Estimates of load rates on the lower limb joints using smartphone accelerometers during physical activity

by

Susan Nazirizadeh

Thesis for the degree of Doctor of Philosophy

Supervisors: Dr Alexander Forrester
Prof Maria Stokes
Prof Nigel Arden

April 26, 2018
Although the causes and pathology of the progression of osteoarthritis are not entirely understood, an active lifestyle avoiding excessive load on the joints can control symptoms of osteoarthritis (e.g. joint pain and stiffness). The aim of this thesis was to develop, validate, and test an algorithm for estimating impact loading through the lower limbs using wearables (smartphones and smartwatches). The viscoelastic nature of articular cartilage means it is susceptible to high load rates, hence, the mean load rate magnitude was estimated from accelerometer recordings of wearables and used as a surrogate for estimating impact loading on the lower limb joints. The validity of the mean load rate magnitude was assessed against the gold standard equipment, the force plate ($R^2 = 0.77$). Further, the mean load rate magnitude was used as a feature in the classification of everyday activities with support vector machine classifiers with an accuracy of 80%. An app was then developed which monitored mean load rate magnitude using Markov chain Monte Carlo methods for testing the reliability of monitoring over a period of seven days. The accumulated mean load rate magnitude was used to estimate the error, $\hat{e}_{\text{smartphone}} = 2.66\%$, for seven-day recordings. Finally, a function to score pain was added to the final version of the app, termed $OApp^{TM}$. A single case study assessed the ability of $OApp^{TM}$ to compare osteoarthritis-related pain to mean load rate magnitude with a low positive correlation of $r = 0.38$. To conclude, this thesis developed, assessed the validation, and tested a load rate magnitude algorithm, which estimated load rate on the lower limb joints with the accelerometer sensors of wearables. These results form the basis for further research to develop a clinical tool for monitoring load rate and supporting patients to maintain an active lifestyle by avoiding excessive load on their lower limb joints.
# Contents

Contents ix  
List of figures xiii  
List of tables xv  
List of algorithms xv  
Nomenclature xvii  
Declaration of authorship xix  
Acknowledgements xxi  
Acronyms and abbreviations xxv  
Glossary xxvii  

1 Introduction 1  
  1.1 Structure of the thesis ..................................... 3  

2 Background and literature review 7  
  2.1 Background ................................................. 7  
    2.1.1 The Mechanical properties of cartilage as a material ........ 8  
    2.1.2 Load rate during locomotion measured with the gold standard equipment ........................................ 11  
  2.2 Literature review: clinical use of physical activity monitoring .... 12  
    2.2.1 Method of literature review .................................. 13  
    2.2.2 Subjective measurements ..................................... 13  
    2.2.3 Objective measurement ....................................... 17  
    2.2.4 Commercially available devices used for activity monitoring .... 21  
  2.3 Literature review: estimation of loading on the joints with accelerometers 21  
    2.3.1 Method of literature review .................................. 22  
    2.3.2 Estimations of loading on the joints with accelerometers .... 22  
  2.4 Literature review: signal processing within physical activity monitoring 27  
    2.4.1 Method of literature review .................................. 27  
    2.4.2 Overall sequences of signal processing ......................... 28  
    2.4.3 Data acquisition ........................................... 29  
    2.4.4 Data windowing ............................................. 29
2.4.5 Features extraction ........................................ 29
2.5 Research rationale ........................................... 30
2.6 Overall aim ..................................................... 31
2.7 Summary ......................................................... 32

3 General methodology ........................................... 33
3.1 Study designs .................................................. 33
3.2 Equipment ....................................................... 34
3.2.1 Microelectro mechanical systems (MEMS) ............... 34
3.2.2 The Sony® Xperia™ Z5 Compact and SmartWatch 3 .... 34
3.3 Participants ..................................................... 35
3.4 Ethical approval ................................................ 36
3.5 Data collection .................................................. 36
3.6 Data analysis ................................................... 36
3.6.1 Load rate ..................................................... 36
3.6.2 Concept of uncertainty ..................................... 37
3.6.3 Cross-validation ............................................. 39
3.6.4 Bootstrapping ............................................... 41
3.7 Summary ........................................................ 42

