Mechanical testing of bones: the positive synergy of finite–element models and in vitro experiments

BY LUCA CRISTOFOLINI1,2,*, ENRICO SCHILEO2, MATEUSZ JUSZCZYK1,2, FULVIA TADDEI2, SAULO MARTELLI1,2 AND MARCO VICECONTI2

1 Facoltà di Ingegneria, Università di Bologna, Italy
2 Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Bologna, Italy

Bone biomechanics have been extensively investigated in the past both with in vitro experiments and numerical models. In most cases either approach is chosen, without exploiting synergies. Both experiments and numerical models suffer from limitations relative to their accuracy and their respective fields of application. In vitro experiments can improve numerical models by: (i) preliminarily identifying the most relevant failure scenarios; (ii) improving the model identification with experimentally measured material properties; (iii) improving the model identification with accurately measured actual boundary conditions; and (iv) providing quantitative validation based on mechanical properties (strain, displacements) directly measured from physical specimens being tested in parallel with the modelling activity. Likewise, numerical models can improve in vitro experiments by: (i) identifying the most relevant loading configurations among a number of motor tasks that cannot be replicated in vitro; (ii) identifying acceptable simplifications for the in vitro simulation; (iii) optimizing the use of transducers to minimize errors and provide measurements at the most relevant locations; and (iv) exploring a variety of different conditions (material properties, interface, etc.) that would require enormous experimental effort. By reporting an example of successful investigation of the femur, we show how a combination of numerical modelling and controlled experiments within the same research team can be designed to create a virtuous circle where models are used to improve experiments, experiments are used to improve models and their combination synergistically provides more detailed and more reliable results than can be achieved with either approach singularly.

Keywords: skeletal biomechanics; human bones; numerical modelling; mechanical in vitro testing; sensitivity analysis; experimental validation

1. Introduction

Physiome research lies at the centre of a crossroad between two different epistemological traditions: that of natural sciences (and biology in particular) and physics. In natural sciences, empiricism is predominant, while mathematical

*Author for correspondence (cristofolini@tecno.ior.it).

One contribution of 13 to a Theme Issue ‘The virtual physiological human: computer simulation for integrative biomedicine I’.

This journal is © 2010 The Royal Society
modelling is traditionally limited to inductive models that extrapolate from repeated experimental observations. Conversely, physics is rooted in the scientific method where deductive and abductive mathematical modelling plays a central role in the formulation and falsification of hypotheses. This dichotomy makes it difficult to define a clear methodological relationship between experiments and mathematical models that is accepted by the vast majority of practitioners. Some researchers see with great suspicion any deductive model, trusting only empirical evidence from experiments; others consider mathematical modelling as a research domain on its own, without the need for any direct empirical evidence. In addition, the extreme specialization of research has slowly separated mathematical modelling skills from experimental skills in most research groups, and it is not rare to see groups where only one of these skills is truly developed. This is a pity: the complexity involved with understanding the systemic behaviour of human physiology is overwhelming, and one should in principle be ready to use every empirical technique available to advance comprehension.

Although we believe that the relationships between mathematical modelling and controlled experiments are generic in nature, to sustain our arguments with concrete evidence, we shall refer in this work to a specific problem of great relevance for musculoskeletal physiology: the assessment of the functional capacity of bone segments to withstand physiological, para-physiological or pathological loading (bone strength). Bone strength has been widely investigated with controlled experiments or numerical models, but only rarely with a combination of both.

Investigations on the mechanics of bony structures date as far back as the seventeenth century, but these were mainly theoretical observations (Galilei 1638). The first modern studies addressing mechanics and biomechanics of whole bones were theoretical considerations based on anatomical and mechanical observations (e.g. Wolff 1892; Koch 1917; Paul 1966–1967). Only more recently were in vitro experiments performed both on intact bones and on bones operated with implantable devices (osteosyntheses, prostheses, etc.). Such experiments were aimed at assessing the following.

— The structural behaviour of bones under load: for instance, the stiffness of bones was assessed in Simkin & Robin (1973), Stein (1976) and Cristofolini et al. (2010).

— The strain distribution on the bone surface can be investigated either using strain gauges (Huiskes et al. 1981; Field & Rushton 1989; Cristofolini et al. 2009b), or photoelasticity (Zhou et al. 1990; Slemmon et al. 1991; Cotton et al. 1994; Cristofolini et al. 1994; Peindl et al. 2004), or alternative optical techniques (Duncan 1992).

— The strength and/or mode of failure of whole bones (Martens et al. 1986; Yang et al. 1996; Lochmüller et al. 2002; Cristofolini et al. 2007a).

The loading conditions replicated in vitro generally follow two different philosophies that reflect the complexity of the human musculoskeletal system. In the first case, individual load components are applied to the bone, with no direct connection to the loading condition found in vivo. A review of this approach was published recently (Sharir et al. 2008). This approach was originally applied in
studies of whole bones in the 1970s (Simkin & Robin 1973; Stein 1976). Such loading conditions are still frequently used (e.g. Heiner & Brown 2001; Dunlap et al. 2008; Cristofolini et al. 2010) for a number of reasons.

— While for some bones the loading conditions are quite limited (e.g. the proximal femur; Bergmann 2001), many other bones in vivo undergo a number of quite different loading conditions during a variety of motor tasks. Rather than replicating a large number of loading conditions, in some cases, it is preferable to separately apply the main load components to the bone (bending in different planes, torsion and in some cases axial loading).

— In many cases, no details are available about the magnitude and direction of the loads applied in vivo to the bones. For instance, there is still debate on the weight bearing function of the slender human fibula, which according to different authors could range from zero to around 30 per cent of that of the adjacent, massive tibia (Lambert 1971; Takebe et al. 1984; Wang et al. 1996). When information is scarce or inaccurate, it may be preferable to bypass the problem by shifting to a simplified, but better controlled, loading condition.

— Applying individual load components, in general, enables improved control of the testing conditions, reducing bias and noise.

In the second case, when sufficient information is available and it is necessary to include the complexity of in vivo loading in the in vitro simulation, experimental studies aim at replicating the load components applied during selected motor tasks (O’Connor 1992; Cristofolini 1997). Different motor tasks need to be simulated to represent the physiological range of loading configurations (O’Connor et al. 1996; Heller et al. 2001; Cristofolini et al. 2007b). In general, such in vitro simulations involve a more complex set-up, often including the action of relevant muscle groups (Cristofolini et al. 1995; Duda et al. 1998; Cristofolini & Viceconti 1999b; Szivek et al. 2000; Stolk et al. 2001; Britton et al. 2003).

Alternatively, the mechanical behaviour of bone structures has been investigated with mathematical models. Given the complexity of bone structures, approximated solutions of the mathematical relationships describing their behaviour are accomplished by means of numerical (or computational) models (Babuška & Oden 2004). Perhaps the most commonly used numerical models in biomechanics are finite-element (FE) models (Zienkiewicz 1967). An FE model is a numerical model that enables calculation of a field of selected physical quantities (e.g. stress, strain, but also temperature or flow) based on discretization of the domain of interest into elements of relatively simple geometry and finite dimensions. Similar to experimental studies, FE models also have been extensively used for the determination of the mechanical stresses that physiological activities, pathological conditions or surgical modifications induce in human bones (Orne & Young 1976; Rybicki & Simonen 1977; Simon et al. 1977; Hayes et al. 1978). The knowledge of bone stresses is in fact of great importance in research (e.g. to investigate mechano-biological phenomena; Ruimerman et al. 2005), prosthesis design (Long & Bartel 2006; Martelli et al. 2008) and clinical practice (e.g. to plan the individual rehabilitation after

Phil. Trans. R. Soc. A (2010)
massive skeletal reconstruction procedures; Taddei et al. 2003). FE models, when compared with most experimental techniques, offer the advantage of estimating the stress/strain field over the whole region of interest rather than in a few selected points, and permit a time-effective and virtually infinite variation of study parameters. Moreover, image processing and FE model generation procedures developed in the last 20 years (Keyak et al. 1990; Lotz et al. 1991; Viceconti et al. 2004; Taddei et al. 2007) facilitate patient-specific FE models to be derived from in vivo diagnostic data, while the mechanical stress in bones cannot be measured in vivo without the use of an invasive, and in most cases unethical, surgical procedure (Aamodt et al. 1997). Subject-specific modelling procedures (see Viceconti & Taddei (2003) for a review on this topic) enable the creation of an FE model of a bone segment from computed tomography (CT) images. Such procedures currently represent the best source of information on long bone morphology and mechanical properties applicable in vivo. As a result of the continuous technical improvements in image processing and the widespread use of CT imaging, the number of subject-specific FE modelling studies is exponentially growing, exceeding 100 papers in the biomechanics literature (Schileo et al. 2008b).

