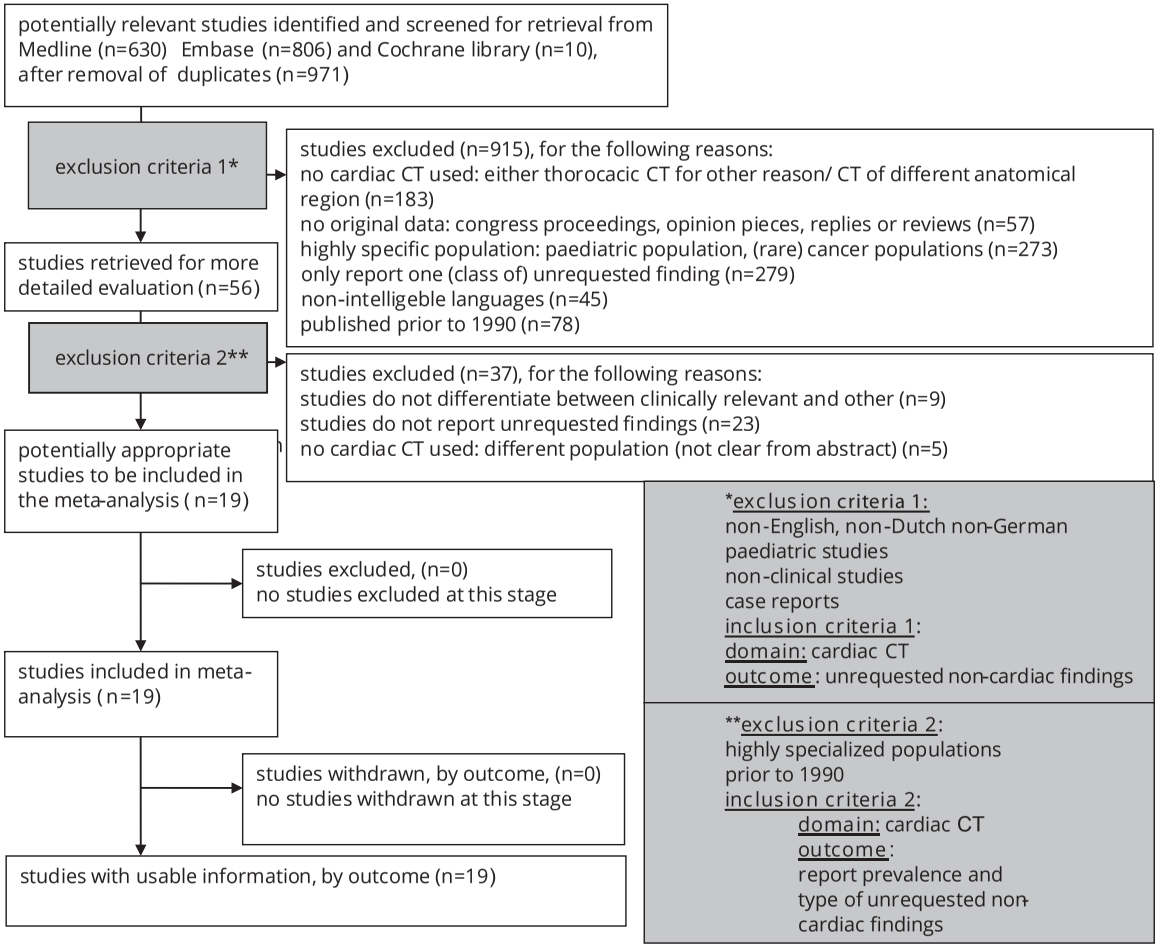


Table 1. Query syntax for MEDline, EMBASE, and the Cochrane Library

database	search strategy
MEDline	(("computed tomography"[tiab] OR CT[tiab]) AND (thora*[tiab] OR chest[tiab] OR cardiac[tiab])) AND (incidental[tiab] OR accidental[tiab] OR ancillary[tiab] OR extra-coronary[tiab] OR non-coronary[tiab] OR extracardiac[tiab] OR extra-cardiac[tiab] OR non-cardiac[tiab])
EMBASE	'computed tomography':ab,ti OR ct:ab,ti AND (thora*:ab,ti OR chest:ab,ti OR cardiac:ab,ti) AND (incidental:ab,ti OR accidental:ab,ti OR ancillary:ab,ti OR 'extra coronary':ab,ti OR 'non coronary':ab,ti OR 'extracardiac':ab,ti OR 'extra cardiac':ab,ti OR 'non-cardiac':ab,ti) AND [embase]/lim
The Cochrane Library	(((computed tomography):ti,ab,kw or (CT):ti,ab,kw) AND ((thora*):ti,ab,kw or (chest):ti,ab,kw or (cardiac):ti,ab,kw)) AND ((incidental):ti,ab,kw or (accidental):ti,ab,kw or (ancillary):ti,ab,kw or (extra-coronary):ti,ab,kw or (non-coronary):ti,ab,kw or (extracardiac):ti,ab,kw or (extra-cardiac):ti,ab,kw or (non-cardiac):ti,ab,kw)

Materials and Methods

Systematic Review: Search and Inclusion

A systematic review method was employed to ensure comprehensive coverage of the available evidence. The Meta-analysis of observational studies (MOOSE) checklist[10] was consulted during the writing of this article (Supplement S1). A systematic electronic search was performed on 15-09-2011 using the MEDline, EMBASE and Cochrane databases. Synonym lists were generated to describe our intended domain and outcome: adult patients undergoing routine cardiac CT and (overall) prevalence of unrequested findings. These were subsequently used to build the search (Table 1).

The titles and abstracts from the different databases resulting from this search were combined and duplicates were manually filtered. The remaining articles were then subjected to the selection procedure further outlined in Figure 1. Briefly, the titles and the abstracts were screened by two experienced medical researchers independently (CFB and MJAG) on the basis of predefined exclusion and inclusion criteria (exclusion and inclusion criteria 1, Figure 1), largely to ensure general applicability of the articles. Briefly, we assessed whether the abstracts retrieved by the search reported on extra-cardiac findings on cardiac CTs met the inclusion criteria 1 (Figure 1). Studies published before

Table 2. Selected items from STROBE checklist, together with percentage and number of articles in which items were scored 'yes'.

STROBE item number	description of STROBE item (verbatim from STROBE checklist)	number of articles reporting (%)	Corresponding specification of STROBE item	number of articles reporting (%)
5	describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	11 (58)	referral source (clarifies which portion of included patients are self-referral, from primary care, from emergency care, from intramural specialist care, screening)	5 (26)
6(a)	give the eligibility criteria, and the sources and methods of selection of participants	7 (37)		
14a	give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	13 (68)	reports the prevalence of CVD risk factors and comorbidities in included patients (e.g. smoking, hypertension, hypercholesterolemia)	7 (37)
15	report numbers of outcome events or summary measures	19 (100)	results for individual (types of) findings given in absolute numbers as well as prevalences	2 (11)

1990 were excluded due to the non-comparability in access to and quality of CT-scanning between recent years and the 1980s. Full text papers meeting these criteria were screened using the second set of inclusion and exclusion criteria (exclusion and inclusion criteria 2, Figure 1). This second set of inclusion and exclusion criteria were intended to discriminate between articles containing truly useful information and those that were less relevant to routine clinical practice. This included studies investigating incidental CT findings in highly specialized subpopulations (e.g. only patients with cardiac tumours, patients with sarcoidosis), studies only reporting on one (class of) unrequested/incidental finding (e.g. breast lesions[11] or cardiac abnormalities[12]) and studies that turned out not to report on unrequested cardiac findings after full-text review.

Systematic Review: Data Extraction

The 'STrengthening the Reporting of OBservational studies in Epidemiology' (STROBE)[13,14] checklist for cross-sectional studies was used as a framework to assess the quality of the reporting in the included articles. We selected those items pertaining to the reporting of

study population and (completeness of) reporting of results and adapted them so that they would more specifically address the prevalence of incidental findings on cardiac CTs.

Briefly, we deemed the items concerning the setting, the sources/eligibility criteria of participants, and the description of study participant characteristics and the reporting of results to be the most germane and these were further specified to our research question. The resulting specified items, alongside the original STROBE items from which they were derived (Table 2), were scored as present or absent by two authors independently. The items were scored as present if the item was reported adequately anywhere in the assessed article. In the case of referral source, an item was scored as reported if it was clear how the study population came to be referred for cardiac scanning. For the prevalence of CVD risk factors, we required that the prevalence of the major risk factors (smoking, hypertension and some form of CVD history) be reported. Finally, we assessed whether the absolute numbers of unrequested findings as well as their prevalence could be delineated.

Data on study parameters and the prevalence of unrequested findings were extracted from the included papers by two authors independently, with consensus sought in cases of disagreement. Table 1: The primary outcome of interest was clinically relevant unrequested findings, defined as those unrequested findings which required short-term follow-up, either with further diagnostic procedures or therapeutic interventions.

Meta-analysis and Meta-regression

All statistical analyses were carried out using the R statistical program[15] version 2.13.1. Meta-analysis and meta-regression were carried out using the metafor[16] package version 1.6.0. We pooled the reported prevalences of Clinically Relevant Unrequested findings in order to come to more meaningful conclusions (Table 2). By then assessing heterogeneity and performing univariate meta-regression we sought to assess the degree of 'differentness' and to then explain it using easily extracted study parameters, such as the age of the patient group. The proportions of clinically relevant unrequested findings were logit transformed to improve approximate normality. These were used in the analyses and meta-regression, with the results being back-transformed before presentation here. Heterogeneity was assessed by computing the proportion of unexplained variance using the I2 and Tau2 statistics.[17] Pooled estimates were generated using restricted maximum likelihood estimator random effect approach when the I2 was found to be higher than 25%[18]; this random effects approach makes allowances for the excess heterogeneity the I2 statistic reflects. Funnel plots were generated and visually inspected for approximate symmetry to assess the risk of publication bias.

For the meta-regression, mixed effects regression using unrestricted maximum likelihood estimator method was employed to estimate the effects of potentially explanatory variables that could be abstracted from the articles. The reported mean age, proportion of smokers, years since publication and use of full Field Of View (FOV; whether or not all available anatomical regions were assessed) were considered as potentially explanatory for differences in the levels of clinically relevant unrequested findings reported. Where an explanatory variable was not reported, we imputed it using simple median imputation (only relevant for the proportion of smokers).

Results

Systematic Review

The majority of the nineteen papers reviewed routine cardiac CTs were conducted in convenience samples of patients with suspected Coronary Artery Disease (CAD) to determine the prevalence and significance of any unrequested findings. Three studies only retrospectively reviewed the radiology reports[19-21]; the prevalence of unrequested findings in these studies was not substantially different from that of studies prospectively evaluating the presence of unrequested findings. A number restricted their investigation to narrow cardio-centric FOVs[22-25] while others assessed reconstructions based on the maximally available FOV.

Several studies also drew a direct comparison between the unrequested findings detectable on full thoracic FOV and smaller, cardiac FOV. Kim et al.[26] compared the prevalence on LDCT scout views with a narrower cardiac-focused reconstructed FOV and found a very large discrepancy between the two, with the overwhelming majority of clinically relevant unrequested findings being missed in the narrower FOV. Aglan et al.[27] similarly compared the prevalence of unrequested findings observed with a narrow FOV with a full FOV using a split-sample approach. They also found far more unrequested finding on full 'thoracic' FOV. An indirect comparison between the prevalences reported in those articles based upon a restricted FOV and those based upon a full FOV did not show the same trend. This was confirmed quantitatively (see meta-analysis results below).

All studies distinguished between clinically relevant unrequested extra-cardiac findings and clinically irrelevant findings by classifying the former as those that require further action or follow-up and the latter as those that do not (non-relevant). Some also opted for a multimodal classification into mild, moderate and severe, with the latter two requiring some form of clinical action.[19,23,28,29]

This classification did not seem systematically pre-specified in any of the papers and was typically described pragmatically and briefly in the methods as based upon the attendant need for further follow-up or action according to the insights of the evaluating radiologists and cardiologists (with one exception, where raters simply filled in premade worksheets[26]). Some articles did explain how select, specific findings were handled, such as the criteria used to assess coronary artery aneurysms.[30] This is most notably the case for lung nodules, which two papers explicitly classified them according to the Fleischner[31] criteria,[29,32] whilst one paper[33] chose to classify all visible nodules as potentially relevant.

Four papers also reported whether the detected (potentially) relevant unrequested findings actually led to therapeutic or diagnostic consequences, chiefly through chart-review. Machaalany et al.[28] found an overall prevalence of 8.2% of potentially relevant unrequested findings, of which 7% were indeterminate. They performed telephone and chart-review follow-up and found that no indeterminate findings had converted to relevant findings. Lehman et al.[34] investigated the number of unrequested findings observed in the course of an ongoing study conducted amongst patients presenting to their emergency room with acute chest pain. Whilst newly detected unrequested findings were detected in 20.5% of patients, patient management was only actually changed in 4.4% of cases overall. Similarly, Onuma et al.[35] found a prevalence of clinically relevant unrequested findings of 22.7% in patients suspected of CAD with 3.6% of the total population eventually having therapeutic consequences. In post CABG-patients, Mueller et al.[24] found 19.7% unrequested findings with documented follow-up in 9.6% of patients.

Meta-analysis

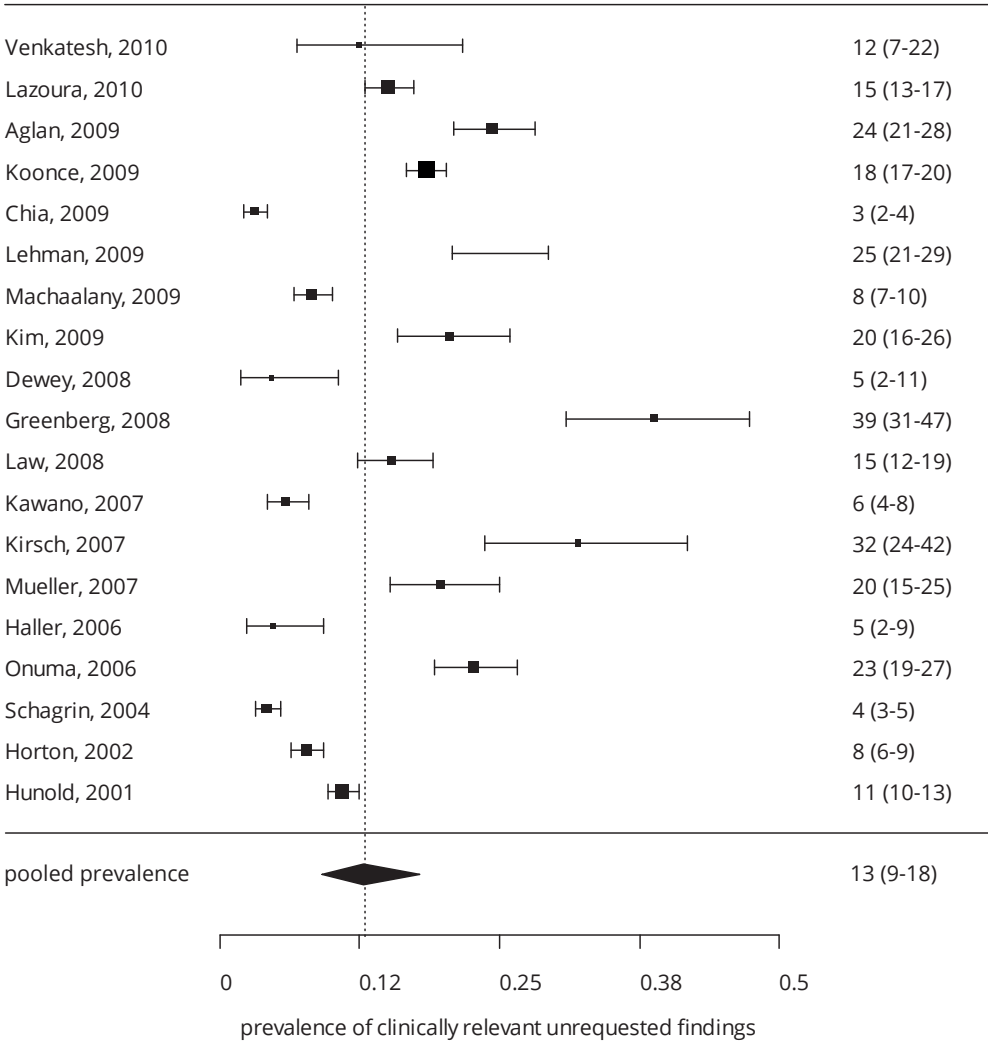
The nineteen cardiac CT studies, incorporating 12922 patients, showed a pooled prevalence of clinically relevant unrequested findings of 13% (95% confidence interval: 9-18%, Figure 2, Figure 3). We found an overall I2 statistic of 98%. This suggests excess inter-study variability and correspondingly random effects were employed to generate pooled estimates. We found that the random-effects pooled prevalence estimates differed substantially from the fixed effects estimates, in keeping with the degree of heterogeneity suggested by the I2. The funnel plot was approximately symmetrical, suggesting a low risk of publication bias (not reproduced here).

Univariate meta-regression for variables that could potentially explain this heterogeneity did not yield any significant associations between the parameters assessed and the prevalence of clinically relevant unrequested findings in the dataset. We found no significance in mixed effects meta-regression for the proportion of smokers ($p=0.33$), mean

Figure 3. Overview of included articles with abstracted parameters: demographic and scan parameters

							number of findings**									
study	publication year	no. of patients	FOV	mean age (yrs)	smokers (%)	male gender (%)	great vessals	lung	mediastinum	subthoracic	other	total	clinically relevant	prevalence clinically relevant unrequested findings (%)	number of detectors	slice thickness assessed
							4	23	2	38	0	67	10	12.5 (7-22)	64 slice	0.5 mm
Lazoura[29]	2010	1044	Full FOV	61	72	74	135	278	?	222	94	729	151	15 (13-17)	128-slice	0.6 mm
Aglan[27]	2009	542	full FOV	58	?	57	98	121	18	28	0	391	139	24.3 (21-28)	64 slice	5 mm
Koonce[19]	2009	1764	full FOV	58	?	61	24	346	18	59	69	507	325	18.4 (17-20)	64-slice	5 mm or 3 mm or 0.6 mm*
Chia[33]	2009	1061	full FOV	56	19	65	?	54	11	33	5	103	48	3.1 (2-4)	64-slice	5 mm
Lehman[34]	2009	395	full FOV	53	48	63	5	142	?	52	4	205	98	20.5 (21-29)	64-slice	3 mm
Machaalany[28]	2009	966	full FOV (32-50 cm)	58	61	55	?	?	?	?	?	401	80	8.2 (7-10)	64-slice	2.5 mm
Kim[26]	2009	254	whole thorax	59	33	56	2	205	4	65	6	282	52	20.4 (16-26)	64-slice	?
Dewey[41]	2008	108	32cm FOV	63	?	78	1	7	2	6	0	16	5	5 (211)	16-slice	4 mm
Greenberg[22]	2008	134	16-35 cm	54	41	78	6	85	20	16	61	188	52	39 (31-48)	40-slice	?
Law[20]	2008	435	30 cm FOV	56	25	63	?	33	4	27	1	72	72	15.4 (12-19)	16-slice	3 mm or 1.5 mm*
Kawano[42]	2007	617	full FOV	66	?	56	0	62	32	50	5	149	36	5.8 (4-8)	64-slice	0.5 mm
Kirsch[32]	2007	100	25 cm FOV	63	41	68	13	43	36	42	7	142	32	32 (24-42)	64-slice	0.75 mm
Mueller[30]	2007	259	small FOV (+-25 cm)	64	?	74	3	29	1	2	1	51	34	13.1 (15-25)	16-slice	?
Haller[43]	2006	166	full FOV	64	?	74	3	23	5	2	3	36	8	4.8 (2-9)	16-slice	1 mm
Onuma[35]	2006	503	whole thorax	66	52	76	9	246	4	72	14	345	114	22.7 (19-27)	16-slice or 64-slice	5 mm
Schragin[21]	2004	1356	not specified	53	?	69	?	?	?	?	?	278	57	4.2 (3-5)	EBT	3 mm
Horton[44]	2002	1326	35 cm FOV	55	25	64	?	89	?	10	4	103	103	7.8 (7-9)	EBT	3 mm
Hunold[25]	2001	1812	26 cm FOV	59	?	78	?	?	72	92	99	583	191	11 (10-13)	EBT	?

Figure 2. Forest plot of the included study showing the prevalence of clinically relevant unrequested findings and pooled prevalence estimate



age of included subjects (p=0.87), gender (p=0.82), FOV (p=0.59) and the number of years since publication (p=0.26).

The STROBE items included and specified to our research question show the frequent absence of reporting of the referral source and the prevalence of other CVD risk-factors; only 37% of articles mention the proportions of study subjects included from different sources (i.e. primary care, specialist care, self-referral). We found that 14 articles mentioned cursory study patient characteristics but that these were usually limited to age and gender,

with parameters such as smoking status missing in 8/19 (42%) studies (not shown in table 2) and only 35% mentioning the cardinal CVD risk factors (smoking and hypertension and CVD history). We also observed that whilst all 19 studies reported numbers of unrequested findings (also an eligibility criterion); only 11% articles reported these data in such a way that the prevalence and absolute numbers of each (class of) unrequested finding could each be calculated. Many authors chose to report the absolute numbers of each finding in detail, but the possibility that single patients may have had multiple findings prevented accurate calculations of prevalence.

Discussion

Unrequested findings were found to occur in approximately 13% of patients undergoing cardiac CT. This high overall prevalence is largely in line with what has been reported in screening settings.[36]

Surprisingly, the high level of heterogeneity in prevalence on unrequested findings (i.e. 3-39%) was not explained by likely study and population characteristics, such as smoking and age. Similarly, imaging technique (i.e. FOV) did not explain the heterogeneity between studies.

More detailed imaging and population characteristics that could have explained the heterogeneity were not systematically reported, as shown by the results of the STROBE quality check, with a only a third of articles fully describing the referral population source and reporting their risk profiles.

Differences in definition and classification of the endpoint, i.e. clinically relevant unrequested findings, is probably the largest contributor to the high level of heterogeneity. In each article the clinical relevance of the unrequested findings were defined based upon prevailing local insights and the expert opinion of the evaluating radiologists rather than any systematic evidence of prognostic significance, making it very difficult to begin to assess the nature of the criteria.

Consensus on the definition and classification of relevant unrequested findings is impossible in the absence of evidence concerning the prognostic and diagnostic value of such findings. Evidence supporting the wider prognostic value of (types of) unrequested findings might support more systematic reporting and acting-upon unrequested findings observed on cardiac CT and other scan-types by demonstrating their value and raising awareness. It is plausible that further research could also differentiate between findings

with higher value and those with little or none. Such evidence would improve studies in this field, which until now have treated unrequested findings as large undifferentiated groups and assigned significance according to individual author's insights.

The growing acceptance of more and earlier cardiovascular CT screening[37-39] amongst pre-symptomatic patients introduces further complication. Amongst these patients there is little precedent supporting the prognostic significance of unrequested findings in routine care settings. Whilst accurate risk stratification is more difficult amongst these patients, due to the longer time horizons and more subtle defects involved, the benefits of earlier targeted preventative measures might be correspondingly larger.

We found large discrepancies between the prevalence of clinically relevant (i.e. requiring follow-up) findings and the number of findings that actually led to therapeutic or diagnostic interventions.[24,28,34,35] This suggests either a lack of communication between radiologists and clinicians, differences in the perceived clinical relevance of certain findings between these groups, or both. This seeming lack of consensus may have also contributed to the unexplained heterogeneity between the studies found in the meta-analysis.

Limitations

We acknowledge that our study suffers from several limitations, including language limited to English, Dutch and German. Furthermore by choosing to limit our analysis to only cardiac CT, the prevalences we found may not be representative of the prevalence of unrequested findings in other anatomical regions or using other modalities. We examined the effects of study parameters on the prevalence of clinically relevant unrequested findings using the aggregate level data reported in the included studies and did not pursue the individual patient data (i.e. used mean patient age instead of the actual ages of all the individual patients). The latter approach is likely to have been more sensitive to subtle variations between the population.[25] Furthermore, one of the parameters included in the meta-regression (smoking) was missing in almost half of the studies. Consequently we imputed this parameter using median imputation, further reducing its variability and hence the sensitivity of our analysis.

Recommendations

In the absence of a standard definition of clinically relevant unrequested and to facilitate comparison between studies, we recommend that future studies transparently report the nature of the unrequested findings detail their absolute numbers and prevalence,

the impact on patient care and outcome (if applicable), and the demographic and clinical characteristics of the source population. This is in lieu of reporting detailed criteria for clinical relevance, which may be impossible to pre-specify at this stage. In addition to adhering to the STROBE checklist, we suggest authors further specify the exact referral sources of patients included and report the prevalence of relevant risk factors, as specified in Table 2.

Conclusion

We found a high prevalence of clinically relevant unrequested findings among published studies. The large range of prevalences could not be satisfactorily explained in this analysis. Further research to assess the true prognostic value of individual (sets of) unrequested findings that incorporates follow-up to measure associated patient outcomes would be desirable to inform an evidence-based response to the high prevalence of potentially clinically relevant unrequested information on thoracic CT scans.

References

1. Garvey CJ, Hanlon R. Computed tomography in clinical practice. *BMJ*. 2002; 324: 1077–1080.
2. Scrutinio D, Giannuzzi P. Comorbidity in patients undergoing coronary artery bypass graft surgery: impact on outcome and implications for cardiac rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2008; 15: 379–385.
3. Smith-Bindman R. Is computed tomography safe? *N Engl J Med*. 2010; 363: 1–4.
4. Brenner DJ, Hricak H. Radiation exposure from medical imaging: time to regulate? *JAMA*. 2010; 304: 208–209.
5. Hillman BJ, Goldsmith JC. The uncritical use of high-tech medical imaging. *N Engl J Med*. 2010; 363: 4–6.
6. Mettler FA, Bhargavan M, Faulkner K, Gilley DB, Gray JE, Ibbott GS et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources--1950-2007. *Radiology*. 2009; 253: 520–531.
7. Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'Byrne MM, Port JD et al. Incidental findings in imaging research: evaluating incidence, benefit, and burden. *Arch Intern Med*. 2010; 170: 1525–1532.
8. Budoff MJ, Gopal A. Incidental findings on cardiac computed tomography. Should we look? *J Cardiovasc Comput Tomogr*. 2007; 1: 97–105.
9. Sosnouski D, Bonsall RP, Mayer FB, Ravenel JG. Extracardiac findings at cardiac CT: a practical approach. *J Thorac Imaging*. 2007; 22: 77–85.
10. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000; 283: 2008–2012.
11. Hussain A, Gordon-Dixon A, Almusawy H, Sinha P, Desai A. The incidence and outcome of incidental breast lesions detected by computed tomography. *Ann R Coll Surg Engl*. 2010; 92: 124–126.
12. Quentin M, Kröpil P, Steiner S, Lanzman RS, Blondin D, Miese F et al. [Prevalence and clinical significance of incidental cardiac findings in non-ECG-gated chest CT scans]. *Radiologe*. 2011; 51: 59–64.
13. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiol Camb Mass*. 2007; 18: 805–835.

14. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008; 61: 344–349.

15. The R Foundation: A Language and Environment for Statistical Computing. Vienna, 2012.

16. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010; 36: 1–48.

17. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21: 1539–1558.

18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986; 7: 177–188.

19. Koonce J, Schoepf JU, Nguyen SA, Northam MC, Ravenel JG. Extra-cardiac findings at cardiac CT: experience with 1,764 patients. *Eur Radiol*. 2009; 19: 570–576.

20. Law YM, Huang J, Chen K, Cheah FK, Chua T. Prevalence of significant extracoronary findings on multislice CT coronary angiography examinations and coronary artery calcium scoring examinations. *J Med Imaging Radiat Oncol*. 2008; 52: 49–56.

21. Schragin JG, Weissfeld JL, Edmundowicz D, Strollo DC, Fuhrman CR. Non-cardiac findings on coronary electron beam computed tomography scanning. *J Thorac Imaging*. 2004; 19: 82–86.

22. Greenberg-Wolff I, Uliel L, Goitein O, Shemesh J, Rozenman J, Di Segni E et al. Extra-cardiac findings on coronary computed tomography scanning. *Isr Med Assoc J*. 2008; 10: 806–808.

23. Kirsch J, Araoz PA, Steinberg FB, Fletcher JG, McCollough CH, Williamson EE. Prevalence and significance of incidental extracardiac findings at 64-multidetector coronary CTA. *J Thorac Imaging*. 2007; 22: 330–334.

24. Mueller J, Jeudy J, Poston R, White CS. Cardiac CT angiography after coronary bypass surgery: prevalence of incidental findings. *AJR Am J Roentgenol*. 2007; 189: 414–419.

25. Hunold P, Schmermund A, Seibel RM, Grönemeyer DH, Erbel R. Prevalence and clinical significance of accidental findings in electron-beam tomographic scans for coronary artery calcification. *Eur Heart J*. 2001; 22: 1748–1758.

26. Kim JW, Kang E-Y, Yong HS, Kim YK, Woo OH, Oh Y-W et al. Incidental extracardiac findings at cardiac CT angiography: comparison of prevalence and clinical significance between precontrast low-dose whole thoracic scan and postcontrast retrospective ECG-gated cardiac scan. *Int J Cardiovasc Imaging*. 2009; 25 Suppl 1: 75–81.

27. Aglan I, Jodocy D, Hiehs S, Soegner P, Frank R, Haberfellner B et al. Clinical relevance and scope of accidental extracoronary findings in coronary computed tomography angiography: a cardiac versus thoracic FOV study. *Eur J Radiol*. 2010; 74: 166–174.

28. Machaalany J, Yam Y, Ruddy TD, Abraham A, Chen L, Beanlands RS et al. Potential clinical and economic consequences of noncardiac incidental findings on cardiac computed tomography. *J Am Coll Cardiol*. 2009; 54: 1533–1541.

29. Lazoura O, Vassiou K, Kanavou T, Vlychou M, Arvanitis DL, Fezoulidis IV. Incidental non-cardiac findings of a coronary angiography with a 128-slice multi-detector CT scanner: should we only concentrate on the heart? *Korean J Radiol*. 2010; 11: 60–68.

30. Mueller J, Jeudy J, Poston R, White CS. Cardiac CT angiography after coronary bypass surgery: prevalence of incidental findings. *AJR Am J Roentgenol*. 2007; 189: 414–419.

31. MacMahon H, Austin JHM, Gamsu G, Herold CJ, Jett JR, Naidich DP et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005; 237: 395–400.

32. Kirsch J, Araoz PA, Steinberg FB, Fletcher JG, McCollough CH, Williamson EE. Prevalence and significance of incidental extracardiac findings at 64-multidetector coronary CTA. *J Thorac Imaging*. 2007; 22: 330–334.

33. Chia P-L, Kaw G, Wansaicheong G, Ho K-T. Prevalence of non-cardiac findings in a large series of patients undergoing cardiac multi-detector computed tomography scans. *Int J Cardiovasc Imaging*. 2009; 25: 537–543.

34. Lehman SJ, Abbara S, Cury RC, Nagurney JT, Hsu J, Goela A et al. Significance of cardiac computed tomography incidental findings in acute chest pain. *Am J Med*. 2009; 122: 543–549.

35. Onuma Y, Tanabe K, Nakazawa G, Aoki J, Nakajima H, Ibukuro K et al. Noncardiac findings in cardiac imaging with multidetector computed tomography. *J Am Coll Cardiol*. 2006; 48: 402–406.

36. Burt JR, Iribarren C, Fair JM, Norton LC, Mahbouba M, Rubin GD et al. Incidental findings on cardiac multidetector row computed tomography among healthy older adults: prevalence and clinical correlates. *Arch Intern Med*. 2008; 168: 756–761.

37. Elias-Smale SE, Proença RV, Koller MT, Kavousi M, van Rooij FJA, Hunink MG et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. *J Am Coll Cardiol*. 2010; 56: 1407–1414.

38. Johnson KM, Dowe DA. The detection of any coronary calcium outperforms Framingham risk score as a first step in screening for coronary atherosclerosis. *AJR Am J Roentgenol*. 2010; 194: 1235–1243.

39. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest

pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). *Circulation*. 2007; 115: 402–426.

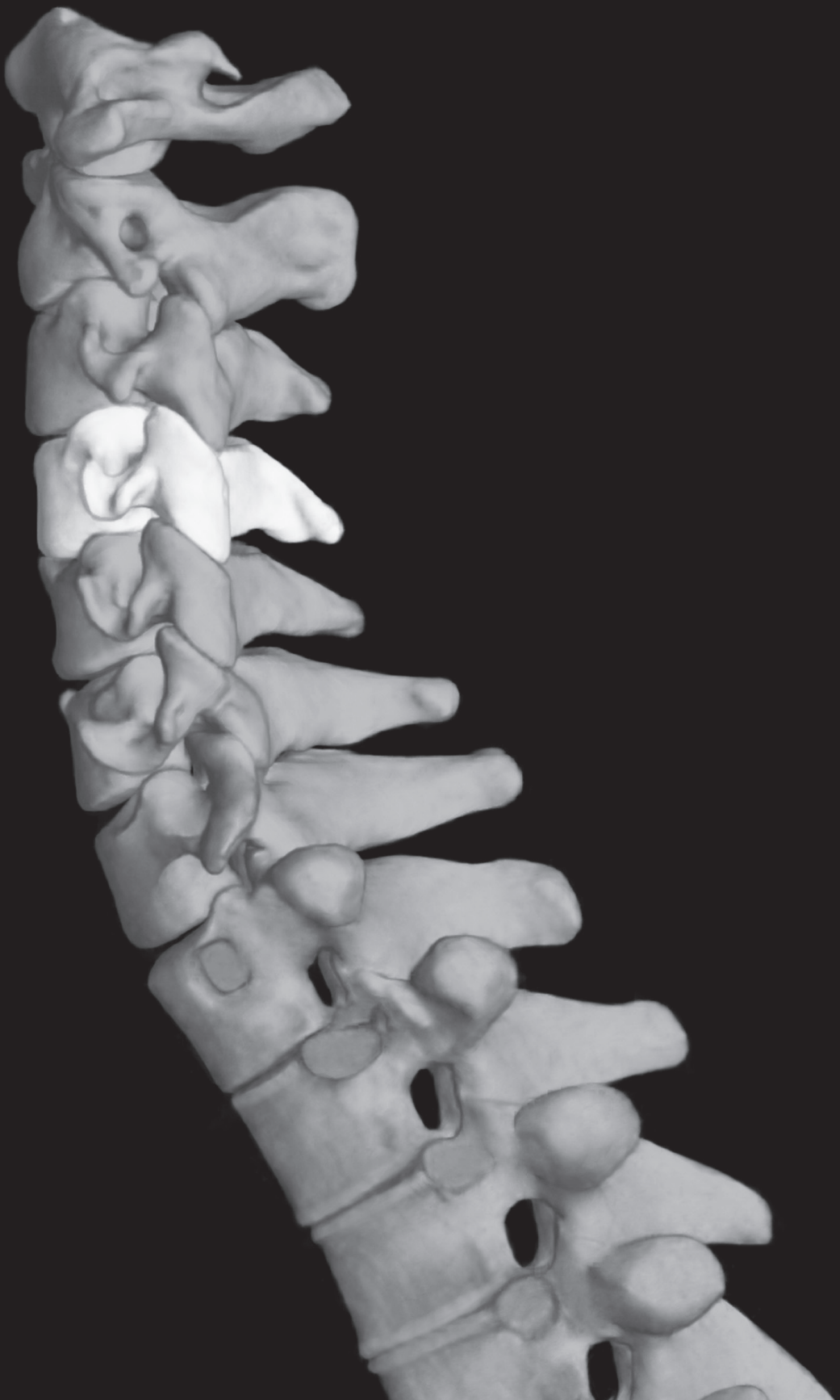
40. Venkatesh V, You JJ, Landry DJ, Ellins ML, Sheth T. Extracardiac findings in cardiac computed tomographic angiography in patients at low to intermediate risk for coronary artery disease. *Can Assoc Radiol J*. 2010; 61: 286–290.

41. Dewey M, Schnapauff D, Teige F, Hamm B. Non-cardiac findings on coronary computed tomography and magnetic resonance imaging. *Eur Radiol*. 2007; 17: 2038–2043.

42. Kawano Y, Tamura A, Goto Y, Shinozaki K, Zaizen H, Kadota J. Incidental detection of cancers and other non-cardiac abnormalities on coronary multislice computed tomography. *Am J Cardiol*. 2007; 99: 1608–1609.