4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill 45
4.1 Introduction .................................................... 45
4.1.1 Previous work in the literature ............................. 45
4.1.2 Rationale ..................................................... 50
4.2 Aim and objectives ............................................ 50
4.2.1 Objectives ................................................... 50
4.2.2 Hypothesis ................................................... 50
4.3 Pilot single case study to refine feasibility for monitoring load rate with wearables ........................................... 51
4.3.1 Introduction of pilot study ................................ 51
4.3.2 Aim of pilot study ........................................ 51
4.3.3 Methods of pilot study ................................... 51
4.3.4 Results of pilot study ..................................... 51
4.3.5 Discussion and conclusion of pilot study ............... 52
4.4 Methods ........................................................ 52
4.4.1 Study design ............................................... 52
4.4.2 Participants and recruitment ............................... 52
4.4.3 Ethical approval ............................................ 53
4.4.4 Equipment ................................................... 53
4.4.5 Procedure .................................................... 55
4.4.6 Data analysis ............................................... 55
4.5 Results .......................................................... 61
4.6 Discussion ....................................................... 65
4.7 Limitations and strengths of the study ....................... 70
4.8 Conclusion ....................................................... 71
5 Study 2: Classification and quantification of load rate estimates of everyday activities using wearable accelerometers 73
5.1 Introduction ......................................................... 73
  5.1.1 Previous work in the literature ......................... 74
  5.1.2 Rationale ...................................................... 76
5.2 Aim and objectives ............................................... 76
  5.2.1 Aim .............................................................. 76
  5.2.2 Objectives ...................................................... 77
  5.2.3 Hypothesis ...................................................... 77
5.3 Methods ............................................................. 77
  5.3.1 Study design .................................................... 77
  5.3.2 Participants and recruitment .............................. 77
  5.3.3 Ethical approval .............................................. 78
  5.3.4 Equipment ...................................................... 78
  5.3.5 Procedure ...................................................... 78
  5.3.6 Data analysis ................................................... 79
5.4 Results .............................................................. 88
5.5 Discussion .......................................................... 92
5.6 Limitations and strengths of the study ....................... 95
5.7 Conclusions .......................................................... 95

6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers 97
6.1 Introduction .......................................................... 97
  6.1.1 Previous work in the literature ......................... 97
  6.1.2 Rationale ...................................................... 99
6.2 Aim and objectives ............................................... 99
  6.2.1 Aim .............................................................. 99
  6.2.2 Objectives ...................................................... 99
6.3 Methods ............................................................. 100
  6.3.1 Study design .................................................... 100
  6.3.2 Participants and recruitment .............................. 100
  6.3.3 Ethical approval .............................................. 100
  6.3.4 Equipment ...................................................... 101
  6.3.5 Procedure ...................................................... 101
  6.3.6 Data collection ................................................... 102
  6.3.7 Data analysis ................................................... 103
6.4 Results .............................................................. 104
6.5 Discussion .......................................................... 112
6.6 Limitations and strengths of the study ....................... 115
6.7 Conclusions .......................................................... 115

7 Participant and public consultation: user interface for OApp™ for people with lower limb joint osteoarthritis 117
7.1 Introduction .......................................................... 117
  7.1.1 Rationale ...................................................... 118
7.2 Purpose ............................................................... 118
List of Figures

1.1 Left: a normal knee joint; right: a knee joint with OA. (Lavelle, 2007)  

2.1 The overall structure of articular cartilage (Oei et al., 2014) consists of three layers: superficial, middle and deep. The layers include proteoglycan (PG), chondrocytes, and collagen fibrils. Left: a normal histology with an intact articular surface, a high concentration of PG in all layers, and tightly ordered collagen fibrils. Middle: degradation of PG in the superficial and middle layers which leads to either the disorganisation or reorientation of the superficial layer’s collagen network and water influx. Right: an additional loss of PG and, in turn, a fibrillation of the cartilage’s surface which thereby leads to advanced OA.

2.2 The stress ($\sigma$) and strain ($\epsilon$) curve for a viscoelastic material at different loading levels (Luo et al., 2014). The hysteresis between point OB shows a continuous line for the loading and a dotted line for the unloading process. Between points O and B, the material deforms during the loading process and goes back to its original shape at point O. The area between the continuous line and the dotted line is the absorbed energy. Passing point B initiates the cracking of the material. The hysteresis between point OM shows that, when the material passes the initial cracking point, the material will not go back to its original shape after unloading. It will reach point N rather than point O. The crack initiation means that the material is then in the stable crack growth phase until it reaches point C, at which point it transforms into an unstable crack growth. From this point, the material fails. The ratio of stress ($\sigma$) and strain ($\epsilon$) can vary with time in terms of the load which is applied to the viscoelastic material. The faster the load is applied to the viscoelastic material, the faster point B is reached.

2.3 Stress ($\sigma$) to strain ($\epsilon$) curve for a viscoelastic material at different loading rates (figure inspired by Özkaya et al. (2017)).

