Open Access

Nicholas Ohs, Fabian Keller, Ole Blank, Yuk-Wai Wayne Lee, Chun-Yiu Jack Cheng, Peter Arbenz, Ralph Müller and Patrik Christen*

Towards *in silico* prognosis using big data

DOI 10.1515/cdbme-2016-0016

Abstract: Clinical diagnosis and prognosis usually rely on few or even single measurements despite clinical big data being available. This limits the exploration of complex diseases such as adolescent idiopathic scoliosis (AIS) where the associated low bone mass remains unexplained. Observed low physical activity and increased RANKL/OPG, however, both indicate a mechanobiological cause. To deepen disease understanding, we propose an *in silico* prognosis approach using clinical big data, i.e. medical images, serum markers, questionnaires and live style data from mobile monitoring devices and explore the role of inadequate physical activity in a first AIS prototype. It employs a cellular automaton (CA) to represent the medical image, micro-finite element analysis to calculate loading, and a Boolean network to integrate the other biomarkers. Medical images of the distal tibia, physical activity scores, and vitamin D and PTH levels were integrated as measured clinically while the time development of bone density and RANKL/OPG was observed. Simulation of an AIS patient with normal physical activity and patient-specific vitamin D and PTH levels showed minor changes in bone density whereas the simulation of the same AIS patient but with reduced physical activity led to low density. Both showed unchanged RANKL/OPG and considerable cortical resorption. We conclude that our integrative *in silico* approach allows to account for a variety of clinical big data to study complex diseases.

***Corresponding author: Patrik Christen,** ETH Zurich, Institute for Biomechanics, Leopold-Ruzicka-Weg 4, 8093 Zurich, Switzerland, E-mail: patrik.christen@hest.ethz.ch
Nicholas Ohs, Fabian Keller, Ole Blank and Ralph Müller: ETH Zurich, Institute for Biomechanics, Leopold-Ruzicka-Weg 4, 8093 Zurich, Switzerland, E-mail: nicholas.ohs@hest.ethz.ch (N. Ohs), kellerf@student.ethz.ch (F. Keller), blanko@student.ethz.ch (O. Blank), ram@ethz.ch (R. Müller)
Yuk-Wai Wayne Lee and Chun-Yiu Jack Cheng: Department of Orthopaedics & Traumatology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, NT, Hong Kong SAR, China, E-mail: waynelee@ort.cuhk.edu.hk (Y.-W. Wayne Lee), jackcheng@cuhk.edu.hk (C.-Y Jack Cheng)
Peter Arbenz: ETH Zurich, Computer Science Department, Universitätstrasse 6, 8092 Zurich, Switzerland, E-mail: arbenz@inf.ethz.ch

Keywords: adolescent idiopathic scoliosis; Boolean network; cellular automaton; clinical big data; micro-finite element analysis.

1 Introduction

Understanding the physiology of the human body is a tremendous challenge, understanding and finally treating diseases is arguably an even more difficult task. Many diseases exhibit alterations in complex signalling pathways that lead to distinct phenotypes used in diagnosis and prognosis. However, both usually rely on few or even single clinical measurements despite clinical big data, i.e. medical images, serum markers, questionnaires and live style data from mobile monitoring devices being available. Exploiting these clinical big data would potentially enable one explore and thus further understand complex diseases such as adolescent idiopathic scoliosis (AIS) where many alterations in the signalling cascade and other factors such as a patient's lifestyle play an essential role.

AIS is a prevalent spinal deformity associated with systemic low bone mass (Z-score ≤ -1 with the reference to local matched population), which can persist into adulthood predisposing to osteoporosis in later life. If not treated properly, the curve will deteriorate leading to cardiopulmonary compromise, back pain, degenerative spine disease, and psychosocial disorder [1]. Low bone mass has been identified as a risk factor for curve progression. The underlying mechanism of low bone mass in AIS, however, remains unknown although it would potentially provide a basis for counteracting the associated complications [2]. It has been suggested that low bone mass is related to inadequate physical activity [3] and an increase in RANKL/OPG [1], which might lead to increased bone resorption. Most recently, low vitamin D levels were observed in AIS patients [4] indicating that multiple factors play a role in AIS.