Both numerical models and in vitro experiments are models of the physical event under investigation. Therefore, both their relevance and their reliability cannot be taken for granted. The present authors are convinced that a synergistic use of numerical models and in vitro experiments can provide at the same time corroboration to both types of models, and eventually a deeper insight into the physical event being investigated. There are a few studies where numerical modelling and controlled in vitro experiments are combined in a single study, mostly to the specific scope of corroborating/falsifying the numerical models with the experiments, what is usually called validation. Validation is a crucial aspect since it is the only acknowledged procedure to ensure model reliability for a clinical application (Viceconti et al. 2005). After the pioneering work of Huiskes et al. (1981), several studies investigated the accuracy with which FE models are able to predict experimentally measured values of surface strains (Lotz et al. 1991; Keyak et al. 1993; Ota et al. 1999; Gupta et al. 2004; Anderson et al. 2005; Taddei et al. 2006, 2007; Bessho et al. 2007; Schileo et al. 2007; Yosibash et al. 2007b) and bone-implant micro-motions (Taddei et al. 2010). Other studies concentrated on the prediction of lumped parameters such as structural stiffness for a given component of loading (Cody et al. 2000) or failure load in a given loading configuration (Crawford et al. 2003). However, the combination of the two approaches, numerical and experimental, is most times limited to validation purposes, without any contribution going in the opposite direction, from FE models to in vitro experiments: therefore, in most cases it appears as a one-way process.

The scope of this paper is to explain how, and to what extent, the synergistic use of in vitro experiments and FE models can improve physiome research. In particular, the authors want to demonstrate that the combination of numerical modelling and controlled experiments within the same research team can be designed to create a virtuous circle where models are used to improve experiments, experiments are used to improve models, and their combination synergistically increases the knowledge we have on a given musculoskeletal problem.

Phil. Trans. R. Soc. A (2010)
2. Limitations of FE models

By definition, a model cannot account for something that is totally unknown to the modeller. In the context of this paper, using numerical models one can investigate only known (or at least suspected) scenarios. To a certain extent, this is also true for experimental models: controlled experiments are designed with some specific failure scenarios in mind. However, it is not impossible that during a controlled experiment something totally expected is observed, which might suggest other potential failure scenarios; this cannot happen to a model. Mathematical models are entirely fabrications of the human mind, and as such they can only know what is already known.

Besides this primary limitation, there are others related to the specific nature of numerical models. The biggest limitation comes from the process at the root of each model: idealization. We observe the physical reality, and from this observation we develop an idealized representation of the phenomenon of interest, which we describe in mathematical terms (Frigg & Hartmann 2008). Such an idealization can be achieved either by neglecting certain aspects (Aristotelian idealization) or by assuming true something we know as being false (Galilean idealization: e.g. a mass-less object). In both cases, this process is associated with some limitations of the validity of the model, which cannot be overcome. For instance, a model where contact is assumed to be frictionless will never be able to elucidate anything useful about frictional abrasion.

The second limitation resides in the numerics we use to solve the numerical model (Babuska & Oden 2004). Every numerical solution is somehow approximated, and this approximation somehow defines a ‘resolution’ for our model. Details and events ‘finer’ than this resolution cannot be discriminated with that model. Some numerical approximations, such as those that come from finite-precision calculus used by computers, are usually very small and thus of lesser concern. But in methods like the FE method, there are other aspects such as the discretization of the integration domain that might play a much bigger role. So for example, if we model a bone using a 10-node tetrahedral element with a parabolic displacement shape function (and thus linear strain shape function), every variation over the element edge length that is more than linear will not be predicted by the model. Thus, the contact stresses predicted at the tip of a relatively sharp object, like the tip of a hip stem, are not reliable up to two to three elements distance, unless special care is applied in modelling such tip contact adequately.

The third limitation is related to the so-called identification of the model, i.e. the determination of the values to be assigned to the model parameters in order to complete the calculus. In general, these values derive from experimental measurements or from some estimation, and in both cases have finite precision. Such uncertainty in the model parameters unavoidably propagates to the model predictions (prediction uncertainty or model sensitivity). Any attempt to discriminate phenomena whose differences fall inside such prediction uncertainty is by definition impossible. A typical problem is the definition of the boundary condition in validation experiments. For example, if no special care is taken, it is easy to commit errors of a few millimetres on the point of application of the...
Figure 1. Examples of transducers used for *in vitro* mechanical testing of bones. (a) Displacement transducers (linear variable displacement transducers (LVDTs) in this instance) are used to measure microscopic displacements of bone extremities under load (also visible are the lead wires of the strain gauges). (b) Detail of strain gauges used for point-wise measurement of strain on the bone surface: triaxial stacked rosettes of different sizes are bonded in the different regions of the proximal femur.

force with respect to the bone. This uncertainty propagates in terms of applied moments and therefore can produce fairly big errors in the stress and strain predictions.

### 3. Limitations of *in vitro* experiments

*In vitro* experiments are time consuming and expensive. Inclusion of any additional detail in the experimental set-up (e.g. simulation of the action of a muscle group) requires additional experimental equipment and more sophisticated controls. Moreover, in most cases, *in vitro* experiments involve costly transducers such as strain gauges and displacement transducers (figure 1), and their conditioning and data acquisition units.

*In vitro* measurements are affected by experimental error (figure 2; Taylor 1997). Random errors can be induced by a number of factors including the following.

— Noise of the measurement systems: all measurement systems, including mechanical ones, are affected by ‘noise’, including mechanical vibration, electrical noise, interference from neighbouring devices, etc. (Doyle & Phillips 1989; Dally & Riley 2005).

— Scarce repeatability of the applied loads: in most cases material testing machines or dedicated simulators are used. In all such cases, actuators and control systems are used that unavoidably introduce some random error (Dally & Riley 2005). In addition, biomechanical testing of bones requires the use of dedicated loading fixtures that in many cases introduce an additional error (Cristofolini *et al.* 1997; Dally & Riley 2005).
Figure 2. Types of error affecting experimental measurements. (a) An ideal accurate measurement has good repeatability and the average measured value is close to the ‘true’ value; (b) systematic errors lead to biased measurements; (c) large experimental noise leads to scarcely repeatable measurements.

— Uncertainty in the alignment and positioning of the test specimens: holding and applying loads to bone segments can be difficult, as bones lack flat surfaces or connectors that are normally used to position and load specimens in classic experimental mechanics. This can result in a large variability between the loads applied at different test repetitions or between specimens (O’Connor 1992; Cristofolini et al. 1997).

— Uncertainty in the positioning and alignment of the transducers: in most cases strain gradients are present in bone testing, and relative displacements between two objects vary from point to point. Moreover, as strain is described by a tensor, and displacement by a vector, strain or displacement measurements according to a given direction (which is determined by the sensor itself) are affected by alignment. Therefore, if a transducer is randomly misplaced or misaligned, the readout will suffer from an unpredictable error (Cristofolini & Viceconti 1997; Cristofolini et al. 1997, 2007b).

In addition, systematic errors can be induced by the following.

— Poor information about in vivo loads. Only recently, and only for a limited number of anatomical regions, have accurate forces been measured in vivo (e.g. Kotzar et al. 1991; Taylor & Walker 2001; Bergmann 2008). This also includes strain rate, as bone, to some extent, is viscoelastic. In most cases, in vitro simulations must rely on partial information, based on a limited number of subjects, and often limited by major assumptions (O’Connor 1992).