2.4 (A) shows a gait pattern with an impact peak at the moment the foot begins to have contact with the ground. (B) shows a gait pattern with a foot strike that does not cause an impact peak. (Daoud et al., 2012)

2.5 Pre-processing sequences of the raw data.

3.1 The data set is repeatedly separated into training set (marine blue) and the test set that contains only one observation (light blue). The test result is then estimated by averaging the $n$ results. This figure was inspired by Friedman et al. (2001).
3.2 The data set is randomly separated into $k = 5$ non-overlapping groups. Each group will be a testing set (light blue), and the remaining points are the training set (marine blue). The test result is then estimated by averaging the five results. This figure was inspired by Friedman et al. (2001).

3.3 The original data set is $Z$, which has $n = 3$ observations. Each bootstrap data set has $n$ observations, resampled with replacement from the original data set. Each bootstrap data set is used to estimate the parameter $S$. This figure was inspired by Friedman et al. (2001).

4.1 Alter-G® anti-gravity treadmill (Photo: Willem Eerland)

4.2 The active phases were walking of $5\text{km h}^{-1}$ (green), jogging of $8\text{km h}^{-1}$ (yellow) and running of $12\text{km h}^{-1}$ (pink).

4.3 Four gait cycles of load rate magnitude of Participant 4 during running on the anti-gravity treadmill.

4.4 Left: Whole dataset of all participants for Smartphone 1; right: same data set with a linear regression line.

4.5 Left: Whole dataset of all participants for Smartwatch 1; right: same data set with a linear regression line.

4.6 Left: Residuals vs the fitted values. Right: fitted response vs. observed response in Model 1.

4.7 Left: Residuals vs the fitted values. Right: fitted response vs. observed response in Model 2.

4.8 Left: Residuals vs the fitted values. Right: fitted response vs observed response in Model 3.

4.9 The results from the one-way ANOVA test, comparing the $1.000 R^2$ from Model 1, Model 2 and Model 3.

4.10 The results from the one-way ANOVA test, comparing the $1.000 \text{RMSER}$ from Model 1, Model 2 and Model 3.

4.11 Left: Boxplot of bootstrapped $R^2$ values between four devices; right: pairwise comparison of the mean $R^2$ of the four devices.

5.1 Smartphone attached to the participant’s right upper leg, smartwatch strapped onto the wrist of their non-dominant hand (left in this case), and a GPS device strapped onto their other wrist.

5.2 Different options for separating lines exists for a linearly separable set of 2D-data. This figure was inspired by Smola and Schölkopf (1998).

5.3 The support vector machine algorithm finds the optimal hyperplane that gives the largest minimum distance to the training examples. This figure was inspired by Smola and Schölkopf (1998).

5.4 Transforming nonlinear data into a high-dimensional feature space gains a linear separation. This figure was inspired by Smola and Schölkopf (1998).

5.5 Distance to hyperplanes. This figure was inspired by Smola and Schölkopf (1998).

5.6 The support vectors are lying on the margin with the KKT conditions. This figure was inspired by Smola and Schölkopf (1998).

5.7 The three-layered load rate algorithm consists of a feature generator, support vector machine classifier, and categorisation. L= low (blue); M= moderate (green); I=intensive (red).
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8</td>
<td>True (left) and predicted (right) proportions of time spent walking (top),</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>jogging (middle) and running (bottom) for all participants. The load rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>was divided into three categories: low (blue), moderate (green), and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>intensive (yellow).</td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>True (left) and predicted (right) proportions of time spent on asphalt</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>(top), grass (middle) and gravel (bottom) for all participants. The load</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rate was divided into three categories: low (blue), moderate (green), and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>intensive (yellow).</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>An example of data, which is independent from each other and is identically</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>distributed. This figure was inspired by Bishop (2006).</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>A first-order Markov chain of observation ( {x_i} ) in which the</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>distribution ( p(x_i</td>
<td>x_{i-1}) ) of a particular observation ( x_i )</td>
</tr>
<tr>
<td></td>
<td>is dependent on the previous observation ( x_{i-1} ). This figure was</td>
<td></td>
</tr>
<tr>
<td></td>
<td>inspired by Bishop (2006).</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 1.</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 2.</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 3.</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 4.</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 5.</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 6.</td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 7.</td>
<td></td>
</tr>
<tr>
<td>6.10</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 8.</td>
<td></td>
</tr>
<tr>
<td>6.11</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 9.</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>The moving average load rate magnitude with 10 minute windows throughout a</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>seven-day data collection period depicting Participant 10.</td>
<td></td>
</tr>
<tr>
<td>6.13</td>
<td>Mean load rate compared between smartphone and smartwatch on a log-log scale.</td>
<td>110</td>
</tr>
<tr>
<td>6.14</td>
<td>A comparison of the accumulated mean load rate magnitude with increasing</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>sample size and convergence which was measured during the seven days of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>data collection for all participants.</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>OApp’s user interface. Left: No pain and right: tolerable.</td>
<td>120</td>
</tr>
<tr>
<td>7.2</td>
<td>OApp’s user interface. Left: intensive and right: excruciating/ unbearable.</td>
<td>121</td>
</tr>
<tr>
<td>8.1</td>
<td>Structure of the procedure of Study 4.</td>
<td>130</td>
</tr>
<tr>
<td>8.2</td>
<td>The moving average load rate magnitude with 10 minutes windows throughout a</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>14-day data collection period pre- and post-injection depicting.</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>The mean load rate magnitude compared between the smartphone and</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>smartwatch data on a log-log scale both pre- and post-injection.</td>
<td></td>
</tr>
</tbody>
</table>
8.4 The mean load rate magnitude compared between the smartphone and the smartwatch and the momentary pain for both the pre- and post-injection. 134
8.5 The accumulated mean load rate magnitude with increasing sample size and convergence, which was measured during the 14 days of data collection both pre- and post-injection. 135
List of Tables