Therefore, we propose a novel prognosis approach using clinical big data for the *in silico* simulation of complex diseases with the ultimate goal to improve clinical diagnosis and prognosis of these diseases as well as the

evaluation of treatment options. Medical images containing bone tissue density are integrated into a cellular automaton (CA) model where the tissue densities define local states. These states are updated according to a local rule that takes into account the biological cells present in each CA cell. The cell number or activity is normalised representing a normal healthy value by default but with the possibility to either initialise patient-specific or experimental values. It allows to integrate serum markers of specific cells if they are available and at the same time does not require the definition of a value if it is not known. The cells itself are updated through mechanical tissue loading as calculated with micro-finite element (micro-FE) analyses and a molecular factor determined in a Boolean network, which models the interaction of cells, molecules, and mechanical loading. The entire procedure is applied to obtain cell and tissue densities as well as molecular states at different time points that describe the time evolution of the system. Here, we describe and formulate the theoretical basis of this integrative *in silico* prognosis approach and present a first prototype demonstrating the integration of medical image, serum marker, and lifestyle data for AIS.

The purpose of this study is to formulate a theory for integrating varying clinical big data, and implement and employ it to simulate low bone mass typical of AIS.

2 Material and methods

We performed *in silico* simulations over a period of 6 months integrating high-resolution peripheral quantitative computed tomography (HR-pQCT) images of the distal tibia, physical activity scores, and vitamin D and PTH levels of an AIS patient. Two simulations were run including an AIS control case with normal physical activity score, and vitamin D and PTH levels as measured in the patient's serum, and an AIS case with reduced physical activity as observed in other studies but again with the same patient-specific measurements as in the control simulation.

2.1 Clinical patient assessment

To validate the proposed *in silico* prognosis, a 17-year-old AIS girl (patient number 666; major Cobb angle 37°) was randomly selected from the database of a prospective randomised controlled trial who received observation alone. Anthropometry, curve severity, pubertal assessment and bone parameters were assessed at baseline and at the 12-month time-point. Bone parameters were measured with dual-energy X-ray absorptiometry (DXA, XR-36; Norland Medical Systems, Fort Atkinson, WI, USA) and HR-pQCT (XtremeCT, Scanco Medical AG, Switzerland). Blood was collected on the same date of clinical visit for the measurement of serum bone metabolic markers with standard protocols conducted by the University Pathology Unit. Physical activity was assessed by a standardised quantitative questionnaire adapted from Slemenda et al. [5]. The amount of weight bearing activity is expressed as number of hours per day. The sum of hours spent on various types of weight bearing activity was used to summarise daily activities for each subject. Ethical approval was obtained from the University Clinical Research Ethics Committee (Ref. No. CRE-2008.054-T) [6].

2.2 Cellular automaton

The medical image of a bone, e.g. a HR-pQCT scan, is a 3D representation of the bone density in a cubic domain at a certain voxel resolution. It can be regarded as the regular grid, G, of cells of a CA. For each CA cell and a given time point, t, a density value, a strain energy density (SED) value, the number or activity of the different biological cells and the presence or absence of several molecules usually found in bone are defined:

$$\begin{aligned} g \in G,\ g(t) &= (\rho(t), \epsilon(t), n_0(t), n_1(t), \ldots, n_N(t), \\ &= (m_0(t), m_1(t), \ldots, m_M(t)), \end{aligned} \tag{1}$$

with ρ being the tissue density value, ϵ being the SED value, N the number of cell types considered, n_i the number or activity of biological cells of type i inside each voxel, M the number of molecule types considered and m_j indicating whether the molecule of type j is present or absent. This tuple makes up the state of this CA cell at time t. Our first prototype uses three cell types: osteoblasts (n_{OBL}) forming new bone tissue and osteoclasts (n_{OCL}) removing old bone tissue resulting in a continuous remodelling of the tissue, and osteocytes (n_{OCY}) considered to respond to mechanical tissue loading and orchestrate the action of osteoblasts and osteoclasts accordingly.

In order to update the CA, rules are defined that are applied to each CA cell simultaneously and considering a certain neighbourhood. For each $g \in G$, $N(g)$ describes the set of all considered neighbours of g. In the presented prototype, a 3D von Neumann neighbourhood is defined. Finally, an update rule f that generates the new bone cells and tissue densities in the CA is given by

$$g(t+1) = f(g(t), N(g(t))). \tag{2}$$

In our prototype AIS CA, the update rule is the same for every CA cell. Our prototype employs the following update rule: The new SED values for each voxel computed with the micro-FE method is an input. In a first update step, the cell numbers based on a function that transforms SED values to an increase or decrease of OBL/OCL are modified according to the SED-bone remodelling relationship found earlier [7]. To avoid infinite increase/decrease of cells, they are saturated at a minimum and maximum level for both, OBL and OCL. In a second step, the Boolean network for each voxel is evaluated, which will again result in modifications to the OBL/OCL counts. Finally, the tissue density is updated based on the OBL/OCL counts. If the density values are higher than bone, they are distributed to the neighbouring voxels. The initial bone cell count and distribution are determined in a preceding iteration where more OBLs are placed in regions of high SED and more OCLs in regions of low SED. The CA is only updated a few layers away from the top and bottom where the boundary conditions of the micro-FE analysis are applied to exclude boundary artefacts.