43. Haller S, Kaiser C, Buser P, Bongartz G, Bremerich J. Coronary artery imaging with contrast-enhanced MDCT: extracardiac findings. *AJR Am J Roentgenol*. 2006; 187: 105–110.

44. Horton KM, Post WS, Blumenthal RS, Fishman EK. Prevalence of significant noncardiac findings on electron-beam computed tomography coronary artery calcium screening examinations. *Circulation*. 2002; 106: 532–534.



Chapter 4

Intra- and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography

Buckens CF
de Jong PA
Mol CP
Bakker E
Stallman HP
Mali WP
van der Graaf Y
Verkooijen HM

PLoS One. 2013 Aug 5;8(8):e71204

Abstract

Intra- and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography

Objectives

To evaluate the reliability of semiquantitative Vertebral Fracture Assessment (VFA) on chest Computed Tomography (CT).

Methods

Four observers performed VFA twice upon sagittal reconstructions of 50 routine clinical chest CTs. Intra- and interobserver agreement (absolute agreement or 95% limits of agreement) and reliability (Cohen's kappa or intraclass correlation coefficient(ICC)) were calculated for the visual VFA measures (fracture present, worst fracture grade, cumulative fracture grade on patient level) and for percentage height loss of each fractured vertebra compared to the adjacent vertebrae.

Results

Observers classified 24-38% patients as having at least one vertebral fracture, giving rise to kappas of 0.73-0.84 (intraobserver) and 0.56-0.81 (interobserver). For worst fracture grade we found good intraobserver (76-88%) and interobserver (74-88%) agreement, and excellent reliability with square-weighted kappas of 0.84-0.90 (intraobserver) and 0.84-0.94 (interobserver). For cumulative fracture grade the 95% limits of agreement were maximally ±1.99 (intraobserver) and ±2.69 (interobserver) and the reliability (ICC) varied from 0.84-0.94 (intraobserver) and 0.74-0.94 (interobserver). For percentage height-loss on a vertebral level the 95% limits of agreement were maximally ±11.75% (intraobserver) and ±12,53% (interobserver). The ICC was 0.59-0.90 (intraobserver) and 0.53-0-82 (interobserver). Further investigation is needed to evaluate the prognostic value of this approach.

Conclusion

In conclusion, these results demonstrate acceptable reproducibility of VFA on CT.

Introduction

Osteoporosis is a growing problem in the aging population, affecting up to one in three women and one in five men over 50 years of age,[1] leading to millions of fractures annually and contributing substantially to morbidity and mortality,[2,3] particularly in the developed world. Subclinical vertebral fractures are an early sign of osseous fragility and their prevalence among adults is approximately 25%, increasing with age.[4] Subclinical vertebral fractures may precede overt osteoporosis and may predict future fractures, independently of dual-energy X-ray absorptiometry, which is currently the standard modality used to diagnose osteoporosis but has only modest predictive value for future fractures.[5]

Vertebral fractures are visible on much routine clinical imaging that happens to visualize the spine, including chest Computed Tomography (CT). Despite being visible on chest CT, vertebral fractures are seldom assessed or reported unless this is specifically requested. Systematically reporting vertebral fractures and deformities on imaging that happens to visualize the spine would not require any additional imaging and could opportunistically identify patients who would benefit from preventative care. this is not currently common practice.

One of the most widely used methods for vertebral fracture assessment (VFA) is Genant's semiquantitative method,[6] which assesses the shape of the deformity and its severity. Previously this method has been shown to have fair to good reproducibility and reliability on lateral CT scout views, radiographs or spinal densitometry.[6–10] Vertebral fractures may be even more readily detectable on CT than on conventional radiography.[11] To the best of our knowledge, The intra- and interobserver variability of vertebral fracture assessment of Genant's VFA method has not been studied on multislice CT. Knowledge on reproducibility and reliability is a necessary prerequisite for further investigations into the potentially substantial prognostic value of vertebral fractures on routine chest CT.

In this study, we determine the intra- and interobserver reliability and agreement of VFA on sagittal reformats of chest CT.

Methods

Analysis and reporting of the study was performed according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[12]

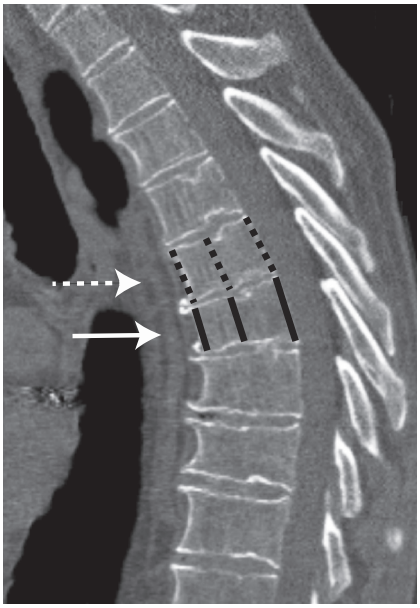
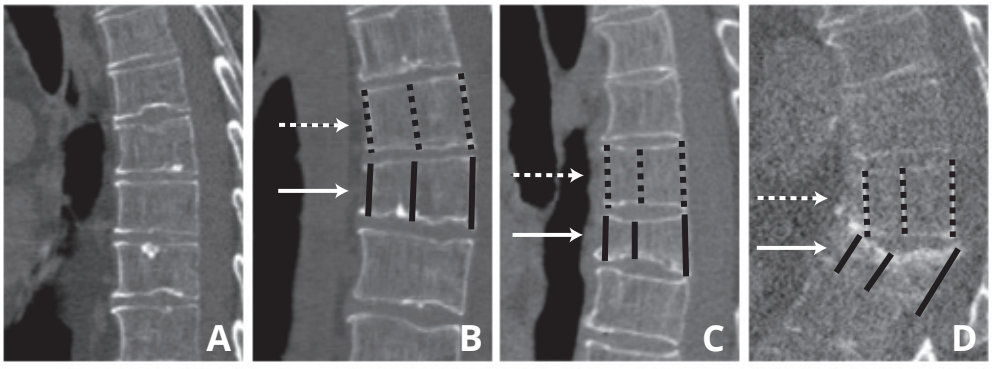


Figure 1. Moderate fracture. Degenerative spine showing a moderate (grade 2) wedge-shaped fracture (solid arrow), with a reference vertebra immediately cranial (dashed arrow). Anterior, middle and posterior height measurement lines drawn on both.

Measurements:
For this patient, there is a fracture present, the worst fracture grade is 2, the cumulative fracture grade is 2 and the worst height loss of the fractured vertebra is the anterior height at 25%.

Figure 2. Sagittal reformats showing examples of all possible fracture stages



A=grade 0 (unfractured); B=grade 1 (mild); C=grade 2 (moderate); D=grade 3 (severe). Also shown are the anterior, middle and posterior height measurement calliper placements of the fractured vertebra (solid white arrows) and an adjacent reference vertebra (dashed white arrows).

Source population and sampling

The present study was conducted in the context of the PROVIDI study, a study on the Prognostic Value of unrequested Information on Diagnostic Imaging. This multicenter study aims to establish the prognostic value of unrequested findings on thoracic CT and was described elsewhere.[13] Briefly, it includes all patients above forty years of age who

underwent chest CT in one of eight participating Dutch hospitals between 2002 and 2005 (making it retrospective in nature), with exclusion of patients with a primarily oncological indication on radiological referral form.[13] As such it contains a heterogeneous range of protocols and reconstruction formats, representing routine practice. CTs from two academic centers and one peripheral center were deemed to be of sufficient quality to allow sagittal reconstruction. In the other hospitals the slice thicknesses of the stored CTs was >3 mm limiting for multi-planar reconstructions.

A random sample of 45 subjects was drawn from the available 6010 anonimyzed CT scans. The sample was 'enriched' with five subjects with moderate to severe vertebral fractures by a researcher who was not among the observers. The average age of the patients was 64 years (range: 54–79 years) and 34 (75%) patients were male.

Vertebral fracture assessment

Semiquantitative vertebral fracture assessment was performed by four observers with different levels of experience: one board certified chest radiologist with 10 years of experience, two radiology residents with 3 years and 4 years of experience and a research physician with less than one year of experience. For each individual patient, CTs were rated twice and in a different random order more than one week after the first VFA session. Raters received a brief introductory training prior to the first rating session. Observers assessed the vertebral body morphology of each visible vertebral body at or around the mid-sagittal slice for that level in bone settings (Figure 1). Observers recorded whether the visible vertebrae appeared to be fractured and graded the fractures according to Genant's semiquantitative VFA.[6] This method identifies and categorizes fractures according to the worst height loss relative to a normal unfractured vertebrae as height loss of 20-25% (mild), height loss of 25-40% (moderate) or height loss more than 40% (severe) (Figure 2).

In addition to the semiquantitative visual assessment we quantified the anterior, posterior and mid-body heights of the fractured vertebra and the adjacent normal vertebra using electronic calipers (Figure 1 and 2). Observers were instructed to use the vertebrae above (cranial to) the fractured vertebra as the 'reference' vertebrae when two equally distant vertebra were available (Figure 2). The height loss percentage was then calculated by taking the difference in the anterior, middle and posterior heights of the fractured and reference vertebra divided by the reference heights (and multiplying it by 100). For each fractured vertebra, the greatest percentage height loss (either anterior, middle or posterior) was used. The observers were not able to revise the subjective fracture grades based on these quantitative measurements.

Table 1. Outcome measures, their level of measurement (patient or vertebral), definition and the statistical methods applied to analyze intra- and interobserver agreement and reliability.

			measure	
level	**outcome**	**definition**	**intra-/inter-observer agreement**	**intra-/inter-observer reliability**
patient	fracture present	fracture present (yes/no)	% absolute agreement	Cohen's kappa
	worst fracture grade	grade 0=<20% height loss grade 1=20-25% height loss grade 2=25-40% height loss grade 3=>40% height loss	% absolute agreement	weighted kappa*
	cumulative fracture grade	sum of all grades for all fractures, continuous scale	95% limits of agreement	intraclass correlation coefficient
vertebral	height loss	measured height loss, expressed as percentage **	95% limits of agreement	intraclass correlation coefficient
	fracture present	fracture present (yes/no)	% absolute agreement	Cohen's kappa

* square weighted Cohen's kappa.

** the fractured vertebra is compared to the nearest unfractured vertebra, with preference given to vertebrae cranial to (above) the fractured vertebra; the percentage of the worst height loss of each fractured vertebra (either anterior, middle or posterior part of the vertebral corpus), is given (see Figure 1)

Analysis

Intra and interobserver agreement and reliability were estimated for five measures likely to hold prognostic relevance[14]: three patient-level measures (presence of a fracture, worst fracture grade and cumulative fracture grade) and two vertebral level measures (quantitative percentage height loss and presence of fracture) (Table 1). Note that the cumulative fracture grade, which is computed simply by summing up all fracture grades for each patient (i.e. two mild (grade 1) fractures and moderate (grade 2) one give a cumulative grade of four), is also known as the spinal deformity index.[14]

On an intra- and an interobserver level, we assessed agreement and reliability (Table 1). Agreement indicates the absolute closeness of repeated measurements[15] and is particularly important when assessing the utility of a measure to track health status-changes over time using repeated measurements. For categorical measures (presence of fracture on both

vertebral and patient levels, worst fracture grade) we computed absolute agreement[12] (i.e. the proportion of cases in which the first rating was exactly similar as the second). On the interobserver level, values were calculated for the first set of each observer only. Agreement of continuous measures (percentage height loss and cumulative fracture grade) was assessed using the Bland-Altman 95% limits of agreement,[12] which can be interpreted as the maximum magnitude by which repeat measurements would be expected to differ in each direction, in 95% of repetitions. Reliability indicates whether a test can effectively distinguish between study objects (in our case either vertebrae or patients), despite observer error. The reliability of a measure is critically important in diagnostic practice, where distinguishing between affected and non-affected persons at a single time-point is the principle goal.

For the dichotomous measure (presence of a fracture on both patient and vertebral levels) we calculated Cohen's kappas.[12] For the ordinal measure (the worst recorded fracture grade), reliability was assessed using square-weighted Cohen's kappa. Weighted kappa allows for the ordering in fracture grade assignment (mild–moderate–severe). Reliability is rated as 'moderate' for values between 0.41-0.60, as 'substantial' for values between 0.61-0.8 and as 'excellent' for values above 0.80.[16] To investigate the reliability of continuous measurements (cumulative fracture grade and vertebral height loss) the Intra-Class Correlation Coefficient (ICC) was used. ICCs can be interpreted as the percentage of the variability between the ratings which is due to differences between the patients, and not due to observer error.[12] The two-way ICC(2,1) was computed for interobserver ICCs, to reflect the fact that a sample or patients and a sample of raters was observed, whilst a one-way ICC(1,1) was computed for the intra-observer ICCs. ICCs exceeding 0.7 are considered good and ICCs exceeding 0.8 excellent, with observer error having a negligible effect on observed correlations between two (sets of) measurements.[17]

All analyses were performed using the R statistical software package (version 3.0.1,[18] with use of the 'IRR' package (version 0.83[19]) for calculating the ICCs, kappas and absolute agreement. For all reliability and agreement measures we present the values for the four observers as well as the ranges for the values observed. 95% confidence intervals were generated for the reliability measures using 2000 bootstrap replications.

Ethics statement

This study was approved by the ethical review board of the University Medical Center Utrecht (decision number 06/193), which waived the need for written informed consent.

Table 2. Description of patient population: frequencies and proportions or medians and ranges for each outcome based on first measurement session

level	outcome	observer 1	observer 2	observer 3	observer 4
patient	fracture present n (%)	15 (30%)	12 (24%)	16 (32%)	14 (28%)
	worst fracture grade n (%)				
	grade 0	35 (70%)	38 (76%)	34 (68%)	36 (72%)
	grade 1	5 (10%)	6 (12%)	8 (16%)	9 (18%)
	grade 2	4 (8%)	2 (4%)	5 (10%)	2 (4%)
	grade 3	6 (12%)	4 (8%)	3 (6%)	3 (6%)
	cumulative fracture grade (median, range)*	2 (1-14)	2 (1-8)	2 (1-13)	2 (1-9)
vertebral	height loss % (median, range)*, **	35.6 (3.2-72.3)	38.2 (19.3-74.2)	29.3 (6.1-79.1)	33.0 (4.5-72.3)
	fracture present: n (%)	25 (3.7%)	16 (2.3%)	29 (4.3%)	24 (3.6%)

* including only fractured vertebrae of patients classified as fractured

**note that some vertebrae, classified as fractured on visual assessment,showed an absolute height loss of less than 15% upon caliper measurement.

Results

The observers scored between 12 and 19 (24 to 38%) of the included patients as having at least one vertebral fracture. The worst fracture grade observed was mild in 5 to11 patients, moderate in 2 to10 patients and severe in 2 to 6 patients. The median cumulative fracture grade for all four observers was 2 (range 0 to 14). Observers reported median height loss amongst the fractured vertebrae ranging between 29.3 to 35.6% (Table 2).

Agreement

For patient-level fracture presence, the intraobserver agreement was between 88 and 94%, indicating that the observers classified the same patients similarly (i.e. unfractured or fractured) (Table 3). The interobserver agreement was lower, but still good, ranging from 82 to 92%. The worst fracture grade showed an intraobserver agreement of 76 to 88% and an interobserver agreement of 74 to 88%. For the cumulative fracture grade, the intraobserver and interoberserver 95% limits of agreement ranged from ±1.22 to ±1.99 and ±1.60 to ±2.69, respectively. This indicates that if the same or a different radiologist was to re-assess the

Table 3. Intra- and interobserver agreement for fracture presence, worst fracture grade, cumulative fracture grade and vertebral height loss

				agreement			
level	**outcome**	**measure**	**obs**	**1**	**2**	**3**	**4**
patient	fracture present	absolute agreement (%)*	**1**	88	90	86	82
			2		94	84	84
			3			90	92
			4				90
	worst fracture grade	absolute agreement (%)*	**1**	76	82	78	74
			2		88	76	80
			3			84	88
			4				84
	cumulative fracture grade	95% limits of agreement**	**1**	±1.99	±2.69	±1.6	±2.15
			2		±1.8	±2.58	±1.7
			3			±1.8	±1.84
			4				±1.22
vertebral	height loss (%)	95% limits of agreement**	**1**	±5.97	±7.25	±8.26	±11.71
			2		±8,29	±9.77	±11.31
			3			±8.36	±12.53
			4				±11.75
	fracture present	absolute agreement (%)*	**1**	98	98	98	97
			2		99	98	97
			3			98	97
			4				98

obs=observer; intra-observer agreement indicated by grey background

* percentage of absolute agreement in the first session of each observer for the interobserver and between the first and second sessions for the intraobserver

** the 95% limits of agreement are the range of observer variation; this indicates that differences beyond this range cannot be ascribed to observer error alone

same patient more than once, a change in fracture grade of 2 may be due to observer error alone but a change of 3 or more would be unlikely due to measurement error alone. The intraobserver and interobserver limits of agreement of the vertebral height loss ranged from ±5.97 to 11.75% and ±7.25 to 12.53%, respectively (Table 3). These values indicate that differences of up to 12.53% can be considered as measurement and observer error, upon repeat measurement of the same vertebra. The agreement for vertebral-level presence of fracture ranged from 97 to 99%, perhaps reflecting the low incidence of fractures on a vertebral level.

Table 4. Intra- and interobserver reliability for fracture presence, worst fracture grade, cumulative fracture grade and vertebral height loss

				agreement			
level	**outcome**	**measure**	**obs**	**1**	**2**	**3**	**4**
patient	fracture present	kappa	**1**	0.73 (0.52-0.91)	0.75 (0.50-0.91)	0.67 (0.42-0.88)	0.56 (0.29-0.79)
			2		0.84 (0.63-1.00)	0.61 (0.34-0.83)	0.59 (0.29-0.82)
			3			0.78 (0.58-0.96)	0.81 (0.61-0.96)
			4				0.76 (0.52-0.95)
	worst fracture grade	weighted kappa*	**1**	0.84 (0.68-0.93)	0.85 (0.67-0.95)	0.79 (0.56-0.92)	0.73 (0.45-0.88)
			2		0.89 (0.75-0.98)	0.82 (0.58-0.92)	0.87 (0.65-0.94)
			3			0.90 (0.78-0.96)	0.88 (0.67-0.97)
			4				0.89 (0.71-0.96)
	cumulative fracture grade	ICC**	**1**	0.91 (0.54-0.97)	0.75 (0.61-0.96)	0.94 (0.57-0.98)	0.87 (0.45-0.93)
			2		0.84 (0.71-0.94)	0.74 (0.57-0.93)	0.87 (0.48-0.96)
			3			0.91 (0.76-0.95)	0.89 (0.56-0.94)
			4				0.94 (0.65-0.98)
vertebral	height loss (%)	ICC**	**1**	0.90 (0.81-0.95)	0.82 (0.62-0.92)	0.81 (0.65-0.90)	0.56 (0.27-0.75)
			2		0.75 (0.48-0.91)	0.7 (0.42-0.85)	0.53 (0.21-0.73)
			3			0.82 (0.68-0.90)	0.55 (0.29-0.73)
			4				0.59 (0.33-0.77)
	fracture present	kappa	**1**	0.72 (0.58-0.85)	0.63 (0.43-0.79)	0.58 (0.41-0.73)	0.43 (0.24-0.59)
			2		0.74 (0.55-0.88)	0.57 (0.37-0.73)	0.39 (0.18-0.58)
			3			0.71 (0.57-0.82)	0.51 (0.32-0.67)
			4				0.56 (0.38-0.71)

obs=observer; intra-observer agreement indicated by grey background; * square-weighted kappa; ** intraclass correlation coefficient (ICC) (95% confidence intervals based on 2000 bootstrap replicates)

Reliability

For fracture presence the intraobserver reliability was good to excellent (kappa 0.73 (0.52-0.91) to 0.84 (0.63-1)) (Table 4). The interobserver kappas ranged from 0.56 (0.29-0.79) to 0.81 (0.61-0.96), indicating fair to excellent interobserver reliability. For worst fracture grade, intraobserver reliability (weighted kappa) ranged from 0.84 (0.68-0.93) to 0.9 (0.78-0.96) whilst the interobserver scores ranged from 0.73 (0.45-0.88) to 0.88 (0.67-0.97), indicating substantial to very good reliability. For the cumulative fracture grade, the intraobserver reliability was excellent (ICCs: 0.84 (0.71-0.94) to 0.94 (0.65-0.98)), and the interobserver was good to excellent (0.74 (0.57-0.93) to 0.94 (0.57-0.98)). The interobserver reliability of vertebral height loss was moderate to good (ICC: 0.59 (0.33-0.77) to 0.9 (0.81-0.95)), as was the interobserver (ICC: 0.53 (0.21-0.73) to 0.82 (0.62-0.92)). This rather large range was attributable to one of the four observers, without which the intra- and interobserver minimums would have been 0.75 (0.48-0.91) and 0.7 (0.42-0.85), respectively. A similar pattern repeated itself in the presence of fracture measure on a vertebral level: the intra-observer reliability ranged from 0.56 (0.38-0.71) to 0.74 (0.55-0.88) and the inter-observer reliability ranged from 0.39 (0.18-0.58) to 0.63 (0.43-0.79).

Discussion

Vertebral fracture assessment on routine chest CT scans in an adult population shows generally good reliability and agreement. Specifically, for fracture presence and worst fracture grade we found excellent reliability and agreement. For cumulative fracture grade we found good reliability but modest agreement. For vertebral height loss we found good agreement but modest reliability, largely attributable to one of the observers.

Reliability indicates the ability of a test to distinguish between different individuals in spite of measurement error, whilst agreement indicates the absolute closeness of repeated measurements. For example, a weighing scale may be able to accurately and reproducibly measure the body weight of patients with a low margin of error, thus having good agreement. The reliability however also depends in part on the variability of the body weight between the patient sample of interest. If they have body weights very close together (low variability), even the scale's small margin of error will confound its reliability and the reliability values associated with it will be low.

Our findings are in line with the reported interobserver reliability for the semiquantitative method on conventional radiography (interobserver kappa values ranging from 0.60 to

0.80[6,9]) and demonstrate that semiquantitative vertebral fracture assessment method can reliably be applied on sagittal reconstructions of chest CTs. The participating observers represent a range of of different levels of radiological experience, including a relative novice with less than one year of experience, two intermediate observers with several years of experience each and a highly experienced board certified radiologist. This range is representative of clinical practice.

Since the majority of vertebral fractures are clinically silent and underreported, the diagnosis is often delayed. Presumably this underreporting is due to a number of reasons, including the extra time involved in creating and assessing the necessary sagittal reformats, the tendency of radiologists to focus on requested pathologies, unfamiliarity with the application of vertebral fracture assessment to CT and a general uncertainty surrounding the prognostic implications of subclinical vertebral fractures. By showing the reliability of well-established vertebral fracture assessment schemes on sagittal CT, the willingness to consider vertebral fracture assessment on CTs may increase. The detection of subclinical vertebral fractures on routine imaging that happens to visualize the spine has the potential to be a useful and cost-effective means of identifying patients at risk for future osteoporotic fractures, who can then be treated preventatively with fall prevention, lifestyle advice, hormonal supplementation and mainly antiresorptive medication; interventions that are proven to reduce fracture risk. There is growing momentum to this end; current guidelines already list these fractures as an indication for treatment.[20]

For cumulative fracture grade, the modest limits of agreements we found may be acceptable in practice. A previous study[14] has shown that the cumulative fracture grade is predictive for future fracture risk, particularly when the grade >3 and especially when >7. Therefore the maximal limits of ±2.69 do not necessarily preclude the prognostic utility of this measure. However, further research is needed to determine which cut-offs are most prognostically useful.

The vertebral level height loss measurement performed additionally to the standard visual assessment as proposed by Genant showed limits of agreement very close to the minimum height difference which a trained observer is likely able to detect (i.e. 13%[21]). The reliability values (ICCs) however varied widely across the observers. This suggests that the reliability of this VFA measure may also fluctuate similarly in clinical practice. This variability was also repeated for the vertebral level presence of fracture (although not the patient-level presence of any fracture variable). Furthermore, some low (>20%) height loss values were found in vertebrae that were classified as (usually mildly) fractured upon visual inspection. This may be due to unfractured vertebrae being misclassified as fractured and/or due to

incorrect calliper measurement. These problems that observers had with quantification, which was also the most time-consuming part of the study, may be overcome by automated vertebral body height measurement on CT in the future.

Limitations

Since PROVIDI scans were acquired and stored between 2002-2005, and were retrospectively reconstructed, prospective reconstruction with new scanner generations would presumably result in better image quality and in theory non-comparability to our findings. Whilst a prospective study with the attendant better quality of stored reconstructions could result in better reliability and agreement, we feel our dataset provides a realistic assessment for how VFA might perform across a spectrum of scanner generations currently in use in a variety of settings and locales.

Inherent to our study design, we lack an external reference standard with which to compare the observers' ratings. Additional imaging performed in PROVIDI patients, such as lateral chest X-rays, which might have been used for this purpose, were not included in the original study design and are also beyond the scope of this paper. Demonstrating the reliability and agreement does not require such an external 'gold standard' as comparisons are done between and within observers, rather than with an external reference standard, as in a diagnostic accuracy study. we enriched our sample to ensure an adequate number of higher fracture severities would be present. We believe that this is unlikely to have influenced agreement and reliability measures,[22] as previous studies investigating the prevalence of vertebral fractures on routine clinical CT showed prevalence of vertebral fractures similar to ours, ranging from 10-35%.[23–26] Finally, the clinical histories of the included patients were not available within the PROVIDI cohort. Consequently other causes of vertebral fracture such as past major trauma were not known, nor was it known which proportion of the patients identified with fractures were already receiving fracture prevention. Prior literature suggests that only a minority of fractures will have been known and a minority will have been under treatment.[27]

Conclusion

In conclusion, We found that semiquantitave vertebral fracture assessment can be applied on standard sagittal reconstructions of routine clinical chest CTs with acceptable reliability and agreement. Future research to evaluate the prognostic value of these VFA measures on routine clinical CTs should elucidate which of the four VFA is the strongest predictor of future fractures.

References

1. Keen RW. Burden of osteoporosis and fractures. *Curr Osteoporos Rep*. 2003; 1: 66–70.
2. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos Int*. 2006; 17: 1726–1733.
3. Kanis JA, on behalf of the World Health Organization Scientific Group. Assessment of osteoporosis at the primary health-care level. Technical Report. World Health Organization Collaborating Centre for Metabolic Bone Diseases, University of Sheffield, UK.
4. Van der Jagt-Willems HC, van Hengel M, Vis M, van Munster BC, van Campen JPCM, Tulner LR et al. Why do geriatric outpatients have so many moderate and severe vertebral fractures? Exploring prevalence and risk factors. *Age Ageing*. 2012; 41: 200–206.
5. Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. *Postgrad Med J*. 2007; 83: 509–517.
6. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993; 8: 1137–1148.
7. Buehring B, Krueger D, Checovich M, Gemar D, Vallarta-Ast N, Genant HK et al. Vertebral fracture assessment: impact of instrument and reader. *Osteoporos Int*. 2010; 21: 487–494.
8. Binkley N, Krueger D, Gangnon R, Genant HK, Drezner MK. Lateral vertebral assessment: a valuable technique to detect clinically significant vertebral fractures. *Osteoporos Int*. 2005; 16: 1513–1518.
9. Kim YM, Demissie S, Eisenberg R, Samelson EJ, Kiel DP, Bouxsein ML. Intra-and inter-reader reliability of semi-automated quantitative morphometry measurements and vertebral fracture assessment using lateral scout views from computed tomography. *Osteoporos Int*. 2011; 22: 2677–2688.
10. Takada M, Wu CY, Lang TF, Genant HK. Vertebral fracture assessment using the lateral scoutview of computed tomography in comparison with radiographs. *Osteoporos Int*. 1998; 8: 197–203.
11. Wintermark M, Mouhsine E, Theumann N, Mordasini P, van Melle G, Leyvraz PF et al. Thoracolumbar spine fractures in patients who have sustained severe trauma: depiction with multi-detector row CT. *Radiology*. 2003; 227: 681–689.
12. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011; 64: 96–106.
13. Gondrie MJA, Mali WPTM, Buckens CFM, Jacobs PCA, Grobbee DE, van der Graaf Y. The PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) Study: rationale and design. *Eur J Epidemiol*. 2010; 25: 751–758.
14. Siris ES, Genant HK, Laster AJ, Chen P, Misurski DA, Krege JH. Enhanced prediction of fracture risk combining vertebral fracture status and BMD. *Osteoporos Int*. 2007; 18: 761–770.

15. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; 59: 1033–1039.

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33: 159–174.

17. Nunnally JC. Psychometric theory. 3rd ed. McGraw-Hill: New York, 1994.

18. The R Foundation: A Language and Environment for Statistical Computing. Vienna, 2012.

19. Gamer M. irr: Various Coefficients of Interrater Reliability and Agreement. 2012.

20. Singer AJ, Boonen S. Osteoporosis management: translating research into optimal fracture protection II. *Curr Med Res Opin*. 2008; 24: 1789–1796.

21. Kleerekoper M, Nelson DA, Peterson EL, Tilley BC. Outcome variables in osteoporosis trials. *Bone*. 1992; 13 Suppl 1: S29–34.

22. Reed WM, Ryan JT, McEntee MF, Evanoff MG, Brennan PC. The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology*. 2011; 258: 938–943.

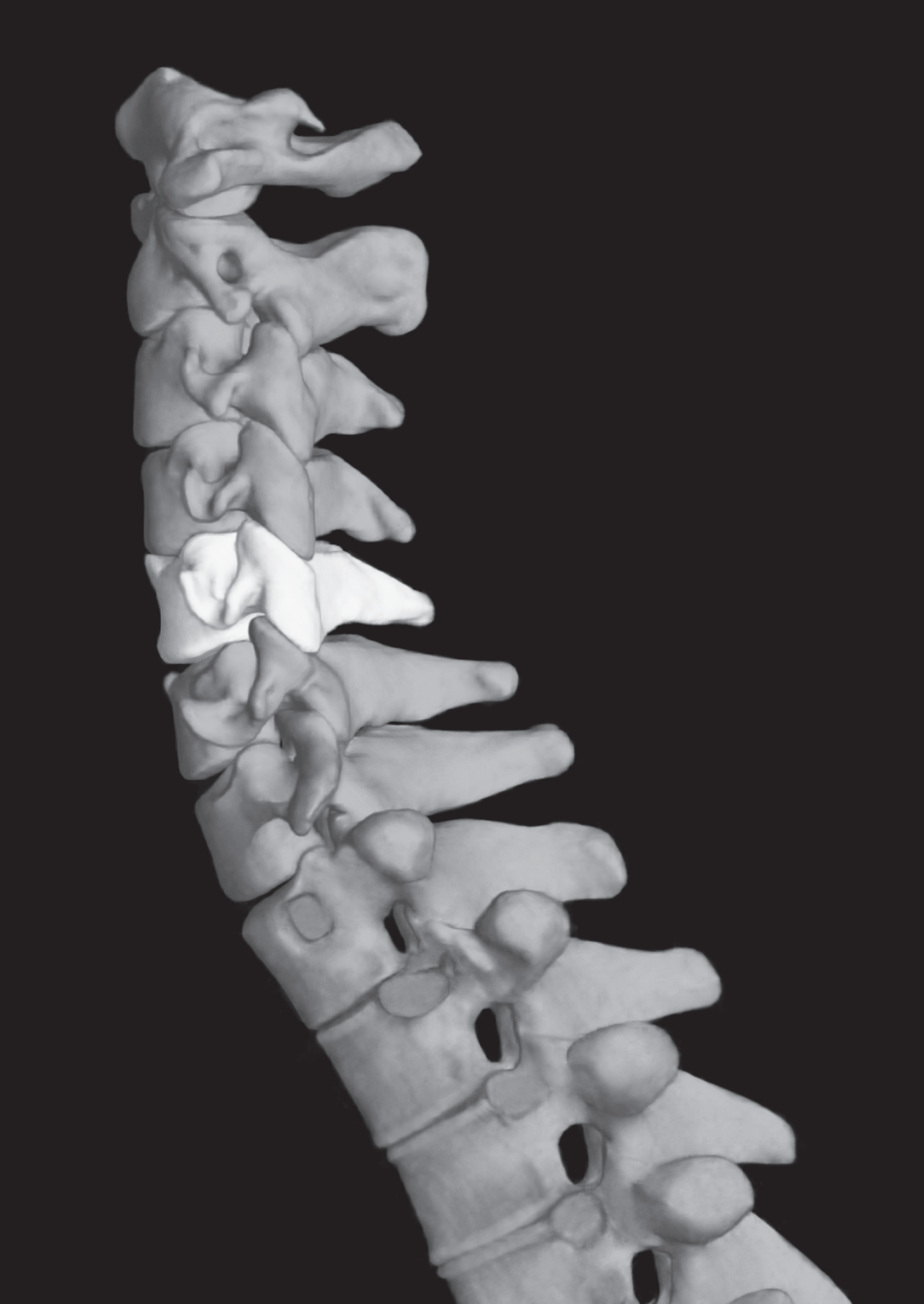
23. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009; 69: 179–183.

24. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N et al. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009; 69: 555–559.

25. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008; 63: 160–164.

26. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008; 18: 1696–1702.

27. Bazzocchi A, Spinnato P, Albisinni U, Battista G, Rossi C, Guglielmi G. A careful evaluation of scout CT lateral radiograph may prevent unreported vertebral fractures. *Eur J Radiol*. 2012; 81: 2353–2357.



Chapter 5

Prevalent vertebral fractures on chest CT: higher risk for future hip fracture

Buckens CF
de Jong PA
Mali WP
Verhaar HJ
van der Graaf Y
Verkooijen HM

J Bone Miner Res. 2014 Feb;29(2):392-8

Abstract

Prevalent vertebral fractures on chest CT: higher risk for future hip fracture

Subclinical or undiagnosed vertebral fractures on routine chest computed tomography (CT) may be useful for detecting patients at increased risk of future hip fractures who might benefit from preventive interventions.

We investigated whether prevalent vertebral fractures on routine chest CT are associated with future hip fractures. From a source population of 5679 patients ≥40 years old undergoing chest CT in one of three Dutch hospitals between 2002 and 2005, patients hospitalized for hip fractures (n=149) during a median follow-up of 4.4 years were identified. Following a case-cohort design, a random sample of 576 patients was drawn from the source population and added to the cases. In this group, the presence and severity of vertebral fractures was determined using semiquantitative vertebral fracture assessment and multivariate case-cohort appropriate Cox modeling.

We found that cases were older (69 versus 63 years) and more often female (48% versus 38%) than the source population. Compared with those with no fracture, patients with any vertebral fracture had triple the risk of future hip fracture (age- and gender-adjusted hazard ratio (HR)=3.1, 95% confidence interval (CI) 2.1-4.7). This HR rose to 3.8 (CI 2.6-5.6) if mild fractures were discounted. Future fracture risk increased significantly with increasing severity of vertebral fracture status: from mild (HR=2.4, CI 1.5-3.7) and moderate (HR=4.8, CI 2.5-9.2) to severe (HR=6.7, CI 2.9-15.5). The same was true for having higher cumulative fracture grades: 1 to 3 (HR=2.7, CI 1.8-4.1), 4 to 6 (HR=4.8, CI 2.2-10.5), or ≥7 (HR=11.2, CI 3.7-34.6).

In conclusion, prevalent vertebral fractures on routine clinical chest CT are associated with future hip fracture risk.