2.1 Outcome and quality of relevant studies for measuring physical activity in clinical use (Subjective measurements, part 1) ........................................... 15
2.2 Outcome and quality of relevant studies for measuring physical activity in clinical use (Subjective measurements, part 2) ................................. 16
2.3 Outcome and quality of relevant studies for measuring physical activity in clinical use (Objective measurements, part 1) ................................. 19
2.4 Outcome and quality of relevant studies for measuring physical activity in clinical use (Objective measurement, part 2) .............................. 20
2.5 The outcome and quality of studies measuring impact loading (part 1). ........................................... 25
2.6 The outcome and quality of studies measuring impact loading (part 2) ........ 26

4.1 Outcome and quality of relevant studies validating acceleration sensors against the golden standard equipment. ........................................... 48
4.2 Outcome and quality of relevant studies validating acceleration sensors against the golden standard equipment. ........................................... 49
4.3 The $R^2_{Model}$ and RMSER values for all participants using all of the smartphone and smartwatch data which was collected. ......................... 64
4.4 Pairwise comparison with the Bonferroni correction. All differences in the models were significant ($p < 0.0001$, $\alpha = 0.05/4 = 0.0125$) ....................... 65
4.5 The $R^2_{speed}$ values for each speed for the smartphone between the shoulder blades. ................................................................. 65
4.6 The $R^2_{device}$ values are compared with the different positions which those devices were placed on the body from 6 of the participants ......... 66
4.7 Pairwise comparison of four devices with the Bonferroni correction ($\alpha = 0.05/6 = 0.0083$). ................................................................. 66

5.1 Trial 1-13 ................................................................. 79
5.2 Feature matrix .......................................................... 80
5.3 Four features with the highest accuracy for the smartphone ................. 85
5.4 Activity accuracy with LOSOCV. Both feature sets had no significantly different means ($p > 0.05$). ............................................... 88
5.5 Terrain accuracy with LOSO cross validation. No significant difference between both feature sets ($p > 0.05$). .................................. 88
5.6 Activity and terrain accuracies with 10-fold cross validation. No significant difference between both feature sets ($p > 0.05$). ............... 89

6.1 Pearson’s correlation coefficients between smartphone and smartwatch. All values were significant ($p << 0.001$). ..................................... 111
6.2 Error estimation for seven days ........................................... 112
8.1 Outcome and quality of relevant studies to measure pain and physical activity ................................................................. 126
8.2 Mean load rate and mean pain for each seven day period both pre- and post-injection. ......................................................... 132
8.3 The Pearson’s correlation coefficients between the smartphone and smartwatch. ............................................................... 133
8.4 The Spearman’s correlation coefficients between smartphone/smartwatch and pain score. .................................................. 133

A.1 Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 1) ................. 148
A.2 Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 2) ............... 149
A.3 Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 3) ............... 150
List of algorithms

4.1 Pseudocode for the data analysis of Study 1 ........................................ 60
5.1 Pseudocode for the data analysis of Study 2 ........................................ 86
Nomenclature