— Poor identification of the anatomic reference frames. Reference frames are needed to enable consistent loading on different bone specimens, with respect to the alignment and position of the applied loads (Cristofolini in press). As such reference frames often rely upon subjective identification of bone landmarks, different operators might tend to align the specimens (and the applied loads) in a different way.
— **Ill-designed loading systems.** In some cases, loading systems are designed in a way that prevents control of the applied load. An example is that of over-constrained loading systems, where additional load components (other than the intended one(s)) are generated within the loading system (Cristofolini & Viceconti 1999a). If such additional load components are neither controlled nor measured, they bias the applied load in an unknown way.

— **Poor preparation of the transducers.** In many cases, transducers (e.g. strain gauges) must be bonded onto the bone surface. Poor preparation can result in extremely inaccurate readouts (Viceconti et al. 1992; Vishay Micro-Measurements 2005, 2009).

— **Perturbation induced by strain measurement systems.** When a sensor is bonded to the surface of a structure, it contributes to the load-bearing capacity. Therefore, the presence of a strain sensor (strain gauge or photoelastic coating) tends to reinforce the surface, leading to a systematic underestimate of the actual strain distribution. It has been shown that such a perturbation induced by strain gauges (Perry 1985, 1986; Ajovalasit & Zuccarello 2005) or photoelastic coatings (Zandman et al. 1962; Cristofolini et al. 1994) can be significant in relatively soft materials such as bone tissue.

A further limitation of experimental testing is that the addition of any new measured parameter, or replication of measurements in additional locations, is associated with the need of using more transducers, together with more powerful data loggers. Therefore, any new measurement location is associated with a significant cost increase. The inclusion of additional transducers is also associated with an increased complexity (and possible errors) of the testing set-up.

It must also be considered that, if an *in vitro* experiment needs to be performed repeatedly under similar conditions (for instance, to test different implantable devices or different loading configurations), most of the experimental cost and effort need to be replicated. In fact, in general, there is little economy of scale for *in vitro* testing of bones. Therefore, *in vitro* experiments are not ideal candidates for performing comparative studies on a large number of conditions or for performing sensitivity analysis where a large number of scenarios need to be explored.

In addition, availability of human tissue specimens is limited, both for practical and for ethical reasons. For these reasons, in many cases, the sample size of *in vitro* experiments is statistically underpowered: in very few cases are sample sizes larger than 10 found, while in many cases the sample is limited to a single-bone specimen (Cristofolini et al. 2010). This prevents the drawing of sound conclusions from a statistical point of view, unless a large effect exists that can be statistically detected even with a small sample size (Montgomery 2005). Also tissue preservation poses serious problems. In fact, it has been shown that both embalming and freezing can significantly alter the mechanical tissue properties, unless dedicated protocols are followed (Linde & Sorensen 1993; Currey et al. 1995; Yosibash et al. 2007a; Ohman et al. 2008).
Synergy of FE models and in vitro experiments

A final, but critical, limitation of experimental techniques is that measurements often focus only on the external surface, which can be accessed by contact sensors or by optical methods. Only recently have radiographic techniques provided sufficient resolution to enable measuring displacements and strains within a bone structure (Voide et al. 2009).

4. Basic prerequisites of FE models

Following the approach described in §2, we can divide the prerequisites for accurate FE modelling into three parts: mathematics, numerics and identification. The main steps to build an FE model that meets the basic prerequisites are the following.

(a) Definition of the modelling scope

The scope of the model should be defined as clearly and unambiguously as possible. Currently, we have identified the following general scope categories: representation, correlation, quantification, prediction and simulation. We model reality to conceive it; so the first scope of modelling is representational. Representations are useful to many purposes, but probably the first one is ‘remembering’: we create mental models of reality to remember it. The exact mechanism that provides memory is still under investigation, but it is known that we remember a concept by connecting it to a network of other concepts; so as soon as we memorize a model, this gets connected to other models. This is the second scope of modelling: correlating portions of reality. This makes it possible to compare two portions of reality: so models are used also to compare. Models make it possible to quantify and measure. Another scope of modelling is predicting. As soon as we discover regularities in the world, we are able to make predictions. The last scope of modelling is simulation. Through prediction, we can mentally explore which actions will transform portions of reality in the way we want. Representational models are not relevant in this context; correlation models are usually statistical; quantification models are used in exercises such as inverse problems; prediction is the most relevant category here, while simulation involves a replacement of the reality with the model. These are general categories, but the modelling scope must be defined with greater detail: assuming we pursue a predictive modelling scope, we should define which portion of reality we want to capture, which quantities we expect to predict, under which general conditions.

(b) Idealization

From observation, we notice how the quantities identifying the state of the system are organized in space and time, and how they correlate with each other. From this, we create a cognitive artefact which depicts that portion of reality by idealizing it (for instance, we assume that the contact between two bodies is frictionless or that an empirical law regulates the relation between two quantities). In scientific modelling, this cognitive artefact should be expressed in logical terms: depending on the type of logical reasoning we use, models can be divided into inductive models (i.e. regression models, data models), deductive models (i.e. models based on the laws of physics) or abductive models (i.e. Bayesian models).
(c) Deployment

The model is now transformed into something that can be practically used to absolve the modelling scope. Typically, the idealization is captured into mathematical form, which is then solved for a set of initial values either analytically or numerically. Owing to the complexity of the models involved in biomedicine, most models are solved numerically.

(d) Existence and uniqueness

If we deploy a model into a mathematical form, we must verify that this mathematical problem provides one and only one solution for a given set of initial values. This step is frequently taken for granted, especially in deductive models, because the mathematical forms used have already been extensively tested from this point of view. However, if we use a new mathematical form, this is an important step.

(e) Verification

If the mathematical model is solved numerically, we need to determine the accuracy with which the equations are solved. For linear models, it is generally possible to estimate the errors produced by the numerical solution. More specifically, post hoc indicators like the stress error indicator, or full convergence tests (figure 3) on parameters such as the potential energy of the entire bone, displacements and strains at the points of interest (e.g. strain gauge locations) can estimate the error owing to one of the most delicate aspects of FE numerics, which is the spatial discretization of the domain.

For instance, in a recent paper (Helgason et al. 2008b) where a femur was modelled with quadratic tetrahedral elements, the authors used as a general convergence metric the global strain energy density, and then, since the paper was focused on strain prediction, they also verified the convergence of the cumulative relative error (each mesh refinement compared with the densest refinement) of the absolute values of maximum and minimal strain at the points of interest.

The threshold for achieving convergence was set at a 1 per cent rate of change (provided a steady decreasing trend on five progressively more refined meshes), and was met with an average element edge length of 3.3 mm.

(f) Sensitivity analysis

A predictive model estimates the state of a system from a set of initial values. It is important to verify how sensitive are the model estimates to variations of these initial values. The reason is twofold: firstly, the initial values used for the identification of the model are always associated with an uncertainty, and we need to ensure that this uncertainty does not affect the conclusions we draw from the model. Secondly, if we notice that one model estimate has an abnormal sensitivity to small variation of some initial values, this might be an indication that idealization or its mathematical or numerical deployments have some problems. Assuming one has a reliable estimate of the uncertainty associated with each parameter to be used in the model, it is recommended to run a sensitivity analysis to estimate how these uncertainties interact and propagate in the model and affect the predictions of interest.
Synergy of FE models and in vitro experiments

Figure 3. Example of a convergence plot to check the adequacy of the mesh used to model a human femur: in this instance, the displacement under load at a given point of the bone (the centre of the femoral head, in this instance) asymptotically tends to a certain value. Similar plots should be inspected for a number of other indicators such as strain values at most relevant points and strain energy density. Mesh refinement is stopped when a further increase in the number of elements involved affects the predicted output by less than an assigned threshold. (Reproduced with permission. Copyright © VPHOP consortium.)