Introduction

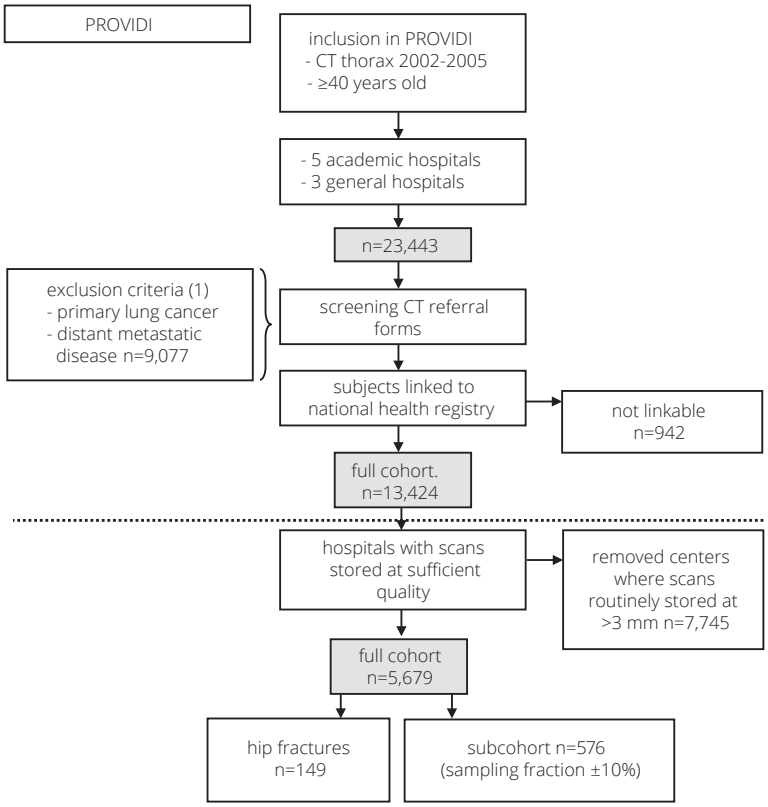
Vertebral fractures are the most common type of osteoporosis-associated fragility fractures[1] and are associated with increased morbidity such as restrictive pulmonary disease, chronic back pain, loss of independence, and reduced quality of life. Moreover, patients with vertebral fractures have an increased mortality[2] and are more likely to suffer future vertebral and nonvertebral fractures.[3,4] This supports their role as early manifestations of the decreased bone mineral density (BMD) and deteriorated bone microarchitecture that characterize osteoporosis,[5] something reflected in current osteoporosis guidelines.[6]

Osteoporosis is currently diagnosed using dual-energy X-ray absorptiometry (DXA), which is used to estimate areal bone mineral density (BMD). However, the majority of osteoporotic fractures occur in patients with BMDs above the threshold for osteoporosis, making it an imperfect measure on which to base treatment decisions.[7] Currently available interventions for osteoporosis and increased fracture risk are fall prevention, lifestyle guidance, calcium/vitamin D supplementation, and (most importantly) antiresorptive medications, such as bisphosphonates, denosumab, selective estrogen-receptor modulators, hormonal therapy, and other bone anabolic agents such as strontium ranelate and exogenous PTH analogs.[8] These interventions are effective mainly at slowing disease progression rather than reversing it, arguably making early diagnosis and prevention of osteoporosis the best route toward reducing its mounting disease burden.[6,9]

Routine diagnostic imaging that happens to include the spine may provide opportunities for detection of subclinical or asymptomatic vertebral fractures in early stages of osteoporotic disease, without requiring any additional imaging. Chest computed tomography (CT), which is widely used, visualizes the spine and could be used to detect subclinical vertebral fractures,[10] complementing other risk-assessment tools by opportunistically identifying patients at risk of future fracture. Previous studies have shown that the prevalence of radiological vertebral fractures on chest CT is relatively high and that fewer than 15% of these vertebral fractures are reported,[11-15] the majority of which are clinically unknown.[15] Furthermore, recent research has shown that there is a high correlation between simple trabecular vertebral attenuation values on abdominal CT and DXA scores, and found that attenuation values may correlate better than DXA with prevalent fractures to boot.[16] This suggests that there may be potential gain in systematically assessing vertebral fractures on routine chest CT to identify patients who might benefit the most from early intervention.[9]

In this study, we investigate whether subclinical vertebral fractures on sagittal reconstruction of routine chest CT are associated with future hip fractures.

Figure 1. PROVIDI flowchart



Materials and Methods

PROVIDI

The present research was conducted in the context of the Prognostic Value of Unrequested Information on Diagnostic Imaging (PROVIDI) study. This multicenter study aims to establish the prognostic value of unrequested findings on chest CT in routine clinical care and was described in detail elsewhere.[17] Briefly, it includes all patients older than 40 years who underwent chest CT in one of eight participating Dutch hospitals between 2002 and 2005. As such, it contains a heterogeneous range of protocols and reconstruction formats, representing routine practice. Patients with primary lung cancer or distant metastatic disease of other origin were excluded because of their poor prognosis and the

very low likelihood that unrequested findings would alter their medical management. The institutional ethics committees of the participating hospitals approved the PROVIDI study and waived the need for informed consent.

Study population

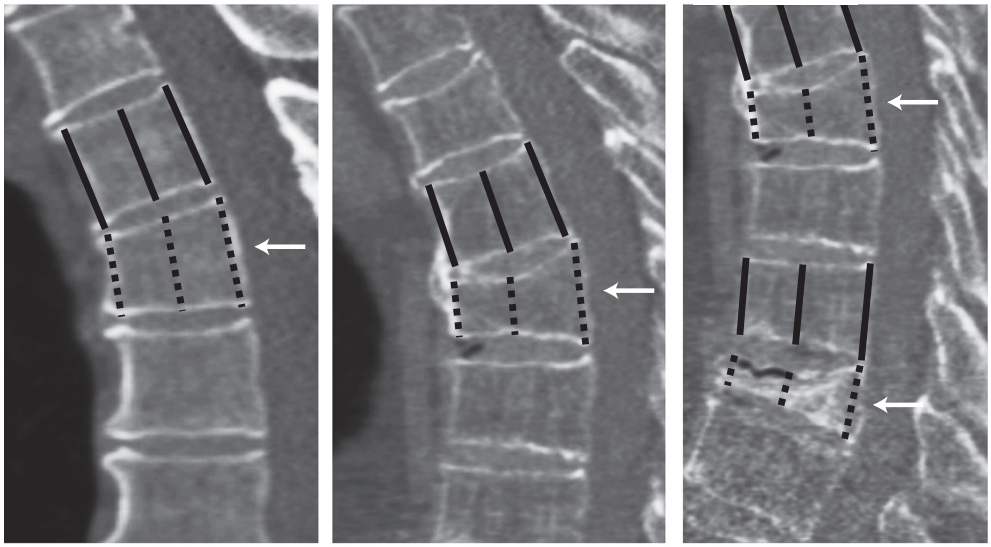
Of the 13,424 patients in the PROVIDI study, only patients from three of eight participating centers were included, as only the CTs from two academic centers and one peripheral center were deemed to be of sufficient quality to allow sagittal reconstruction. In the other hospitals, the slice thicknesses of the stored CTs was routinely higher than the <3 mm necessary for multiplanar reconstructions (reconstructions made from CT scans with a slice thickness <3 mm are not interpretable). For the current study, we further excluded patients with a musculoskeletal CT indication (n=174) from the overall cohort to ensure that the vertebral deformity observed was truly unrequested. The remaining source population consisted of 5679 patients, 3315 from the tertiary centers and 2364 from the peripheral hospital (Figure 1).

PROVIDI follows a case-cohort design. This efficient approach involves taking a completely random sample from the source population (termed the "subcohort") from which the distribution of variables of interest (vertebral fractures in this case) can be estimated. After a follow-up period, the case status is ascertained in all members of the source population. All cases identified are also assessed for the variables of interest. Correctly applied, this approach has been shown to reliably yield valid results,[18] comparable to complete cohort analysis, at a fraction of the effort and cost.

The subcohort was sampled so that it would be at least 10% of the source population (n=599, 10.5%) because sampling fractions above this proportion do not yield any appreciable increase in accuracy.[18] Twenty-three (3.8%) of these scans were not retrievable or had been stored in reconstructions above 3 mm or with a very restricted field of view (e.g. cardiac reconstruction only) and were excluded from analysis, resulting in a subcohort of 576 subjects (10% of the source population). Of these, 16 happened to be cases. Additional cases identified in the remainder of the cohort were added to this subcohort to form the case-cohort data set of 709.

When study subjects had more than one chest CT during follow-up, only the first CT was evaluated. CTs were made with a range of different systems from different vendors and were stored at a range of slice thicknesses, depending on local protocols. Patient characteristics, slice thickness, and the use of a contrast agent were extracted from chest CT reports by a

Figure 2. Illustration of vertebral deformity gradation on sagittal reformats of routine chest CT



A=mild fracture or deformity, grade 1 (arrow); B=moderate fracture, grade 2 (arrow); C=severe fracture in the lower part of the figure, grade 3 (arrow), a moderate fracture in the upper part of the figure, grade 2 (arrow); height measurements shown for fractured vertebrae (dashed lines) and for adjacent unfractured reference vertebrae (solid lines)

research physician, who also recorded the diagnostic chest CT indication.

Outcome

Cases were defined as individuals hospitalized for a hip fracture after the CT scan. Incident hip fractures were ascertained through linkage of the PROVIDI cohort with the national death registry and the national registry of hospital admittance and discharge diagnoses from January 2002 to the end of December 2008. This national database collects information from all hospitals in the Netherlands on every admittance, recording admittance and discharge diagnoses as well as other demographic information. This linkage was performed using a validated probabilistic method.[19,20] These databases employ the ICD-9 CM and ICD-10 classification systems and include entries for each hospital admission, readmission, and transfer. We included codes ICD-9-CM 820 (fracture of the neck of the femur), 821.0, and 8.21.1 (unspecified femur fracture) and ICD-10 S72.0-S72.1 (hip fracture). In cases of multiple valid endpoints, the date of first hospitalization was used.

Table 1. Baseline characteristics of hip fracture cases and the random subcohort from the included three hospitals

variable	subcohort (n=576)	cases (n=149)
gender (% female)	38%	48%
age (mean in years; range)	63 (40-92)	69 (41-92)
follow-up (median in days; interquartile range)	1622 (1478-1800)	1567 (979-1792)
academic center (% from a tertiary center)	58%	52%
indication		
lung indication	153 (27%)	41 (28%)
cardiovascular indication	178 (31%)	47 (32%)
malignancy related indication	89 (15%)	35 (23%)
other indication	156 (27%)	26 (17%)
percentage of patients with vertebral fracture	152 (26%)	84 (56%)
fracture (worst)		
no fracture	425(74%)	65 (44%)
worst fracture grade: 1 (mild)	104 (18%)	46 (31%)
worst fracture grade: 2 (moderate)	30 (5%)	23 (15%)
worst fracture grade: 3 (severe)	17 (3%)	15 (10%)
fracture (cumulative)		
no fracture	425 (74%)	65 (44%)
cumulative fracture grade: 1-3	127 (22%)	60 (40%)
cumulative fracture grade: 4-6	17 (3%)	16 (11%)
cumulative fracture grade: ≥7	7 (1%)	8 (5%)

Vertebral measurements

Genant's semiquantitative vertebral fracture assessment (VFA) method[10] was used to identify and classify vertebral fractures. This method identifies fractures according to the height loss of the vertebral body (viewed sagitally) and categorizes them according to worst height loss (either anterior–middle or posterior), relative to adjacent normal unfractured vertebrae. The three fracture grades are: mild grade 1 fractures (height loss 20% to 25%), moderate grade 2 fractures (25% to 40%), and severe grade 3 fractures with height loss of more than 40% (Figure 2). Deformities that seemed nonfractural in origin (e.g. Schmörl's nodes or congenital anomalies) were not counted as fractures. The vertebrae were assessed on sagittal reconstructions of the axially stored chest CTs, at or around the midsagittal

Table 2. Hazard ratios for a hip fracture associated with vertebral fracture measures

determinant	crude	adjusted*
presence of vertebral fracture		
no (n=65)	1	1
yes (n=84)	3.8 (2.6-5.6)	3.1 (2.1-4.7)
worst fracture grade		
none (n=65)	1	1
mild (20% to 25%) (n=46)	2.9 (1.9-4.5)	2.4 (1.5-3.7)
moderate (25% to 40%) (n=23)	5.2 (2.8-9.5)	4.8 (2.5-9.2)
severe (>40%) (n=15)	10.2 (4.6-22.9)	6.7 (2.9-15.5)
cumulative fracture grade		
0 (n=65)	1	1
1 to 3 (n=60)	3.2 (2.1-4.8)	2.7 (1.8-4.1)
4 to 6 (n=16)	6.2 (3.0-13.0)	4.8 (2.2-10.5)
7 (n=8)	17.0 (5.3-54.8)	11.2 (3.7-34.6)

data presentd as: hazard ratio (95% confidence interval); * adjusted for age and gender

point, with the rater free to adjust the window level, orientation, and slice thickness of the reconstruction as desired.

The following patient-level VFA parameters were calculated from these basic fracture definitions: fracture presence, worst fracture grade, and cumulative fracture grade. The presence of fracture was scored yes if there was any fracture visible at all. The worst fracture grade was defined as the grade of the worst fracture visible (either 0, 1, 2, or 3). The cumulative fracture grade was the sum of all recorded fracture grades. This sum was then categorized into one of three categories: 1 to 3, 4 to 6, or ≥7[21]. Figure 2c shows an illustrative example. This patient did have at least one fracture, had a worst fracture grade of 3, and adding up the fracture grades of the two visible fractures results in a cumulative fracture grade of 5 (placing in the 4 to 6 category). In a sample of 50 PROVIDI CTs, we found good agreement between four observers of varying experience for these three measures (Cohen's kappa 0.56 to 0.81).

A second fracture classification was constructed that discounted mild fractures, as these have been found to have a lower reproducibility in some other studies.[22] In this scheme, fracture was scored as present only if at least one moderate or severe fracture was present, and mild fractures were recoded as nonfractured. Cumulative fracture grade was omitted from this analysis as redefining.

Table 3: Hazard ratios for a hip fracture associated with moderate and severe fractures only

determinant	crude	adjusted*
presence of vertebral fracture		
no (n=111)	1	1
yes (n=38)	4.7 (3.3-6.9)	3.8 (2.6-5.6)
worst fracture grade		
none (n=111)	1	1
moderate (25% to 40%) (n=23)	3.8 (2.5-6.1)	3.5 (2.2-5.5)
severe (>40%) (n=15)	7.5 (4.3-12.9)	4.6 (2.6-8.0)

data presentd as: hazard ratio (95% confidence interval); * adjusted for age and gender

Statistical analysis

Cox proportional hazard modeling was used to determine the association between our VFA parameters and the risk of future hip fractures. To account for the case-cohort approach, weighting was applied to the analysis as described by Prentice.[23] Univariate analysis was used to estimate the crude, unadjusted association between VFA parameters and hip fracture. To identify potentially relevant covariates, we evaluated age, scan indication, gender, and the slice thickness together in a multivariate model. Age (in years) and slice thickness of the stored axial reconstructions (in millimeters) were modeled continuously, whereas gender and scan indication (either lung, cardiovascular, malignancies, or other) were handled as factors. Using Wald statistic p-values with a threshold of 0.15, we selected the covariates to be included in calculating the multivariate-adjusted associations between VFA measures and hip fracture risk. All analyses were performed using R 2.12.2 (the R foundation, Vienna, Austria).

Results

Of the 5679 patients in the source population, 149 patients suffered a hip fracture during a median follow-up of 4.4 years (interquartile range 3.6 to 5.0). Cases were older than the source population (mean age 69 versus 63 years) and more often female (48% versus 38%). A majority (56%) of cases had at least one prevalent vertebral fracture; in the subcohort, prevalent vertebral fractures were detected in 26% of the individuals. There were more severe (grade 3), moderate (grade 2), and mild (grade 1) fractures in the case group (10% versus 3%, 15 versus 5%, and 31% versus 18%, respectively). Similarly, higher cumulative fracture grades were more prevalent among cases (Table 1).

During the assessment on which covariates to include in the adjusted model, neither slice thickness (p=0.60) nor CT indication (lowest p=0.46) were associated with future fracture risk, whereas gender (p=0.03) and age (p < 0.01) were. Hence, age- and gender-adjusted hazard ratios (HR_{adj}) are presented alongside the crude HRs.

All three VFA measures (presence of a fracture, worst fracture grade, and cumulative fracture grade) were significantly associated with future fracture risk after adjustment for age and gender. After adjustment for age and gender, the presence of any vertebral fracture was associated with future hip fracture by a HR_{adj} of 3.1 (95% CI 2.1–4.7) (Table 2). For men, the HR_{adj} was 2.8 (CI 1.7–4.8), whereas it was 3.5 (CI 1.9–6.4) for women. According to this parameter, the patient presented in Figure 2c, with a fracture present, would thus have a 3.1-fold increased fracture risk compared with a patient of the same gender and age without a vertebral fracture on chest CT.

Future fracture risk increased with increasing vertebral fracture severity. Having a mild fracture grade was associated with a 2.4-fold increased future fracture risk (HR_{adj} 2.4, 95% CI 1.5–3.7), a moderate fracture grade with a 4.8-fold increased risk (HR_{adj} 4.8, 95% CI 2.5–9.2), and a severe fracture with a 6.7-fold increased risk (HR_{adj} 6.7, 95% CI 2.9–15.5). For example, the patient from Figure 2c, with a severe worst fracture grade, would have a 6.7-fold higher risk of suffering a hip fracture compared with a patient of similar age and gender with no visible vertebral fractures (worst fracture grade 0).

For cumulative fracture grade, compared with a score of 0, a score of 1 to 3 conferred a HR_{adj} of 2.7 (95% CI 1.8–4.1), a score of 4 to 6 conferred a HR_{adj} of 4.8 (95% CI 2.2–10.5), and a score of ≥7 conferred a HR_{adj} of 11.2 (95% CI 3.7–34.6). For example, the patient in Figure 2c would have a cumulative fracture grade of 5, implying a risk 4.8 times greater for future hip fracture compared with a patient of the same gender and similar age with no vertebral fractures on chest CT (cumulative fracture grade 0).

If mild fractures were discounted and only moderate and severe fractures considered, 26% of the cases had at least one moderate or severe fracture compared with 8% of the subcohort. The hazard rate for hip fractures was higher among this group, both before (HR=4.7; CI 3.3–6.9) and after adjustment for age and gender (HR_{adj}=3.8; CI 2.6–5.6, Table 3).

Discussion

In this study, we demonstrated that vertebral fracture assessment on sagittal reformats of routine clinical chest CT are associated with future hip fractures and that this risk

increases with increasing vertebral fracture severity and cumulative fracture burden. This finding could be put to use by opportunistically assessing routine clinical chest CT on vertebral fracture status to identify patients who could potentially benefit from fracture prevention.[24,25]

We found similar HRs for the different worst-fracture grades and the cumulative-fracture grade categories. The cumulative-fracture grade is distinct from the worst-fracture grade measure in that a high value may consist of several mild fractures, making it more reflective of the overall fracture burden across the whole visible spine, but this distinction does not seem to result in a markedly different future fracture risk, implying that severe fractures have a similar influence on fracture risk to multiple mild fractures.

We found higher HRs for fracture presence when mild fractures were discounted. Mild fractures are considered by many to be somewhat more subjective and their clinical relevance has been questioned; it has been suggested that mild fractures may represent degenerative changes rather than frank fractures.[22] As such, they may represent an earliest detectable stage of bone demineralization, preceding more severe deformity. Here, we have shown that even mild fractures are significantly and independently associated with future hip fracture. Computer-assisted VFA, such as is already extensively implemented in lateral spinal densitometry,[26] could help to reduce the inter-observer variability of mild fractures, making them more reliably identifiable. The high proportion of patients scanned with a mild fracture we observed here might result in a low specificity of any fracture-prediction models including them. Incorporating mild fractures into existing fracture-risk assessment models such as FRAX (a patient questionnaire used to assess fragility fracture risk) would introduce more variables (beyond age and gender, which we included here) on which to base the final risk stratum, hopefully reducing this potential disadvantage.

Previous studies have reported on the predictive potential of vertebral morphometry using either radiographs or densitometry acquired in selected (typically high-risk) populations, and their outcomes are largely in line with our findings.[21,27-36] Siris and colleagues[21] found relative risks for incident nonvertebral fractures between 1.6 and 2.1 for a similar set of VFA measures to our own, amongst osteoporotic and osteopenic postmenopausal women. In a general population, Van Staa and colleagues[4] have shown that clinically apparent vertebral fractures increase the risk of future fractures elsewhere, with a risk ratio of 2.9. In contrast to these important studies, we investigated a routine clinical population for prevalent vertebral fractures, the majority of which are presumably clinically silent rather than apparent,[15] showing that these subclinical fractures are also associated with future hip fracture risk outside tightly controlled experimental settings. We also limited our

outcome definition to only hip fractures, rather than fractures at any nonvertebral site, perhaps explaining the higher fracture risk observed in this manuscript.

There are indications that increasing vertebral fracture severity correlates directly with an impaired bone microarchitecture.[37] As an osteoporosis biomarker, radiological vertebral fractures may thus be independent from areal BMD (DXA) or patient history (such as FRAX), a fact reflected in their increasingly central role in fracture-prevention guidelines.[6] By serendipitously identifying a different group of patients at risk of hip fracture (those coming for unrelated chest CT) and visualizing a suitably early manifestation of osteoporosis may make it able to complement the aforementioned approaches[38] at minimal additional cost. This could lead to more health utility being derived from routine clinical CT and a reduction of the high excess costs imposed by prevalent fractures.[39] Conceivably, opportunistically identified, clinically silent vertebral fractures could be used as an indication for further risk assessment using FRAX or DXA, or a more complete risk-assessment tool that incorporates radiological fracture status into risk-prediction models could be developed.

Recent work correlating simple attenuation values of a trabecular region of high lumber vertebra on abdominal CT has not only found a strong association between this simple measure and DXA but also found that this measure might be better able to distinguish between patients with prevalent vertebral fractures than DXA.[16] In particular, they noted that many patients with prevalent fractures had degenerative osseous changes that may have resulted in spurious DXA scores. This line of thinking has already shown the potential of underutilized information in routine imaging.

Determining the efficacy of using routine clinical chest CT to identify at-risk patients in terms of fractures prevented cannot be addressed directly with this data but deserves further investigation. Available evidence seems to indicate that the benefits of treatment amongst patients with a vertebral fracture might be comparable in magnitude to the established efficacy observed amongst patients selected for low BMD.[24,25] Intervention in patients who have suffered a minimal trauma fracture has also been shown to be effective in preventing further fractures.[40]

A limitation of this study may be that there is no additional clinical information (e.g. a FRAX score) available beyond age and gender that would allow CT-based VFA to be compared with clinical risk assessment, or for incremental value assessment. However, this CT-only setting is in keeping with radiological practice, where access to comprehensive patient histories is often limited and seldom sought.

Similarly, the design of PROVIDI precluded any correlation with prior DXA scanning that a subset of patients may have coincidentally undergone. There was also no information available on which patients were already receiving fracture-risk reduction treatment and which proportion of the prevalent vertebral fractures had already been identified. Available literature indicates that the former may be the case for around half of patients with vertebral fractures.[41] whereas the latter applied only to a minority of patients.[15] Such a proportion suggests that the potential for gain exists, but the impact of identifying at-risk patients through this approach remains to be quantified. By using stored chest CTs, our analysis was limited to the thoracic and upper lumbar spine. Because these are considered the principal sites for osteoporotic vertebral fractures,[42] the risk that additionally visualizing the lower lumbar spine would substantially change the observed associations is minimal.

In conclusion, vertebral fracture assessment measures obtained from sagittal reformats of routine clinical chest CT are associated with future hip fractures, with increasing risk accompanying increasing fracture severity and cumulative fracture burden.

References

1. Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet*. 2002 May 18;359(9319):1761–7.
2. Kado DM, Browner WS, Palermo L, Nevitt MC, Genant HK, Cummings SR. Vertebral fractures and mortality in older women: a prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med*. 1999 Jun 14;159(11):1215–20.
3. Van Brussel MS, Lems WF. Clinical relevance of diagnosing vertebral fractures by vertebral fracture assessment. *Curr Osteoporos Rep*. 2009 Sep;7(3):103–6.
4. Van Staa TP, Leufkens HGM, Cooper C. Does a fracture at one site predict later fractures at other sites? A British cohort study. *Osteoporos Int*. 2002 Aug;13(8):624–9.
5. Osteoporosis prevention, diagnosis, and therapy. *NIH Consens Statement*. 2000 Mar 27;17(1):1–45.
6. Lim LS, Hoeksema LJ, Sherin K. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. *Am J Prev Med*. 2009 Apr;36(4):366–75.
7. Siris ES, Chen Y-T, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE, Berger ML. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med*. 2004 May 24;164(10):1108–12.
8. Cranney A, Guyatt G, Griffith L, Wells G, Tugwell P, Rosen C. Meta-analyses of therapies for postmenopausal osteoporosis. IX: Summary of meta-analyses of therapies for postmenopausal osteoporosis. *Endocr Rev*. 2002 Aug;23(4):570–8.
9. Kern LM, Powe NR, Levine MA, Fitzpatrick AL, Harris TB, Robbins J, Fried LP. Association between screening for osteoporosis and the incidence of hip fracture. *Ann Intern Med*. 2005 Feb 1;142(3):173–81.
10. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993 Sep;8(9):1137–48.
11. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009 Jan;69(1):179–83.
12. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N, Gavelli G. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009 Mar;69(3):555–9.
13. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008 Feb;63(2):160–4.
14. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008 Aug;18(8):1696–702.
15. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ. Unreported Vertebral Body Compression Fractures at Abdominal Multidetector CT. *Radiology*. 2013 Feb 28;

16. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med*. 2013 Apr 16;158(8):588–95.

17. Gondrie MJA, Mali WPTM, Buckens CFM, Jacobs PCA, Grobbee DE, van der Graaf Y. The PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) Study: rationale and design. *Eur J Epidemiol*. 2010 Oct;25(10):751–8.

18. Onland-Moret NC, van der A DL, van der Schouw YT, Buschers W, Elias SG, van Gils CH, Koerselman J, Roest M, Grobbee DE, Peeters PHM. Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol*. 2007 Apr;60(4):350–5.

19. De Bruin A, Karduan J, Gast F, De Bruin E, van Sijl M, Verweij G. Record linkage of hospital discharge register with population register: experiences at Statistics Netherlands. *Stat J UN Econ Comm Eur*. 2004;1(21):23–32.

20. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011 May;64(5):565–72.

21. Siris ES, Genant HK, Laster AJ, Chen P, Misurski DA, Krege JH. Enhanced prediction of fracture risk combining vertebral fracture status and BMD. *Osteoporos Int*. 2007 Jun;18(6):761–70.

22. Jiang G, Eastell R, Barrington NA, Ferrar L. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int*. 2004 Nov;15(11):887–96.

23. Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986;73(1):1–11.

24. Kanis JA, Barton IP, Johnell O. Risedronate decreases fracture risk in patients selected solely on the basis of prior vertebral fracture. *Osteoporos Int*. 2005 May;16(5):475–82.

25. Delmas PD, Genant HK, Crans GG, Stock JL, Wong M, Siris E, Adachi JD. Severity of prevalent vertebral fractures and the risk of subsequent vertebral and nonvertebral fractures: results from the MORE trial. *Bone*. 2003 Oct;33(4):522–32.

26. Kim YM, Demissie S, Genant HK, Cheng X, Yu W, Samelson EJ, Kiel DP, Bouxsein ML. Identification of prevalent vertebral fractures using CT lateral scout views: a comparison of semi-automated quantitative vertebral morphometry and radiologist semi-quantitative grading. *Osteoporos Int*. 2012 Mar;23(3):1007–16.

27. Crimi A, Loog M, de Bruijne M, Nielsen M, Lillholm M. Shape-based assessment of vertebral fracture risk in postmenopausal women using discriminative shape alignment. *Acad Radiol*. 2012 Apr;19(4):446–54.

28. Diacinti D, Pisani D, Barone-Adesi F, Del Fiacco R, Minisola S, David V, Aliberti G, Mazzuoli GF. A new predictive index for vertebral fractures: the sum of the anterior vertebral body heights. *Bone*. 2010 Mar;46(3):768–73.

29. Lillholm M, Ghosh A, Pettersen PC, de Bruijne M, Dam EB, Karsdal MA, Christiansen C, Genant HK, Nielsen M. Vertebral fracture risk (VFR) score for fracture prediction in postmenopausal women. *Osteoporos Int*. 2011 Jul;22(7):2119–28.

30. Loog M, de Bruijne M. Discriminative shape alignment. *Inf Process Med Imaging*. 2009;21:459–66.

31. Pettersen PC, de Bruijne M, Chen J, He Q, Christiansen C, Tankó LB. A computer-based measure of irregularity in vertebral alignment is a BMD-independent predictor of fracture risk in postmenopausal women. *Osteoporos Int*. 2007 Nov;18(11):1525–30.

32. Ruyssen-Witrand A, Gossec L, Kolta S, Dougados M, Roux C. Vertebral dimensions as risk factor of vertebral fracture in osteoporotic patients: a systematic literature review. *Osteoporos Int*. 2007 Sep;18(9):1271–8.

33. Baker-LePain JC, Luker KR, Lynch JA, Parimi N, Nevitt MC, Lane NE. Active shape modeling of the hip in the prediction of incident hip fracture. *J Bone Miner Res*. 2011 Mar;26(3):468–74.

34. Gregory JS, Stewart A, Undrill PE, Reid DM, Aspden RM. Bone shape, structure, and density as determinants of osteoporotic hip fracture: a pilot study investigating the combination of risk factors. *Invest Radiol*. 2005 Sep;40(9):591–7.

35. Wustrack R, Seeman E, Bucci-Rechtweg C, Burch S, Palermo L, Black DM. Predictors of new and severe vertebral fractures: results from the HORIZON Pivotal Fracture Trial. *Osteoporos Int*. 2012 Jan;23(1):53–8.

36. Black DM, Arden NK, Palermo L, Pearson J, Cummings SR. Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures. Study of Osteoporotic Fractures Research Group. *J Bone Miner Res*. 1999 May;14(5):821–8.

37. Genant HK, Delmas PD, Chen P, Jiang Y, Eriksen EF, Dalsky GP, Marcus R, San Martin J. Severity of vertebral fracture reflects deterioration of bone microarchitecture. *Osteoporos Int*. 2007 Jan;18(1):69–76.

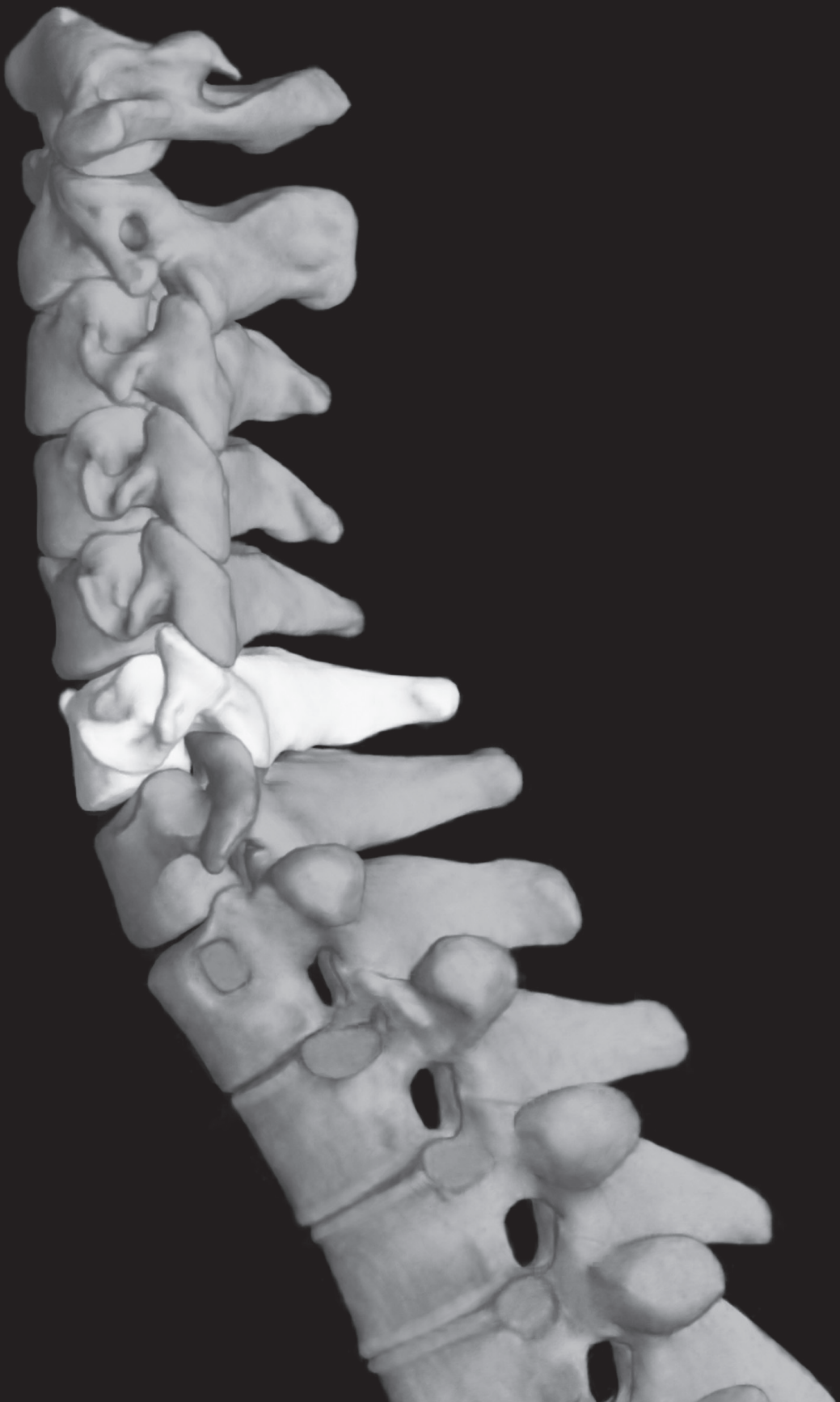
38. Steurer J, Haller C, Häuselmann H, Brunner F, Bachmann LM. Clinical value of prognostic instruments to identify patients with an increased risk for osteoporotic fractures: systematic review. *Plos One*. 2011;6(5):e19994.

39. Hopkins RB, Tarride JE, Leslie WD, Metge C, Lix LM, Morin S, Finlayson G, Azimaee M, Pullenayegum E, Goeree R, Adachi JD, Papaioannou A, Thabane L. Estimating the excess costs for patients with incident fractures, prevalent fractures, and nonfracture osteoporosis. *Osteoporos Int*. 2013 Feb;24(2):581–93.

40. Lih A, Nandapalan H, Kim M, Yap C, Lee P, Ganda K, Seibel MJ. Targeted intervention reduces refracture rates in patients with incident non-vertebral osteoporotic fractures: a 4-year prospective controlled study. *Osteoporos Int*. 2011 Mar;22(3):849–58.

41. Bazzocchi A, Spinnato P, Albisinni U, Battista G, Rossi C, Guglielmi G. A careful evaluation of scout CT lateral radiograph may prevent unreported vertebral fractures. *Eur J Radiol*. 2012 Sep;81(9):2353–7.

42. Skowrońska-Jóźwiak E, Płudowski P, Karczmarewicz E, Lorenc R, Lewiński A. Identification of vertebral deformities in the Polish population by morphometric X-ray absorptiometry - results of the EPOLOS study. *Endokrynol* Pol. 2009 Apr;60(2):68–75



Chapter 6

Opportunistic screening for osteoporosis on routine computed tomography? An external validation study

Buckens CF
Dijkhuis G
de Keizer B
Verhaar HJ
de Jong PA

Eur Radiol. 2015 Jan 17 (Epub ahead of print)

Abstract

Opportunistic screening for osteoporosis on routine computed tomography? An external validation study

Objectives

Opportunistic screening for osteoporosis using computed tomography (CT) examinations that happen to visualise the spine can be used to identify patients with osteoporosis. We sought to verify the diagnostic performance of vertebral HU measurements on routine CT examinations for diagnosing osteoporosis in a separate, external population.

Methods

Consecutive patients who underwent a CT examination of the chest or abdomen and who had also received a Dual Energy X-ray Absorptiometry (DXA) test were retrospectively included. CTs were evaluated for vertebral fractures and vertebral attenuation (density) values were measured. Diagnostic performance measures and the area under the receiver operator characteristics curve (AUC) for diagnosing osteoporosis were calculated.

Results

302 patients with a mean age of 57.9 years were included of which 82 (27%) had osteoporosis according to DXA and 65 (22%) had vertebral fractures. The diagnostic performance for vertebral HU measurements was modest, with a maximal AUC of 0.74 (0.68-0.80). At that optimal threshold the sensitivity was 62% (51-72%) and the specificity was 79% (74-84%).

Conclusions

We confirmed that simple trabecular vertebral density measurements on routine CT contain diagnostic information related to bone mineral density as measured by DXA, albeit with substantially lower diagnostic accuracy than previously reported

Introduction

It has recently been suggested that opportunistic screening for osteoporosis using routine computed tomography (CT) examinations that happen to visualise the spine can be used to identify patients with osteoporosis.[1] Osteoporotic fractures are a major contributor to late life morbidity and mortality and impose a substantial societal cost. Despite the availability of treatments of proven efficacy, there is room for improving the treatment rates at the time of and after major fragility fractures.[2-5]

Dual Energy X-ray Absorptiometry (DXA) is widely used tool in assessing osteoporosis. The widespread employment of CT scanning in the course of routine care can be used to opportunistically screen populations for osteoporosis outside of any existing screening programs. Risk assessment tools that incorporate multiple clinical parameters such as the World health Organization's FRAX tool are increasingly commonly used to identify patients who may be at an increased fracture risk. Most CT examinations include all or part of the spine, providing the opportunity to measure vertebral attenuation values of the trabecular regions of vertebral bodies expressed by Hounsfield Units (HU) potentially providing a free source of information reflecting bone mineral density (BMD) in a distinct population compared to that are currently being considered for DXA, but only partially overlapping with that population.