\( F_L \)  Force that is applied on the joints, so called ‘load’  \([\text{kg m}^{-2} \text{s}^{-2}]\)
\( t \)  Time  \([\text{s}]\)
\( m \)  Body mass  \([\text{kg}]\)
\( x \)  Position  \([\text{m}]\)
\( v \)  Velocity  \([\text{m s}^{-1}]\)
\( a \)  Acceleration  \([\text{m s}^{-2}]\)
\( j \)  Jerk  \([\text{m s}^{-3}]\)
\( X \)  First observed variable  \([]\)
\( Y \)  Second observed variable  \([]\)
\( n \)  Number of observations  \([]\)
\( f(X) \)  Unknown function of \( X \), which is related to \( Y \)  \([]\)
\( \hat{Y} \)  Prediction for \( Y \)  \([]\)
\( \hat{f}(X) \)  Estimation for \( f(X) \)  \([]\)
\( E \)  Expected value  \([]\)
\( \epsilon \)  Error term  \([]\)
\( Z \)  Original data set  \([]\)
\( S(Z) \)  Any parameter computed from original data set  \([]\)
\( \hat{S} \)  Estimate from original data set  \([]\)
\( B \)  Number of bootstrapping sample sets  \([]\)
\( y_{m,i} \)  Responds variable of the linear model (Estimated load rate from force plate) with observation \( m \) and participant \( i \)  \([]\)
\( \alpha_{\text{wear}} \)  Intercept of the load rate estimated by the wearables  \([]\)
\( \beta_{\text{wear}} \)  Slope of the load rate estimated by the wearables  \([]\)
\( a_{\text{participant}} \)  Intercept of the load rate magnitude of the wearables  \([]\)
\( b_{\text{participant}} \)  Slope of the load rate of the force plate  \([]\)
\( \text{RMSE}_R \)  Bootstrapped estimation of \( \text{RMSE}_R \)  \([]\)
\( \hat{R}^2 \)  Bootstrapped estimation of \( R^2 \)  \([]\)
\( x_k \)  Linearly separably dataset  \([]\)
\( y_k \)  Labels of linearly separably dataset  \([]\)
\( g(x) \)  Decision function for separating hyperplane  \([]\)
\( w \)  Orthogonal vector  \([]\)
\( b \)  Threshold for separating hyperplane  \([]\)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>Margin to obtain optimal hyperplane</td>
</tr>
<tr>
<td>$H$</td>
<td>Hyperplane</td>
</tr>
<tr>
<td>$d_k$</td>
<td>Distance between a point with the hyperplane</td>
</tr>
<tr>
<td>$d_{ok}$</td>
<td>Distance between the origin with the hyperplane</td>
</tr>
<tr>
<td>$w'$</td>
<td>Orthogonal vector divided through the margin</td>
</tr>
<tr>
<td>$b'$</td>
<td>Threshold for separating hyperplane divided through the margin</td>
</tr>
<tr>
<td>$\tau(w)$</td>
<td>Objective function, which has to be minimised in order to obtain optimal hyperplane</td>
</tr>
<tr>
<td>$\alpha_k$</td>
<td>Lagrangian multiplier</td>
</tr>
<tr>
<td>$L$</td>
<td>Lagrangian: minimised with respect to $w$ and $b$ and maximised with respect to $\alpha_k$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Significance level, the probability of rejecting the null hypothesis when it is true</td>
</tr>
<tr>
<td>$z_{\alpha/2}$</td>
<td>Signed number of standard deviations by which the value of an observation is above the mean value, here helps to find confidence intervals</td>
</tr>
<tr>
<td>$f(x)$</td>
<td>General function of an observation $x$</td>
</tr>
<tr>
<td>$p(x)$</td>
<td>Probability distribution of general function</td>
</tr>
<tr>
<td>$\hat{f}$</td>
<td>Estimation of $f$ with Monte Carlo Method</td>
</tr>
<tr>
<td>$\text{var}[\hat{f}]$</td>
<td>Variance of estimation of $f$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Unknown population standard deviation</td>
</tr>
<tr>
<td>$\hat{s}$</td>
<td>Estimated standard deviation</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Unknown population mean</td>
</tr>
<tr>
<td>$e$</td>
<td>Error term</td>
</tr>
</tbody>
</table>
Declaration of authorship

I, Susan Nazirizadeh, declare that the thesis entitled Estimates of load rates on the lower limb joints using smartphone accelerometers during physical activity and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

• this work was done wholly or mainly while in candidature for a research degree at this University;

• where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

• where I have consulted the published work of others, this is always clearly attributed;

• where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

• I have acknowledged all main sources of help;

• where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

• none of this work has been published before submission

Signed:.........................................................................................................................

Date:....................................................................................................................................
Acknowledgements

First of all, I would like to thank my supervisors for their expertise and their hard work. I learned a lot during my PhD with the help of my patient supervisors. Without them, I would have never finished this work.

Further, I would like to thank Arthritis Research UK (Grant reference 20194) for funding parts of my PhD.