2.3 Boolean network

The interaction of the local molecules with the local biological cells under the influence of mechanical loading in each CA cell is modelled with a Boolean network. Each molecule is either present in a CA cell or not. The number of CA cells in which a specific molecule is present, is governed by the measurement of the appropriate serum marker. In the present AIS network, RANKL, OPG, vitamin D, PTH, OCL, OBL and mechanical tissue loading (Mech) are integrated. The corresponding Boolean functions are as follows:

$$\begin{aligned} \text{RANKL} &= \text{OBL} \wedge (\neg\text{Mech} \vee \text{VitD} \vee \text{PTH}) \\ \text{OPG} &= \text{OBL} \wedge \text{Mech} \wedge (\neg\text{PTH} \vee \neg\text{VitD}) \end{aligned} \tag{3}$$

All molecules are either present or not. They are represented by a Boolean value. Mech is a thresholded Boolean value of ϵ. The cell numbers/activities of OBL/OCL are also thresholded. The molecule states are initialised randomly with a spatially uniform distribution and an average cover rate of 50%. Molecules that are never updated in the network are redistributed before every time iteration to avoid bias due to the specific configuration that was generated for the first iteration. OBL is then increased if PTH is true whereas OCL is increased if RANKL is true.

2.4 Computational implementation

The presented modelling approach is well suited for parallel implementation since each CA cell only depends on its nearest neighbourhood but is propagated independently with each time step. We therefore implemented the prototype for the execution on a graphics processing unit (GPU) using OpenCL. This allows to exploit the large amount of computing units available on a GPU compared to only a few cores of a CPU on a normal computer workstation. The analysis is performed with the dedicated micro-FE solver ParOSol [8] on the CPU. Boundary conditions are defined according to a bone loading estimation algorithm [9], providing physiological *in vivo* loading for this particular patient. They include compression (zz-direction) and shear strains (zx- and zy-direction). For the CA, biological cell and tissue density, molecular states as well as the mechanical tissue loading are stored for the full CA domain and thus a so called full-space approach is followed. Tissue density and loading are stored as floating point numbers. Since GPU memory is very limited, the biological cell count/activity per voxel is restricted to 8-bit characters, which still provides 256 quantification levels. For each molecule, only 1 bit is needed in the Boolean network, thus, for each voxel all molecular information can be encoded in a single byte.

3 Results

In the control simulation, bone density stayed constant with only a minor change of 0.03% between start and end point, representing normal healthy bone remodelling with 1.40% formation and 1.43% resorption. Reducing physical activity led to a decrease of 20.65% in bone density with 1.40% and 22.05% bone formation and resorption, respectively (Figure 1). In both cases, RANKL/OPG did not change during the simulated time.

Calculating tissue loading with micro-FE analysis required 494 s on a supercomputer (96 CPU cores, CSCS Cray XC40) while the CA computation including the Boolean network required 45 s on a workstation GPU (AMD Radeon HD 6750M 1024 MB, MacBook Pro 2011).

4 Discussion

We here propose an *in silico* prognosis approach to combine varying clinical measurements and thus clinical big data to explore complex diseases. The theoretical basis is outlined as well as a first prototype simulating AIS related