The vertebral attenuation values of the trabecular (non-cortical) regions of vertebral bodies, as expressed by the Hounsfield Units (HU) can be extracted from CT.[1] Trabecular bone is preferentially affected by osteoporosis, particularly in the early phases of the disease process.[6] Additionally, vertebral compression fractures can also be visualized on CT.[7,8] Following data from pragmatic prognostic studies, these now contribute directly to treatment decisions in new guidelines and form part of the indication to begin treatment, along with clinical history, DXA T-score and laboratory evaluations.[9,10] This has led to the inclusion of a lateral view to standard DXA assessment so that vertebral fracture assessment is also possible on DXA.

A threshold of 110 HU for lumbar vertebra 1 (L1) or 115 HU for thoracic vertebra 12 (Th12) was proposed as a cut-off yielding high specificity[11] for identifying patients at risk of osteoporosis and potentially in need of further screening and treatment.

Before further large outcome studies in clinical or screening settings can be pursued it is crucial to determine the external validity of this approach.[12] The objective of this study was to evaluate the diagnostic performance of vertebral HU measurements on routine

CT examinations for diagnosing osteoporosis in an external population. For this we have included a population separate to the initial sample in which vertebral attenuation was investigated,[1] but one which remains within routine radiological practice: we included a population of convenience who underwent a CT exam of the chest or abdomen and who had also received a DXA test within 90 days.

Materials and Methods

Setting and patients

The University Medical Centre Utrecht Ethical Review Board approved this study and the need for informed consent was waived. Consecutive patients who underwent a CT examination of the chest or abdomen between 2005 and 2012 and who had also received a DXA test within 90 days (before or after the CT) were retrospectively included. CT examinations were acquired in the course of routine care using multidetector CT scanners (16-256 detector rows, Philips Medical Systems, Best, the Netherlands). Dual-energy X-ray absorptiometry (DXA, Hologic Discovery A, Hologic Inc, Bedford MA) was performed in the spine (L2-L4) and hips in the course of routine care. As per common practice, the cut-off for osteoporosis was set as having a T-score ≤-2.5 at any measured location, either in L2-L4 and/or a hip. The cut-off for osteopenia was ≤-1.

Measurement of vertebral fractures on CT

Sagittal CT reformats were evaluated for vertebral fractures (height loss ≥25% compared to an adjacent normal vertebra) according to Genant's semiquantitative Vertebral Fracture Assessment (VFA) method.[13,14] The reconstructions were assessed at or around the mid-sagittal point. Observers were free to scroll, adjust the orientation and window. This method on CT has previously been shown to have excellent reliability for the presence of a vertebral fracture.[7] One of two observers, either one board certified radiologist or one senior radiology resident with a special interest in musculoskeletal imaging performed the measurements. They were blinded to the DXA results.

Measurements of bone density on CT

CT attenuation values were measured by the same two observers on stored axial images within trabecular regions of the bodies of L1 or the nearest visible, unfractured, visually normal vertebra, as previously described.[1] If L1 was not visualized, Th12 was measured

instead. A single click and drag region of interest was placed in the anterior, upper portion of the body of the vertebra, inside trabecular bone. The region of interest was drawn so that it was as large as possible without intersecting the vertebral cortex and without including dense bone islands, hemangiomas or traversing vessels. The CT attenuation was measured using Hounsfield Units, with lower values representing lower bone mineral density. (Figure 1)

Statistical analysis

Sensitivity, specificity, positive and negative predictive value, accuracy and area under the receiver operator characteristics curve (AUC) was calculated for three different pre-selected thresholds (80, 110 and 160 HU) of mean HU for trabecular vertebral bone density for DXA-defined osteoporosis, as previously described.[1] We also determined the optimal HU threshold in this cohort, defined as the threshold yielding the maximum proportion

Figure 1. Axial image through the superior part of the vertebral body of L1 showing placement of a measurement region of interest placed within the trabecular bone

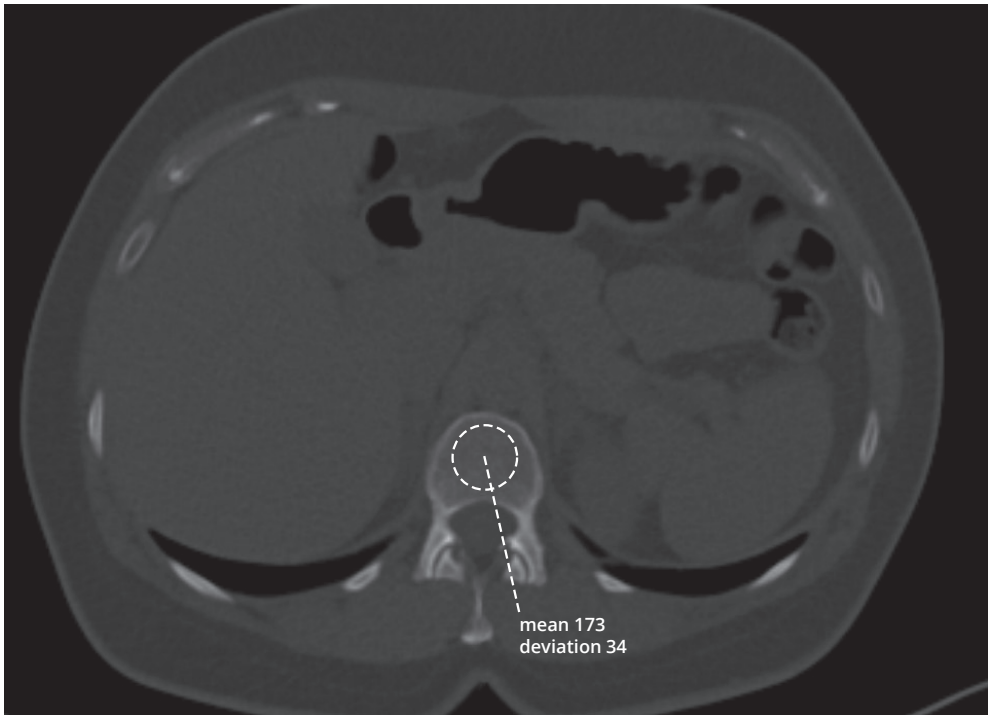


Table 1: Characteristics of patients included with and without osteoporosis, as defined by a DXA T-score ≤2.5

variable	description	osteoporosis	no osteoporosis
n	number	82	220
age	years (SD)	61 (16)	57 (15)
sex	male (%)	62 (32%)	72 (33%)
vertebra measured*	L1 (%)	72 (88%)	198 (90%)
anatomical area**	abdomen (%)	53 (65%)	136 (62%)
DXA area***	LWK(%)	58 (71%)	119 (54%)
T-score	mean (SD)	-3.2 (0.62)	-1.1 (0.96)
days	mean (SD)	30 (25)	40 (26)
fracture	fracture (%)	27 (33%)	38 (17%)
HU	mean (SD)	97 (40)	138 (50)
indication	inflammatory/autoimmune	10 (12%)	41 (19%)
	endocrine disorder	12 (15%)	63 (29%)
	fracture	28 (34%)	44 (20%)
	steroid therapy	9 (11%)	25 (11%)
	transplantation	10 (12%)	17 (8%)
	malignancy	3 (4%)	7 (3)
	other	10 (12%)	23 (10%)
$HU_{L1}80/HU_{Th12}/85$		33 (40%)	21 (10%)
$HU_{L1}110/HU_{Th12}/115$		56 (68%)	74 (34%)
$HU_{L1}160/HU_{Th12}/165$		75 (91%)	156 (71%)

* if first lumbar (L1) was not visualized or was fractured, the twelfth thoracic (Th12) was measured instead

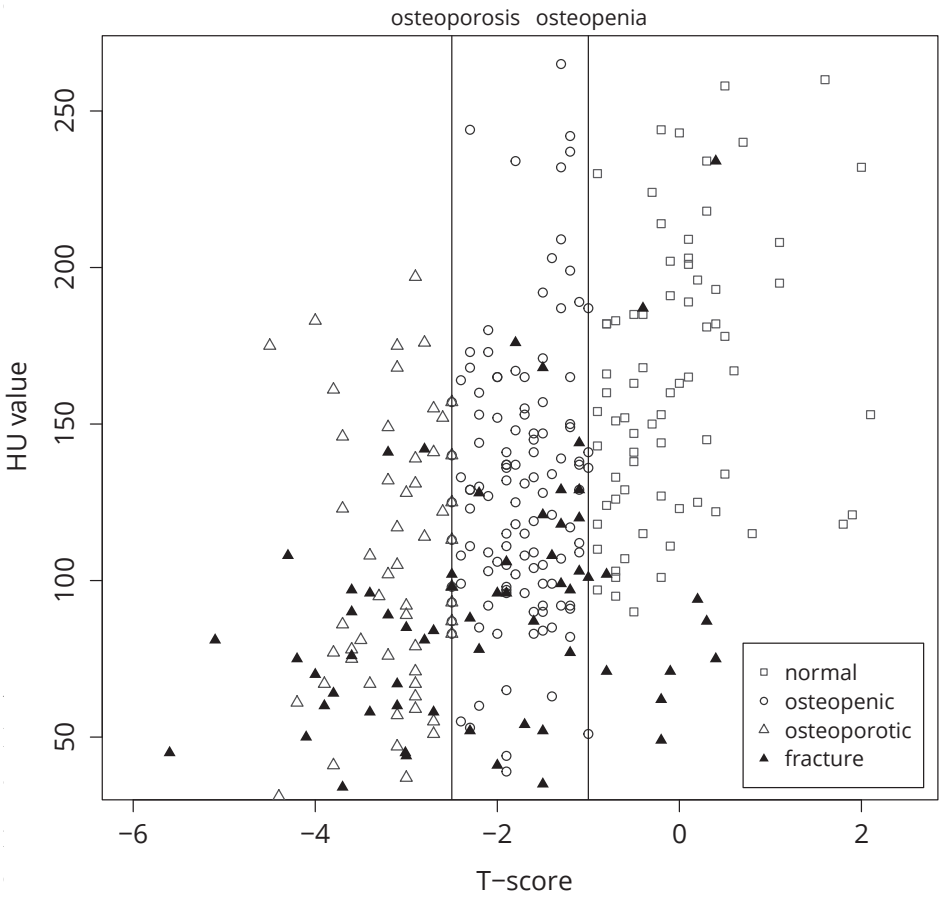
** thoracic and abdominal CTs were included

*** lowest areal bone density value obtained value used, either derived from lumbar spine or hip

of correct classifications (i.e. true positives and true negatives), from a univariate logistic regression model, where CT attenuation was set as the independent variable and DXA-defined osteoporosis the dependant variable. A multivariate logistic regression model was fitted to assess the value of adding age and gender to vertebral density on CT when predicting DXA-defined osteoporosis.

Subsequently these analyses were repeated after adding the variable presence of one or more vertebral fractures to the attenuation classifier at each threshold. In effect patients with a vertebral fracture were thus classified as having CT defined 'osteoporosis' regardless of their HU value. For example, a patient with a HU value at L1 of 150 HU would not be

Figure 2. Scatter plot of the vertebral densities in Hounsfield Units (HU) and DXA T-scores of the sample population



Normal (squares), osteopenic (circles) and osteoporotic (hollow triangles) individuals are displayed, along with those who had a vertebral fracture (solid triangles) on CT

DXA T-scores and vertebral density on CT were in turn set as the independent variables in two univariate models.

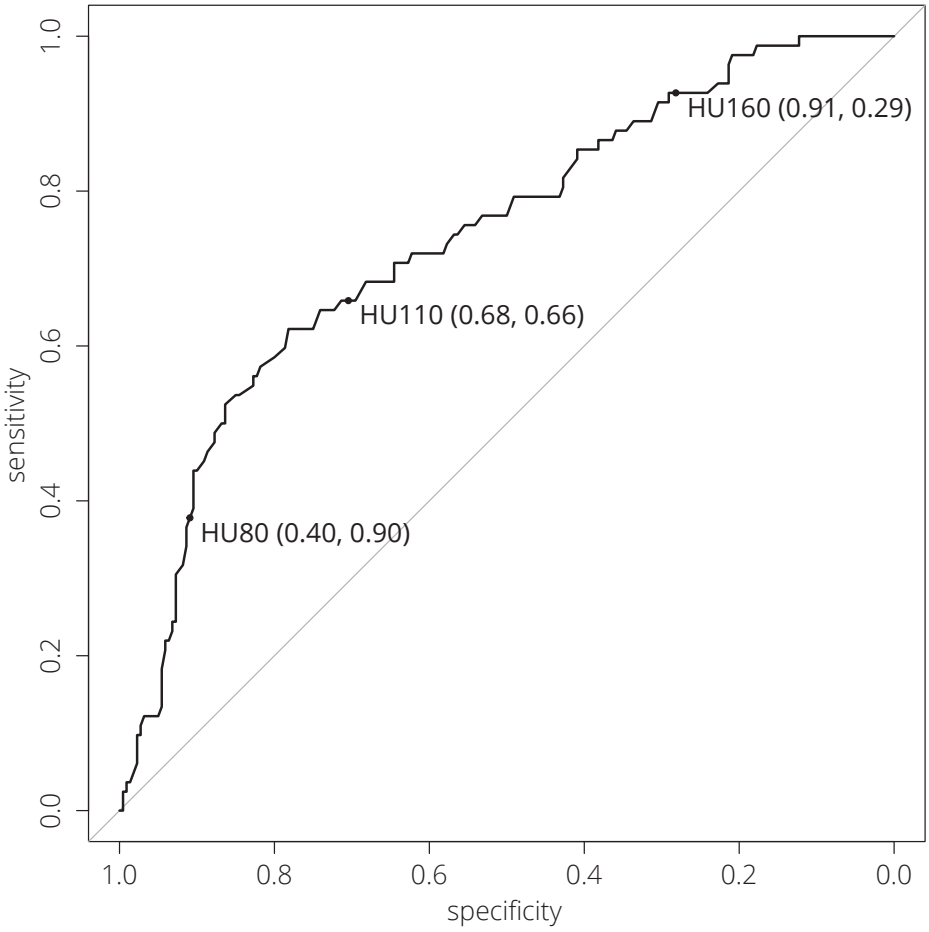
Logistic regression models were fit to ascertain the optimal performance and threshold of vertebral HU in our sample, and to examine the incremental diagnostic value of fracture status, age and gender in a multivariate model.

Table 2. Diagnostic accuracy of routine computed tomography for osteoporosis defined as DXA T-score <-2.5

outcome	determinants	threshold	sens	spec	accuracy	PPV	NPV	AUC
osteoporosis	vertebral density	$HU_{L1}80/HU_{Th12}85$	40% (30-51)	90% (86-94)	77% (71-82)	60% (44-76)	80% (77-84)	0.65 (0.59-0.71)
		$HU_{L1}110/HU_{Th12}115$	68% (59-78)	66% (60-73)	67% (60-74)	43% (35-52)	85% (80-90)	0.67 (0.61-0.73)
		$HU_{L1}160/HU_{Th12}165$	91% (84-98)	29% (23-35)	46% (39-52)	32% (29-36)	90% (80-98)	0.60 (0.56-0.65)
		optimal* $HU_{L1}99/HU_{Th12}104$	62% (51-72)	79% (73-84)	74% (67-81)	52% (41-62)	85% (80-89)	0.74 (0.68-0.80)
osteoporosis	vertebral density and vertebral fracture	$HU_{L1}80/HU_{Th12}85$ and/or fracture	56% (45-67)	79% (74-84)	73% (66-79)	50% (39-61)	83% (78-87)	0.68 (0.62-0.74)
		$HU_{L1}110/HU_{Th12}115$ and/or fracture	71% (61-80)	61% (55-67)	64% (57-71)	40% (33-47)	85% (79-90)	0.66 (0.60-0.72)
		$HU_{L1}160/HU_{Th12}165$ and/or fracture	91% (85-96)	27% (21-33)	44% (38-50)	32% (28-35)	89% (79-96)	0.59 (0.55-0.64)

sens=sensitivity; spec=specificity; PPV=positive predictive value; NPV=negative predictive value; AUC=area under the receiver operator characteristics curve; data are given with 95% confidence interval between brackets. Confidence intervals generated using 2000 stratified bootstrap replicates;
*optimal threshold is the threshold where the largest proportion of patients is correctly classified

Figure 3. Receiver operating characteristics curve for vertebral bone density model predicting DXA-assigned osteoporosis



Three pre-defined thresholds are marked along with the corresponding sensitivity and specificity.

Results

302 patients (98 males) with a mean (SD) age of 57.9 (15.2) years were included of which 82 (27%) had osteoporosis according to DXA, whilst 132 (44%) were osteopenic and 82 (27%) had a normal BMD. The mean interval between DXA and CT was 37.1 (26.3) days. Vertebral fractures were found in 65 (22%) of the patients, 27 (33%) amongst those with DXA defined osteoporosis and in 38 (17%) patients without osteoporosis (Table 1, Figure 2). Amongst women with DXA-defined osteoporosis, 19/56 (34%) had a vertebral fracture, compared to

28/148 (19%) amongst those without osteoporosis. Men showed similar proportions, with 8/26 (31%) with at least one fracture amongst osteoporotic men and 10/72 (14%) amongst non-osteoporotic. Vertebral fractures were most frequent amongst patients with lower vertebral density values on CT, with all but 61 (94%) occurring in those with L1 values below 160 HU (Figure 2).

The diagnostic performance for vertebral HU measurements was modest, as measured by the AUC (Table 2, Figure 3). We found an AUC of 0.74 (0.68-0.080), with an optimal diagnostic performance (the threshold of HU where the proportion of correctly classified patients was the greatest) at the HU threshold of L1≤99/Th12≤104. At that threshold the sensitivity was 62% (51-72) and the specificity was 79% (74-84). At the threshold defined by Pickhardt as achieving the best balance between sensitivity and specificity (L1≤160/Th12≤165), we found a high sensitivity of 91% (84-98) with a specificity of 29% (23-35).

Including prevalent vertebral fractures as a dichotomous classifier did not improve the diagnostic performance (Table 2). The optimal AUC was virtually unchanged at 0.74 (0.676-0.804), and was slightly lower than the CT vertebral density only model at the three preselected thresholds, reflecting the lower specificity associated with this approach. At the lowest HU threshold (80/85 HU) the addition of fracture presence reclassified a total of 38 patients, of which 13 had osteoporosis. For the middle and the highest thresholds, fracture presence reclassified 14 (2 with osteoporosis) and 5 of the participants (none with osteoporosis), respectively. This represents a negative net reclassification index at all three thresholds due to lower specificity (table 2). In a multivariate logistic regression model age (p=0.23), gender (p=0.66) and fracture status (p=0.63) did not significantly improve the model. The AUC of this model was also 0.74 (0.68-0.80).

With vertebral fractures as an outcome and DXA T-score as the independent variable, we found an AUC of 0.611 (0.53-0.69). When vertebral density on CT was set as the independent variable we found a higher AUC of 0.765 (0.70-0.83).

Discussion

We confirmed that simple trabecular vertebral density measurements on routine CT contain diagnostic information related to bone mineral density as measured by DXA, albeit with substantially lower diagnostic accuracy than previously reported.

Pickhardt et al.[1] had found a higher diagnostic performance, with an AUC around 0.83 and a sensitivity of 76% paired with specificity of 75% at a 135 HU threshold. These performance

measures was substantially higher than the AUC of 0.74 and optimal sensitivity of 62% and specificity of 79% we observed (at an optimal L1≤99/Th12≤104 threshold). This discrepancy might suggest that the diagnostic performance of vertebral density on CT for DXA-defined osteoporosis is relatively variable across different populations and different settings. Differences in scanning equipment from different manufacturers and differences in scanning protocols might explain a part of the difference, along with differences in the population demographics under study. Although the age (mean overall age 58 years) and gender (68% female overall) distributions in our sample did not differ substantially from those of Pickhardt (59 years and 81%, respectively), more subtle differences in referral patterns to DXA and CT are likely to exist. Although this limited study cannot fully explain the differences in diagnostic accuracy, our data suggest some caution is warranted when considering the performance of CT for the opportunistic screening of osteoporosis. Recent work has shown that bone density measurements on CT at other sites, such as the hip,[15] are also highly correlated to DXA measurements. Further research on technical of clinical determinants that influence CT bone density measurements seems warranted.

References

1. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med*. 2013; 158: 588–595.
2. Bessette L, Ste-Marie L-G, Jean S, Davison KS, Beaulieu M, Baranci M et al. The care gap in diagnosis and treatment of women with a fragility fracture. *Osteoporos Int*. 2008; 19: 79–86.
3. Metge CJ, Leslie WD, Manness L-J, Yogendran M, Yuen CK, Kvern B et al. Postfracture care for older women: gaps between optimal care and actual care. *Can Fam Physician*. 2008; 54: 1270–1276.
4. King AB, Fiorentino DM. Medicare payment cuts for osteoporosis testing reduced use despite tests' benefit in reducing fractures. *Health Aff*. 2011; 30: 2362–2370.
5. Leslie WD, Giangregorio LM, Yogendran M, Azimaee M, Morin S, Metge C et al. A population-based analysis of the post-fracture care gap 1996-2008: the situation is not improving. *Osteoporos Int*. 2012; 23: 1623–1629.
6. Riggs BL, Melton LJ, Robb RA, Camp JJ, Atkinson EJ, McDaniel L et al. A Population-Based Assessment of Rates of Bone Loss at Multiple Skeletal Sites: Evidence for Substantial Trabecular Bone Loss in Young Adult Women and Men. *J Bone Miner Res*. 2007; 23: 205–214.
7. Buckens CF, de Jong PA, Mol C, Bakker E, Stallman HP, Mali WP et al. Intra and Interobserver Reliability and Agreement of Semiquantitative Vertebral Fracture Assessment on Chest Computed Tomography. *PLoS ONE*. 2013 Aug 5;8(8):e71204.
8. Buckens C, de Jong P, Mali W, Verhaar H, van der Graaf Y, Verkooijen H. Prevalent vertebral fractures on chest CT: Higher risk for future hip fracture. *J Bone Miner Res*. 2014 Feb;29(2):392-8.
9. Raisz LG. Clinical practice. Screening for osteoporosis. *N Engl J Med*. 2005; 353: 164–171.
10. The National Osteoporosis Foundation (NOF). Clinician's Guide to Prevention and Treatment of Osteoporosis. The National Osteoporosis Foundation: Washington, DC, 2013.
11. Majumdar SR, Leslie WD. Conventional computed tomography imaging and bone mineral density: opportunistic screening or 'incidentaloporosis'? *Ann Intern Med*. 2013; 158: 630–631.
12. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000; 19: 453–473.
13. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993; 8: 1137–1148.
14. Oei L, Rivadeneira F, Ly F, Breda SJ, Zillikens MC, Hofman A et al. Review of radiological scoring methods of osteoporotic vertebral fractures for clinical and research settings. *Eur Radiol*. 2013; 23: 476–486.
15. Gruber M, Bauer JS, Dobritz M, Beer AJ, Wolf P, Woertler K et al. Bone mineral density measurements of the proximal femur from routine contrast-enhanced MDCT data sets correlate with dual-energy X-ray absorptiometry. *Eur Radiol*. 2013; 23: 505–512

Table 3 (supplement). Diagnostic accuracy of routine computed tomography for having a vertebral fracture on CT or DXA-defined osteoporosis (T-score ≤2.5)

determinants	threshold	sens	spec	accuracy	PPV	NPV	AUC
vertebral density on CT	$HU_{L1}80/HU_{Th12}85$	38% (30-47)	96% (92-98)	73% (67-78)	86% (71-94)	70% (67-74)	0.67 (0.62-0.72)
	$HU_{L1}110/HU_{Th12}115$	68% (60-77)	74% (66-80)	72% (64-79)	63% (54-72)	78% (71-84)	0.71 (0.66-0.76)
	$HU_{L1}160/HU_{Th12}165$	90% (85-95)	32% (26-39)	55% (49-61)	47% (43-51)	83% (72-92)	0.61 (0.57-0.66)
	optimal* $HU_{L1}103/HU_{Th12}108$	63% (55-72)	81% (75-87)	74% (67-81)	69% (59-79)	77% (72-82)	0.77 (0.72-0.83)

sens=sensitivity; spec=specificity; PPV=positive predictive value; NPV=negative predictive value; AUC=area under the receiver operator characteristics curve; data are given with 95% confidence interval between brackets. Confidence intervals generated using 2000 stratified bootstrap replicates;
*optimal threshold is the threshold where the largest proportion of patients is correctly classified

Table 4 (supplement). Characteristics of male and female patients with and without osteoporosis, as defined by a DXA T-score ≤2.5

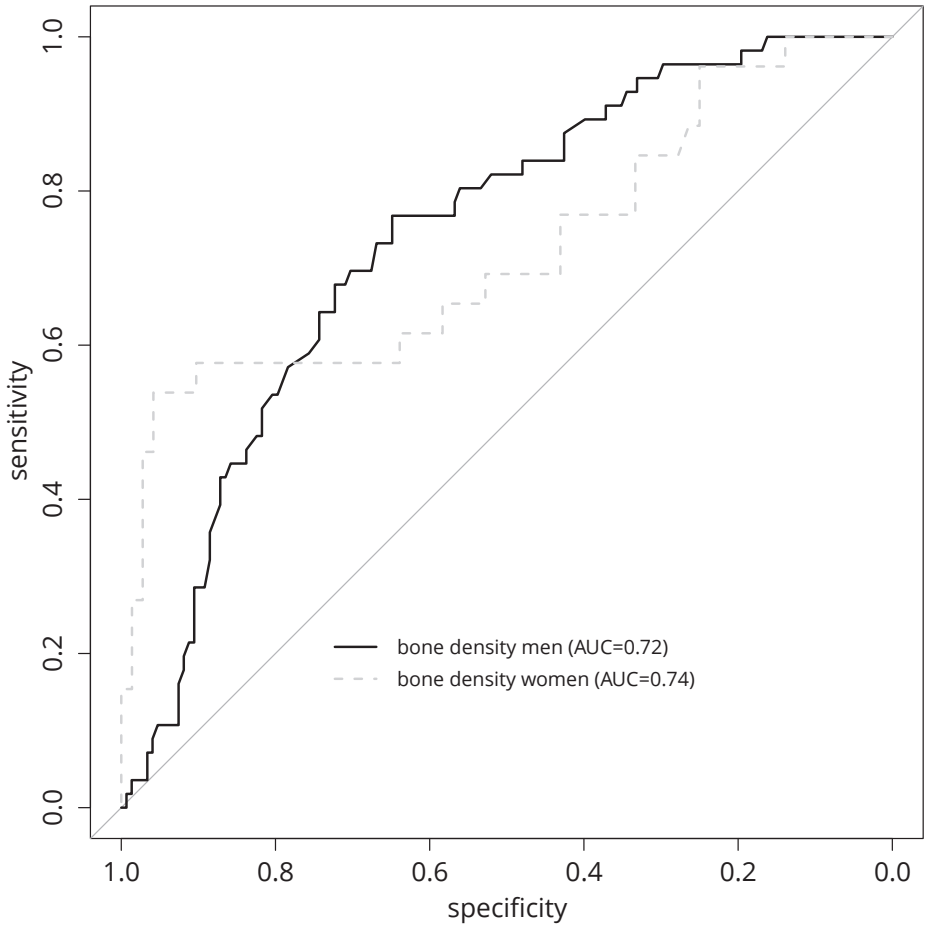
		women		men	
variable	**description**	**osteoporosis**	**no osteoporosis**	**osteoporosis**	**no osteoporosis**
n	number	56	148	26	72
age	years	64	56	56	58
vertebra measured*	L1 (%)	46 (82%)	134 (91%)	26 (100%)	64 (89%)
anatomical area**	abdomen (%)	38 (68%)	87 (59%)	15 (58%)	49 (68%)
DXA area	lumbar spine (%)	37 (66%)	81 (55%)	21 (81%)	38 (53%)
T-score	mean	-3.3	-1,2	-3.1	-0.9
days	mean	31	37	29	46
fracture	fracture (%)	19 (34%)	28 (19%)	8 (31%)	10 (14%)
HU	mean	95	134	102	145
indication	inflammatory/ autoimmune	6 (11%)	25 (17%)	4 (15%)	16 (22%)
	endocrine disorder	10 (18%)	53 (36%)	2 (8%)	10 (14%)
	fracture	20 (36%)	25 (17%)	8 (31%)	19 (26%)
	steroid therapy	6 (11%)	15 (10%)	3 (12%)	10 (14%)
	transplantation	4 (7%)	9 (6%)	6 (23%)	8 (11%)
	malignancy	3 (5%)	5 (3%)	0 (0%)	2 (3%)
	other	7 (12%)	16 (11%)	3 (12%)	7 (10%)
HU_{L1}≤80/ HU_{Th12}≤85		23 (41%)	19 (13%)	10 (38%)	2 (3%)
HU_{L1}≤110/ HU_{Th12}≤115		41 (73%)	50 (34%)	15 (58%)	24 (33%)
HU_{L1}≤160/ HU_{Th12}≤165		54 (96%)	108 (73%)	21 (81%)	48 (67%)

* if first lumbar (L1) was not visualized or was fractured, the twelfth thoracic (Th12) was measured instead

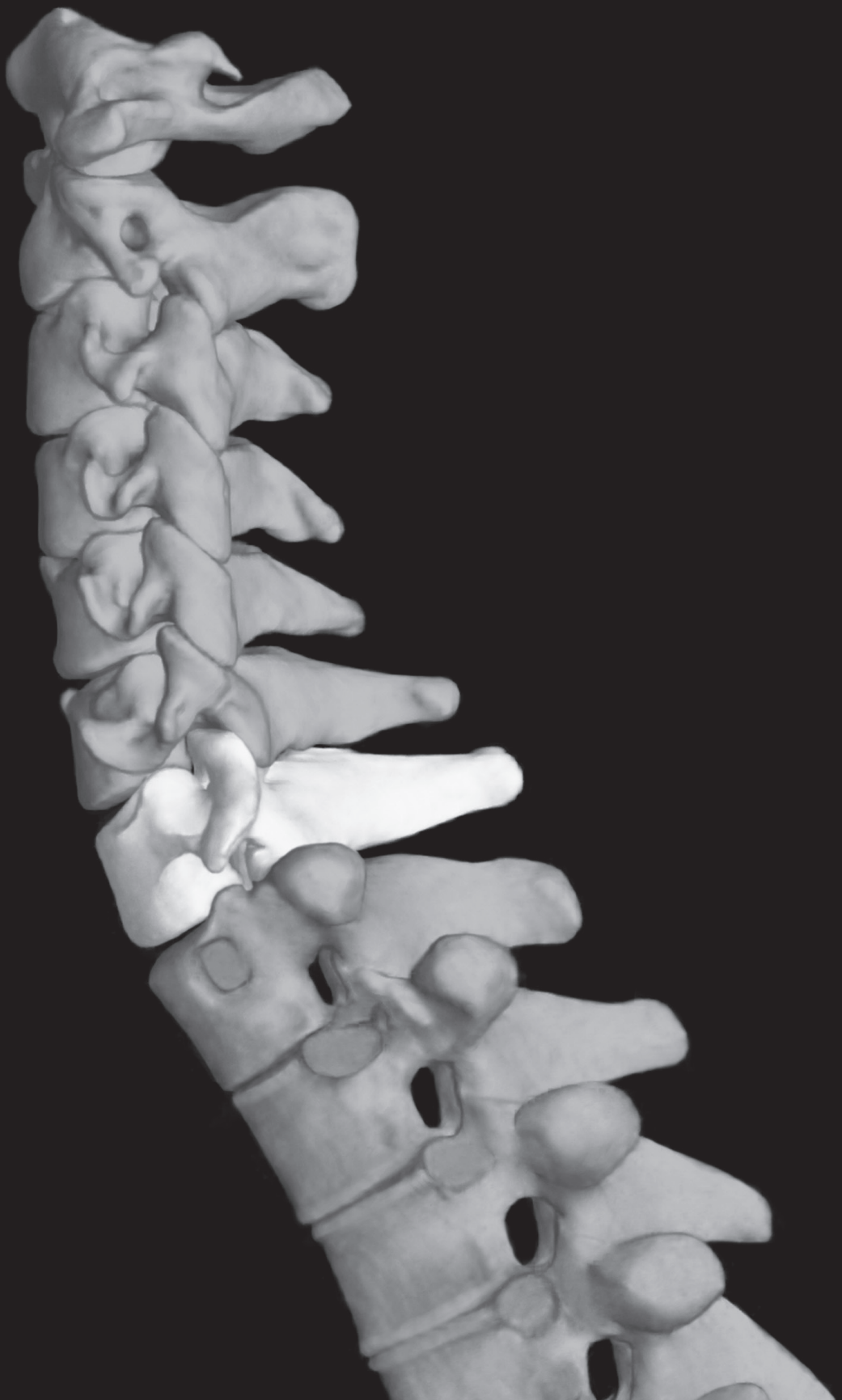
** thoracic and abdominal CTs were included

*** lowest areal bone density value obtained value used, either derived from lumbar spine or hip

Figure 4 (supplement). **Receiver operating characteristics curve for vertebral bone density model predicting DXA-assigned osteoporosis**



Three pre-defined thresholds are marked along with the corresponding sensitivity and specificity



Chapter 7

Osteoporosis markers on low-dose lung cancer screening chest computed tomography scans predict all-cause mortality

Buckens CF
van der Graaf Y
Verkooijen HM
Mali WP
Isgum I
Mol CP
Verhaar HJ
Vliegenthart R
Oudkerk M
van Aalst CM
de Koning HJ
de Jong PA

Eur Radiol. 2015 Jan;25(1):132-9

Abstract

Osteoporosis markers on low-dose lung cancer screening chest computed tomography scans predict all-cause mortality

Objectives

Further survival benefits may be gained from low-dose chest computed tomography (CT) by assessing vertebral fractures and bone density. We sought to assess the association between CT-measured vertebral fractures and bone density with all-cause mortality in lung cancer screening participants.

Methods

Following a case-cohort design, lung cancer screening trial participants (n=3673) who died (n=196) during a median follow-up of 6 years (inter-quartile range: 5.7–6.3) were identified and added to a random sample of n=383 from the trial. We assessed vertebral fractures using Genant's semiquantative method on sagittal reconstructions and measured bone density (Hounsfield Units, HU) in vertebrae. Cox proportional hazards modeling was used to determine if vertebral fractures or bone density were independently predictive of mortality.

Results

The prevalence of vertebral fractures was 35% (95% confidence interval 30–40%) among survivors and 51% (44–58%) amongst cases. After adjusting for age, gender, smoking status, pack years smoked, coronary and aortic calcium volume and pulmonary emphysema the adjusted HR for vertebral fracture was 2.04 (1.43–2.92). For each 10 HU decline in trabecular bone density the adjusted HR was 1.08 (1.02–1.15).

Conclusions

Vertebral fractures and bone density are independently associated with all-cause mortality.

Introduction

The National Lung cancer Screening Trial has shown that timely screening can reduce mortality; that trial showed a 20% reduction in lung cancer mortality and a remarkable 7% all-cause mortality reduction over a median followup of 6.5 years, following computed tomography (CT) based screening for lung cancer only.[1] These findings have sparked renewed interest in screening of heavy smokers, a group at particular risk for serious diseases, with debate revolving mainly around cost-effectiveness and harm-benefit ratios.[2] Currently, screening chest CTs are only considered for evaluation for pulmonary nodules, as markers of early lung cancer. The benefits associated with screening could be maximized by screening for multiple diseases on the acquired imaging, only marginally increase costs[3] through increased read times, without increasing radiation dose. Chest CT could additionally reveal pulmonary emphysema – one of the manifestations of COPD and an independent risk factor for lung cancer[4,5] – and arterial calcifications in the coronary arteries and aorta, which are strong and independent predictors of CVD-related mortality.[6,7]

Besides lung and other cancers, cardiovascular diseases (CVD) and chronic obstructive pulmonary disease (COPD), smokers are also at an increased risk for osteoporosis.[8] All these diseases often coexist in smokers and they may share common pathophysiological mechanism.[9] The increased risk for multiple diseases and the fact that biomarkers for these conditions can all be visualized on available chest CT exams invites investigation of a multi-disease approach[3] to screening chest CTs.