I would like to thank Mo Gimpel (Director of Medical & Science Performance Support Southampton FC) for providing access to the anti-gravity treadmill for Study 1. A big thanks goes to the team in Oxford, who was a great help in organising and recruiting participants for Study 4. Further, I would like to thank all my participants, who took part in each study. Without you, this PhD would have no meaning. I am grateful that I found so many enthusiastic volunteers, who were willing to sacrifice their time for science.

My colleges in the CED group made this hard journey to a sweet experience. I would like to thank each of them for their expertise and their friendship. The atmosphere in the group was very special and I am missing these people already. Further, I would like to thank all my friends around the world, who always believed in me and were patient with my abstinence.

I would like to thank my parents a lot, who always believed in me. Without their love, I would not stand here. I would like to give special thanks to my wonderful siblings Yasmin, Yousef and Dawood, who are all my role models. You guys might not know it, but you inspired and made me stronger for each step of my life. Thank you!

And last but not least I would like to thank my partner Nawar. You rock my world!
Für Viktoria, Niek und die Vielen die noch kommen werden. Auf dass ihr die Welt mit euren eigenen Augen und Verstand entdeckt.
Acronyms and abbreviations

**CSI** intra-articular corticosteroids injections. 2, 5, 35, 123, 127, 128, 136, 137

**HAGOS** Hip and Groin Outcome Score. 127, 129–131, 136

**HMM** hidden Markov model. 74

**IID** independent and identically distribution. 80, 100, 102, 103

**IPAQ** International Physical Activity Questionnaire. 127, 129–131, 135

**k-fold CV** k-fold cross-validation. 39, 40, 75

**KKT** Karush-Kuhn-Tucker. x, 83, 84

**LOOCV** leave-one-out cross-validation. 39, 75

**LOSOCV** leave-one-subject-out cross-validation. xiii, 39, 40, 75, 77, 85, 86, 88, 89, 92–95

**MEMS** micro electro-mechanical systems. 17, 27, 30, 34, 42, 45, 52, 144

**MET** Metabolic Equivalent of Task. 131


**OS** operation system. 31, 35

**PG** proteoglycan. ix, 8

**PPI** patient and public involvement. 4, 117–121, 128, 140

**PRA** peak resultant acceleration. 23

**PVA** peak vertical acceleration. 23

**RMSER** root mean squared error ratio. x, xiii, 56, 59–61, 63, 64, 67, 68, 70
SP 1 Smartphone 1: between the shoulder blades. 53, 61, 63, 64, 66, 68

SP 2 Smartphone 2: right hip. 53, 68

SVM support vector machine. 74, 75, 80–82, 84

SW 1 Smartwatch 1: right wrist. 53, 61, 66, 68, 69

SW 2 Smartwatch 2: left wrist. 53, 68
Glossary

**Markov chain** In probability theory, a data set is called first-order Markov chain if a current observation is dependent on the previous observation in the dataset. 98, 100, 103, 114, 115

**Monte Carlo** In probability theory, the Monte Carlo method is known as random sampling method. 98, 100, 102, 103, 114, 115

**accuracy** In science, engineering and statistics the accuracy of a measurement is the degree of how close the measurement is from its true value. According to ISO 5725-1, accuracy consists of trueness (proximity of measurement results to the true value) and precision (repeatability or reproducibility of the measurement). The classification accuracy is defined as a percentage of correct predictions. It describes how well the classifier categorises the test set of the data correctly. It is the ratio between correct classified data points and the complete data set multiplied by 100 to obtain a percentage. 14, 17, 28, 29, 75

**ActiGraph** A common used accelerometer sensor in research, which monitors physical activity and sedentary time. The sensors are suitable to wear on the wrist or waist. The data is send to a software and can be analysed on the software afterwards. The method is validated (Sasaki et al., 2011). 14

**anti-gravity treadmill** An anti-gravity treadmill lifts the body of the users so that lower limbs bear a lower body weight per percentage. It is normally used in the rehabilitation process, since it reduces the weight on the lower limb joints. In this thesis the M320 from Alter-G® was used (Alter-G, 2017). x, 4, 31, 50–54, 70

**capacitive** In electronics, the capacitance is the ability of a body to store an electric charge. 34

**central limit theorem** In probability theory, the centre limit theorem says, that the distribution of the sample is approximately normal if the sample size is large. 41

**classification** In machine learning, classification is a method to categorise previously unseen data into groups. The groups are defined by a labelled data set. 29, 36, 148

xxvii
**classifier** In machine learning, a classifier is a tool that categorises new previously unseen data into groups. The method to this tool is classification. 4, 29, 81, 148, 150

**compliance** In medicine, compliance is the degree to which a patient correctly follows medical advice. In this thesis, it refers to whether a participant is correctly using the devices and the app that was provided by the researcher. 14, 137