Depending on the cases, this can be done with a simpler exploratory analysis such as the design of the experiment and related simplified Taguchi strategies (Montgomery 2005) or using a full-blown Monte Carlo-based statistical FE modelling approach. Sensitivity analysis is important not only because it shows the effect of the input uncertainties on the output, but also because it shows how sensitive the model predictions are to the input parameters. While in some cases a high sensitivity is unavoidable, in many cases the finding of an unexpectedly high sensitivity of a predicted quantity over small variations of an input parameter suggests that the model itself (e.g. its mathematics or its numerics) might have a problem. Sensitivity analysis is frequently the best way to discover truly unforeseeable errors in the model.

(g) Validation

Once we know what is the accuracy of the numerical solution, we can determine the predictive accuracy of the model by comparing its predictions with the equivalent quantities measured in a controlled experiment. As a general principle, one should never forget that a statement like ‘bone is viscoelastic’ is completely false. What we mean is that if bone behaviour is modelled using viscoelastic equations, usually more accurate results are obtained than using other types of
mathematical models. But also this second statement is somehow inaccurate, as it neglects the aspect of the limit of validity. Every mathematical model is reasonably accurate only within certain limits: no material behaves elastically indefinitely, and most materials under sufficiently high temperature become viscoelastic. It must be noted that validation (in the sense of determining whether a computational model represents the actual physical event with sufficient accuracy; Babuska & Oden 2004) is perhaps not even possible. In fact, from an epistemological point of view, an agreement between physical observations and model predictions does not validate a theory (such as a model, in our case), while a single exception is sufficient to falsify the same theory (Popper 2003). Strictly speaking, a model can never be invalidated: it can only be falsified. According to Popper, however, a model can be ‘corroborated’ if it withstands severe tests and is not superseded by another theory. In this perspective, validation with respect to a specific series of tests and preset tolerances may be a legitimate basis for decision making (Babuška & Oden 2004). There is not a single possible approach for validation that applies to all problems. When the decision of which mathematical model is the best one is not self-evident, we recommend whenever possible to use the strong inference approach (Platt 1964; see §9), i.e. selecting two or more candidate mathematical models, and comparing them with each other with respect to the results of controlled experiments. One should always remember Ockham’s razor: if two models show the same predictive accuracy, everything else being the same, we should choose the simplest one.

5. Basic prerequisites of in vitro experiments

Although this is sometimes not sufficiently appreciated, in vitro experiments are just models of the physical event under investigation. Therefore, the closeness of an in vitro experiment to the physical event under investigation should not be taken for granted. This should be assessed by comparing the experimental output against some evidence of the physical event under investigation: in the case of bone testing, this could be for instance some clinical data of bone fractures or in vivo recording of load.

The design of a set-up for in vitro experiments must be based on a compromise between the desire to include as many details as possible to make the experiment represent physiological conditions and the need to keep the experiment simple to enable accurate control of the inputs (Currey 2008). In vitro experiments are models by all means. When designing an experiment, one must bear in mind that the experiment, like any other model, is capable of capturing only some details of a reality that is much more complex. Therefore, in vitro experiments must be designed with a specific research question in mind. Details that are not relevant to such a research question should be omitted, as they would only reduce the overall control we have on the experiment (Cristofolini 1997). In other words, quoting Albert Einstein: we should make things as simple as possible, but not simpler.

Before data from in vitro tests can be used, the reliability of such experiments must be assessed, so that each measurement is associated with an estimate of its uncertainty. Repeated measurements on the same specimen are crucial to estimate the measurement repeatability. At the same time, repetitions on more specimens are necessary to estimate the inter-specimen variability.
Synergy of FE models and in vitro experiments

(Cristofolini 1997; Taylor 1997). Extensive work is needed to minimize sources of experimental error (§3) so as to ensure that an in vitro experiment provides sufficiently accurate results to be used in clinically relevant applications, or to support FE models (Cristofolini 1997). While such refinement must be carried out with painstaking care in the laboratory, significant support can be provided by FE models for some aspects (§7).

6. How in vitro experiments can improve FE models

Many of the limitations of FE models described in §2 relate to the numerics and post-processing strategies, and therefore the experiment has no role in solving them. However, other limitations relate to the importance of providing FE models with reliable input data (identification) and for checking FE models against reliable reference measurements (validation). Below is listed a number of roles that experiments can play in improving FE models of whole bones (figure 4). The key point is that the most accurate FE models must be built of bone segments (in a subject-specific fashion) that are available also for in vitro mechanical testing. Such mechanical testing must include a multi-scale approach, so as to make available information at the different dimensional scales, from the whole body down to the tissue and subtissue levels.

(a) Preliminary identification of most relevant scenarios and indicators

As noted above, a numerical model can only address known scenarios: if the details of an FE model (e.g. loading configurations, type of elements, mesh refinement) are chosen to provide the most accurate information for a chosen indicator (e.g. bone stress), they might be unsuitable for predicting other magnitudes (e.g. implant-bone displacements, or fracture). In a certain way, this is also true for experiments (which are based on in vitro models of reality). However, as in vitro experiments rely upon physical specimens, to some extent they are closer to reality and therefore are better capable of capturing some aspects of reality that might be completely missed by FE models. Once an in vitro experiment has shown that a specific scenario or a specific mode of failure is potentially relevant, FE models can be designed or tuned to address such a scenario or mode more specifically.

(b) Tissue level and subtissue level: constitutive equations and failure criteria

In the literature, FE models aimed at predicting organ-level strength, structural behaviour or bone-implant response usually adopt a continuum-level assumption for the modelling of bone tissue. As outlined in §1, subject-specific FE models from CT data have recently become the elective choice to evaluate possible clinical applications. Alternative approaches have been presented where an organ-level model containing information at the tissue level (and therefore explicitly modelling the trabecular architecture) has been generated (Van Rietbergen et al. 2003). However, such models cannot currently be associated with clinical applications, as they imply the use of imaging techniques that are far from clinical application and rely upon massive parallel supercomputing. Continuum-level models require the definition of continuum-level bone tissue mechanical properties.

Phil. Trans. R. Soc. A (2010)
Bone tissue is inhomogeneous, anisotropic and to some extent viscoelastic in nature (Lakes & Katz 1979a,b; Fung 1980; Fritsch & Hellmich 2007). For the majority of biomechanical studies simulating bone stresses as a consequence of physiological loads, bone viscoelasticity can be neglected (Helgason et al. 2008a). Where strain rate is an important factor (e.g. impact), bone viscoelasticity can be partially accounted for by incorporating the strain-rate effect in the definition of elastic modulus (Carter & Hayes 1977). The modulus of elasticity (the Young modulus) of bone tissue has been reported to vary by over 50 per cent for cortical bone and by over 500 per cent for cancellous bone, depending on anatomical site, direction of loading and donor details (Reilly et al. 1974; Rohl et al. 1991; Hodgkinson & Currey 1992; Keller 1994). Most of this variability can be explained in terms of inhomogeneity. However, in early FE models, bone tissue was modelled as a fully homogeneous and isotropic material (Verdonschot
Synergy of FE models and in vitro experiments

et al. 1993; Villarraga et al. 1999), distinguishing just the histological bone type (cortical versus trabecular). Average elastic constants were thus assigned to the entire bone structure. This is associated with large errors of the same order of magnitude of the uncertainty associated with the Young modulus. Errors of the order of 10–50% are to be expected on average, with local peaks largely exceeding 100 per cent. More recently, bone inhomogeneity was introduced in FE studies (Keaveny & Bartel 1993; Viceconti et al. 1998), exploiting the capability of deriving bone tissue density from CT scan data (Kalender 1992) and using the acknowledged relationship that exists between the density and elasticity of bone tissue, when tested under the hypothesis of continuum mechanics (Carter & Hayes 1977; Keller 1994; Wirtz et al. 2000; Morgan et al. 2003).

The above cited density–elasticity relationship is, however, still associated with a very wide confidence band (Helgason et al. 2008a). There is room for significant improvements driven by experimental testing. First of all, it has been highlighted that a significant part of this uncertainty can be removed by following adequate guidelines for testing of bone tissue specimens (Keaveny et al. 1997; Ohman et al. 2007; Helgason et al. 2008a; Lievers et al. 2010). However, a random error still remains, which is associated with the inter-subject variability of the density–elasticity dependency.