Recent research has shown that simple CT attenuation values in the trabecular region of vertebral bone are highly correlated with DXA values, making them a viable marker for osteoporosis,[10] particularly amongst COPD patients.[11] Bone mineral density predicts mortality in other settings,[12] but it is not known if CT attenuation is also associated with mortality in a lung cancer screening population. Vertebral fractures are visible on sagittal reformats of screening chest CT exams and can be scored with good reliability and agreement.[13] These fractures have previously been shown to predict all-cause mortality and hip fractures in the general population.[14-20] Vertebral fractures are prevalent in more than 10% of the general adult population[21-25] but less than 15% come to clinical attention. Bisphosphonate therapy aiming to reduce fracture risk has additionally been found to significantly reduce all-cause mortality, suggesting that excess mortality risk may be modifiable through targeted intervention.[26]

The objective of this study is to assess the prevalence and severity of vertebral fractures and low bone density on chest CT and their association with all-cause mortality in smokers participating in lung cancer screening.

Materials and methods

Study Population

This study was performed within the framework of the Dutch Belgian Lung Cancer Screening Trial (NELSON-trial; ISRCTN 63545820).This trial was ethically approved by the Dutch Ministry of Health and informed consent was obtained from all participants. The design of the trial has been described more extensively elsewhere.[27] Briefly, participants were current or former (cessation <10 years) smokers between the ages of 50 and 75 years with a smoking history of at least 10 cigarettes/day >30 years, or >15 cigarettes a day for >25 years.

Design

For the present study, 3679 participants from the screening arm from two participating centers in the Netherlands were included: the University Medical Center in Groningen (UMCG) and the University Medical Center in Utrecht (UMCU). In order to efficiently calculate the prevalence and the absolute risks without manually scoring this entire cohort, a case-cohort approach was used. This entails that a random sample of ≈10% (n=383) of the full cohort was drawn (the "subcohort"). The subcohort is used to estimate the distribution of covariates (here bone density, vertebral fracture status, cardiovascular calcification and emphysema) in the full cohort and derive the base line hazard function. Due to the random nature of the subcohort, 4 patients in the subcohort died during the followup are also present in the subcohort. This is appropriate, as these individuals also contribute to the baseline hazard function. All cases outside the subcohort (defined as those who died during follow-up) are then added to the subcohort to form the complete case-cohort dataset.[28] Only this resulting case-cohort dataset then needs to be manually scored.

CT scanning protocol

Low-dose volumetric CT scans were acquired between January 2004 and December 2007. CTs were obtained on 16-slice MDCT scanners with a collimation of 16 x 0.75 mm. The UMCG participants were scanned using a Sensation-16 CT (Siemens Medical Solutions, Forchheim, Germany), whereas UMCU participants were scanned on either Mx8000 or Brilliance-16P CT (Philips Medical Systems, Cleveland, OH, USA). Exposure settings were adjusted according to body weight; either 120 kVp when participants weighed <80 kg or 140 kVp, both at 30 mAs. This yielded an effective dose of <0.9 and <1.6 mSv respectively. Axial reconstructions with a slice thickness of 1-mm at 0.7-mm increment and using a smooth reconstruction filter (Siemens B30f, Philips B-filter) were stored at acquisition.

CT emphysema quantification

The emphysema quantification method has previously been described.[4] Briefly, the lungs were automatically segmented (separated from the chest wall, diaphragm, mediastinum, and airways)[29] and a noise reduction filter was applied to decrease the influence of noise on the quantitative measurements.[30] Within the segmented lung volume, the attenuation measured by the Hounsfield units (HU) of each voxel was assessed to quantify emphysema severity. Computed tomographic emphysema was defined as the percentage of voxels in inspiratory CT with an attenuation below -950 Houndsfield Units (HU).[31] As an additional secondary analysis, the HU value at the 15th percentile of the attenuation distribution curve as a measure of CT emphysema[32] was calculated.

CT arterial calcification quantification

Full details of the calcium quantification have previously been described.[33] Briefly, calcifications were quantified on 3.1 mm thick chest CT reconstructions with 1.4 mm increment. A threshold of 130 HU in combination with three-dimensional connected component labeling was used to identify potential calcifications. Subsequently, aortic calcifications were detected on multi-atlas-based segmentation of the aorta, followed by a supervised pattern recognition system identifying aortic calcifications based on spatial, size and texture features. Given the excellent agreement between automatic and manual calcium scoring of aortic lesions in low-dose CT,[33,34] only extreme values in aortic calcium scores were inspected and manually corrected if needed. Coronary calcifications are extracted based on a probabilistic coronary calcium map providing an a priori probability for spatial appearance of coronary calcifications on a chest CT scan, followed by a supervised pattern recognition system detecting calcifications based on spatial- and texture features. All automatically detected coronary calcifications were inspected and manually corrected if needed. Total aortic and coronary calcification burdens were quantified in terms of total calcium volume (mm^3).

CT vertebral fracture assessment

Vertebral fractures were visually identified and graded on sagittal CT reconstructions according to Genant's semi-quantitative vertebral fracture assessment method.[35] This method identifies fractures according to the amount of height loss of the vertebral body compared to an adjacent non-fractured vertebra by comparing the anterior, middle and/or posterior aspect of the vertebral body with its immediate neighbors. The fracture severity grades are defined as: grade 1 deformities (mild height loss of 20–25%), grade 2 fractures (moderate height loss of 25–40%) and severe grade 3 fractures with height loss of more than

40% (Figure 1). Vertebrae were assessed at or around mid-sagittal slices, with the rater being free to adjust slice-thickness, orientation and window level as needed. The following participant-level measures were calculated: fracture presence (yes/no), worst fracture grade and cumulative fracture category. The worst fracture grade was defined as the grade of the worst fracture visible in a patient (maximum 3). For data analysis worst fracture grade 2 and 3 were combined into a single category due to the low number of moderate and severe fractures, leading to the categories no fracture, mild deformity, vertebral fracture. The cumulative fracture category was the sum of all recorded fracture grades in a participant and was categorized into one of three categories: 0 (no fractures), 1–3 (mild cumulative fracture burden), or ≥436 (moderate to severe cumulative fracture burden). In a separate sample of 50 chest CT scans we found good interobserver reproducibility between four observers of varying experience for the three measures (Cohen's kappa 0.56–0.81).[13] In that study the observers included one PhD candidate, two radiological residents with a musculoskeletal interest (one with 2 years' experience and one with 4 years' experience), and an experienced board certified chest and musculoskeletal radiologist with more than 10 years' experience.

Adjacent, morphologically normal vertebrae were used for comparison to determine the grade of deformity. Figure 1a: grade 1, mild deformity (arrow) of a vertebra with an estimated height loss of 20-25%, as compared to the vertebra above. Figure 1b: moderate, grade 2 fracture (arrow) with an estimated height loss of 25-40%, compared to the vertebra above. Figure 1c: severe, grade 3 fracture (arrow) with an estimated height loss of >40% when compared to the vertebra above.

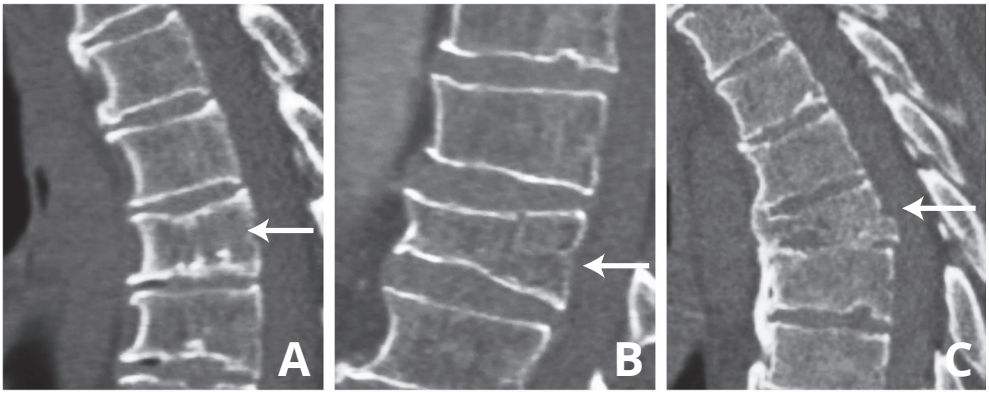
Vertebral bone density measurement

CT attenuation values were measured on axial CT reformats in trabecular regions of the bodies of Lumbar 1 or nearest visible, unfractured, visually normal vertebra, as previously described.[10] If Lumbar 1 was not visualized or if it was fractured Thoracic 12 was measured instead. A single click and drag region of interest was placed in the anterior upper portion of the body of the vertebra inside trabecular bone. The region of interest was drawn so that it did not intersect the vertebral cortex and did not include dense bone islands, hemangiomas or traversing vessels. The CT attenuation was measured in HU, with lower values representing less dense bone.

Cases

Cases were defined as participants who died during the follow-up period. Mortality data were obtained through linkage with the national death registry from January 2004 to the

Figure 1. Vertebral deformity gradation on sagittal reformats of low dose lung cancer



Adjacent, morphologically normal vertebrae were used for comparison to determine the grade of deformity
Fractured vertebrae are indicated by white arrow
A=Grade 1, mild deformity of vertebra with estimated height loss of 20–25%, compared to vertebra above
B=Grade 2, moderate fracture with an estimated height loss of 25–40%, compared to vertebra above
C=Grade 3, severe fracture with an estimated height loss of >40%, compared to vertebra above

end of December 2010. This linkage was performed using validated probabilistic method.[36, 37]

Statistical analysis

Cox-proportional hazard modeling was used to determine the association between the vertebral fracture measures and mortality risk. To account for the case-cohort approach, a validated[28] weighting method was applied to the analysis, as described by Prentice.[38] Crude hazard ratios (HR) were calculated accordingly for the three vertebral fracture measures (dichotomous presence of fracture, worst fracture grade and cumulative fracture grade) and bone density. These crude HRs were first adjusted for age (in years at inclusion) and gender in a basically adjusted model, and then additionally for smoking status (current or former), pack years smoked, coronary artery calcifications (in mm^3), aortic calcifications (in mm^3) and emphysema. The vertebral density model was further adjusted for the presence of vertebral fracture, to ensure that bone density was independent of vertebral fracture presence. The vertebral fracture measures were similarly adjusted for vertebral density in the fully adjusted model. To improve the interpretability of the hazard ratios of vertebral density, the raw HU values were transformed by inverting them (as we expected an inverse association, with lower bone density having a higher hazard for all-cause mortality) and dividing them by 10. For each model, we estimated the discriminatory performance (the ability of the model to distinguish between cases and non-cases) by calculating an over-

Table 1. Characteristics of survivors and non-survivors

description	survivors (n=397)	deceased (n=196)
age (median in years (IQR))	60 (56-64)	63 (59-67)
gender (female)	11 (3%)	10 (5%)
follow-up (median in years (IQR))	2176 (2092-2298)	2162 (2034-2297)
smoking status (current smokers)	216 (57%)	114 (58%)
pack years (median (IQR))	38 (29-49)	39 (31-55)
aortic calcium in mm^3 (median (IQR))	560 (87-2017)	1742 (444-4845)
coronary calcium in mm^3 (median (IQR))	143 (10-612)	546 (106-1396)
emphysema score* (mean (SD))	1.03 (1.1)	1.14 (1.1)
15th percentile HU value (mean (SD))	-921 (21)	-923 (23)
trabecular vertebral bone density in HU (mean (SD))	100 (32)	92 (31)
presence of vertebral fracture (n with fracture)	130 (35%)	94 (51%)
worst fracture grade (mild)	102 (27%)	75 (38%)
worst fracture grade (moderate or severe)	31 (8%)	24 (12%)
cumulative fracture grade (1-3)	110 (29%)	80 (41%)
cumulative fracture grade (≥4)	23 (6%)	19 (10%)

* logarithmically transformed number of voxels below 950 Hounsfield units; IQR=interquartile range

optimism corrected Harrell's concordance statistic (c-statistic). This value is analogous to the more widely used receiver operating characteristic curve, adapted for time-to-event data. We corrected for over-optimism using 2000 bootstrap replicates, reporting the median. These replicates were also used to estimate the non-parametric 95% confidence intervals (by taking the 0.025 and 0.975 quartiles) for the reported c-statistices. The random subcohort sample was drawn using the random number generator function in SPSS (SPSS Inc. SPSS for Windows, Version 15, Chicago, Illinois, US). All further analyses were performed using R 3.0.1.[39]

Results

Subject characteristics

Of the 3679 participants, 196 (5.3%) died during a median follow-up of 5.93 years (range: 4.1–6.7) (Table 1), resulting in 9 deaths/1000 person-years. Cases were on average 63 and survivors 60 years of age at inclusion. An equal proportion of cases and survivors were current smokers (57 and 58%) and both groups had smoked the same number of pack years

(38 and 39). There was a higher volume of aortic and coronary calcifications in cases but emphysema scores were similar in both groups (Table 1). Vertebral density was assessed at the L1 level in 83% of survivors and 82% if cases, with the remainder being assessed in the Th12 vertebra.

Vertebral fractures and bone density

Vertebral fractures were common in this population of smokers (Table 1). In the randomly sampled subcohort, the prevalence of fractures was 35% (95% confidence interval: 30–40), compared to 51% (43–58) amongst cases. Moderate and severe fractures were more common in cases (13%, 9–18) compared to the survivors (8%, 6–11) and the cumulative fracture grades were higher in the cases. Bone density values were found to be lower amongst cases (Table1).

Association of low bone density with mortality

Vertebral bone density had a small but statistically significant negative association with mortality. Inverted vertebral density had a crude hazard rate of 1.087 (1.027–1.15) per 10 HU decrease. After adjustment for age and gender this was 1.08 (1.018–1.146) and after correction for cardiovascular calcifications, smoking status and emphysema it was 1.076 (1.006–1.151). The univariate optimism-corrected c-statistic was modest 0.587 (0.522–0.647), whilst the fully corrected model (excluding only vertebra fracture status) achieved 0.698 (0.644–0.747). Without the attenuation variable, the combination of age, gender, smoking details, arterial calcifications and emphysema achieved a c-statistic of 0.681 (0.627–0.736), indicating that vertebral bone density has a modest incremental value for predicting mortality.

Association of vertebral fractures with all-cause mortality

Vertebral fractures were associated with all-cause mortality in this population (Table 2 and Figure 2). Without adjustment, the presence of any fracture or deformity was associated with an almost doubled mortality risk, with a HR of 1.93 (95% CI 1.37–2.73). Similar associations were present between mortality and the worst and cumulative fracture grade. A mild vertebral fracture resulted in a crude HR for death of 1.93 (95% CI 1.33–2.80) and a moderate or severe fracture in a crude HR of 1.97 (95% CI 1.11–3.48). Having a cumulative fracture grade of 1–3 was associated with a crude HR of 1.89 (95% CI 1.31–2.72) and ≥4 with a HR of 2.16 (95% CI 1.38–4.09). HRs for vertebral fracture presence, severity and cumulative burden remained virtually unchanged after adjustment for age and gender (Table 2). Full

Table 2. Effect of prevalent vertebral fractures and vertebral bone density on all-cause mortality in hazard ratios

variable	crude	adjusted for age	fully adjusted*
fracture yes/no	1.93 (1.37-2.73)	2.04 (1.42-2.95)	2.18 (1.47-3.22)
worst fracture grade 1	1.93 (1.33-2.80)	2.01 (1.35-2.99)	2.12 (1.38-3.25)
worst fracture grade 2 -3	1.97 (1.11-3.48)	2.14 (1.17-3.92)	2.37 (1.28-4.39)
cumulative fracture grade 1-3	1.89 (1.31-2.72)	1.99 (1.35-2.94)	2.13 (1.40-3.24)
cumulative fracture grade ≥4	2.16 (1.38-4.09)	2.28 (1.15-4.51)	2.37 (1.18-4.77)
vertebral bone density**	1.087 (1.027-1.15)	1.080 (1.018-1.146)	1.076 (1.006-1.151)

* adjusted for age, gender, bone density, smoking status, pack years, coronary calcium volume, aortic calcium volume, and emphysema

** note that HU values were inverted and divided by 10 for the purposes of this analysis; the HR represents the increase in risk of death for each 10 HU drop in vertebral bone density, adjusted for age, gender, vertebral fracture presence, smoking status, pack years, coronary calcium volume, aortic calcium volume, and emphysema

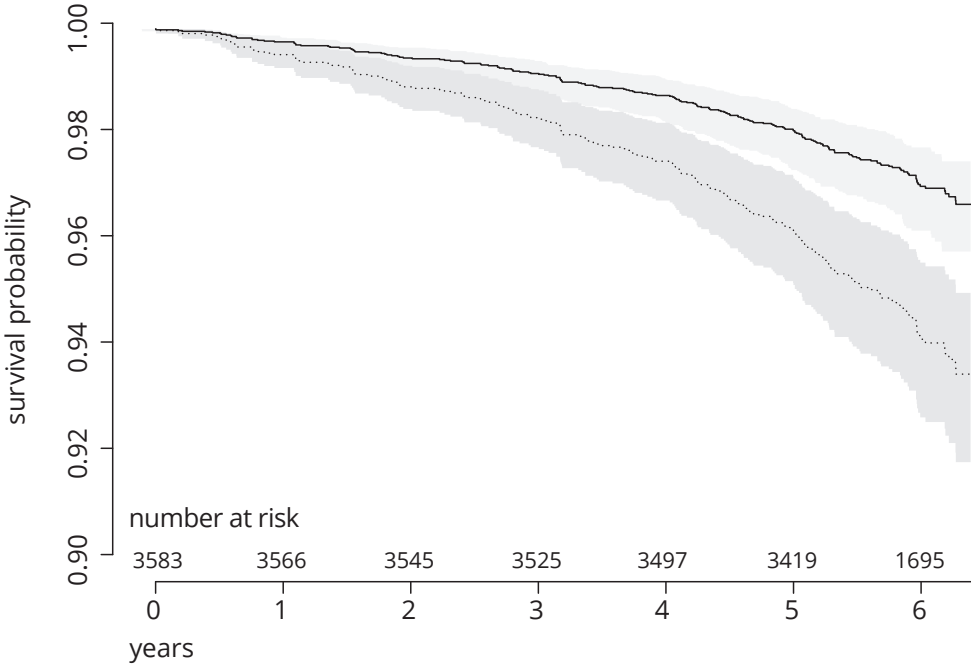
adjustment for age, gender, bone density, smoking status, pack years smoked, aortic and coronary calcifications and emphysema did not substantially modify the HRs (Table 2).

Bone density and vertebral fracture status together achieved an optimism-corrected c-statistic of 0.705 (0.650–0.755) for the presence of vertebral fracture, and a marginally lower c-statistic for the worst and cumulative fracture grades (0.704 (0.649–0.754) and 0.703 (0.648–0.755), respectively), reflecting the penalization for complexity in the optimism correction and their low incremental value over the simple, binary, fracture presence variable. This is slightly higher than the 0.681 (0.627–0.736) achieved without vertebral fracture or attenuation. These HR values suggest that vertebral fractures are an independent biomarker of all-cause mortality in this setting, and that it has a modest incremental value for discriminating cases from non-cases.

Discussion

In this study we demonstrated a high prevalence of vertebral fractures on chest CT in a predominantly male lung cancer screening population and these fractures and lower bone density are associated with a doubling of the risk for all-cause mortality independent of age, smoking status, pack-years smoked, coronary and aortic calcification volume and pulmonary emphysema severity. These easily assessable chest CT findings could serve as biomarkers both for personalizing lung cancer screening intensity and for multiple disease screening, a putative role that deserves further exploration.

Figure 2. Survival curve for participants with vertebral fracture (solid line) and without vertebral fracture (dashed line)



On the x-axis: years of available follow-up with, above x-axis, number at risk remaining after each year. On the y-axis: cumulative survival probability for patient with median values for all covariates. Shaded areas show 95 % confidence interval estimate. Note that the lower limit of the y-axis has been moved up to 0.90 to improve the interpretability of the figure.

The prevalence of vertebral fractures in this study is somewhat higher than in other settings, where the prevalence among (predominantly male) subjects of similar age ranged from 9–31%.[16,19,40] This may be due to the fact that we used CT scans instead of plain imaging techniques, and CT may better detect mild deformities or fractures. However, our population of smokers may also suffer from an inherently increased risk of vertebral fractures, as both smoking itself and other common smoking related diseases (principally COPD and CVD) are associated with vertebral fractures.[11,41–45] The attendant medication use associated with these co-morbidities quite likely play a role in this. Unfortunately, there was no clinical information on co-morbidities or medication usage in this cohort. We also found relatively low bone density as compared to other studies, although this may also be related to the use of low-dose CT acquisition protocol, with lower tube voltages affecting the linear attenuation coefficients. Threshold values from non-low dose scanning protocols may not carry over to low-dose screening protocols.

In other settings studies have investigated the association between vertebral fractures and all-cause mortality using vertebral densitometry and conventional radiography. Most studies found a significant association between vertebral fractures and mortality, also after adjustment for a wide range of covariates.[15–18,40,46] The adjusted HR for prevalent vertebral fractures ranged from 1.5 to 2.9, which is in line with the 2.04 found in our study. Two other studies found a significant crude association of prevalent vertebral fractures with mortality that became non-significant after adjustment.[14,47] Height loss (a consequence of accumulated vertebral deformity and fracture) has also been found to independently predict mortality, with adjusted HRs ranging between 1.3 to 1.8.[48–50] Although the lung cancer screening participants we investigated were predominantly male, within a limited age-range and were selected based on current or former smoking, the associations that we observed are generally in line with findings in other settings.

At present it remains unclear to what extent the excess mortality risk among screening participants with vertebral fractures is modifiable through targeted interventions in this population. The mortality reduction associated with bisphosphonate administration[26] suggests that it might be possible to reduce mortality risk through timely intervention. This possibility deserves to be explored in the context of lung cancer screening as it could be used to compound the significant all-cause mortality reduction already demonstrated in the NLST, by identifying patients at an increased risk of other disease clusters, who may benefit from early treatment. Other disease bio-markers could potentially be employed to mitigate the high financial cost of CT-based screening by personalizing the screening intensity according to the predicted risk of lung cancer and other outcomes.[51–54] For instance, screening intensity may be reduced in subgroups with a predicted lower lung cancer risk, calculated on the basis of all available information. While our results are promising, whether fractures and bone density will gain a role in personalizing lung cancer screening and optimization of cost-effectiveness requires further research. Off-setting any additional benefit derived from detecting vertebral fractures is of course the increased read times necessary to evaluate the sagittal reconstructions. Our experience in this study suggests that the additional time required is limited to about one minute and this may be further reduced in the future when computer assisted vertebral fracture assessment becomes more mature.

Our study has several limitations. Firstly, since we studied a predominately male population of current and former smokers from the general population. Our findings may not be generalizable to women participating in lung cancer screening. Also caution is needed

with extrapolating our results to routine clinical chest CT scans. Secondly, not all vertebral fractures are visible on chest CT as the scan is necessarily limited to thoracic and mostly lumbar 1. This limitation is unavoidable consequence of assessing vertebral fractures on already-acquired chest CTs obtained in the course of lung cancer screening. In routine practice assessing the lateral scout views for vertebral fractures can extend this to the lower lumbar levels that routinely include the thoracolumbar region and help to detect additional fractures.[55,56] Thirdly, as stated above, there was no clinical information on co-morbidities or medication usage in this cohort.

In conclusion, prevalent vertebral fractures and lower vertebral bone density on low-dose chest CT in lung cancer screening participants are independently associated with all-cause mortality. How these findings may be used to benefit current and former smokers participating in screening remains to be determined.

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011 Aug 4;365(5):395–409.
2. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA*. 2012 Jun 13;307(22):2418–29.
3. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. *JAMA*. 2012 Oct 10;308(14):1433–4.
4. Mets OM, Buckens CFM, Zanen P, Isgum I, van Ginneken B, Prokop M, et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. *JAMA*. 2011 Oct 26;306(16):1775–81.
5. Mets OM, Vliegenthart R, Gondrie MJ, Viergever MA, Oudkerk M, de Koning HJ, et al. Lung Cancer Screening CT-Based Prediction of Cardiovascular Events. *JACC Cardiovasc Imaging*. 2013 Jun 8;
6. McEvoy JW, Blaha MJ, Rivera JJ, Budoff MJ, Khan AN, Shaw LJ, et al. Mortality rates in smokers and nonsmokers in the presence or absence of coronary artery calcification. *JACC Cardiovasc Imaging*. 2012 Oct;5(10):1037–45.
7. Xie X, Zhao Y, de Bock GH, de Jong PA, Mali WP, Oudkerk M, et al. Validation and Prognosis of Coronary Artery Calcium Scoring inNon-Triggered Thoracic Computed Tomography: Systematic Review and Meta-Analysis. *Circ Cardiovasc Imaging*. 2013 Jun 11;
8. Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001. *MMWR Morb Mortal Wkly Rep*. 2005 Jul 1;54(25):625–8.
9. Gan WQ, Man SFP, Senthilselvan A, Sin DD. Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. *Thorax*. 2004 Jul;59(7):574–80.
10. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med*. 2013 Apr 16;158(8):588–95.
11. Romme EAPM, Murchison JT, Phang KF, Jansen FH, Rutten EPA, Wouters EFM, et al. Bone attenuation on routine chest CT correlates with bone mineral density on DXA in patients with COPD. *J Bone Miner Res*. 2012 Nov;27(11):2338–43.
12. Johansson C, Black D, Johnell O, Odén A, Mellström D. Bone mineral density is a predictor of survival. *Calcif Tissue Int*. 1998 Sep;63(3):190–6.
13. Buckens CF, de Jong PA, Mol C, Bakker E, Stallman HP, Mali WP, et al. Intra and Interobserver Reliability and Agreement of Semiquantitative Vertebral Fracture Assessment on Chest Computed Tomography. *PLoS ONE*. 2013 Aug 5;8(8):e71204.
14. Jalava T, Sarna S, Pylkkänen L, Mawer B, Kanis JA, Selby P, et al. Association between vertebral fracture and increased mortality in osteoporotic patients. *J Bone Miner Res*. 2003 Jul;18(7):1254–60.

15. Hasserius R, Karlsson MK, Nilsson BE, Redlund-Johnell I, Johnell O. Prevalent vertebral deformities predict increased mortality and increased fracture rate in both men and women: a 10-year population-based study of 598 individuals from the Swedish cohort in the European Vertebral Osteoporosis Study. *Osteoporos Int*. 2003 Jan;14(1):61–8.

16. Trone DW, Kritz-Silverstein D, von Mühlen DG, Wingard DL, Barrett-Connor E. Is radiographic vertebral fracture a risk factor for mortality? *Am J Epidemiol*. 2007 Nov 15;166(10):1191–7.

17. Puisto V, Rissanen H, Heliövaara M, Knekt P, Helenius I. Mortality in the presence of a vertebral fracture, scoliosis, or Scheuermann's disease in the thoracic spine. *Ann Epidemiol*. 2008 Aug;18(8):595–601.

18. Ikeda Y, Sudo A, Yamada T, Uchida A. Mortality after vertebral fractures in a Japanese population. *J Orthop Surg Hong Kong*. 2010 Aug;18(2):148–52.

19. Waterloo S, Ahmed LA, Center JR, Eisman JA, Morseth B, Nguyen ND, et al. Prevalence of vertebral fractures in women and men in the population-based Tromsø Study. *BMC Musculoskelet Disord*. 2012;13:3.

20. Buckens C, de Jong P, Mali W, Verhaar H, van der Graaf Y, Verkooijen H. Prevalent vertebral fractures on chest CT: Higher risk for future hip fracture. *J Bone Miner Res*. 2013 Jul 2;

21. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008 Feb;63(2):160–4.

22. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N, et al. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009 Mar;69(3):555–9.

23. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009 Jan;69(1):179–83.

24. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008 Aug;18(8):1696–702.

25. Chan PL, Reddy T, Milne D, Bolland MJ. Incidental vertebral fractures on computed tomography. *N Z Med J*. 2012 Feb 24;125(1350):45–50.

26. Bolland MJ, Grey AB, Gamble GD, Reid IR. Effect of osteoporosis treatment on mortality: a meta-analysis. *J Clin Endocrinol Metab*. 2010 Mar;95(3):1174–81.

27. Van Iersel CA, de Koning HJ, Draisma G, Mali WPTM, Scholten ET, Nackaerts K, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer*. 2007 Feb 15;120(4):868–74.

28. Onland-Moret NC, van der A DL, van der Schouw YT, Buschers W, Elias SG, van Gils CH, et al. Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol*. 2007 Apr;60(4):350–5.

29. Van Rikxoort EM, de Hoop B, Viergever MA, Prokop M, van Ginneken B. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Med Phys*. 2009 Jul;36(7):2934–47.

30. Schilham AMR, van Ginneken B, Gietema H, Prokop M. Local noise weighted filtering for emphysema scoring of low-dose CT images. IEEE Trans Med Imaging. 2006 Apr;25(4):451–63.

31. Gevenois PA, De Vuyst P, de Maertelaer V, Zanen J, Jacobovitz D, Cosio MG, et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. *Am J Respir Crit Care Med*. 1996 Jul;154(1):187–92.

32. Newell JD Jr, Hogg JC, Snider GL. Report of a workshop: quantitative computed tomography scanning in longitudinal studies of emphysema. *Eur Respir J*. 2004 May;23(5):769–75.

33. Isgum I, Rutten A, Prokop M, Staring M, Klein S, Pluim JPW, et al. Automated aortic calcium scoring on low-dose chest computed tomography. *Med Phys*. 2010 Feb;37(2):714–23.

34. Isgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging*. 2012 Dec;31(12):2322–34.

35. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993 Sep;8(9):1137–48.

36. Siris ES, Genant HK, Laster AJ, Chen P, Misurski DA, Krege JH. Enhanced prediction of fracture risk combining vertebral fracture status and BMD. *Osteoporos Int*. 2007 Jun;18(6):761–70.

37. De Bruin A, Karduan J, Gast F, De Bruin E, van Sijl M, Verweij G. Record linkage of hospital discharge register with population register: experiences at Statistics Netherlands. *Stat J UN Econ Comm Eur*. 2004;1(21):23–32.

38. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011 May;64(5):565–72.

39. Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986;73(1):1–11.

40. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org/

41. Pongchaiyakul C, Nguyen ND, Jones G, Center JR, Eisman JA, Nguyen TV. Asymptomatic vertebral deformity as a major risk factor for subsequent fractures and mortality: a long-term prospective study. *J Bone Miner Res*. 2005 Aug;20(8):1349–55.

42. Nasto LA, Fusco A, Colangelo D, Mormando M, Di Giacomo G, Rossi B, et al. Clinical predictors of vertebral osteoporotic fractures in post-menopausal women: a cross-sectional analysis. *Eur Rev Med Pharmacol Sci*. 2012 Sep;16(9):1227–34.

43. Adler RA. Osteoporosis in men: insights for the clinician. *Ther Adv Musculoskelet Dis*. 2011 Aug;3(4):191–200.

44. Kiyokawa H, Muro S, Oguma T, Sato S, Tanabe N, Takahashi T, et al. Impact of COPD exacerbations on osteoporosis assessed by chest CT scan. *COPD*. 2012 Jun;9(3):235–42.

45. Sennerby U, Melhus H, Gedeborg R, Byberg L, Garmo H, Ahlbom A, et al. Cardiovascular diseases and risk of hip fracture. *JAMA*. 2009 Oct 21;302(15):1666–73.

46. Nuti R, Siviero P, Maggi S, Guglielmi G, Caffarelli C, Crepaldi G, et al. Vertebral fractures in patients with chronic obstructive pulmonary disease: the EOLO Study. *Osteoporos Int*. 2009 Jun;20(6):989–98.

47. Ensrud KE, Thompson DE, Cauley JA, Nevitt MC, Kado DM, Hochberg MC, et al. Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. Fracture Intervention Trial Research Group. *J Am Geriatr Soc*. 2000 Mar;48(3):241–9.

48. Puisto V, Rissanen H, Heliövaara M, Impivaara O, Jalanko T, Kröger H, et al. Vertebral fracture and cause-specific mortality: a prospective population study of 3,210 men and 3,730 women with 30 years of follow-up. *Eur Spine J*. 2011 Dec;20(12):2181–6.

49. Masunari N, Fujiwara S, Kasagi F, Takahashi I, Yamada M, Nakamura T. Height loss starting in middle age predicts increased mortality in the elderly. *J Bone Miner Res*. 2012 Jan;27(1):138–45.

50. Wannamethee SG, Shaper AG, Lennon L, Whincup PH. Height loss in older men: associations with total mortality and incidence of cardiovascular disease. *Arch Intern Med*. 2006 Dec 11;166(22):2546–52.

51. Hillier TA, Lui L-Y, Kado DM, LeBlanc ES, Vesco KK, Bauer DC, et al. Height loss in older women: risk of hip fracture and mortality independent of vertebral fractures. *J Bone Miner Res*. 2012 Jan;27(1):153–9.

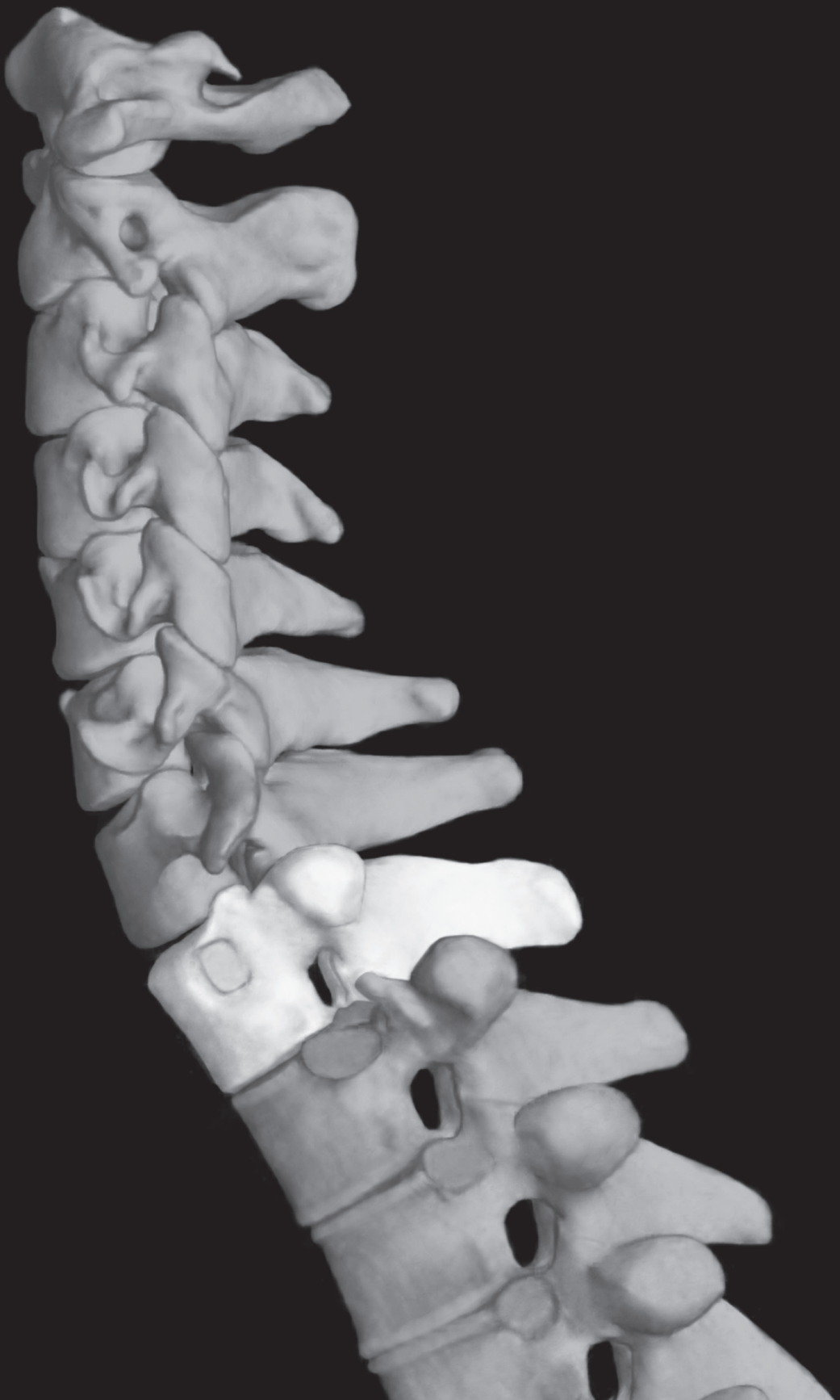
52. Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med*. 2013 Jul 18;369(3):245–54.