**confounder** In statistics, confounders are variables that cannot be controlled or the researcher failed to control, which are directly related to the results. Confounders will damage the validity of an experiment. They lead to bias that distorts the relationship between the variables and the result. 143

**correlation** Correlation is a statistical method which can show whether and how strongly pairs of variables are related. This statistical relation can be influenced by confounder and be biased. 30, 50, 52, 149

**energy** In signal processing, the energy of a signal in a discrete domain is defined as the sum of the square of the signal. It is a measure of the signal strength. 79, 80, 88, 95

**entropy** In signal processing, the entropy of a signal in a discrete domain is defined as the negative sum of the frequency of the signal multiplied by the logarithm of the frequency of the signal. It is the measure of the degree of the prominence of the peaks in the Fast Fourier Transformation of the signal (here in the thesis the load rate signal). 79, 80, 88, 95

**feature** In machine learning, a feature is an observed quantity, which helps to classify or model data. In this thesis, features are extracted from the raw acceleration data, collected from the wearables. x, 23, 28–30, 81, 148, 150

**Fast Fourier transformation** In signal processing, a Fast Fourier transformation calculates the discrete Fourier transform of a signal. Here it converts the signal from its time-domain into the frequency-domain. 80

**frequency-domain features** In machine learning, frequency-domain features are based on the frequency of the signal. Usually, Fourier transformation will be applied to the raw accelerometer to transform the data from time-domain to frequency-domain. To monitor activities with the same intensity but different periodicity it is useful to use frequency-domain features. Every activity has a specific cycle frequency, consequently identifiable with the frequency. Examples are energy of the signal, entropy of the signal and Fourier transformation components. 30

**generalisation error** In machine learning, generalisation errors are measurement of how accurate an algorithm can predict values for previously unseen data. 80
ground reaction force In biomechanics, the ground reaction force is the force exerted by the ground colliding with a body. 12, 45–47, 50, 54

hyperplane In geometry, a hyperplane consists of more than two independent vectors, which span a plane. x, 80–84

impact loading In mechanics, an impact is a high force over a short time period when two or more bodies collide. Impact loading is the high load over a short time period, which is transmitted through the joints. In this thesis, it is assumed that impact loading is instrumental in the progression of OA. 3, 8, 12, 18, 22, 23, 30–32, 45–47, 76, 99, 115, 139, 140

inference In statistics, inference means to use mathematics to draw conclusions in the presence of uncertainty. 38, 102

intensity The Cambridge dictionary defines intensity for physical quantities as the strength of something that can be measured such as light, sound, etc. 17, 22, 23, 30, 149

jerk In physics, jerk is a physical quantity, which is the rate of change of acceleration over time. In other words it is the first derivative of acceleration, the second derivative of velocity, and the third derivative of position. 37, 99–101, 103, 112, 129

load rate algorithm A self-developed three-layered algorithm consisting of 1) load rate estimation from accelerometer data from wearable device sensors, 2) activity classification, and 3) load rate categorisation into ‘low’, ‘moderate’ and ‘intensive’ activities. 4

load rate magnitude Is the magnitude of the load rate in three spatial directions captured with acceleration sensors from wearables. 4, 37, 42, 50, 51, 55–58, 76, 79, 103–106, 111, 130–134, 139, 143, 144

load rate The rate of change of load with respect to time, which is the mass times the physical quantity jerk. In this thesis, the load rate is calculated from the recorded acceleration by the wearables. The load rate does not refer to the actual load rate through the joints, rather to the estimation of the external load rate, which is well correlating according to Study 1. x, xi, 3–5, 8, 10, 11, 23, 24, 27, 30–32, 34, 37, 45–47, 50–52, 56–58, 61, 65–71, 73, 75–79, 84–95, 97–100, 102, 104, 112–115, 125, 127, 129, 131, 135, 139, 140, 142

load The term ‘load’ describes biomechanical physical stresses which act on the body or anatomical structures within the body (Council et al., 2001). These stresses can be kinetic, kinematic, oscillatory or thermal energy sources. In this thesis, since kinetic energy sources are of interest, the term ‘load’ is strictly applied to weight-bearing forces on the joints. Load transmitted through the joints can be measured
with advanced equipment such as prostheses equipped with pressure sensors or estimated with a force plate. Not to be mistaken with the training load in sports science, which is defined as the duration multiplied by intensity of the exercises over a specific amount of time, e.g. five days. 2, 8, 10, 17, 21–23, 124

**margin** Defined by the closest two data points from each two classes. x, 80–82, 84

**linear mixed regression model** Statistical model that contains a fixed and random effect, typically used for repeated measurements in the same statistical unit. 57, 58, 67