Therefore, a further possible improvement, though limited to in vitro validation studies, could consist of testing bone tissue specimens obtained from the same bone that is modelled with FE models. Such tests at the tissue level should be performed on the bone tissue, after organ-level testing has been completed, following an established paradigm for multi-scale testing of bone structures (Cristofolini et al. 2008b). Such tests may include the following.

— Directly measuring the bone mineral density by ashing (Ohman et al. 2007). This procedure overcomes one of the two steps involved in converting CT density to material properties, as material density is directly measured.

— Tissue specimens of some millimetres in dimension can be extracted at selected regions and tested (Helgason et al. 2008a). In this case, directly measured material properties are available for the selected locations, with a direct control on the quality of the material properties assigned to each region of the FE model.

— Tissue histomorphometry from micro-CT scanning (Perilli et al. 2007) can improve the assessment of bone structure and its associated anisotropy. Information from such assessments would enable the enriching of FE models with reliable anisotropy data.

— Subtissue microstructure from polarized light microscopy (Beraudi et al. 2010) provides clear insight into the microstructural arrangement of bone, including collagen orientation.

— All such experimental measurements at the tissue and subtissue levels would provide better assessment of the model parameters, thus improving the correct identification of the model.

Finally, it must be remembered that the cited density–elasticity relationships can reliably define only one elastic modulus. Featuring subject-specific bone anisotropy would constitute a significant improvement for FE models. The
development of atlases and templates incorporating information from a population of human bones could be one way to proceed. More interestingly, it has been recently proposed that micromechanical multi-scale models can provide the whole set of orthotropic bone tissue constants (yet, not the directions of structural symmetry) from information on the mineral content as given from CT scans of a mandible (Hellmich et al. 2008). The corroboration of this micromechanical model up to the tissue level has its basis in multi-scale mechanical testing (Fritsch & Hellmich 2007; Fritsch et al. 2009). Further experimental tests at the organ level are needed to evaluate whether it can overcome the predictive ability of inhomogeneous but isotropic models under loading conditions that clearly elicit a bone anisotropic response (Trabelsi et al. 2009).

(c) Organ level: boundary conditions

A very critical factor determining the output of an FE model is the position of the applied force relative to the bone. For instance, when the hip joint force is applied to the femoral head in vitro (Cristofolini et al. 1994, 2006, 2007a; Cody et al. 1999; Ota et al. 1999; Keyak et al. 2005; Taddei et al. 2006), it is not possible a priori to determine accurately the position of the resultant force, as: (i) the contact area between the bone surface and the loading device is difficult to measure accurately owing to the large deformation of the bone surface and (ii) even if the contact area was accurately measured, the distribution of the contact pressure (and its resultant) cannot be easily measured experimentally. Additionally, long bones undergo significant deflection when loaded, of the order of some millimetres (Cristofolini & Viceconti 1999a,b; Cristofolini et al. 2006). Consequently, it is not possible to predict, a priori, the changing position of the applied force while the bone specimen is deflecting under the applied load. Also, it is important to take such large deflections into account, as they possibly undermine the linearity assumptions.

Therefore, even assuming that the position of the applied force is defined in principle (based on biomechanical and anatomical considerations), the problem remains how to accurately measure the actual position of force application in the actual in vitro experiment. In some studies, it has been assumed that the position of the applied force defined in an FE model should be the one theoretically identified on an ideal physical bone specimen (i.e. ignoring the local and global bone deformation and postulating an idealized contact between the physical specimen and the constraints; Keyak et al. 1998; Cody et al. 1999; Taddei et al. 2006). In other cases, no information has been provided on the location of the applied force that was used to replicate the experimental forces in the FE models (e.g. Ota et al. 1999). Inaccurate identification of such a position would undermine the accuracy of the FE simulation and the comparison between the in vitro experiment and the FE model.

To locate the actual position of the force applied to the femoral head, the bending moment applied to a femur has been measured using strain gauges directly attached to the femur diaphysis (Villa et al. 2000; Colombo et al. 2002). However, determining the position of the applied force with such an arrangement depends on the pre-identification of the centre of the diaphysis, which cannot be achieved without a means to calculate the neutral axis of complex geometries because of the irregular shape of bones.
More recently, a dedicated transducer has been designed to accurately determine the position of the resultant joint force during in vitro tests (Juszczyk et al. in press b). A strain gauge-based transducer was designed to indirectly measure the position of the force applied to long bones by measuring the reaction moments about two perpendicular axes as generated by the applied force. When included in a typical set-up for long bone testing, the overall accuracy of the transducer for measuring the position of the applied force was 0.85 mm (one order of magnitude better than the typical uncertainty associated with such estimates if no dedicated transducer is used). Therefore, the use of such a transducer can improve the identification of the boundary conditions of FE models. In addition, this more accurate determination of the force application point may improve the in vitro validation of FE models of long bones.

(d) Overall quantitative validation

While we have already remarked that absolute validation is not possible, validation relative to a series of experiments may be acceptable (and possibly the only practical solution). Such experiments obviously represent only a series of individual cases, and cannot cover the possible range of cases applied to the numerical model. In fact, in many cases, dedicated validation experiments are designed. Validation experiments are not necessarily designed to represent a physically relevant condition: in most cases, they are designed so as to achieve the best control on the experimental input parameters and to provide accessible and measurable output to be compared between numerical models and in vitro experiments.

Mechanical testing at the organ level is undoubtedly the elective choice for FE organ-level model validation. It must be clear that the experimental–numerical comparison should be made on quantitative data and should be one-to-one for each specimen and one-to-one for each measurement within each specimen. Pooling of data coming from the application of different loading conditions, as well as pooling of data from different specimens, might help demonstrate the generality of a given FE modelling procedure, but the admissibility of each pooling step should be previously ascertained by performing adequate statistical tests such as a factorial ANOVA in case of sample homoschedasticity (Schileo et al. 2007).

The main pieces of information that can be deduced from experimental tests on bone segments are: principal strain and displacements at selected locations (Cristofolini et al. 1997, 2010), failure load (Cristofolini et al. 2007a) and point of fracture initiation (Juszczyk et al. in press a).

From the aforementioned prerequisites of numerical–experimental comparisons and knowing the characteristics of the experimental measurements, some critical issues to replicate the experimental measurements in the numerical models can be identified.

— One-to-one correspondence. It is mandatory that the FE model being validated corresponds to the same physical specimen being tested in vitro to avoid sources of error (related to inter-subject variability) that are difficult to quantify.
— Spatial registration. In order to replicate boundary conditions and the position of sensors, the relative position of the laboratory reference system with respect to that of the CT dataset from which FE models are derived has to be established. A documented procedure to achieve this aim is: (i) to digitize the bone segment, plus any relevant points, and the experimental reference frame with a digital coordinate measurement system (figure 5) and (ii) to subsequently use an iterative closest point algorithm, as proposed for solving rigid registration problems (Besl & McKay 1992), to perform the registration of the acquired points cloud on the tiled surface extracted from the CT data and also to find the transformation between the experimental laboratory reference system and the one of the FE model. The average accuracy of such a procedure was reported to be less than 0.9 mm (Schileo et al. 2007) and represents the accuracy with which relevant points are replicated in the FE models. For instance, the position of the points of load application and parts of the loading system can be acquired to improve the identification of the FE models; the position of transducers can be used to enable a point-by-point validation of the FE model against experimentally acquired strains or displacements.
Synergy of FE models and in vitro experiments

— Measurement area. Strains and displacements are measured on a finite sensing area during the experiment (e.g. strain gauges used in in vitro bone tests can have a grid length ranging typically from 1 to 5 mm). Given the highly inhomogeneous nature of bone tissue and the possible presence of high strain gradients, the calculated strains/displacements must as well be averaged on the sensing area of each strain gauge.

— Measurement direction. In the case of the application of triaxial strain gauges, two superficial principal strains are obtained for comparison with the FE predictions. As a consequence, a procedure should be developed to determine which FE-model principal strains match those determined experimentally. This can usually be done automatically under the assumption of plane-stress state at the model surface (Taddei et al. 2007). In cases where uniaxial strain gauges are used, the gauge direction should be digitized, spatially registered in the model and the component of calculated strain in that direction extracted (e.g. defining a local reference system aligned to gauge direction in the model). The same procedure should be followed for displacements measured in vitro, taking care to not neglect possible rigid body motion artefacts.