53. Young RP, Hopkins RJ. Chronic obstructive pulmonary disease detection during lung cancer screening. *JAMA*. 2012 Feb 15;307(7):664; author reply 664–665.

54. Young RP, Hopkins RJ, Midthun DE. Computed tomographic screening for lung cancer. *JAMA*. 2012 Oct 3;308(13):1320; author reply 1320–1321

55. Bazzocchi A, Spinnato P, Albisinni U, Battista G, Rossi C, Guglielmi G. A careful evaluation of scout CT lateral radiograph may prevent unreported vertebral fractures. *Eur J Radiol.* 2012 Sep;81(9):2353-7.

56. Bazzocchi A, Fuzzi F, Garzillo G, Diano D, Rimondi E, Merlino B et al. Reliability and accuracy of scout CT in the detection of vertebral fractures. *Br J Radiol*. 2013 Dec;86(1032):20130373.



Chapter 8

Vertebral fractures on routine chest computed tomography: relation with arterial calcifications and future cardiovascular events

Buckens CF
de Jong PA
Verkooijen HM
Verhaar HJ
Mali WP
van der Graaf Y
PROVIDI study group

Int J Cardiovasc Imaging. 2015 Feb;31(2):437-45

Abstract

Vertebral fractures on routine chest computed tomography: relation with arterial calcifications and future cardiovascular events

Purpose

Osteoporosis and cardiovascular disease often coexist. Vertebral fractures incidentally imaged in the course of routine care might be able to contribute to the prediction of cardiovascular events.

Methods

Following a case-cohort design, 5679 patients undergoing chest CT were followed for a median duration of 4.4 years. Cases were defined as patients who subsequently developed a cardiovascular event (n=493). The presence and severity of vertebral fractures, as well as aortic, coronary and valvular calcifications on CT were investigated.

Results

Cases were more likely to be male (69% versus 60%) and older (66 versus 61 years old). Prevalent vertebral fractures conferred an elevated risk of cardiovascular events after adjustment for age and gender (hazard ratio (HR) of 1.28, 95% confidence interval [CI] 1.07 to 1.54). This effect remained moderate after correction for cardiovascular calcifications (HR 1.20, CI 0.99 to 1.44). However, in terms of discrimination, vertebral fractures did not have substantial incremental prognostic value after correction (c-statistic was 0.683 versus 0.682 for models with and without vertebral fractures respectively).

Conclusions

Prevalent vertebral fractures on routine clinical chest CT are related to future cardiovascular events but do not have additional prognostic value to models that already include age, gender and cardiovascular calcifications.

Introduction

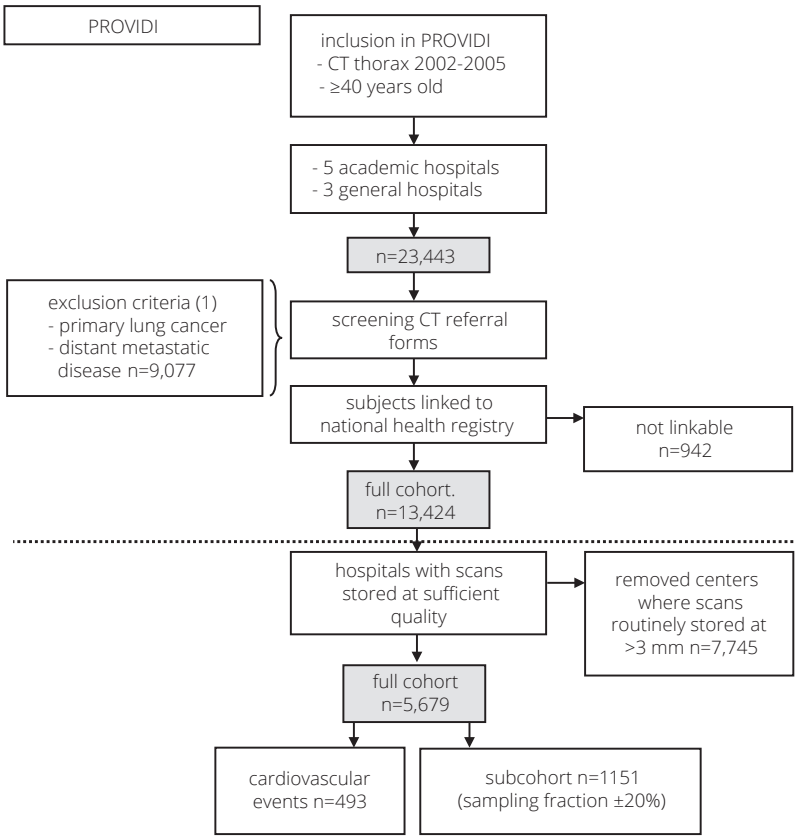
Symptomatic cardiovascular disease (CVD) and osteoporosis are major, preventable and growing health burdens.[1,2] Calcifications in the coronary arteries, aorta and on the cardiac valves are strong predictors for cardiovascular events. As predictors they are independent of other traditional risk factors.[3-5] There is evidence linking osteoporotic osseous demineralization to arterial calcifications -sometimes called the calcification paradox- although the mechanism is poorly understood.[6] Overlapping risk factors including age, smoking, inactivity, estrogen deficiency and chronic inflammation have been suggested as possible explanations,[7] along with shared endogenous calcium metabolism and bone regulatory molecules and pathways.[8] There is evidence that elevated levels of homocysteine and a deficiency in vitamin B12 contribute to osteoporosis[9-12] and to cardiovascular disease, particularly stroke.[13,14] Furthermore injury to the cervical spine can contribute to (vertebrobasilar) stroke.[15-17]

Both CVD and osteoporosis are best managed preventatively to avoid events such as stroke, myocardial infarction and hip fractures.[18] This approach that necessitates timely diagnosis. Interestingly statins, which are commonly prescribed for CVD prevention, may have a positive effect on bone mineralization[19] and prevent vertebral fractures.[20] Conversely, bisphosphonates, prescribed to slow bone demineralization, may inhibit arterial calcification.[21]

The increasingly widespread use of computed tomography (CT)[22] in routine care is providing new avenues for early identification of patients at risk.[23] Opportunistic risk stratification could serve to complement established approaches. Using information contained in available diagnostic imaging (performed for other conditions) does not require additional health care resources and imposes no additional burden or risk on patients. Prevalent vertebral fractures incidentally detected on routine clinical CT may contribute to CVD risk stratification. Prevalent (often asymptomatic) vertebral fractures are a relatively common finding on CT[24-27] and have been associated with all-cause mortality.[28] Together with low bone mineral density, they are also associated with CVD and cardiovascular mortality, especially in populations such as diabetics or patients who recently suffered a coronary event.[29-35] Conversely, cardiovascular calcifications on CT have also been associated with osteoporosis and fractures.[36-38] It remains unknown whether vertebral fractures are independently predictive of future CVD in routine clinical populations after taking account of other, equally accessible, cardiovascular calcifications on CT imaging.

The aim of this study was to determine whether vertebral fractures were predictive of future cardiovascular events and to quantify the incremental prognostic value, if any, of

Figure 1. PROVIDI flowchart



vertebral fractures for predicting CVD events, after accounting for other established risk factors, in a cohort of patients who underwent routine chest CT.

8

Materials and methods

PROVIDI

The present research was conducted in the context of the (Prognostic Value of unrequested Information on Diagnostic Imaging) PROVIDI study. This multicenter study aims to establish the prognostic value of unrequested findings on chest CT in routine clinical care and was described in detail elsewhere.[39] Briefly, it includes adult patients of whom chest CTs were

acquired in participating Dutch hospitals between 2002 and 2005 (Figure 1). Patients with primary lung cancer or distant metastatic disease of other origin were excluded due to their a priori poor prognosis and the attendant low likelihood that unrequested findings would alter their management, but otherwise all patients were included. The institutional ethics committees of the University Medical Center Utrecht approved the PROVIDI study and waived the need for written informed consent (decision number 06/193).

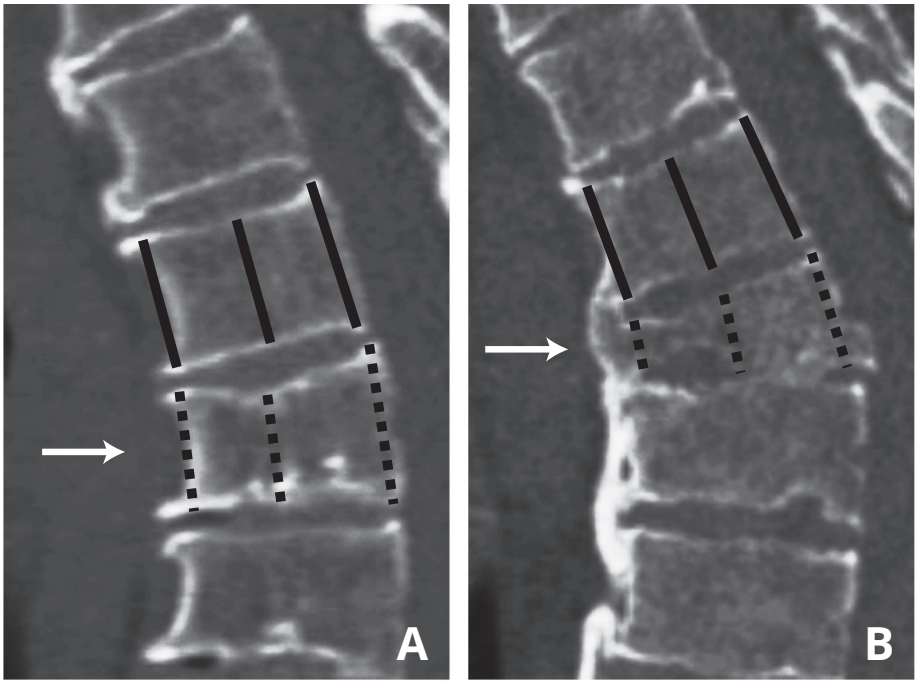
Since vertebral fracture assessment requires multiplanar reconstruction and hence thin CT slices, only patients from three out of eight participating centers were included. The excluded hospitals did not routinely store the thin slices during the study period. The cohort for the present study consisted of 5,679 patients, 3,315 from two tertiary centers and 2,364 from one peripheral hospital (Figure 1). Since the exclusion of the other five participating centers was not plausibly related to either the determinants nor the outcome (missing completely at random) no attempt was made to correct or adjust for the excluded centers.

Case-cohort study population

Following the case-cohort design, a random sample, i.e. 'subcohort', was taken from this study population (n=1151, 20%).[31] The case-cohort design reduces the cost of classic cohort design through the use of a subcohort. This is a completely random sample from the entire cohort at baseline and is usually a small fraction of the full cohort (typically below 10%[46]). cases identified during follow-up are added to the subcohort to form the case-cohort dataset. In this way all the available cases are included in the analysis. Since the number of cases is almost universally the group limiting the precision of the results (i.e. they are the smallest group) it is desirable to include them all. Excluding a portion of the non-cases has little impact on the precision since the group of non-cases in a cohort is so much larger than the group of cases. Since the subcohort is randomly sampled the majority of non-cases not included in the subcohort are missing completely at random. Excluding them in this way does not bias the results (since it is random) and since there are still more non-cases than cases it has a neglible impact on the precision of the results. Expensive and time consuming covariates, such as vertebral fracture and cardiovascular calcification assessment thus only need to be determined in the subcohort and the additional cases identified during follow-up. Cases were defined as all patients from the full study population who experienced a cardiovascular event during follow-up.

Fatal and non-fatal CVD events were obtained through linkage of subjects with the Dutch National Death Registry and the National Registry of Hospital Discharge Diagnoses from

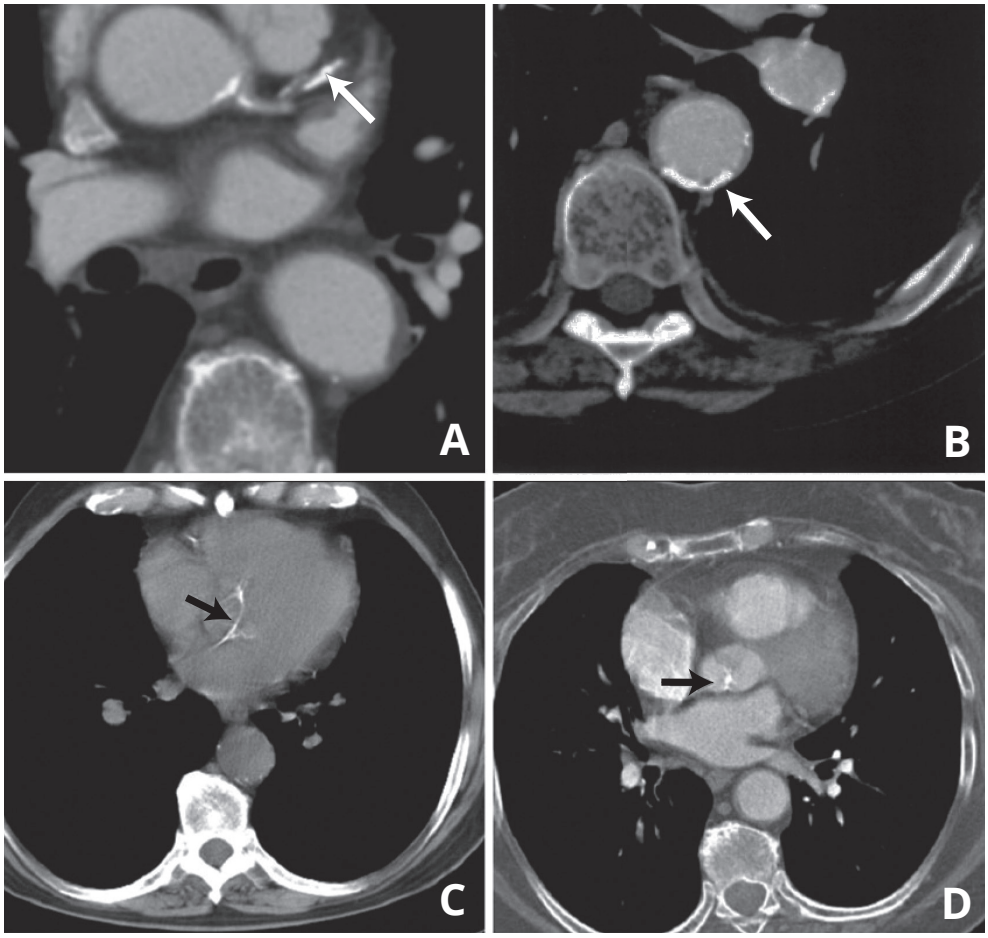
Figure 2. Vertebral fractures on sagittal view



A=sagittal view of portion of thoracic spine showing a moderate (grade 2) fracture
B=sagittal view of portion of thoracic spine showing a severe (grade 3) fracture

baseline to January 2008. Database linkage was performed with a validated probabilistic method.[40] In these national databases, cause of death and the indications for hospitalization are coded by physicians according to the International Classification of Diseases.[41,42] All events were classified using the 9th (discharge diagnoses) and 10th (cause of death) revision of the International Classification of Diseases. Hypertensive disease (codes 401-405), ischemic heart disease (codes 410-414), heart failure (code 428), peripheral vascular (codes 440-448), cerebrovascular disease (codes 430-438), or other heart disease (code 429) were included as cardiovascular events. Cardiovascular death prevailed over hospital admissions, so that when a cardiovascular cause of death was listed, that endpoint code and failure time was used. In cases of multiple valid hospital admissions the first to occur was used. When study subjects had more than one chest CT during follow-up, only the first eligible CT was evaluated. CTs were made with a range of different systems from different vendors and were stored at a maximum slice thickness of 3 mm.

Figure 3. Examples illustrating vascular calcifications



A=coronary arteries; B=descending aorta; C=affecting two leaflets of the aortic valve; D=affecting one leaflet

The full case cohort dataset consisted of the randomly sampled subcohort and additional cardiovascular cases outside the subcohort.

Vertebral measurements

Chest CT scans were scored for vertebral fractures. The reader was blinded for baseline patient characteristics and outcome status. CT scoring was performed at a research workstation (iX Viewer; Image Sciences Institute). Semiquantitative Vertebral Fracture

Assessment (VFA) similar to that widely applied elsewhere[43] was used to identify and classify vertebral fractures. This method identifies fractures according to the height loss of the vertebral body (viewed sagittally), with adjacent normal unfractured vertebrae providing comparison here. The vertebrae were assessed on sagittal reconstructions around the mid-sagittal point, with the rater free to adjust the window level, orientation and slice thickness of the reconstruction as desired. This method has been shown to be reliable for the identification of fractures and their severity.[44] The three fracture grades are: mild grade 1 fractures (height loss 20-25%), moderate grade 2 fractures (25-40%) and severe grade 3 fractures with height loss of more than 40% (Figure 2). Deformities that seemed non-fractural in origin (e.g. Schmorl's nodes, strongly scoliotic deformity or congenital anomalies) were not counted as fractures.

The presence of fracture was scored yes if there was a fracture visible. The worst fracture grade was defined as the grade of the worst fracture visible (either none, mild, moderate, or severe).

Vascular calcification measurements

CT scoring was performed at a research workstation (iX Viewer; Image Sciences Institute). The reader was blinded for patient characteristics and outcome status. Coronary artery calcifications, aortic wall calcifications, mitral valve or annulus calcifications and aortic valve calcifications were scored using a simple visual grading system[5] that has been shown to be reliable on CT in a routine setting.[45] Briefly, calcifications in the four main coronary arteries were categorized as; none, mild (1-2 focal [limited to ≤2 slices] calcifications), moderate (>2 focal calcifications or a single calcification extending for >2 slices) or severe (fully calcified coronary arteries extending over multiple segments). These were then summed (scores of 0-12).

Similarly, the number and size of calcifications in the wall of the descending aorta and ascending aorta were graded as: absent; grade 1, mild (≤3 focal calcifications); grade 2, moderate (4-5 focal calcifications or 1 calcification extending for ≥3 slices) and grade 3, severe (>5 focal calcifications or >1 calcification extending for ≥3 slices). Supra-aortic calcifications were scored as absent, present in one of the three supra-aortic arteries, or present in multiple arteries. Aortic calcifications were then also summed (0-8). Mitral valve calcifications were graded as follows: grade 0, absent; grade 1 single linear calcification; grade 2, two leaflets involved.[3] Aortic valve calcifications were categorized as absent, a single spot, a single line, linear calcifications on 2 cusps and linear calcifications on all 3 cusps (Figure 3).

Statistical analyses

Association of vertebral fractures with future cardiovascular events

A case-cohort appropriate cox modeling was used to assess the association between prevalent vertebral fractures and future cardiovascular events.[46,47] As discussed above, the case cohort dataset (random subcohort plus all cases outside the subcohort) is not biased and has a comparable precision to a full cohort analysis, at a fraction of the cost. However since cases are overrepresented in the dataset (by a factor inversely proportional to the sampling fraction of the subcohort), special adjustment is required to correct for this in the analysis. To this end approaches involving weighting of the different kinds of subjects in the case-cohort dataset (non-cases randomly selected for the subcohort, cases randomly selected for the subcohort and cases added after they had been identified during follow-up) have been developed and extensively validated. These have been shown to yield reliable results and standard errors, comparable to those in a full cohort analysis. essentially the weighting approach uses the subcohort to estimate the distribution of the covariates in the full cohort (as if it had been measured) and then to calculate the resulting baseline hazard. The cases outside the subcohort (those identified after follow-up) are then included in the model to calculate the relevant hazard ratios (without contributing to the baseline hazard function). After crude associations were estimated in univariate cox models, the variables age and gender were then added to generate a basically adjusted model. To investigate whether the relation between prevalent vertebral fractures and future cardiovascular events was explained by cardiovascular calcifications we added these to the cox model on top age and gender for the fully adjusted model.

Added prognostic value of vertebral fractures for cardiovascular events

Finally, in addition to examining the predictive effect of vertebral fractures, the added prognostic value of vertebral fractures, on top of age, gender and cardiovascular calcifications was assessed. These other findings are easily assessable on thoracic CT and are known to be strongly predictive of cardiovascular disease. This would quantify any added value of vertebral fractures would have for (unrequested) cardiovascular risk stratification if they are additionally assessed on a thoracic CT. To assess whole-model discrimination rather than single variables within the model, a survival-appropriate concordance statistic (c-statistic) was computed for each model that was adjusted based on the performance of the model in 100 bootstrap replicates.[47] Comparing the discriminative model performance with and without vertebral fracture status gives an indication of its incremental value when distinguishing patients at a high risk of CVD from those at a lower risk of CVD. The

Table 1. Baseline characteristics of cases and subcohort in the case-cohort sample of patients who underwent routine clinical chest CT

variable	description	cases (n=493)	subcohort (n=1151)
gender	female	152 (31%)	450 (39%)
age	mean in years (interquartile range)	66 (59–74)	62 (54–70)
follow-up	median in years (interquartile range)	4.3 (3.6–4.9)	4.4 (3.6–4.9)
academic center	patients from tertiary center	312 (63%)	818 (71%)
scan indication	lung	118 (24%)	225 (20%)
	cardiovascular	97 (20%)	210 (18%)
	malignancy	110 (22%)	243 (21%)
	other	168 (34%)	484 (42%)
vertebral fracture	patients with vertebral fracture	192 (39%)	366 (32%)
worst fracture grade	0	300 (61%)	785 (68%)
	1 (mild)	135 (27%)	249 (22%)
	2 (moderate)	33 (7%)	81 (7%)
	3 (severe)	25 (5%)	37 (3%)
cumulative fracture grade	0	300 (61%)	785 (68%)
	1-3	155 (31%)	305 (27%)
	4-6	29 (6%)	48 (4%)
	≥7	9 (2%)	13 (1%)
aortic calcification	0	92 (19%)	322 (28%)
	1 (mild)	136 (28%)	302 (26%)
	2 (moderate)	185 (38%)	324 (28%)
	3 (severe)	80 (16%)	205 (18%)
aortic valve calcifcation	none	273 (55%)	654 (57%)
	1 spot	29 (6%)	59 (5%)
	1 line	75 (15%)	111 (10%)
	linear 2 cusps	66 (13%)	119 (10%)
	linear 3 cusps	50 (10%)	208 (18%)
mitral valve calcification	0	321 (65%)	728 (63%)
	1 (mild)	117 (24%)	235 (20%)
	2 (moderate)	55 (11%)	188 (16%)
coronary artery calcification	0	89 (18%)	279 (24%)
	1 (mild)	179 (36%)	404 (35%)
	2 (moderate)	143 (29%)	231 (20%)
	3 (severe)	82 (17%)	237 (21%)

Table 2. Hazard ratios for cardiovascular event per vertebral fracture variable in patients who underwent routine chest CT

	cardiovascular events (n=493)		
hazard ratios (95% CI)	**crude**	**adjusted***	**fully adjusted†**
presence of vertebral fracture			
no	1	1	1
yes	1.37 (1.15-1.64)	1.28 (1.07-1.54)	1.20 (0.99-1.44)
worst fracture grade			
none	1	1	1
mild (20-25%)	1.42 (1.16-1.74)	1.34 (1.09-1.64)	1.27 (1.03-1.57)
moderate (25-40%)	1.04 (0.73-1.49)	1.00 (0.69-1.43)	1.05 (0.72-1.51)
severe (>40%)	1.94 (1.29-2.91)	1.58 (1.05-2.39)	1.29 (0.85-1.97)
cumulative fracture grade			
0	1	1	1
1-3	1.28 (1.05-1.56)	1.19 (0.98-1.45)	1.15 (0.94-1.40)
4-6	1.35 (0.92-1.98)	1.15 (0.79-1.69)	1.02 (0.69-1.51)
≥7	3.20 (1.66-6.17)	2.95 (1.53-5.70)	2.56 (1.30-5.05)

* for age and gender

† for age, gender, aortic calcifications, aortic valve calcifications, mitral valve calcifications and coronary artery calcifications

contribution of vertebral fracture assessment was further quantified using the adjusted Wald chi squared statistic for the analysis of deviance of each vertebral fracture term (fracture presence and worst fracture grade) and their associated p-value.[48]

Results

Study population and CT findings

During a median follow-up of 4.4 (interquartile range 3.6-5.0) years 493 (8.6%) patients suffered a cardiovascular event. These events included 134 myocardial infarctions, 70 strokes and 289 other events. Cases were more likely to be male (69 vs.61%) and older (66 vs. 62 years old), compared to the subcohort. Cases had more vertebral fractures (39 vs. 32%; Table 1) and a higher burden of coronary, aortic and cardiac valve calcifications.

Table 3. Prognostic value of vertebral fractures for cardiovascular events in a population who underwent routine chest CT

model	c-statistic	P-value
fracture yes/no		
crude	0.54 (0.56-0.61)	<0.01
adjusted for age and gender*	0.66 (0.63-0.69)	<0.01
fully adjusted**	0.68 (0.66-0.71)	0.06
worst fracture grade		
crude	0.54 (0.51-0.57)	<0.01
adjusted*	0.66 (0.63-0.69)	0.01
fully adjusted**	0.68 (0.66-0.71)	0.01
cumulative fracture grade		
crude	0.54 (0.51-0.56)	<0.01
adjusted for age and gender*	0.66 (0.63-0.69)	0.01
fully adjusted**	0.68 (0.66-0.71)	0.04
age and gender	0.66 (0.63-0.68)	NA‡
age, gender and calcifications	0.68 (0.66-0.71)	NA‡

data given are the concordance statistic (c-statistic) and the chi-square associated p-values for each variable
* for age and gender
** for age, gender, aortic calcifications, coronary calcifications, mitral valve calcifications and aortic valve calcifications.
‡ chi-squared p-value not applicable to multiple variables simultaneously.

Association of vertebral fractures with future cardiovascular events

Having at least one vertebral fracture was predictive of future cardiovascular events (HR: 1.37, CI: 1.15-1.64). After correction for age and gender this effect weakened, but remained statistically significant (HR: 1.28, 1.07-1.54). After further correction for vascular imaging findings the predictive effect of vertebral fractures was slightly reduced again (HR: 1.20, 0.99-1.44; Table 2) so that it was borderline significant. Appropriately the chi-squared associated p-value for fracture presence was marginally significant in the fully adjusted model (p=0.078).

A similar pattern was seen for worst fracture grade. The HRs associated with the different grades of vertebral fracture diminished somewhat after correction for age and gender, from 1.94 (1.29-2.91) for severe fractures to 1.58 (1.05-2.39). The HR was slightly lower after additional correction for vascular calcifications, to 1.29 (0.85-1.97) (Table 2).

Incremental prognostic value of vertebral fractures for cardiovascular events

In terms of model discrimination, the c-statistic values associated with the progressively more elaborate models also repeated this pattern. When fracture presence was added to a model already containing age, gender and cardiovascular calcifications, the bootstrap corrected c-statistic was only marginally affected, increasing from 0.68 (0.66-0.71) to 0.68 (0.66-0.71, Table 3).

Discussion

We found that prevalent vertebral fractures on routine clinical CT are associated with future cardiovascular events with an adjusted HR of 1.20 (0.99-1.44). This borderline significant result suggests that vertebral fractures are moderately predictive of future cardiovascular events independently of cardiovascular calcifications on CT. However, in terms of model discrimination, the incremental prognostic value of adding vertebral fractures to a model already containing age, gender and cardiovascular calcifications was modest. This suggests that vertebral fractures would not be a useful predictor of cardiovascular events in settings where cardiovascular calcifications may also be assessed.

Epidemiological publications on the association between osteoporosis, fractures and cardiovascular calcifications have yielded conflicting results. While some studies found clear associations,[29,49] others did not.[50-52] Physiologically, nature of the mechanism linking arterial calcification to osseous fragility remains unclear, despite the existence of a large body of basic science in this field.[9-17] There are suggestions that hyperlipidemia may play a role in osteoporosis[19,53,54] and there are also indications that overlapping bone mineralization signaling pathways may be deregulated in both disease clusters.[55] Estrogen deficiency and possibly inflammation and homocysteine have also been suggested.[56] How these factors interplay to cause increased bone formation in arteries and decreased mineralization in the skeleton remains fundamentally unanswered but this study helps to shed some light on the possible prognostic implications of this link. Bone structures, particularly the spine, are visualized on every cardiac CT and on all routine chest CTs, along with cardiovascular calcifications. As such assessing the spine yields gratis extra information with potential prognostic information for both osteoporosis and cardiovascular disease.

Whilst this study provides further evidence of an association between increased arterial calcification both in the aorta, coronary arteries and heart valves and vertebral fractures that is independent of age and gender, it also indicates that the effect is likely to be modest,

at least amongst a routine clinical population. Furthermore, our data suggest that vertebral fractures have at most a marginal incremental prognostic value on top of cardiovascular calcifications. As cardiovascular calcifications are also readily accessible on CT, there seems to be little prognostic gain in additionally assessing vertebral fractures when seeking to predict cardiovascular events. Further research on common pathophysiological mechanisms may in future identify new treatment targets and inform management strategies for both disease clusters.

Whilst the association we observed is statistically significant in this relatively large cohort, the results observed in this study are essentially negative in that they found at most a marginal added value of vertebral fractures in prognostic terms. The strength of this study is the large number of events providing us enough power to examine the independent prognostic value of vertebral fractures for CVD outcomes. By using a routine clinical population, rather than a more specialized and less generalizable study population, the results of this study are more likely to reflect the associations that may be expected in routine clinical care. The marginal gain in prognostic performance observed here thus indicates that systematic assessment of vertebral fractures is unlikely to contribute towards cardiovascular risk stratification in a general clinical setting.

Limitations

A limitation is that our study was unable to correlate the observed vertebral fractures to DXA-defined bone mineral density. DXA is currently still the preferred method for assessing osteoporosis. However, vertebral fractures and bone mineral density measured by DXA are strongly correlated.[57] This suggests that bone mineral density would probably show a similar association to cardiovascular events. Thirdly, as chest CT was used, there was no information on cervical and mid and lower lumbar fractures. This may have caused an underreporting of vertebral fractures in this cohort and the prognostic effect of vertebral fractures may be in fact be larger if the whole spine was included. However, in the clinical setting of dedicated cardiac CT or routine or screening chest CT the lumber spine will also not be visualized and our findings can be expected to apply.

In conclusion, prevalent vertebral fractures have moderate predictive power for future cardiovascular events, after adjustment for age, gender and cardiovascular calcifications also visible on thoracic CT, in a routine clinical population. The incremental improvement offered by vertebral fractures here to the predictive power of cardiovascular prediction models is slight.

Acknowledgments

The PROVIDI Study Group consists of J. Laméris (Department of Radiology, Academic Medical Center, Amsterdam), C. van Kuijk (Department of Radiology, VU University Medical Center Amsterdam), W. ten Hove (Department of Radiology, Gelre Hospitals, Apeldoorn), M. Oudkerk (Department of Radiology, University Medical Center Groningen), Ay L. Oen (Department of Radiology, Elkerliek Hospital, Helmond), J. Wildberger (Department of Radiology, Academic Hospital Maastricht), J. van Heesewijk (Department of Radiology, St Antonius Hospital, Nieuwegein), W. Mali (Department of Radiology, University Medical Center Utrecht), and Y. van der Graaf (Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht).

Sources of Funding: This study was funded by a program grant from the Netherlands Organization for Scientific Research-Medical Sciences (NWO-MW); grant 40-00812-98-07-005. The funder had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Substudies from the PROVIDI cohort have been previously published, including substudies using cardiovascular endpoints. These patient groups partly overlap with the current manuscript. The current manuscript includes a different subset of patients from the previous studies and investigates a different set of covariates, most notably vertebral fractures.

References

1. Alwan A, World Health Organization. Global status report on noncommunicable diseases 2010. World Health Organization: Geneva, Switzerland, 2011.
2. Sambrook P, Cooper C. Osteoporosis. *Lancet*. 2006; 367: 2010–2018.
3. Gondrie MJA, van der Graaf Y, Jacobs PC, Oen AL, Mali WPTM. The association of incidentally detected heart valve calcification with future cardiovascular events. *Eur Radiol*. 2011; 21: 963–973.
4. Gondrie MJA, Mali WPTM, Jacobs PC, Oen AL, van der Graaf Y. Cardiovascular disease: prediction with ancillary aortic findings on chest CT scans in routine practice. *Radiology*. 2010; 257: 549–559.
5. Jacobs PC, Gondrie MJ, Mali WP, Oen AL, Prokop M, Grobbee DE et al. Unrequested information from routine diagnostic chest CT predicts future cardiovascular events. *Eur Radiol*. 2011; 21: 1577–1585.
6. Persy V, D'Haese P. Vascular calcification and bone disease: the calcification paradox. *Trends Mol Med*. 2009; 15: 405–416.
7. Doherty TM, Fitzpatrick LA, Inoue D, Qiao J-H, Fishbein MC, Detrano RC et al. Molecular, endocrine, and genetic mechanisms of arterial calcification. *Endocr Rev*. 2004; 25: 629–672.
8. Lampropoulos CE, Papaioannou I, D'Cruz DP. Osteoporosis--a risk factor for cardiovascular disease? *Nat Rev Rheumatol*. 2012; 8: 587–598.
9. Enneman AW, Swart KMA, Zillikens MC, van Dijk SC, van Wijngaarden JP, Brouwer-Brolsma EM et al. The association between plasma homocysteine levels and bone quality and bone mineral density parameters in older persons. *Bone*. 2014; 63: 141–146.
10. Zhang H, Tao X, Wu J. Association of homocysteine, vitamin B12, and folate with bone mineral density in postmenopausal women: a meta-analysis. *Arch Gynecol Obstet*. 2014; 289: 1003–1009.
11. McLean RR, Jacques PF, Selhub J, Fredman L, Tucker KL, Samelson EJ et al. Plasma B vitamins, homocysteine, and their relation with bone loss and hip fracture in elderly men and women. *J Clin Endocrinol Metab*. 2008; 93: 2206–2212.
12. McLean RR, Jacques PF, Selhub J, Tucker KL, Samelson EJ, Broe KE et al. Homocysteine as a predictive factor for hip fracture in older persons. *N Engl J Med*. 2004; 350: 2042–2049.
13. Spence JD. Homocysteine-lowering therapy: a role in stroke prevention? *Lancet Neurol*. 2007; 6: 830–838.
14. Spence JD. B vitamin therapy for homocysteine: renal function and vitamin B12 determine cardiovascular outcomes. *Clin Chem Lab Med*. 2013; 51: 633–637.
15. Smith WS, Johnston SC, Skalabrin EJ, Weaver M, Azari P, Albers GW et al. Spinal manipulative therapy is an independent risk factor for vertebral artery dissection. *Neurology*. 2003; 60: 1424–1428.
16. Beaudry M, Spence JD. Motor vehicle accidents: the most common cause of traumatic vertebrobasilar ischemia. *Can J Neurol Sci J Can Sci Neurol*. 2003; 30: 320–325.

17. Miley ML, Wellik KE, Wingerchuk DM, Demaerschalk BM. Does cervical manipulative therapy cause vertebral artery dissection and stroke? *The Neurologist*. 2008; 14: 66–73.

18. Krumholz HM, Weintraub WS, Bradford WD, Heidenreich PA, Mark DB, Paltiel AD. Task force #2--the cost of prevention: can we afford it? Can we afford not to do it? 33rd Bethesda Conference. *J Am Coll Cardiol*. 2002; 40: 603–615.

19. Uzzan B, Cohen R, Nicolas P, Cucherat M, Perret G-Y. Effects of statins on bone mineral density: a meta-analysis of clinical studies. *Bone*. 2007; 40: 1581–1587.

20. Schoofs MWCJ, Sturkenboom MCJM, van der Klift M, Hofman A, Pols HAP, Stricker BHC. HMG-CoA reductase inhibitors and the risk of vertebral fracture. *J Bone Miner Res*. 2004; 19: 1525–1530.

21. Santos LL, Cavalcanti TB, Bandeira FA. Vascular effects of bisphosphonates-a systematic review. *Clin Med Insights Endocrinol Diabetes*. 2012; 5: 47–54.

22. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *JAMA*. 2012; 307: 2400–2409.

23. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. *JAMA*. 2012; 308: 1433–1434.

24. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N et al. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009; 69: 555–559.

25. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008; 63: 160–164.

26. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009; 69: 179–183.

27. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008; 18: 1696–1702.

28. Hasserius R, Karlsson MK, Nilsson BE, Redlund-Johnell I, Johnell O. Prevalent vertebral deformities predict increased mortality and increased fracture rate in both men and women: a 10-year population-based study of 598 individuals from the Swedish cohort in the European Vertebral Osteoporosis Study. *Osteoporos Int*. 2003; 14: 61–68.