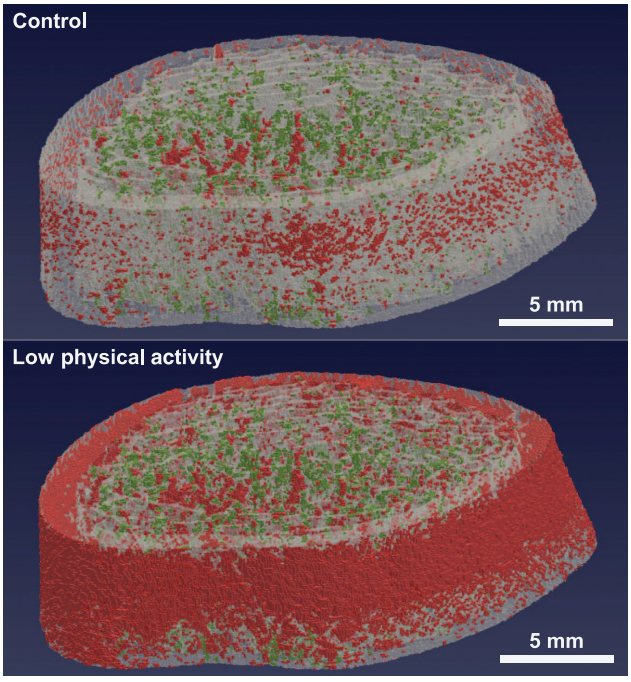

Figure 1: Differences in bone density between the first and last iteration for a control and a low physical activity simulation of an AIS patient. Bone resorption is depicted in red, formation in green, and no change in white transparent.

bone loss due to inadequate physical activity is presented. We successfully demonstrate this effect by feeding clinical big data including medical images, serum markers, and physical activity levels into the *in silico* prognosis. Mobile monitoring data could potentially be included, too.

Our results are in agreement with bone's capability to adapt to mechanical loading since bone density decreased with decreasing physical activity. The *in silico* method furthermore allows a more local analysis of bone loss, showing that resorption occurred predominantly in cortical bone and the centre of trabecular bone indicating alterations in spatial bone remodelling activity that might be linked to the local mechanical loading conditions. Bone formation, in contrast, occurred more evenly distributed throughout the trabecular bone.

The current implementation is very fast with only a few seconds execution time. However, micro-FE calculations were performed on a supercomputer and are currently coupled through a shared results file with the CA requiring additional file reading/writing. Computation time could thus be reduced by including the micro-FE solver in the CA. Although the Boolean networks are modelled for each image voxel, they do not add much to the total computation time and thus, they might be extended to include several quantification levels similar to how the biological cell counts are implemented.

In conclusion, the present *in silico* prognosis method allows to account for a variety of clinical measurements to study complex diseases using the concept of big data.

Funding: This work has been supported by the Holcim Stiftung for the Advancement of Scientific Research and the Swiss National Supercomputing Centre (CSCS).

Author's Statement

Conflict of interest: Authors state no conflict of interest. Material and Methods: Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, et al. Adolescent idiopathic scoliosis. Nat Rev Dis Primers. 2015;1:15068.

[2] Yu WS, Chan KY, Yu FWP, Ng BKW, Lee K-M, Qin L, et al. Bone structural and mechanical indices in adolescent idiopathic scoliosis evaluated by high-resolution peripheral quantitative computed tomography (HR-pQCT). Bone. 2014;61:109–15.

[3] Lee WT, Cheung CS, Tse YK, Guo X, Qin L, Ho SC, et al. Generalized low bone mass of girls with adolescent idiopathic scoliosis is related to inadequate calcium intake and weight bearing physical activity in peripubertal period. Osteoporos Int. 2005;16:1024–35.

[4] Batista R, Martins DE, Hayashi LF, Lazaretti-Castro M, Puertas EB, Wajchenberg M. Association between vitamin D serum levels and adolescent idiopathic scoliosis. Scoliosis. 2014;9(Suppl 1): O45.

[5] Slemenda CW, Miller JZ, Hui SL, Reister TK, Johnston CC Jr. Role of physical activity in the development of skeletal mass in children. J Bone Miner Res. 1991;6:1227–33.

[6] Lam TP, Ng BKW, Cheung LWH, Lee KM, Qin L, Cheng JCY. Effect of whole body vibration (WBV) therapy on bone density and bone quality in osteopenic girls with adolescent idiopathic scoliosis: a randomized, controlled trial. Osteoporos Int. 2013;24:1623–36.

[7] Christen P, Ito K, Ellouz R, Boutroy S, Sornay-Rendu E, Chapurlat RD, et al. Bone remodelling in humans is load-driven but not lazy. Nat Commun. 2014;5:4855.

[8] Flaig C, Arbenz P. A scalable memory efficient multigrid solver for micro-finite element analyses based on CT images. Parallel Comput. 2011;37:846–54.

[9] Christen P, Schulte FA, Zwahlen A, van Rietbergen B, Boutroy S, Melton LJ, et al. Voxel size dependency, reproducibility and sensitivity of an in vivo bone loading estimation algorithm. J R Soc Interface. 2016;13.