To comprehensively assess model predictions, local and global accuracy metrics can be defined. A global accuracy metric can be obtained by plotting the principal strains predicted by the FE model against the corresponding magnitudes recorded experimentally for a number of loading configurations (figure 6). The goodness of the prediction can be expressed by the determination coefficient (usually indicated as $R^2$) and by the slope and intercept of the regression curve. Ideally, one should find a perfectly linear relationship between measurements and predictions ($R^2 = 1$) with unitary slope and zero intercept. A local metric can be also computed, in terms of peak error and of average error (computed as the quadratic norm error, also known as root mean square error).

7. How FE models can improve in vitro experiments

The limitations of in vitro experiments described in §3, in most cases, are related to the need for optimization of the experimental set-up. In this perspective, FE models can fulfil a very important role in assisting with the design of an optimized experimental testing method, which enables: (i) effectively addressing the research question, (ii) optimizing the use of experimental resources, and (iii) minimizing the sources of error. Below is listed a number of roles that the FE model can play in improving experiments (figure 4).

(a) Identification of most relevant loading configuration(s)

One of the key issues in designing a biomechanical simulation (whether in vitro or numerical) is identifying the loading configuration (direction and magnitude of applied forces) that is most relevant to the research question being addressed (O’Connor 1992; Currey 2008). For instance, if one wants to investigate bone remodelling, which is driven by cyclic loads (Lanyon 1980; Fung 1990), there is a need to simulate the motor tasks that are repeated more frequently in daily life. Conversely, if one aims at understanding spontaneous fractures of the hip.
Figure 6. Quantitative assessment of the accuracy of FE predictions in comparison with experimental measurements. The goodness of fit ($R^2$ close to 1) and the closeness of the slope to 1 indicate the ability of the FE model to replicate the strains/displacements observed in the physical specimens. In this instance, eight proximal femurs were investigated using an FE model based on the CT scan of the physical specimens. (a) Principal strains were measured (and compared) at 15 locations for six different loading configurations on all eight femurs (results taken from Schileo et al. (2007)). Ideally, error bars for the measured entities should be plotted. However, for practical reasons (the plot includes 1440 points), they are not included here. Experimentally measured strain was associated with an error of 0.4% (coefficient of variation between test repetitions on the same specimen). (b) Antero-posterior and medio-lateral displacements on the femoral head and femoral diaphysis were measured by four displacement transducers for six different loading configurations (LC1–LC6); in this case, experimental and FE-predicted medio-lateral displacements are compared (L. Cristofolini & E. Schileo 2009, unpublished data). Striped bars, FE model; shaded bars, experimental. (Reproduced with permission. Copyright © VPHOP consortium.)

(i.e. those fractures deriving from physiological or sudden loading, but not from a traumatic event; Jeffery 1974; Rockwood et al. 1991), then occasional overload scenarios such as stumbling or mis-stepping should be simulated.

This question is particularly tricky in orthopaedic biomechanics as bone segments generally undergo a wide variety of loading configurations during daily life, depending on the individual lifestyle, on the motor task being performed and on a number of factors that vary within the same motor task, such as motion, speed, environment, etc. In addition, within a given motor task, the magnitude and direction of the applied forces change over time. For these reasons, if the most relevant motor task is not selected, or if the most relevant instant within that task is not simulated, the in vitro simulation fails to capture the scenario that is relevant to the selected research question.

In principle, one could perform a number of in vitro preliminary experiments to determine which of the possible motor tasks is most relevant, and, for the selected motor task, which instant should be simulated. However, practical performance of such an exploration can be extremely costly and time consuming. FE models in

Phil. Trans. R. Soc. A (2010)
Synergy of FE models and in vitro experiments

Figure 7. Loads applied to the femoral head. (a) The cone covers the range of directions spanned by the hip joint resultant forces during a number of different physiological activities. (b) FE predictions of the strain distribution for five of the load configurations explored: they were used to identify the loading direction that caused the highest risk of fracture. The maximum risk of fracture (first case reported) corresponded to a force tilted by 8° in the frontal plane and neutral in a sagittal plane. (c) The in vitro loading set-up replicated the most critical loading configuration predicted by the FE model. Also visible are the strain gauges used to measure strain on the bone surface, and the LVDTs used to measure deflection of the femur under load. (Reproduced with permission. Copyright © VPHOP consortium.)

this phase can be extremely helpful as a preliminary exploration, even based on a crude FE model, and can provide a good indication in a more effective way. For instance, the most relevant loading configurations to investigate implant stability of hip stems (Stolk et al. 2002b), or spontaneous fractures of the proximal femur (Cristofolini et al. 2007a), were identified using FE models (figure 7): first, the cone covering the range of directions spanned by the hip joint resultant forces recorded with a telemetric prosthesis during a number of different physiological activities (Bergmann et al. 2001) was identified. Then, a validated FE model was used to explore the strain distribution for a number of configurations within such a cone and to identify the loading direction that was most relevant to investigate spontaneous fractures of the femur. The maximum risk of fracture corresponded to a force tilted by 8° in the frontal plane and neutral in a sagittal plane. Finally, the in vitro loading set-up was designed based on the indications from the FE model, where the load was applied in the direction that caused the highest risk of fractures in the region of interest (Cristofolini et al. 2007a).

(b) Identification of an acceptable degree of simplification

Mechanical loading of skeletal bones derives both from bone-to-bone contact and also (and often predominantly) from the action of muscles (Duda et al. 1997; Pedersen et al. 1997; Arjmand et al. 2009). In most cases, a large number of muscles act on the same bones, and many of them are active simultaneously. Replication of each muscle force in vitro requires the use of dedicated actuators
and controllers. As stated above, a compromise must be sought between a very complex experimental set-up that takes into account a large number of factors and the need to achieve accurate control of the testing conditions. With dedicated in vitro experiments, it also is possible to investigate in vitro which muscle groups should be included in a given simulation to provide an adequate loading configuration (e.g. Finlay et al. 1989; Cristofolini et al. 1995). However, such exploratory simulations now can be performed in a more effective way by exploiting FE models (Duda et al. 1998; Stolk et al. 2001; Polgar et al. 2003; Rohlmann et al. 2009).

(c) Optimization of boundary conditions

A large fraction of the error affecting in vitro biomechanical experiments derives from poor control of the boundary conditions and forces applied to the test specimens. Although the boundary conditions and the application of force, in theory, can be defined in a controlled way provided a reference frame is associated with the bone segment (Van Sint Jan & Della Croce 2005; Cristofolini in press), their translation into practice is associated with large errors that propagate to the measured output of the experiment.

The point of application and the direction of the forces in vitro are affected by experimental errors. It is often the case that some components of such errors have little effect on the measured output (e.g. bone strain or failure load), while others propagate to the output in a dramatic way. Sometimes one force component is directly controlled, while others are applied as a consequence, and therefore only indirectly controlled. For example, a cantilever system commonly is used in vitro to simultaneously apply various forces to the bone segment while applying only a single axial load (Finlay et al. 1989; Cristofolini & Viceconti 1999a; Britton et al. 2003). It is difficult to predict a priori how such input uncertainties propagate to the measured output, because of the irregular geometry of bone and its inhomogeneous and anisotropic structure. This problem cannot be addressed in vitro, as it relates to the intrinsic uncertainty of in vitro experiments. The use of FE models enables simulating the uncertainty associated with the position, direction and magnitude of each load component (figure 8a), so as to understand how each propagates to the measured quantities (e.g. strain or failure load). Based on such sensitivity analysis, the experimental set-up can be designed to give the highest priority to accurately controlling those uncertainties (e.g. the position of one of the applied forces) that most severely affect the accuracy of the results (Cristofolini & Viceconti 1999b).