29. Bandeira E, Neves AP, Costa C, Bandeira F. Association between vascular calcification and osteoporosis in men with type 2 diabetes. *J Clin Densitom*. 2012; 15: 55–60.

30. Silva HC, Pinheiro MM, Genaro PS, Castro CHM, Monteiro CMC, Fonseca FAH et al. Higher prevalence of morphometric vertebral fractures in patients with recent coronary events independently of BMD measurements. *Bone*. 2013; 52: 562–567.

31. Mazziotti G, Baracca M, Doga M, Porcelli T, Vescovi PP, Giustina A. Prevalence of thoracic vertebral fractures in hospitalized elderly patients with heart failure. *Eur J Endocrinol*. 2012; 167: 865–872.

32. Lai S-W, Liao K-F, Lai H-C, Tsai P-Y, Lin C-L, Chen P-C et al. Risk of major osteoporotic fracture after cardiovascular disease: a population-based cohort study in Taiwan. *J Epidemiol*. 2013; 23: 109–114.

33. Chung DJ, Choi HJ, Chung Y-S, Lim SK, Yang S-O, Shin CS. The prevalence and risk factors of vertebral fractures in Korean patients with type 2 diabetes. *J Bone Miner Metab* .2013; 31: 161–168.

34. Den Uyl D, Nurmohamed MT, van Tuyl LH, Raterman HG, Lems WF. (Sub)clinical cardiovascular disease is associated with increased bone loss and fracture risk; a systematic review of the association between cardiovascular disease and osteoporosis. *Arthritis Res Ther*. 2011; 13: R5.

35. Von der Recke P, Hansen MA, Hassager C. The association between low bone mass at the menopause and cardiovascular mortality. *Am J Med*. 1999; 106: 273–278.

36. Schulz E, Arfai K, Liu X, Sayre J, Gilsanz V. Aortic calcification and the risk of osteoporosis and fractures. *J Clin Endocrinol Metab*. 2004; 89: 4246–4253.

37. Szulc P, Kiel DP, Delmas PD. Calcifications in the abdominal aorta predict fractures in men: MINOS study. *J Bone Miner Res*. 2008; 23: 95–102.

38. Naves M, Rodríguez-García M, Díaz-López JB, Gómez-Alonso C, Cannata-Andía JB. Progression of vascular calcifications is associated with greater bone loss and increased bone fractures. *Osteoporos Int*. 2008; 19: 1161–1166.

39. Gondrie MJA, Mali WPTM, Buckens CFM, Jacobs PCA, Grobbee DE, van der Graaf Y. The PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) Study: rationale and design. *Eur J Epidemiol*. 2010; 25: 751–758.

40. De Bruin A, Karduan J, Gast F, De Bruin E, van Sijl M, Verweij G. Record linkage of hospital discharge register with population register: experiences at Statistics Netherlands. *Stat J UN Econ Comm Eur*. 2004; 1: 23–32.

41. World Health Organization. International statistical classification of diseases and related health problems. World Health Organization: Geneva, 2011.

42. American Medical Association, Medicode (Firm). ICD-9-CM: international classification of diseases, 9th revision, clinical modification, volumes 1 and 2, color-coded, illustrated, 1999. American Medical Association: Dover, DE, 1998.

43. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993; 8: 1137–1148.

44. Buckens CF, de Jong PA, Mol C, Bakker E, Stallman HP, Mali WP et al. Intra and Interobserver Reliability and Agreement of Semiquantitative Vertebral Fracture Assessment on Chest Computed Tomography. *PLoS ONE*. 2013; 8: e71204.

45. Jacobs PCA, Isgum I, Gondrie MJA, Mali WPTM, van Ginneken B, Prokop M et al. Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement. *AJR Am J Roentgenol*. 2010; 194: 1244–1249.

46. Onland-Moret NC, van der A DL, van der Schouw YT, Buschers W, Elias SG, van Gils CH et al. Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol*. 2007; 60: 350–355.

47. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15: 361–387.

48. Graubard BI, Korn EL. Hypothesis Testing with Complex Survey Data: The Use of Classical Quadratic Test Statistics with Particular Reference to Regression Problems. *J Am Stat Assoc*. 1993; 88: 629–641.

49. Kammerer CM, Dualan AA, Samollow PB, Périssé ARS, Bauer RL, MacCluer JW et al. Bone mineral density, carotid artery intimal medial thickness, and the vitamin D receptor BsmI polymorphism in Mexican American women. *Calcif Tissue Int*. 2004; 75: 292–298.

50. Samelson EJ, Cupples LA, Broe KE, Hannan MT, O'Donnell CJ, Kiel DP. Vascular calcification in middle age and long-term risk of hip fracture: the Framingham Study. *J Bone Miner Res*. 2007; 22: 1449–1454.

51. Sinnott B, Syed I, Sevrukov A, Barengolts E. Coronary calcification and osteoporosis in men and postmenopausal women are independent processes associated with aging. *Calcif Tissue Int*. 2006; 78: 195–202.

52. Mussolino ME, Madans JH, Gillum RF. Bone mineral density and stroke. *Stroke J Cereb Circ*. 2003; 34: e20–22.

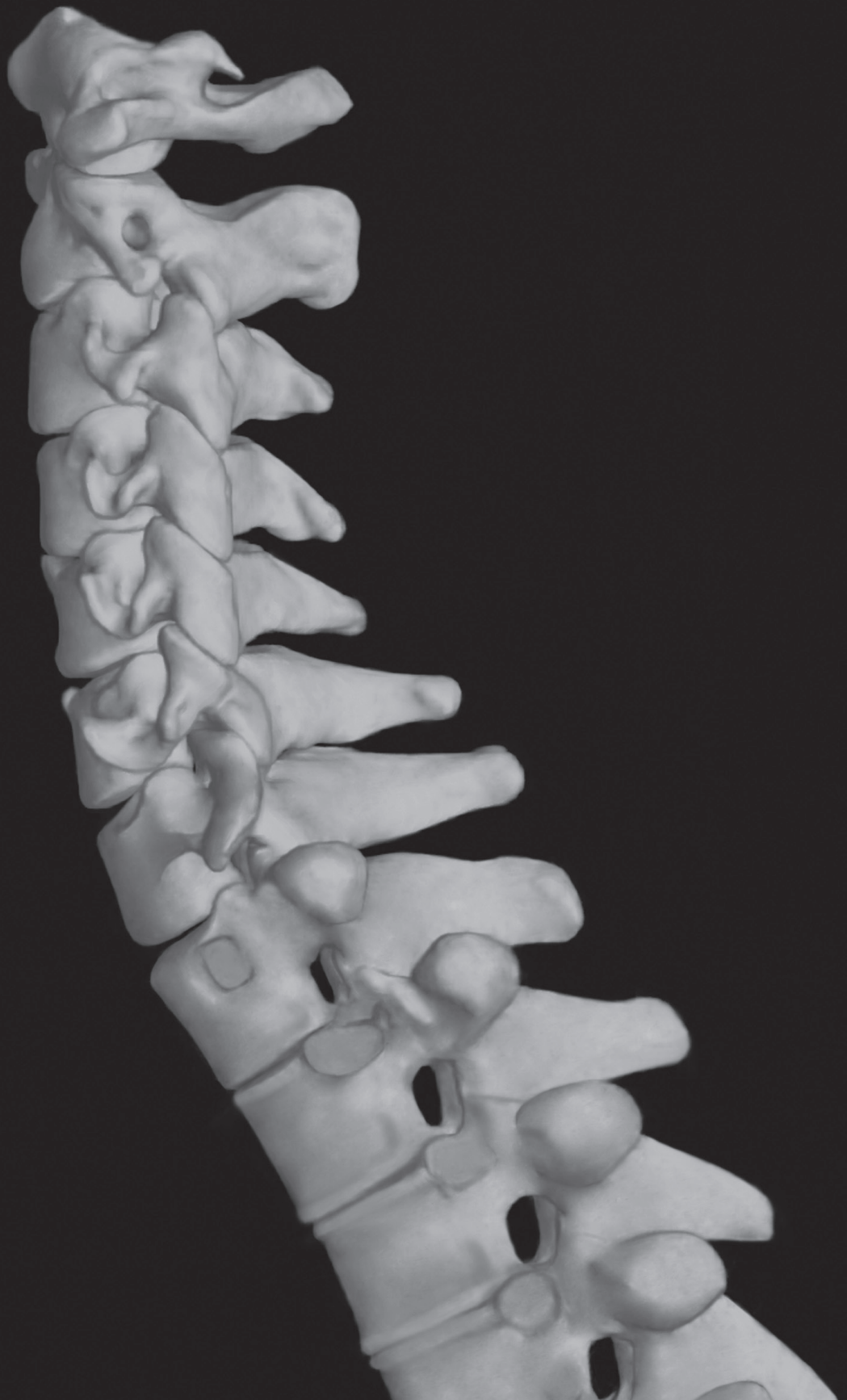
53. Parhami F. Possible role of oxidized lipids in osteoporosis: could hyperlipidemia be a risk factor? *Prostaglandins Leukot Essent Fatty Acids*. 2003; 68: 373–378.

54. Danilevicius CF, Lopes JB, Pereira RMR. Bone metabolism and vascular calcification. *Braz J Med Biol Res*. 2007; 40: 435–442.

55. Hofbauer LC, Schoppet M. Clinical implications of the osteoprotegerin/RANKL/RANK system for bone and vascular diseases. *JAMA*. 2004; 292: 490–495.

56. Koshiyama H, Ogawa Y, Tanaka K, Tanaka I. The unified hypothesis of interactions among the bone, adipose and vascular systems: 'osteo-lipo-vascular interactions'. *Med Hypotheses*. 2006; 66: 960–963.

57. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med*. 2013; 158: 588–595.



Chapter 9

Summary and discussion

Summary

Vertebral fractures are an early sign of the osseous fragility and the micro-architectural deterioration that characterizes osteoporosis.[1] Osteoporosis is a large and growing problem, with aging populations increasingly plagued by osteoporotic fractures and the attendant morbidity and mortality.[2] The disease causes an estimated 8.9 million fractures worldwide annually, 4.5 million in the developed west. The lifetime risk for wrist, hip or vertebral fracture (the cardinal locations for fragility fractures) for populations in developed countries is in the order of 30-40%. The morbidity associated with osteoporosis accounts for an estimated 2.8 million disability adjusted life years annually.[3]

Preventative treatments for osteoporosis are available but a cure is not. This makes early detection important in managing the disease; the sooner that preventative treatment is initiated in the course of the disease, the more bone deterioration and fracture risk that can be prevented. Nonetheless it remains under diagnosed and subsequently undertreated.[4,5] Vertebral fractures and vertebral deformities are an early marker of the micro-architectural deterioration that predates frank demineralization. It is this demineralization that has been historically used to define and detect osteoporosis[6] in the form of the DXA score. Consequently, incidentally found vertebral fractures could contribute towards earlier detection of patients in the early stages of osteoporosis that might benefit the most from preventative and anti-resorptive treatments, complementing the role of DXA. The improving resolution of CT imaging, improved processing and display technologies and its widespread availability and its increased use represent an opportune avenue through which to achieve this.

We approached the prognostic problem of vertebral fractures using the case-cohort design. In **chapter 2** we compare different readily implementable analysis approaches to the more established weighting analyzes used in this manuscript and show that they can yield reliably valid results.

The case-cohort design implemented in the PROVIDI study involves taking a random sample from the entire cohort at baseline, termed the subcohort.[7] This is usually a small fraction of the full cohort. Unlike the better known and more often used case-control study absolute risks of an outcome occurring can be estimated, rather than relative risks. Absolute risks are more informative to patient management and health care policy decisions. Additionally since the selection of the subcohort is not dependent on either the outcome or the covariates under study, the subcohort can be used to study multiple outcomes, as in PROVIDI.

In the case-cohort design all cases that occur during follow-up are then added to the subcohort to make up the case-cohort dataset. Expensive covariates that can be retrospectively determined (such as vertebral fractures in this manuscript) only need to be evaluated in the case-cohort dataset (all cases and the subcohort). Since all the cases are included in the case-cohort dataset and analysis, maximum use is made of their valuable and precision-limiting covariates.

The growing complexity of prognostic research has led to the development of a large number of complex statistical tools to fit, internally validate and assess prognostic models. These tools are almost universally designed with the full cohort in mind and usually need to be individually adapted before they can be applied to a case-cohort dataset with its intentionally large proportion of cases, a very complex and time-consuming task. This reduces the appeal of the case-cohort approach for many researchers. We have compared two pragmatic and easy to implement approaches, one involving simple sampling and one using modern imputation methods to the more established weighting approaches. These can be applied to case-cohort datasets allowing them to be passed through standard prognostic statistical tools without introducing bias in the overall performance of the resulting models. The imputation approach has the added benefit that it makes use of all the baseline covariates that are collected in all cohort members. These are discarded in the standard case-cohort analysis.

In part to demarcate the scope of the problem we investigated the prevalence of clinically relevant unrequested findings on cardiac CT in **chapter 3**. In all we found nineteen radiological studies comprising 12922 patients. Using appropriate random-effects to combine the results, we found that the pooled prevalence of clinically relevant unrequested extra-cardiac findings (variously defined as those requiring immediate action or further follow-up) was high, at 13%, but with a large variation between the different studies: 95% confidence interval 9–18, range: 3–39%. The findings varied widely from lymphadenopathy, aneurysms of the thoracic great vessels, pulmonary nodules. Vertebral fractures were infrequently tabulated separately in most articles. Cardiac calcifications were not included but extra-cardiac calcifications were included in those articles where they were thought to be clinically relevant by the authors. We were unable to explain these differences using reported study population parameters. We posit that these differences may be due to unreported or partly reported factors such as different definitions of clinical relevance and unreported differences between populations. To begin addressing this discrepancy we made suggestions for basic reporting to improve the interpretability and comparability of future research. Specifically we recommended reporting the number of each unrequested finding observed (both number of lesions and number of patients) and

how each was classified. We recommended adhering to the STROBE checklist, reporting the referral sources of the population assessed and reporting the prevalence of the major cardiovascular risk factors.

In **chapter 4** we sought to establish the reliability and agreement of vertebral fracture assessment on routine clinical CT. We found acceptable reproducibility of vertebral fracture assessment with Cohen's kappas of 0.73-0.84 (intraobserver) and 0.56-0.81 (interobserver) for the presence of fracture. Using an established semi-quantitative grading scheme,[7] we also assessed the severity of vertebral fractures present and recorded the worst fracture grade for each patient. We found good intraobserver (76-88%) and interobserver (74-88%) agreement, and excellent reliability with square-weighted kappas of 0.84-0.90 (intraobserver) and 0.84-0.94 (interobserver). These findings show that Genant's vertebral fracture assessment can be reliably and reproducibly applied to CT to assess the presence and severity of fracture.

Hip fracture is the most important osteoporotic fragility fracture.[8] They are associated with major morbidity and mortality.[9] Preventing hip fractures is a major goal for osteoporosis treatment. In **chapter 5** we used the PROVIDI cohort to investigate whether subclinical vertebral fractures on sagittal reconstruction of routine chest CT are associated with future hip fractures. In line with previous literature we found that cases were older and more often female. Compared with those with no vertebral fracture, patients with any vertebral fracture had a roughly tripled the risk of future hip fracture. We found an age- and gender-adjusted hazard ratio of 3.1 (95% CI 2.1-4.7). This hazard ratio rose to 3.8 (CI 2.6-5.6) if mild fractures were discounted, as they sometimes are in the literature. Unsurprisingly more severe fractures increased future hip fracture risk significantly. Mild fractures showed a HR of 2.4 (CI 1.5-3.7), moderate fractures a HR of 4.8 (CI 2.5-9.2) and severe fractures a HR of 6.7 (CI 2.9-15.5). The same was true for having higher cumulative fracture grades: a cumulative grade of 1 to 3 imparted a HR of 2.7, relative to having no fracture (CI 1.8-4.1), a grade of 4 to 6 a HR of 4.8 (CI 2.2-10.5) and a cumulative grade of ≥7 saw patients saddled with an HR of 11.2 (CI 3.7-34.6). These hazard rates are largely in line with previous literature in more specific study settings but had not previously been shown in a general radiological clinical setting. These results show that vertebral fracture status in general radiological practice is related to future risk of hip fracture. Whether radiologists can contribute to reducing this risk through systematic reporting of vertebral fracture status is dependent on a number of externalities; including the role of such reporting is given in the care pathway and to what extent that risk is modifiable in that population (most osteoporosis research has focused more narrowly on post-menopausal women and geriatric populations).

Recent work has shown that vertebral density on CT correlates well with DXA defined WHO bone mineral density classification grades,[10] and indeed DXA scores themselves,[11] The WHO uses bone mineral density as measured by DXA to classify patients as either having normal bone density, being osteopenic (a bone mineral density in the lumbar spine or hip between -1 and -2.5 standard deviations below healthy peak BMD[6]) or osteoporotic (a BMD below 2.5 standard deviations). CT that visualizes the spine contains information on the trabecular bone mineral content of the vertebrae; bone with higher mineral content and denser trabecular architecture absorbs more radiation. This X-ray attenuation measure is expressed using Hounsfield Units (HU)-named for the electrical engineer who played an important part in the initial development of CT.[12] The correlation of vertebral density on CT with future fracture risk is not known but the close correlation with DXA makes it appealing as much previous work has been conducted on the basis of DXA values. Theoretically lowered HU values as a consequence of sparser trabecularization of bone might manifest at a very early stage of osteoporosis, before vertebral fractures occur. This has not yet been demonstrated.

Being able to use routine clinical CT to infer bone density, as well as vertebral fracture, is a potentially valuable addition to improving screening patients who might benefit from early fracture prevention. In **chapter 6** we performed an external validation study of the diagnostic performance of vertebral density and we also assessed the relation of prevalent vertebral fracture with vertebral density and DXA score. We investigated the diagnostic performance of vertebral density for diagnosing DXA measures in the osteoporotic range. For this we included 302 patients with a mean age of 57.9 years of which 82 (27%) had osteoporosis according to DXA and 65 (22%) had vertebral fractures. The diagnostic performance for vertebral HU measurements was modest, with a substantial AUC of 0.74 (0.68–0.80). At that optimal threshold the sensitivity was 62% (51-72%) and the specificity was 79% (74-84%). This confirmed the diagnostic value of vertebral density for low BMD, albeit with a lower AUC than the 0.83 (95% CI: 0.81–0.85) that had been previously reported.[10] Properly conducted external validation studies almost always show a reduced performance compared to the derivation sample, the value of external validation studies lies in their providing an indication for the magnitude of this decrease.

External validation studies in independent populations are crucial in scientific research in general but medical research in particular as they can give a realistic assessment of the expected performance of new insights in populations other than the specific study-population. Particularly in the case of clinical diagnostic and prognostic models, external validation is a prerequisite before implementation of a new diagnostic or prognostic model.[13] Such attempts at external replication and validation are seldom undertaken, and when they are they routinely receive much less academic attention than the initial eye-

catching results. The principle of "trust, but verify" that underpins the scientific method is put into practice too infrequently.[14]

In **chapter 7** we investigated the prognostic value of vertebral fractures for all-cause mortality using the Dutch Belgian Lung Cancer Screening Trial (known by its Dutch-language acronym the NELSON-trial). Results from the National Lung cancer Screening Trial have shown that lung cancer screening using low dose chest CT can reduce lung cancer mortality amongst smokers.[15] These findings have tipped the debate around implementing lung cancer screening towards eventual implementation,[16] with debate hinging mainly upon cost-benefit ratios.[17] The benefits of screening would be amplified by additionally assessing other findings with prognostic value, beyond pulmonary nodules. The smoking behavior that predisposes the intended screening population to higher risks of lung-cancer also predisposes them to COPD, cardiovascular disease and osteoporosis, making the ancillary benefits of additionally screening for non-lung cancer diseases potentially substantial. Indeed, recent work has investigated the feasibility and performance of pulmonary emphysema assessment on low dose lung-cancer screening CT in smokers who are otherwise healthy[18] and found it to hold promise for detecting COPD, whilst arterial and coronary calcifications have been shown to predict mortality due to cardiovascular event.[19,20]

As low bone mineral density[21] and vertebral fracture[22–24] have been identified as predictors of survival in other settings, and smokers are thought to be at an increased risk for both[25,26] we hypothesized that vertebral density assessment and vertebral fracture assessment might meaningfully predict all-cause mortality in lung cancer screening population of smokers.

We assessed vertebral density in Hounsfield Units (HU) and vertebral fracture status in a subset of the NELSON data. As discussed above, vertebral trabecular density is strongly associated with DXA-measured BMD.[10,11] Mortality was recorded during a median follow-up of 6 years. We found that the prevalence of vertebral fractures was high at 35% (CI 30–40%) among survivors but even higher amongst those who had expired during follow-up, at 51% (44–58%). To evaluate whether vertebral fractures were independently predictive of mortality we adjusted for age, gender, smoking status, pack years smoked, coronary and aortic calcium volume and pulmonary emphysema. The adjusted HR we found for vertebral fracture was 2.04 (1.43–2.92). For each 10 HU decline in trabecular bone density the adjusted HR was 1.08 (1.02–1.15). These results showed that vertebral fractures and bone density are associated with all-cause mortality independent of age, gender, smoking characteristics, coronary and aortic calcifications and emphysema amongst lung cancer screening participants.

The value of sub- or pre-clinical findings of prognostic value is even more germane in the setting of CT-based lung cancer screening, where cost-benefit ratios play a central role in deciding whether to implement screening in smokers with an elevated risk or not. These smokers tend to suffer from many co-morbidities, including osteoporosis, cardiovascular disease[27,28] and COPD.[18] The early detection of a higher risk for cardiovascular disease and osteoporosis in this pre-clinical population has great potential for reducing the burden of these chronic, irreversible diseases. This is a central advantage and opportunity of modern imaging: it is able to directly visualize the end organ damage caused by disease processes.

Cardiovascular disease is a major and preventable health burden. The arterial calcium depositions that characterize cardiovascular disease and the osseous demineralization that characterizes osteoporosis often coexist, although the mechanism is poorly understood.[29–31] In **chapter 8** we investigated whether vertebral fractures are independently predictive of future cardiovascular events. With a median follow-up of 4.4 years 4 (interquartile range 3.6-5.0) we were only able to assess the immediate, short term cardiovascular events. Chronic cardiovascular disease falls beyond this follow-up duration. We found that prevalent vertebral fractures conferred an elevated risk of cardiovascular events after adjustment for age and gender with an HR of 1.28 (CI 1.07-1.54). This effect remained moderate after correction for cardiovascular calcifications (HR 1.20, CI 0.99 to 1.44). However, in terms of discrimination, vertebral fractures did not have substantial incremental prognostic value after correction (c-statistic was 0.683 versus 0.682 for models with and without vertebral fractures respectively).

Prevalent vertebral fractures on routine clinical chest CT are related to future cardiovascular events but do not have additional prognostic value to models that already include age, gender and cardiovascular calcifications.

Discussion

Unrequested findings of indeterminate prognostic value are proliferating quickly. This is a side-effect of the continuing successful innovation in medical imaging that has led to increased temporal and spatial resolution as well as better availability. This thesis has focused on unrequested information from vertebrae using the case cohort approach. This approach can also easily be applied to other findings of potential value. The finding in question and the manner in which it may be detected and graded needs to be unambiguously defined, as well as the outcome on which it may potentially have an impact (for instance vertebral fracture, semiquantitatively graded, and related to future hip fracture). Association studies and multivariate models can then be developed from longitudinal data

and externally validated. The crucial follow up step is to then conduct an impact study to investigate the improvement in patient outcomes achieved by systematically reporting and acting on the finding. For such research to be successful it needs to be perceived as being useful for patients and a close cooperation between radiologists and referring clinicians is required.

Communicating vertebral fracture status

Broadly speaking systematically reporting unrequested findings implies a paradigm shift in the thinking about the role of the radiologist in the care-pathway, from one of supporting referring clinicians only by addressing the clinician's direct imaging questions, towards taking a larger stake in maintaining the overall quality of patient care. Whilst radiologists have long reported clinically salient incidental findings on imaging investigations, the aforementioned improvements in imaging techniques has created a new class of findings with long-term clinical implications, such as vertebral fractures, vertebral bone density and cardiovascular calcifications. Currently the systematic reporting of unrequested findings on diagnostic imaging such as prevalent vertebral fractures, that predispose to future disease, but which are usually clinically silent in themselves, remains inconsistent in radiological practice.[32–36] As well as improving the quality of patient care, systematic assessment of unrequested findings can help to make the added value of the radiology service more explicit.

There are three plausible possibilities for integrating such reporting into the care pathway: directing the report towards the referring clinician, the patient's primary care physician or towards the patients themselves. Each possibility brings its own advantages and complications.

Referring clinicians receiving risk stratification information based on unrequested findings does not require the current reporting pathways to be amended- radiologists can simply append the unrequested findings to their reports. The potential complication lies in the fact that clinicians will receive addenda not related to their field of expertise. A pulmonologist receiving a warning about vertebral fractures, for instance, could consider referring the patient to a rheumatologist for DXA screening and fracture prevention management. For these findings to have impact the referring clinicians must be able and willing to act upon the unrequested findings appropriately, by referring their patients to appropriate specialist or performing indicated screening or treatment themselves. Issues pertaining to reimbursement and the division of responsibilities would have to be carefully defined before implementation.[37]

If systematic unrequested findings and any attendant risk stratification was instead sent to the primary care physician some of these complications could be avoided. Primary care physicians are broadly responsible for the overall care of their patients and are not limited to one area of expertise. They are especially involved in preventative care such as smoking cessation and cardiovascular risk management. As such they are more likely to be receptive to receiving unrequested information pertaining to a long-term prognostic factor. Since they maintain overarching oversight over the care of their patients they are better placed to assess whether the new prognostic information across medical specialties warrants new management, further screening or whether treatments that have already been initiated need to expanded. This approach would require that a new reporting pathway be created specifically for the systematic reporting of unrequested findings (currently all findings are communicated to the referring clinician. The referring clinician will then in turn communicate with the primary care physician, choosing to include all, some or none of the imaging findings described in the radiology report).

Communicating the unrequested findings to patients represents a paradigm shift in the reporting pathway. Radiologists could directly report the absence or presence of prognostically important unrequested findings in easily understandable layman's terms. Patients themselves could then decide which action they consider most appropriate. This would make the radiologist more visible to the primary beneficiary of the imaging study and place the responsibility for acting on the prognostic imaging with the patient. Such exposure could help to motivate patients to follow-through with the necessary lifestyle changes. Earlier literature suggests that many patients would prefer direct and immediate reporting of imaging findings.[38] More recent work on patient preferences has yielded contradictory results that suggest that many patients may not appreciate the role of the radiologist, perhaps reflecting the radiologist's relative invisibility to the patient.[39] The content and tone of these direct reports would have to be carefully constructed to make intelligible to patients and make them less anxiety provoking. Furthermore context must be provided if the prognostic information is to be meaningful. Fellow physicians who have enjoyed the same training and may have received thousands of radiological reports share a body of knowledge and a lexicon that facilitates a nuanced communication of findings. Despite growing health awareness and instantaneous availability of medical information on the internet, patients are not generally equipped with similar experience. Their knowledge of treatment and risk abatement options will not be as detailed as that of referring clinicians and primary care providers.

In the context of the unrequested vertebral fractures, recent work on incidentally detected fractures on lateral chest radiographs has suggested that communication about the fracture status, the evidence concerning its prognostic implications and the prevailing treatment

guidelines is effective in increasing osteoporosis treatment levels. This was the case when the communication was directed towards physicians or additionally towards patients as well.[40] Follow-up simulation studies drawing upon the expected reduction in fracture rate resulting from the increased treatment uptake suggest that such communication is highly cost effective.[41]

Implications for radiological practice

On the part of the radiologist systematically reporting on unrequested findings that increase the risk of a disease other than the one that triggered the referral for imaging would require, at most, extra reading time but would not require any additional image acquisition, scanner time or radiation exposure. In the case of vertebral fractures it would require the reading of sagittal reformats of the thorax or abdomen, as sagittal views are crucial to assessing vertebral fractures.[42] Previous literature has shown that the vast majority (approximately 80%) of visible vertebral fractures are not reported in the course of routine care[32,33,35,36,42] Generating and storing sagittal reformats during routine clinical practice using the thin slice data would open the way to systematically assessing reporting vertebral height loss. With further improvements in PACS infrastructure, particularly storage capacity and computing power, a future where thin slice data of each investigation is stored, rather than one or two reconstruction planes, is conceivable. Reconstruction planes to assess the spine, including the sagittal plane, or in the case of scoliotic patients, oblique planes, could then easily be made at any point in the future at will to evaluate (progressive) height loss.

Mild vertebral fractures

Vertebral deformities, or mild fractures with 20-25% height-loss in Genant's semiquantitative grading scheme[7] deserve special consideration here. These are frequently excluded from many vertebral fracture assessments in the literature as being too subjective and showing reproducibility that is substantially lower than that of moderate and severe fractures. We have shown in this manuscript that these mild fractures are in themselves associated with future hip fractures, mortality, and cardiovascular disease. Furthermore, we found that they were more numerous than either the moderate or the severe fractures. Intuitively, milder vertebral deformities (the term fracture is usually reserved for moderate and severe height loss) may also be the first manifestations of early osseous degradation, preceding more severe height loss. This makes their exclusion doubly unfortunate as they could identify at risk individuals at a more early stage of disease progression. The disadvantages of including them in vertebral fracture analysis – in the form of a lower specificity of any putative opportunistic screening tool – are also likely to be limited. The

risk exists that including mild deformities will flag a large number of people who would not benefit much from fracture prevention treatment for further screening and treatment, driving up costs and reducing the cost-effectiveness of opportunistic fracture screening. The subtlety of mild deformities might also lead to false positives due to the difficulty of definitively identifying them. On the other hand a finding of increased fracture risk through opportunistic screening is unlikely to cause excessive patient burden or trigger large numbers of follow-up studies and/or invasive procedures of indeterminate value, as is frequently the case for pulmonary nodules.[37] At most a patient would be referred for clinical evaluation of fracture risk (with or without DXA). Since they might represent the earliest manifestation of bone demineralization they also represent an opportunity to intervene in the disease progression at an earlier stage and simply disregarding them may be throwing the baby out with the bathwater. Further research to elucidate their relation to osteoporotic disease and to determine which role they should play in osteoporosis detection and screening is warranted.

Future directions

The eventual implementation of the systematic reporting of vertebral deformities and fractures hinges on securing acceptance of the principle on the part of radiologists and clinicians. Longitudinal outcomes research showing a reduction in fracture rates after early intervention that is initiated on the basis of unrequested vertebral fractures or low vertebral density would conclusively demonstrate its value to both groups. Existing work on incidentally detected vertebral fractures on lateral chest radiographs suggests that more extensive communication can improve uptake of osteoporosis screening and treatment but stops short of showing a reduction in fracture rate and remains to be demonstrated for routine clinical CT. Demonstrating lower fracture rates in a trial setting would be ideal. In the meantime further research showing the equivalence of vertebral density measured on CT and BMD on DXA could help to form a link between these CT findings and the very large existing body of work carried out in patients selected on the basis of low DXA. Such research has already begun[10,11] but much work remains before HU values on CT can be substituted for DXA T-scores. Such equivalence research and research into intermediate outcomes such as treatment uptake cannot substitute for a well-designed trial demonstrating a reduction in fracture rates.

Automatic segmentation would be very helpful in implementing systematic vertebral fracture assessment. Automatic segmentation of vertebrae in CT datasets could potentially reduce the amount of extra reading time required to reliably and systematically assess vertebral fracture status. This is so far the most completely realized future direction in systematic unrequested vertebral assessment. In the future automated segmentation might

even obviate the need for routine sagittal reconstructions of all scans, by automatically screening all CTs and flagging a subset of patients as having a possible deformity. That subset could then be reviewed by reading radiologists. Currently automatic unsupervised segmentation of the spine remains challenging, with the automatic detection and recognition remaining imperfect, particularly in cases where atypical anatomy or degenerative vertebrae form a particular stumbling block.[43] Recent advances that have augmented existing pixel-based approaches (i.e classifying each voxel separately) with more advanced object-based approaches (detecting and dealing with whole organs and structures) have made automated recognition more viable than before, but they remain firmly confined to a research setting.[44,45] The prospect of automatic segmentation would also permit more nuanced quantitative values to be easily extracted from 3D CT datasets, such as vertebral volumes, subtle misalignments and vertebral density sampling from multiple regions of the vertebral corpus or indeed the whole corpus. Such measures could further expand the role of CT in detecting mild deformities and early degenerative changes and make risk stratification more accurate.

Finally, improvements in the resolution, availability and musculoskeletal techniques of MR imaging is leading to unprecedentedly detailed MR images of bone and the peri-osseous supportive structures in both the axial and extra-axial skeleton.[46] The number and range of unrequested findings on MR which might hold clinically relevant prognostic information in the context of osteoporosis is very exciting as it could complement the CT findings in many areas.

References

1. Osteoporosis prevention, diagnosis, and therapy. *NIH Consens Statement.* 2000; 17: 1–45.
2. Cole ZA, Dennison EM, Cooper C. Osteoporosis epidemiology update. *Curr Rheumatol Rep*. 2008; 10: 92–96.
3. Beaglehole R, Irwin A, Prentice T, World Health Organization. Changing history. World Health Organization: Geneva, 2004.
4. Bessette L, Ste-Marie L-G, Jean S, Davison KS, Beaulieu M, Baranci M et al. The care gap in diagnosis and treatment of women with a fragility fracture. *Osteoporos Int*. 2008; 19: 79–86.
5. Metge CJ, Leslie WD, Manness L-J, Yogendran M, Yuen CK, Kvern B et al. Postfracture care for older women: gaps between optimal care and actual care. *Can Fam Physician*. 2008; 54: 1270–1276.
6. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. *World Health Organ Tech Rep Ser*. 1994; 843: 1–129.
7. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993; 8: 1137–1148.
8. Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet*. 2002; 359: 1761–1767.
9. Kado DM, Browner WS, Palermo L, Nevitt MC, Genant HK, Cummings SR. Vertebral fractures and mortality in older women: a prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med*. 1999; 159: 1215–1220.
10. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med*. 2013; 158: 588–595.
11. Schwaiger BJ, Gersing AS, Baum T, Noël PB, Zimmer C, Bauer JS. Bone Mineral Density Values Derived from Routine Lumbar Spine Multidetector Row CT Predict Osteoporotic Vertebral Fractures and Screw Loosening. *AJNR Am J Neuroradiol*. 2014 Aug;35(8):1628-33.
12. Wells PNT. Sir Godfrey Newbold Hounsfield KT CBE. 28 August 1919 - 12 August 2004: Elected F.R.S. 1975. *Biogr Mem Fellows R Soc*. 2005; 51: 221–235.
13. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014; 14: 40.
14. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009; 338: b375.
15. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365: 395–409.
16. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. *JAMA*. 2012; 308: 1433–1434.

17. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA*. 2012; 307: 2418–2429.

18. Mets OM, Buckens CFM, Zanen P, Isgum I, van Ginneken B, Prokop M et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. *JAMA*. 2011; 306: 1775–1781.

19. McEvoy JW, Blaha MJ, Rivera JJ, Budoff MJ, Khan AN, Shaw LJ et al. Mortality rates in smokers and nonsmokers in the presence or absence of coronary artery calcification. *JACC Cardiovasc Imaging*. 2012; 5: 1037–1045.

20. Xie X, Zhao Y, de Bock GH, de Jong PA, Mali WP, Oudkerk M et al. Validation and Prognosis of Coronary Artery Calcium Scoring inNon-Triggered Thoracic Computed Tomography: Systematic Review and Meta-Analysis. *Circ Cardiovasc Imaging*. 2013 Jul;6(4):514-21.

21. Johansson C, Black D, Johnell O, Odén A, Mellström D. Bone mineral density is a predictor of survival. *Calcif Tissue Int*. 1998; 63: 190–196.

22. Jalava T, Sarna S, Pylkkänen L, Mawer B, Kanis JA, Selby P et al. Association between vertebral fracture and increased mortality in osteoporotic patients. *J Bone Miner Res*. 2003; 18: 1254–1260.

23. Hasserius R, Karlsson MK, Nilsson BE, Redlund-Johnell I, Johnell O. Prevalent vertebral deformities predict increased mortality and increased fracture rate in both men and women: a 10-year population-based study of 598 individuals from the Swedish cohort in the European Vertebral Osteoporosis Study. *Osteoporos Int*. 2003; 14: 61–68.