(d) Optimized use of transducers

Most of the transducers used in vitro provide point-wise measurements: strain gauges measure strain at the point of application; displacement transducers measure relative motion between two well-defined points. These types of transducers do not provide any information about the distribution of the measured variable beyond the points of application. Therefore, it is important that they are applied to the points where the variable of interest is most relevant (typically where it reaches its maximum value). FE models can provide a preliminary distribution of the measured magnitude so as to suggest where the transducers should be deployed to provide the most relevant information.
Synergy of FE models and in vitro experiments

Figure 8. Position and alignment errors affecting in vitro experiments. (a) The position, direction and magnitude of the forces applied in vitro to a bone segment are affected by errors (grey lines) with respect to the intended position (black line). (b) The position and alignment of the strain gauges used in vitro on a segment are affected by errors (grey) with respect to the intended position (black). FE models can be used to estimate how each of such components of error propagate to the measured output (Cristofolini et al. 1997; Cristofolini & Viceconti 1999b).

Although not frequent, instances where the use of strain gauges (Cristofolini et al. 1997) and displacement transducers (Cristofolini et al. 2007b, c, 2008a) were optimized using FE models are found in the literature.

The position and orientation of transducers are affected by errors (figure 8b). Even though such uncertainty can be estimated in vitro with repeated applications, this error cannot completely be avoided. If a transducer is placed in a region where the measured quantity has a steep gradient, malpositioning or misalignment of the transducer will propagate to a large uncertainty on the output measurement. FE models can be used to assess the distribution of the measured variable so as to ensure that transducers are placed in regions where such gradients are not excessively steep (Cristofolini et al. 1997; Stolk et al. 2002a).

Some transducers, such as strain gauges, provide an output that is some average over the area covered by the transducer itself (typically a few square millimetres). This effect can be both beneficial (it operates as an averaging filter, reducing the effect of the position errors described above) and detrimental (when the average is taken, the peak values are underestimated). FE models can help in defining the optimal size of the strain gauge taking into account the steepness of gradients and the presence of local peaks (Cristofolini et al. 1997).
(e) Sensitivity analysis

There are cases where the research question being investigated requires assessing the sensitivity of an output to the variation of one or more input parameters. Some examples are the following.

— The effect of bone quality on bone strength is extremely crucial for investigating bone pathologies such as osteoporosis (WHO 1994; NIH 2000).
— Different loading conditions (e.g. in relation to different motor tasks, or different patient activity levels) are known to affect bones or bone-implant constructs in different ways (Dorey & Amstutz 2002; Stolk et al. 2002b; Polgar et al. 2003).
— While developing a prosthetic device, the designer needs to know if different prosthetic materials (having different stiffness) provide better or worse performance when coupled to the host bone (e.g. Cheal et al. 1992; Lengsfeld et al. 1992; McMinn & Daniel 2006).
— Interface conditions (osteointegration as opposed to low-friction sliding) between an implantable device and the host bone or between an implantable device and the fixation media (bone cement) can determine the success or failure of a device (Collier et al. 1988; Goodman 1994; Morita et al. 1997).

Exploring the effect of such variables experimentally would require a large number of experiments and specimens. FE models are definitely more effective in performing this type of sensitivity analysis and greatly complement in vitro experiments (Martelli et al. 2006; Schileo et al. 2008c).

8. A success story: integrated investigation of the proximal human femur

In this section, a brief summary is provided to describe how, in our group, FE models and experiments synergistically improve each other, and enable an extremely accurate, detailed and reliable investigation of a number of aspects related to the proximal femur.

(a) Refinement of in vitro testing

FE models were used to optimize the experimental set-up following the approach described in §7.

— The inter-femur and intra-femur variance were reduced in steps (figure 9). The first big improvement was achieved with the introduction of the composite femur models, which enable a significant reduction of the inter-specimen variability in cases when composite bones are a viable substitute to cadaveric specimens (Cristofolini et al. 1996; Cristofolini & Viceconti 2000; Heiner & Brown 2001; Dunlap et al. 2008). Later on, with repeated testing and the support of an FE model that replicated the experimental testing conditions, sources of error were assessed. Each source of variability was examined and kept under control as much as possible, focusing first
on the strain measurement method, and later on the loading set-up. The optimal positions for strain measurement and the optimal size and type of strain gauges were identified based on the strain gradients predicted by FE models both for the femoral diaphysis (Cristofolini & Viceconti 1997; Cristofolini et al. 1997) and for the proximal metaphysis (Cristofolini et al. 2009b). This enabled a significant improvement of test repeatability (figure 9). Both the inter-specimen repeatability and, most of all, the intra-specimen repeatability that can be achieved with such a support from FE simulations are better than any other value found in the literature.

— The most suitable locations for measuring in vitro implant-bone micromotions in cementless hip stems (Monti et al. 1999) and in resurfacing hip prostheses (Cristofolini et al. 2007c) were chosen based on FE-predicted plots of micromotions.

— Based on a sensitivity analysis supported by FE models, details of the loading set-up were modified so as to improve the repeatability of some settings (alignment and magnitude of some load components) that most severely affected the test output (Cristofolini & Viceconti 1999a,b). This enabled further improving the test repeatability (figure 9).
The most relevant loading configurations to replicate spontaneous fractures (Cristofolini et al. 2007a) and to measure implant-bone micromotions in resurfacing hip prostheses (Cristofolini et al. 2007c) were chosen based on dedicated FE analyses (figure 7).

An indication about the need for simulating (or not simulating) the action of the abductor muscles during in vitro replication of spontaneous fractures was obtained from FE models (Cristofolini et al. 2007a).

(b) Improved and validated FE model

FE models were improved exploiting the synergy with in vitro experiments following the approach described in §5. Figure 10 shows how the strain prediction accuracy of FE models of intact bone segments developed by our group improved in the past years through a combined numerical–experimental approach.

The replication in the FE models of relevant points, such as strain gauge locations and anatomical reference frames, was improved by the adoption of progressively more refined instruments for the digital coordinate measurements (Taddei et al. 2006; figure 5).

The replication, in the FE models, of the point of load application was improved by the development and use of a dedicated transducer (Juszczyk et al. in press b).

The first major improvement in strain prediction accuracy (Taddei et al. 2007; figure 10) was due to a modification of the material mapping procedure from the CT grid to the FE models. This modification was in turn driven by a comparison of the mechanical properties attributed in the FE models with those obtained from mechanical tests on the same tissue specimens (Ohman et al. 2007).

The second major improvement in strain prediction accuracy (Schileo et al. 2007; figure 10) was due to a modification in the adopted density–elasticity relationship. This change, besides being corroborated with a strong inference approach (§9), reflected the results of an internal analysis and a review of the procedures used for mechanical testing at the continuum level (Helgason et al. 2008a).

A final improvement in the ability of our FE models for accurately predicting the strain distribution (Schileo et al. 2008a; figure 10) arose from a controlled tissue-level experiment aimed at evaluating basic relationships between different density measures for bone tissue.

To the authors’ knowledge, the most recent values reported in figure 10, from Schileo et al. (2007, 2008a), are, together with analogous figures reported in two other recent works (Bessho et al. 2007; Yosibash et al. 2007b), the most accurate reported so far in the literature.

(c) Results achieved

We started the process of integrating in vitro experiments and FE models back in the 1990s. More recently, with the most recent progress of FE modelling and computational power, huge steps were taken. This integrated approach recently
Synergy of FE models and in vitro experiments

First of all, such an integrated approach enabled an extremely detailed investigation of the strain pattern in the proximal femur. A combination of experimental and FE methods enabled measuring strains on the surface of the intact femur with a repeatability that was better than most of the literature (coefficient of variation for principal strain of 0.4% between test repetitions and 11–50% between paired specimens; Cristofolini et al. 2009b), and predicting such strains with an accuracy of 7 per cent (Schileo et al. 2008a).

At the same time, this synergistic approach provided an improved understanding of proximal femur fractures (figure 11). A method was devised that enabled replicating spontaneous fractures in vitro while identifying the point of fracture initiation with a spatial resolution of better than 1 mm (Cristofolini et al. 2007a), and a resolution in time of microseconds (Juszczyk et al. in press a). The femur was tested to fracture in vitro based on the set-up in figure 7 (Cristofolini et al. 2007a). A high-speed movie (15 000 frames per second) was taken to identify the point where fracture started. In parallel, an FE model of the same femur specimen was built based on past experience (Schileo et al. 2008b). The model was capable of predicting failure in 10 femurs: the point of fracture initiation was predicted with an accuracy of 10 mm, while the failure load was predicted with an accuracy of 17 per cent (Malandrino et al. 2008). More recently, failure load was predicted with an accuracy of better than 10 per cent for a single femur in a side-fall configuration (Grassi et al. 2010).
Figure 11. Investigation of fracture of a proximal femur. (a–c) A high-speed movie (15,000 frames per second) was taken while the femur was tested to fracture in vitro. The image in the centre of each picture is a direct view of the femoral neck from superior-lateral; the ones on the left and right (antero-medial and postero-medial views of the neck, respectively) are reflected images obtained from two mirrors placed next to the femur and suitably oriented. Frame (a) was taken 0.3 ms before fracture became visible (frame b). Frame (b) shows the instant when the crack starts opening on the lateral part of the neck (indicated by black pointers). Frame (c) shows a later stage of fracturing (2.8 ms after frame b). (d) An FE model of the same femur specimen was capable of predicting the failure load and the point of fracture initiation. A map of the regions (dark grey) where fracture was predicted based on a principal strain criterion is shown, where the femur is viewed from the top. (e) For comparison, an image of the fractured specimen is shown. (Reproduced with permission. Copyright © VPHOP consortium.)

This approach was also extended to the design, optimization and pre-clinical validation of hip implants. A detailed investigation was carried out in vitro of load transfer (Cristofolini et al. 2009a) and implant-bone micromotions (Cristofolini et al. 2007c) of contemporary resurfacing hip prostheses. In parallel, FE models were developed that are capable of accurately predicting bone strains and implant micromotion in implanted conditions (Taddei et al. 2010).

The studies quoted above related to a deterministic assessment of a number of in vitro specimens and FE-simulated cases. More recently, we developed a comprehensive risk analysis for the biomechanical performance of epiphyseal prostheses. This experimental–numerical combined approach supported successful optimization of the design of a proximal epiphyseal replacement prosthesis (Martelli et al. 2008) and the definition of the clinical indication by means of a probabilistic FE study (Schileo et al. 2008c).

(d) How could we do this?

The authors believe that the success reported by our group for a number of problems related to the proximal human femur derives from the synergistic use of in vitro experiments and FE models. In fact, the combination of FE modelling
and controlled experiments within the same research team was exploited to create a virtuous circle where models were used to improve experiments, experiments were used to improve models and their combination synergistically increased the knowledge of the proximal femur (figure 4).

— A basic requisite is the presence of a computational group and an experimental group within the same research team (in association with a multi-disciplinary team that is typical of bioengineering laboratories, including physicians, biologists, physicists, engineers, etc.).

— During the past years, integration and collaboration between the computational group and the experimental group within our laboratory steadily grew, with computational researchers spending more and more time in the experimental facilities and vice versa.

— In vitro experiments were designed in collaboration with the computational group to ensure that the loading conditions and measured variables were suitable for integration also into FE models. FE researchers assisted during the key phases of the in vitro experiments to get a practical sense and have a clearer view of the experimental details.

— FE models were analysed to provide the experimental researchers with as much information as possible to optimize the experimental set-up. Experimental researchers often sat next to the FE researchers both when FE models were built and during the final steps of post-processing.

9. The role of FE models and in vitro experiments in a strong inference perspective

It has been suggested that scientific research would progress more effectively if a ‘strong inference’ approach was taken (Platt 1964). In its separate elements, strong inference is just based on the classic scientific method and inductive inference that relies upon formulation of hypotheses, and on the execution of experiments aimed at confuting such hypotheses. The difference comes in its systematic application: strong inference consists of applying the following steps:

— devising multiple alternative hypotheses;
— devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
— carrying out the experiment so as to get a clean result; and
— recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain and so on.

It is suggested that the classical method based on a single hypothesis is likely to induce a confirmation bias. An approach based on multiple hypotheses enables overcoming this limitation (Platt 1964). More recently, suggestions to improve the strong inference method, renamed ‘strong inference plus’
(Jewett 2005), included the addition of two preceding phases before the actual hypothesis testing.

— An exploratory phase, with the goal of clarifying which factors have a role in the problem being addressed. In this phase, intuitions, serendipitous observations and yet unproven theories are acceptable. It is here that scientific creativity can play a crucial role.

— A pilot phase should start when the exploratory phase has provided reasonably reliable observations. In this phase, a small number of experiments are performed. This is the experiential test of the exploratory phase. This phase enables optimizing the subsequent experiments (testing procedures, sample size, etc.).

In this context, we believe that the synergic use of in vitro experiments and FE models will enable an effective implementation of the strong inference approach in this field. The respective roles of in vitro experiments and FE models should be clearly outlined.

— In the exploratory phase, all possible tools can be useful: in vivo evidence, anecdotal cases, simplified in vitro experiments, provisional FE models. As the scope is to use intuition to identify possibly relevant research questions, the most creative combination of methods is useful.

— The pilot phase involves a large number of possible scenarios to be explored. In this phase, in vitro experiments can provide a feeling of the mechanics underlying the problem. At the same time, FE models can provide preliminary sensitivity analyses (even based on crude and simplified models) to get an idea of possible influencing factors. As an example, if one is concerned about the mechanical failure of a bone implanted with a given device, an in vitro test can show whether the failure is driven by notching, by interface failure, by bone fracture, etc. Subsequently, a sensitivity analysis performed with an FE model can give an idea of which factors affect the output most significantly (bone quality, body weight of the subject, etc.).

— The hypothesis testing phase can rely on both in vitro tests and FE models: based on the information gained in the pilot phase, the research team now has a clearer idea of what experiments are needed to negotiate each of the alternative hypotheses. As discussed in the preceding sections, a combined use of in vitro tests and FE models provides the most powerful tool to address each hypothesis.

10. Conclusions

Biomedical research and biophysical research are moving towards a convergence. This requires a general reflection of the epistemological and methodological differences of these two domains of knowledge (or better to say these two artificial partitions of the same knowledge). In this article, we explored, through literature review and our own experience in the specific goal of predicting bone strength, how experimental and numerical methods can be combined synergistically in

Phil. Trans. R. Soc. A (2010)
order to obtain more than the sum of what we would get with each approach in isolation. We also showed how such a combination makes the concrete implementation of strong inference much more practical as a core methodological approach to our daily research efforts.

It is clear that the type of approach we propose requires that each group develop a broad array of skills, and ensures that researchers with a different background work effectively together. We all know these are very difficult goals to achieve. But in our opinion, in the long run, this is a major opportunity to truly advance the mechanistic understanding of human pathophysiology.

The European Community (grants IST-2004-026932 ‘Living Human Digital Library—LHDL’ and no. 223865 ‘The Osteoporotic Virtual Physiological Human—VPH-OP’) and Regione Emilia-Romagna (Region-University Research Programme 2007–2009) co-funded this study. The authors wish to thank Melinda Harman for revising the script, and Luigi Lena for the artwork. This study was supported by public institutions only (European Community and Regione Emilia-Romagna). The authors declare that they do not have any financial or personal relationships with other people or organizations that could have inappropriately influenced this study.

References


*Phil. Trans. R. Soc. A* (2010)


*Phil. Trans. R. Soc. A* (2010)


Galilei, G. 1638 *Discorsi e dimostrazioni matematiche intorno a due nuove scienze*. (Reprinted by P. Boringhieri, Turin 1958.)


*Phil. Trans. R. Soc. A* (2010)
Synergy of FE models and in vitro experiments


Lochmüller, E.-M., Groll, O., Kuhn, V. & Eckstein, F. 2002 Mechanical strength of the proximal femur as predicted from geometric and densitometric bone properties at the lower limb versus the distal radius. Bone 30, 207–216. (doi:10.1016/S8756-3282(01)00621-4)


Phil. Trans. R. Soc. A (2010)


Synergy of FE models and in vitro experiments


Phil. Trans. R. Soc. A (2010)
Synergy of FE models and in vitro experiments