24. Puisto V, Rissanen H, Heliövaara M, Knekt P, Helenius I. Mortality in the presence of a vertebral fracture, scoliosis, or Scheuermann's disease in the thoracic spine. *Ann Epidemiol*. 2008; 18: 595–601.

25. Gan WQ, Man SFP, Senthilselvan A, Sin DD. Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. *Thorax*. 2004; 59: 574–580.

26. Romme EAPM, Murchison JT, Phang KF, Jansen FH, Rutten EPA, Wouters EFM et al. Bone attenuation on routine chest CT correlates with bone mineral density on DXA in patients with COPD. *J Bone Miner Res*. 2012; 27: 2338–2343.

27. Jacobs PC, Gondrie MJA, van der Graaf Y, de Koning HJ, Isgum I, van Ginneken B et al. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. *AJR Am J Roentgenol*. 2012; 198: 505–511.

28. Jacobs PCA, Isgum I, Gondrie MJA, Mali WPTM, van Ginneken B, Prokop M et al. Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement. *AJR Am J Roentgenol*. 2010; 194: 1244–1249.

29. Persy V, D'Haese P. Vascular calcification and bone disease: the calcification paradox. *Trends Mol Med*. 2009; 15: 405–416.

30. Doherty TM, Fitzpatrick LA, Inoue D, Qiao J-H, Fishbein MC, Detrano RC et al. Molecular, endocrine, and genetic mechanisms of arterial calcification. *Endocr Rev*. 2004; 25: 629–672.

31. Lampropoulos CE, Papaioannou I, D'Cruz DP. Osteoporosis--a risk factor for cardiovascular disease? *Nat Rev Rheumatol*. 2012; 8: 587–598.

32. Bartalena T, Giannelli G, Rinaldi MF, Rimondi E, Rinaldi G, Sverzellati N et al. Prevalence of thoracolumbar vertebral fractures on multidetector CT: underreporting by radiologists. *Eur J Radiol*. 2009; 69: 555–559.

33. Müller D, Bauer JS, Zeile M, Rummeny EJ, Link TM. Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. *Eur Radiol*. 2008; 18: 1696–1702.

34. Obaid H, Husamaldin Z, Bhatt R. Underdiagnosis of vertebral collapse on routine multidetector computed tomography scan of the abdomen. *Acta Radiol*. 2008; 49: 795–800.

35. Williams AL, Al-Busaidi A, Sparrow PJ, Adams JE, Whitehouse RW. Under-reporting of osteoporotic vertebral fractures on computed tomography. *Eur J Radiol*. 2009; 69: 179–183.

36. Woo EK, Mansoubi H, Alyas F. Incidental vertebral fractures on multidetector CT images of the chest: prevalence and recognition. *Clin Radiol*. 2008; 63: 160–164.

37. Lee CI, Forman HP. What we can and cannot see coming. Radiology 2010; 257: 313–314.

38. Schreiber MH, Leonard M, Rieniets CY. Disclosure of imaging findings to patients directly by radiologists: survey of patients' preferences. *AJR Am J Roentgenol*. 1995; 165: 467–469.

39. Kuhlman M, Meyer M, Krupinski EA. Direct reporting of results to patients: the future of radiology? *Acad Radiol*. 2012; 19: 646–650.

40. Majumdar SR, McAlister FA, Johnson JA, Bellerose D, Siminoski K, Hanley DA et al. Interventions to increase osteoporosis treatment in patients with 'incidentally' detected vertebral fractures. *Am J Med*. 2012; 125: 929–936.

41. Majumdar SR, Leslie WD. Conventional computed tomography imaging and bone mineral density: opportunistic screening or 'incidentaloporosis'? *Ann Intern Med*. 2013; 158: 630–631.

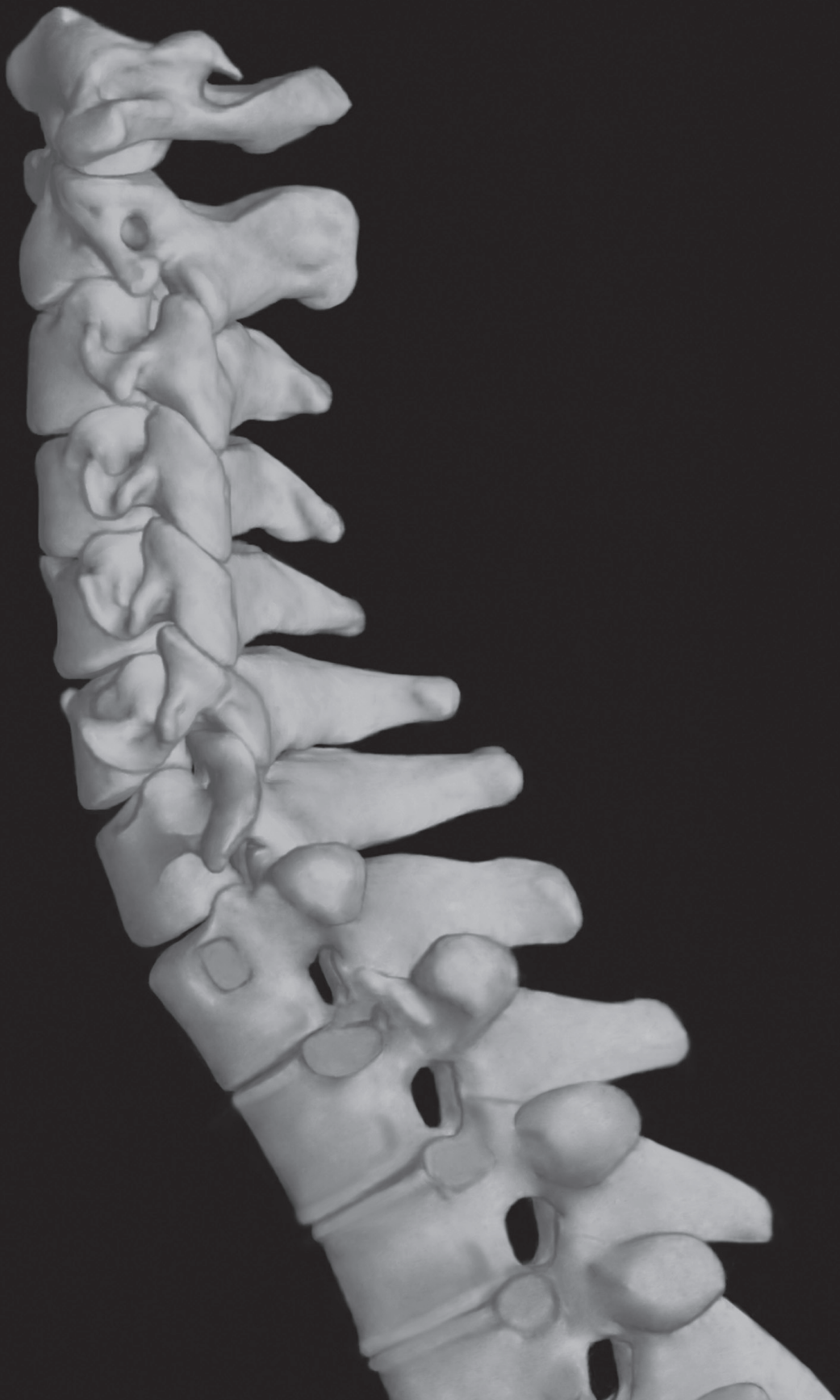
42. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ. Unreported Vertebral Body Compression Fractures at Abdominal Multidetector CT. *Radiology*. 2013 Jul;268(1):120-6.

43. Huang J, Jian F, Wu H, Li H. An improved level set method for vertebra CT image segmentation. *Biomed Eng Online*. 2013; 12: 48.

44. Ghebreab S, Smeulders AWM. Combining strings and necklaces for interactive three-dimensional segmentation of spinal images using an integral deformable spine model. *IEEE Trans Biomed Eng*. 2004; 51: 1821–1829.

45. Schwier M, Chitiboi T, Hülnhagen T, Hahn HK. Automated spine and vertebrae detection in CT images using object-based image analysis. *Int J Numer Methods Biomed Eng*. 2013; 29: 938–963.

46. Trattnig S, Zbýň S, Schmitt B, Friedrich K, Juras V, Szomolanyi P et al. Advanced MR methods at ultra-high field (7 Tesla) for clinical musculoskeletal applications. *Eur Radiol*. 2012; 22: 2338–2346.



Nederlandse samenvatting

Samenvatting

Beeldvormende onderzoeken worden steeds vaker ingezet in de klinische praktijk en ook de kwaliteit hiervan neemt toe. Daardoor zijn er vaker onverwachte nevenbevindingen waarvan sommige toekomstige aandoeningen mede kunnen voorspellen. Indien op tijd ontdekt, zouden deze aandoeningen beter en relatief goedkoper behandeld kunnen worden en zouden de ernstigere symptomen van deze aandoeningen voorkomen kunnen worden.

In de PROVIDI-studie worden zulke mogelijk voorspellende nevenbevindingen op computertomografie (CT)-scans van de borstkas gerelateerd aan ziekenhuisopnames en doodsoorzaken in de navolgende jaren. Deze CT-scans zijn een rijke bron van nevenbevindingen door de vele anatomische structuren die hierop zichtbaar zijn.

Dit proefschrift is gericht op wervelfracturen en -inzakkingen, die goed zichtbaar zijn op CT-scans. Eerder onderzoek laat zien dat wervelfracturen samenhangen met osteoporose (botontkalking), hart- en vaatziekten en vroegtijdig overlijden.

Osteoporose is een groot en groeiend gezondheidsprobleem in de ontwikkelde wereld en leidt tot veel botbreuken. Vroege ontdekking van deze aandoening en preventieve behandeling is daarom cruciaal om fracturen te voorkomen.

In de PROVIDI-studie wordt de case-cohortonderzoeksopzet gebruikt. In **hoofdstuk 2** gebruiken we een statistische simulatie om twee gebruiksvriendelijke en efficiënte alternatieve benaderingen van case-cohortdata te vergelijken in de context van prognostisch (voorspellend) onderzoek. De prognostische prestaties van de alternatieve benaderingen blijken niet wezenlijk te verschillen van het meest gangbare, klassieke wegingsschema. Deze alternatieve benaderingen kunnen derhalve gebruikt worden zonder de uitkomsten te veranderen en onderzoek met case-cohort data vereenvoudigen.

In **hoofdstuk 3** onderzoeken we de bestaande literatuur over nevenbevindingen op CT-scans van de borstkas. We analyseren artikelen over CT-scans van het hart die nevenbevindingen buiten het hart beschrijven. Op 13% van de scans zijn klinisch relevante nevenbevindingen te zien. Dit is een forse hoeveelheid en een belangrijk gegeven voor de radioloog en eventueel medisch vervolgonderzoek.

In **hoofdstuk 4** beoordelen vier artsen herhaaldelijk dezelfde 50 scans. Hieruit blijkt dat de aanwezigheid en ernst van een fractuur reproduceerbaar en betrouwbaar gemeten kunnen worden, wat noodzakelijk is om deze gegevens in de praktijk te kunnen gebruiken.

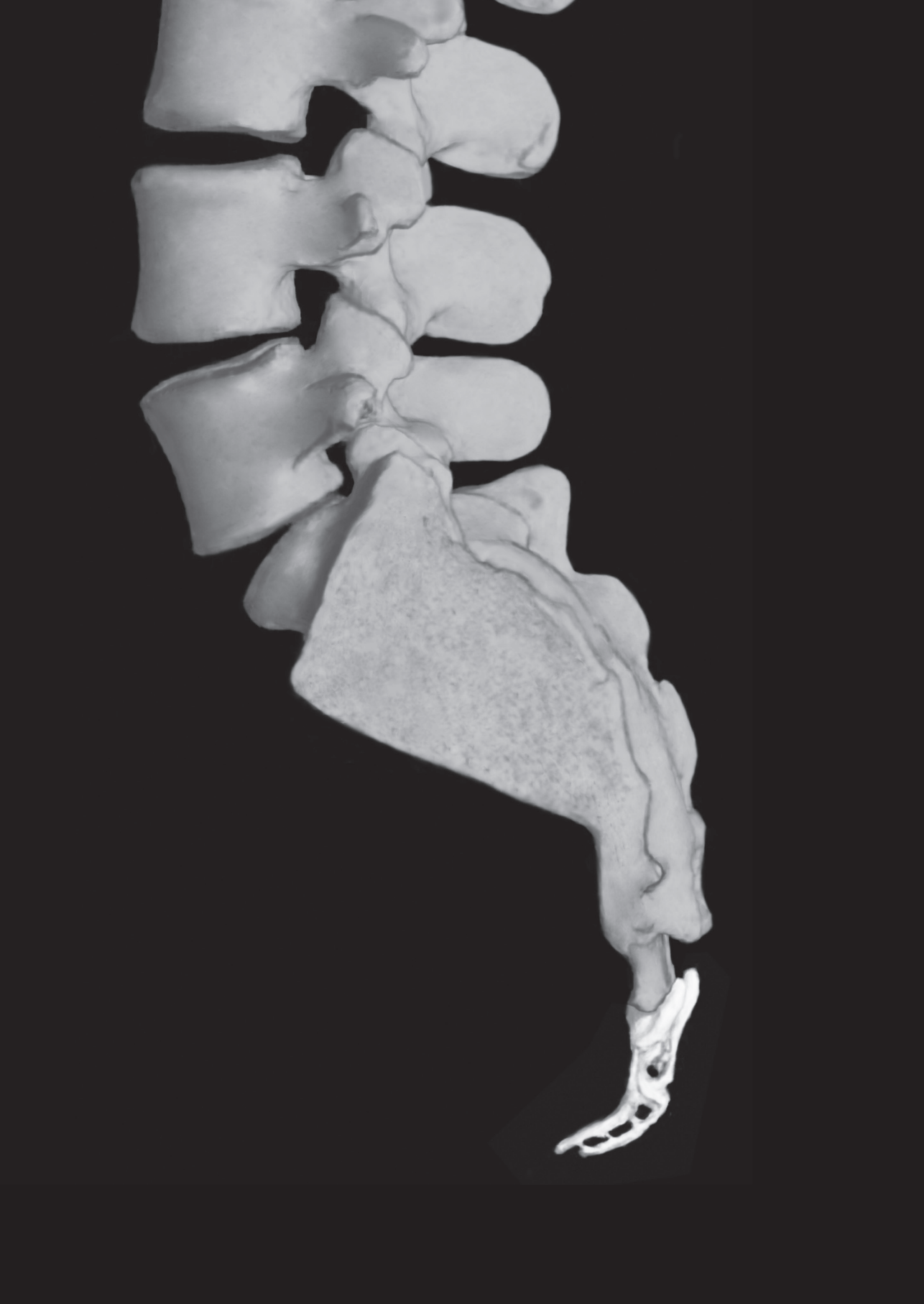
In **hoofdstuk 5** wordt onderzocht of de wervelinzakkingen en -fracturen die op CT-scans zichtbaar zijn heupfracturen kunnen voorspellen. Deze (osteoporotische) heupfracturen zijn van groot belang, omdat ze altijd tot een ziekenhuisopname leiden en gepaard gaan met een kortere levensverwachting. Patiënten met wervelfracturen blijken drie keer zo veel kans te hebben op een heupfractuur als patiënten zonder een fractuur. Bij ernstige fracturen is de kans nog groter.

In **hoofdstuk 6** voeren we een externe validatiestudie uit naar de diagnostische waarde van de klassieke methode voor botdichtheidsmeting met CT (het DXA-onderzoek). Eerder onderzoek liet zien dat deze metingen sterk samenhangen met de werkelijke botdichtheid. Wij vonden een iets kleinere, maar nog steeds grote samenhang (AUC van 0.74 in plaats van 0.83). Extern validatieonderzoek is belangrijk om bevindingen extern te bevestigen en de variatie van deze bevindingen in verschillende contexten in te schatten. Uit ons onderzoek blijkt dat CT-scans wel bruikbare informatie over botdichtheid bevatten, maar dat deze informatie minder goed samenhangt met de werkelijke botdichtheid dan werd gedacht.

In **hoofdstuk 7** onderzoeken we of wervelfracturen vroegtijdig overlijden kunnen voorspellen. Eerder onderzoek heeft aangetoond dat heupfracturen leiden tot vroegtijdig overlijden. Ons onderzoek wijst uit dat dit ook geldt voor wervelfracturen: mensen met een wervelfractuur hebben ongeveer twee keer zo veel kans op vroegtijdig overlijden als mensen zonder fractuur. Ook op zichzelf staand blijkt de botdichtheid op CT voorspellend voor vroegtijdig overlijden.

In **hoofdstuk 8** bekijken we of wervelfracturen hart- en vaatziekten kunnen voorspellen. Bestaande literatuur laat zien dat ze samenhangen, maar het is onduidelijk hoe. In ons onderzoek hangen wervelfracturen weliswaar samen met toekomstige hart- en vaatziekten, maar vaatverkalkingen (die ook goed zichtbaar zijn op CT-scans) blijken een veel betere voorspeller dan wervelfracturen. Hoewel er dus een duidelijk verband bestaat, lijkt het beoordelen van wervelfracturen weinig meerwaarde te hebben.

In de praktijk zal moeten blijken of systematische rapportage en vervolgens behandeling van wervelfracturen de genoemde risico's daadwerkelijk verkleinen.



Appendix

List of affiliations
List of publications
Dankwoord
Biography

List of affiliations

List of affiliations

CM van Aalst	Erasmus Medical Center, Rotterdam, the Netherlands
E Bakker	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
CFM Buckens	Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
TP Debray	Julius Centre for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
G Dijkhuis	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
MJ Gondrie	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
Y van der Graaf	Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht,Utrecht, the Netherlands
I Isgum	Imaging sciences institute, University Medical Centre Utrecht, Utrecht, the Netherlands
P Jairam	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
PA de Jong	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
B de Keizer	Department of Nuclear Medicine, University Medical Center Utrecht, Utrecht, The Netherlands
H de Koning	Erasmus Medical Center, Rotterdam, the Netherlands

WP Mali	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
CP Mol	Imaging sciences institute, University Medical Center Utrecht, Utrecht, the Netherlands
KG Moons	Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands
LM Peelen	Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands
M Oudkerk	University Medical Center Groningen Netherlands, Groningen Netherlands
HP Stallman	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
HJ Verhaar	Department of Geriatric Medicine, the University Medical Center Utrecht, Utrecht, the Netherlands
HM Verkooijen	Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands
R Vliegenthart	University Medical Center Groningen Netherlands, Groningen Netherlands
PROVIDI Study Group	The PROVIDI Study Group consists of J Laméris (Department of Radiology, Academic Medical Center, Amsterdam), C van Kuijk (Department of Radiology, VU University Medical Center Amsterdam), W ten Hove (Department of Radiology, Gelre Hospitals, Apeldoorn), M Oudkerk (Department of Radiology, University Medical Center Groningen), Ay L Oen (Department of Radiology, Elkerliek Hospital, Helmond), J Wildberger (Department of Radiology, Academic Hospital Maastricht), J van Heesewijk (Department of Radiology, St Antonius Hospital, Nieuwegein), W Mali (Department of Radiology, University Medical Center Utrecht), and Y van der Graaf (Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht).

List of publications

Publications in international journals

Buckens CF, Dijkhuis G, de Keizer B, Verhaar HJ, de Jong PA. Opportunistic screening for osteoporosis on routine computed tomography? An external validation study.
Eur Radiol. 2015 Jan 17. [Epub ahead of print]. PMID: 25591750

Buckens CF, de Jong PA, Verkooijen HM, Verhaar HJ, Mali WP, van der Graaf Y; PROVIDI study group. Vertebral fractures on routine chest computed tomography: relation with arterial calcifications and future cardiovascular events.
Int J Cardiovasc Imaging. 2015 Feb;31(2):437-45. doi: 10.1007/s10554-014-0567-9. PMID: 25451815

Buckens CF, van der Graaf Y, Verkooijen HM, Mali WP, Isgum I, Mol CP, Verhaar HJ, Vliegenthart R, Oudkerk M, van Aalst CM, de Koning HJ, de Jong PA. Osteoporosis markers on low-dose lung cancer screening chest computed tomography scans predict all-cause mortality.
Eur Radiol. 2015 Jan;25(1):132-9. doi: 10.1007/s00330-014-3361-0. PMID: 25323601

Diepstraten SC, Sever AR, **Buckens CF**, Veldhuis WB, van Dalen T, van den Bosch MA, Mali WP, Verkooijen HM. Value of preoperative ultrasound-guided axillary lymph node biopsy for preventing completion axillary lymph node dissection in breast cancer: a systematic review and meta-analysis.
Ann Surg Oncol. 2014 Jan;21(1):51-9. doi: 10.1245/s10434-013-3229-6. PMID: 24008555

Buckens CF, de Jong PA, Mol C, Bakker E, Stallman HP, Mali WP, van der Graaf Y, Verkooijen HM. Intra and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography.
PLoS One. 2013 Aug 5;8(8):e71204. doi: 10.1371/journal.pone.0071204. PMID: 23940719

Buckens C, de Jong P, Mali W, Verhaar H, van der Graaf Y, Verkooijen H. Prevalent vertebral fractures on chest CT: Higher risk for future hip fracture.
J Bone Miner Res. 2013 Jul 2. doi: 10.1002/jbmr.2028. PMID: 23821454

Luijkx T, Cramer MJ, **Buckens CF**, Zaidi A, Rienks R, Mosterd A, Prakken NH, Dijkman B, Mali WP, Velthuis BK. Unravelling the grey zone: cardiac MRI volume to wall mass ratio to differentiate hypertrophic cardiomyopathy and the athlete's heart.
Br J Sports Med. 2013 Jun 14. PMID: 23770661

Mets OM, Vliegenthart R, Gondrie MJ, Viergever MA, Oudkerk M, de Koning HJ, Mali WP, Prokop M, van Klaveren RJ, van der Graaf Y, **Buckens CF**, Zanen P, Lammers JW, Groen HJ, Isgum I, de Jong PA. Lung Cancer Screening CT-Based Prediction of Cardiovascular Events.
JACC Cardiovasc Imaging. 2013 Jun 8. doi:pii: S1936-878X(13)00332-X. 10.1016/j.jcmg.2013.02.008. PMID: 23769488

Mets OM, Schmidt M, **Buckens CF**, Gondrie MJ, Isgum I, Oudkerk M, Vliegenthart R, de Koning HJ, van der Aalst CM, Prokop M, Lammers JW, Zanen P, Hoesein FA, Mali WP, van Ginneken B, van Rikxoort EM, de Jong PA. Diagnosis of chronic obstructive pulmonary disease in lung cancer screening Computed Tomography scans: independent contribution of emphysema, air trapping and bronchial wall thickening.
Respir Res. 2013 May 27;14:59. doi: 10.1186/1465-9921-14-59. PMID: 23711184

van Kessel CS, **Buckens CF**, van den Bosch MA, van Leeuwen MS, van Hillegersberg R, Verkooijen HM. Preoperative imaging of colorectal liver metastases after neoadjuvant chemotherapy: a meta-analysis.
Ann Surg Oncol. 2012 Sep;19(9):2805-13. PMID:22396005

Luijkx T, Cramer MJ, Zaidi A, Rienks R, Senden PJ, Sharma S, van Hellemondt FJ, Buckens CF, Mali WP, Velthuis BK. Ethnic differences in ventricular hypertrabeculation on cardiac MRI in elite football players.
Neth Heart J. 2012 Jul 10. PMID:22777563

Luijkx T, Velthuis BK, Backx FJ, **Buckens CF**, Prakken NH, Rienks R, Mali WP, Cramer MJ. Anabolic androgenic steroid use is associated with ventricular dysfunction on cardiac MRI in strength trained athletes.
Int J Cardiol. 2012 Mar 27. PMID:22459398

Luijkx T, Cramer MJ, Prakken NH, **Buckens CF**, Mosterd A, Rienks R, Backx FJ, Mali WP, Velthuis BK. Sport category is an important determinant of cardiac adaptation: an MRI study.
Br J Sports Med. 2012 Jan 4. PMID:22219218

Mets OM, Smit EJ, Mohamed Hoesein FA, Gietema HA, Bokkers RP, Attrach M, van Amelsvoort-van de Vorst S, Scholten ET, **Buckens CF**, Oudkerk M, Lammers JW, Prokop M, de Jong PA. Visual versus Automated Evaluation of Chest Computed Tomography for the Presence of Chronic Obstructive Pulmonary Disease.
PLoS One. 2012;7(7):e42227. PMID:22848747

Buckens CF, Verkooijen HM, Gondrie MJ, Jairam P, Mali WP, van der Graaf Y. Unrequested findings on cardiac computed tomography: looking beyond the heart. PLoS One. 2012;7(4):e32184. PMID:22536315

Gondrie MJ, van der Graaf Y, Jacobs PC, **Buckens SC**, Mali WP; PROVIDI Study Group. The prognostic value of vascular diameter measurements on routine chest computed tomography in patients not referred for cardiovascular indications.
J Comput Assist Tomogr. 2011 Nov-Dec;35(6):734-41. PMID:22082545

Mets OM, **Buckens CF**, Zanen P, Isgum I, van Ginneken B, Prokop M, Gietema HA, Lammers JW, Vliegenthart R, Oudkerk M, van Klaveren RJ, de Koning HJ, Mali WP, de Jong PA. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans.
JAMA. 2011 Oct 26;306(16):1775-81. PMID:22028353

de Jong PA, Gondrie MJ, **Buckens CF**, Jacobs PC, Mali WP, van der Graaf Y; PROVIDI study group. Prediction of cardiovascular events by using non-vascular findings on routine chest CT.
PLoS One. 2011;6(10):e26036. PMID:22022499

Stadhouder A, **Buckens CF**, Holtslag HR, Oner FC. Are existing outcome instruments suitable for assessment of spinal trauma patients?
J Neurosurg Spine. 2010 Nov;13(5):638-47. PMID:21039157

Gondrie MJ, Mali WP, **Buckens CF**, Jacobs PC, Grobbee DE, van der Graaf Y. The PROgnostic Value of unrequested Information in Diagnostic Imaging (PROVIDI) Study: rationale and design.
Eur J Epidemiol. 2010 Oct;25(10):751-8. PMID:20890635

Buckens CF, Saris DB. Reconstruction of the medial patellofemoral ligament for treatment of patellofemoral instability: a systematic review.
Am J Sports Med. 2010 Jan;38(1):181-8. PMID:19966098

FC Oner and **CFM Buckens**. Outcome assessment tools for spinal trauma patients.
In: Spine and Spinal Cord Trauma: Evidence-Based Management. 2010. Thieme, New York. ISBN 978-1-60406-221-2.

Dankwoord

Dankwoord

Ik heb het promotietraject dat tot dit proefschrift geleid heeft ervaren als een leerzame, verrijkende tijd in mijn leven. Voor het proefschrift en voor deze verrijking heb ik veel mensen te danken die dit door hun directe of indirecte bijdrage mogelijk hebben gemaakt. Een aantal van hen wil ik in het bijzonder bedanken.

In the interest of brevity, and due to my own imperfect faculties, I will not attempt to thank everyone. All those who have made a contribution to my life during the halycon era that will become known as 'PhD years' who feel absent from these pages, console yourselves with the knowledge that you have left a lasting imprint on my soul.

Ten eerste mijn promotor, professor Yolanda van der Graaf die met haar visie en inzet altijd verder keek dan ik kon beseffen en hierdoor deze promotie in goede banen heeft geleid. Ik bewonder je grenzeloze energie, de manier waarop jij moeiteloos schakelde tussen multipele taken, rollen en situaties tijdens jouw wandelende managementsessies door het Stratenum. Ik ben jou ontzettend dankbaar voor de manier waarop jij altijd klaar stond om dingen te regelen voor jouw promovendi en voor de manier waarop jij met mij bent omgegaan – door mij alle ruimte te geven om alle lessen van het Julius Centrum te absorberen en in de praktijk te brengen.

Professor Mali, ik ben u zeer dankbaar dat u mij de mogelijkheid heeft gegeven om onder uw supervisie promotieonderzoek te kunnen doen. Door al het goede advies en de met wijsheid doordrongen feedback op de inmiddels talloze manuscriptversies is niet alleen uiteindelijk dit proefschrift ontstaan maar heb ik ook veel geleerd van u.

Pim de Jong, beste Pim, ik bewonder ontzettend jouw energie, drive, pragmatisme, visie, effectiviteit en enthousiasme voor het onderzoek. Jouw plezier in dit werk stimuleerde mij om verder te denken en om ideeën ook daadwerkelijk in de praktijk te brengen. Ik ben zeer dankbaar voor jouw inzet voor deze promotie. Ik hoop nog veel van je te mogen leren.

Helena Verkooijen, beste Lenny, dank voor de intensieve samenwerking. Jouw opmerkzame feedback en scherpe oog voor opmaak en layout zijn de kwaliteit van dit proefschrift sterk ten goede gekomen.

Harald Verhaar, dank voor uw enthousiaste inzet en overleg op geriatrisch en klinisch vlak. Uw kennis en betrokkenheid bij de problematiek van uw patiëntengroep is voor mij een voorbeeld van goede klinische zorg geworden.

Martijn Gondrie, mijn directe voorganger op het PROVIDI-project. Dank voor jouw relaxte en hartelijke introductie tot het onderzoeksleven en voor al jouw adviezen tijdens de promotie. Dankzij ons bijna paralleltraject mag ik nog lang genieten van jou als collega.

Peter Jacobs, de initiërende promovendus op het PROVIDI-project, die veel werk heeft verzet om dit project van de grond te krijgen waardoor wij (de nakomelingen) konden profiteren van de infrastructuur die jij in jouw vooruitziendheid had geschapen. Mijn dank hiervoor is groot.

Tim Luijkx, studiegenoot, mede-promovendus, collega-assistent, paranimf en vriend, jij blijft voor mij een voorbeeld van beter en ik zal je blijven emuleren. Ik hoop dat ik nog lang mag genieten van de gouden combinatie van precisie, intelligentie, gedrevenheid en humor die jij zo moeiteloos handhaaft. Speciale dank voor alle inspanningen die jij naast je fulltime assistentschap en jouw drievoudig vaderschap hebt geleverd om dit boekje tot een geheel te maken.

Thomas Debray, my dear friend and para-nymph, Thank you for the shared jokes, cups of coffee, workout sessions, dinners, foosball games and help with R scripting. I admire your deep knowledge of biostatistics and I hope to continue to derive enjoyment and erudition from your company.

Jan-Willem Gratama en Wim ten Hove, dank voor jullie hartelijke en open onthaal aan een onzekere beginnende assistent in Apeldoorn en voor jullie constante aanmoediging om ook dit onderzoek af te maken.

Mijn dank en bewondering voor de verdere maatschap radiologie in Apeldoorn, Marc Logtenberg, van wie ik zomaar naar een van zijn orthopedische congressen mocht komen en die altijd geduld heeft voor over-eager MSK-assistenten. Patrick Nass, de man die nooit een klinisch relevante bevinding mist (en wiens duim bijna overal in past), Kitty Droogh, die altijd tijd voor je maakt voor uitleg en onderwijs over diffusierestrictie, lumbalewervelanatomie of interpretatie van darmwandverdikking op echografie. Hein Slis, een zeer bedreven supervisor, iPad paraat om dubieuze uitspraken te controleren en altijd een oog voor de beste manier om problemen op te lossen. Jos de Win, een uitermate nette en fatsoenlijke man met een scherp oog voor de juiste manier om afwijkingen te verwoorden en altijd paraat om over geschiedenis of wijn te praten. Fleur van Raamt, een ijzersterk gevoel voor grammatica en oneindig geduld met de assistenten onder haar hoede. En zeker niet de minste, Joost Kardux, de man die al wandelend door de gang alles ziet op alle schermen in het huis en wie bijna moeiteloos een zeer diepe kennis van alle facetten van de radiologie weet bij te houden.

Alle betrokkenen bij de participerende PROVIDI-ziekenhuizen wil ik bedanken. In het bijzonder wil ik prof. dr. C. van Kuijk, dr. J van Heesewijk, prof. dr. J. Wildberger, prof. dr. J. Laméris, prof. dr. M. Oudkerk en drs. A Leenders-Oen bedanken voor hun bijdrage aan het PROVIDI-project.

De collega-assistenten en oud-assistenten in Apeldoorn: Arnoud, Ricardo, Nicki, Anne-Marie, Wilco, Iris, Anja, Pushpa, Maarten Donswijk en Maarten Smits, Arthur, Jesse, Ufuk, Sytse, Saskia, Dalith, Anne, Cécile, Jeroen; ik heb ontzettend veel van jullie geleerd. Mijn dank voor jullie inzet, jullie kennis en jullie gedrevenheid. Ik geniet van jullie gezelligheid en goedheid. Waarde collega's uit Utrecht: ik hoop veel van jullie te mogen leren de komende periode.

De kamergenoten van het gouden promotie tijdperk, Liselotte, Gerdien, Thomas, Paulien, Sjoukje, Floriaan, Putri, Victor, Joris. Mijn dank voor alle gezelligheid en steun.

Elmer dank voor alle inzichten over de Nederlandse cultuur, alle sportsessies en de vele goede verhalen. Hendrik, thanks for always commiserating with my lamentations whenever we caught up and for always letting me be myself. Todor, thank you for sharing the refreshing and unique way that you perceive the world slightly askance. Alisa, Eelco, Angelique, Welling bedankt voor jullie vriendschap. Moge ik nog lang van de gezelligheid en van de gedeelde hobby's genieten. To all my friends near and far; you have enriched my life and I am commensurately grateful.

Lieve ooms en tantes; mijn diepe dank voor alle steun tijdens alle soms moeilijke momenten tijdens deze promotie.

Lieve vader, dear Baba, thank you for inspiring me to fly higher and for always supporting me. My memories of your wisdom and generosity continue to sustain and guide me.

Lieve mama, dank voor alle aanmoediging en liefde. Jouw constante en onvoorwaardelijke steun was ook tijdens deze lange en soms zware promotie voor mij een troost.

JP, je bent mijn voorbeeld door jouw ruimhartigheid en levenslust. Dank voor de fundamentele en onvoorwaardelijke steun die je blijft geven. Louise, thank you for being there for us through thick and thin and always sharing a smile and ready laugh.

Mijn lieve schoonouders, jullie hebben mij met open armen ontvangen in jullie familie en blijven altijd paraat staan voor ons. Ik hoop dat wij nog veel levensgenot met jullie mogen blijven delen. Sarju, thank you for all the deep talks and the wise words.

My dear Sweta, without your support I would not have dared undertake such a herculean task as this thesis. By sharing your life with me and allowing me to share mine with you you have given my life a more profound meaning.

My dear Vida, thank you for brightening my days with your pure joyfulness. Watching you grow and learn continues to be a blessing.

During the course of writing this manuscript I've taken the opportunity to forage widely upon the bountiful fields of knowledge made available to me. Along the way I have made new friends and lost loved ones. I hope that aside from just making me older this multi-year process has made me a more effective and insightful researcher. Sotto voce I also hope that maybe - just maybe - I've become a slightly better and wiser person too. That I have you all to thank for.

Biography

Biography

Stan, as he is commonly known to those around him, began life long before he graduated from the University of Utrecht's Faculty of Medicine: after having been born in the mid-eighties in East-Africa to Dutch parents, he spent his childhood in a multilingual, multicultural and multinational environment, not to mention an exceedingly pleasant climate.

This was immediately and abruptly followed by several teenage years spent in cooler air of Denmark's Copenhagen and, after surviving this immunological challenge and completing high-school, his sights were set on his ancestral Netherlands for higher education.

In keeping with his international orientation he chose to attend a newish English-language liberal arts college (the University College Utrecht), achieving his bachelor of science after three instructive years. During this time he focused mainly on biomedical/biochemical subjects, whilst grounding his sanity in a philosophy side-track.

He then graciously continued to make himself available for pedagogical experimentation by attending the first four year graduate entry medical program in the Netherlands: the SUMMA-program at Utrecht. During these challenging years he somehow found time to dabble in some light medical outcomes research and upon nearing graduation he decided to capitalize on the symbiotic relationship he had developed with his computers by pursuing a career in radiology.

After spending three happy years at Utrecht's most excellent Julius Center for Health Sciences and Primary Care, trying to divine the relationship between those increasingly prevalent incidental radiological findings and real-life patient outcomes and expanding his research toolkit through additional courses and side-projects, Stan is now busily trying to stay afloat as a junior radiology resident at the Gelre hospital in Apeldoorn and the UMC Utrecht, whilst finalizing his PhD. He greatly enjoys his work, especially musculoskeletal imaging, which speaks to his own athletic proclivities.