**Model 1** In Study 1 three models were used. Model 1 is a linear regression model of load rate between wearables and force plate with fixed slop and fixed intercept. x, 56–59, 61, 64, 65, 67, 68

**Model 2** Model 2 is a linear regression model of load rate between wearables and force plate with fixed slop, fixed intercept and random intercept. x, 56–59, 61, 64, 65, 67, 68

**Model 3** Model 3 is a linear regression model of load rate between wearables and force plate with fixed slop, fixed intercept, random intercept and random slope. x, 56–59, 61, 64, 65, 67, 68

**piezoresistive** In electronics, the piezoresistive effect is a change in the electrical resistivity of a semiconductor or metal when mechanical strain is applied. 34

**quality-of-fit** In statistics, the quality-of-fit is the evaluation of the performance of a statistical learning method on a given data set. It can be quantified as the extent to which predictions for a given observation is close to the true value for that observation (Friedman et al., 2001). The two that are used here for a regression setting are root mean squared error ratio (RMSER) and $R^2$. 28, 41, 56, 59, 69, 70

**reliability** In statistics, a high reliability is achieved if the results of a test or experiment yield the same or compatible results if repeated. Reliability can be described by the repeatability and consistency of a test or experiment. 23

**replacement** In statistics, sampling with replacement means that after an observation was randomly drawn from the original data sample set the observation is put back before drawing the next observation. x, 41, 42

**resampling** In statistics, resampling means that the training set of the data is repeatedly drawn to refit a model of interest regarding each sample to obtain additional information about the fitted model. 38, 39, 41

**supervised learning** In machine learning, supervised methods are inferring functions from labelled training data. In supervised learning, each example is a pair consisting of an input object and an output value. 74, 80
time-domain features In machine learning, time-domain features are based on the physical quantity time. Raw accelerometer data is always time-based, hence, can be transformed directly to time-domain features. Examples are mean, variance, correlation, maximum value of the acceleration. 30

unsupervised learning In machine learning, unsupervised methods are functions to describe hidden structure from unlabelled data. 74

validity In statistics, the validity of a measurement tool is the degree to which the tool measures what it is supposed to measure. The validity of a test can be assessed with different types of validity, such as external validity, internal validity, and construct validity to name a few. 13, 23

viscoelastic In material science, viscoelastic is a property of a material that includes viscous and elastic characteristics when a force is applied. Many biological tissues are, out of simplicity reasons, assumed to be viscoelastic based on their structure consisting of solid and fluid components. ix, 8–11, 23
Chapter 1

Introduction

Osteoarthritis is a chronic degenerative joint disease (Litwic et al., 2013), which most often occurs in the lower limbs (Valderrabano and Steiger, 2010). Research in the epidemiology of osteoarthritis (OA) discovered multiple risk factors including, inter alia, age, sex, genetics, obesity, muscle atrophy or local mechanical risk factors, such as joint loading (Litwic et al., 2013). Osteoarthritis is characterised by thinned cartilage with bone spurs and cartilage fragments (Figure 1.1).

![Figure 1.1: Left: a normal knee joint; right: a knee joint with OA. (Lavelle, 2007)](image)

According to the Arthritis Research UK report (Arthritiscentre, 2013), 8.75 million people suffered from OA in the UK in 2013. In other words, around 13.64% of the population or, equivalently, every 11th person in the UK. Furthermore, the same report estimated that, by 2035, 16.6 million people, which is estimated as being around 22.34% of the population, could have OA in the UK. This would mean that every sixth person
is estimated to have OA by 2035. These numbers are alarming and more research has to be conducted to determine how to manage its symptoms, as well as to prevent and understand OA better.

The symptoms of OA are pain, decreased joint range-of-motion, stiffness, muscle weakness and atrophy, joint effusion and swelling, and physical disability (Valderrabano and Steiger, 2010). The reasons behind, and the pathological progression of OA are not entirely understood; thus, there is no cure for OA at present, despite different symptom ameliorating drugs, such as intra-articular corticosteroids injections (CSI) (Arden et al., 2008) or joint replacement surgeries. Besides these symptom ameliorating treatments, Valderrabano and Steiger (2010) believe that moderate exercise delays and alleviates the symptoms of this disorder while also preventing muscle atrophy and stiffness. Consequently, an active lifestyle with a low level of joint loading may control the abovementioned symptoms. The term ‘load’ describes biomechanical physical stresses, which act on the body or on anatomical structures within the body (Council et al., 2001). These stresses can be kinetic, kinematic, oscillatory or thermal energy sources. In the present thesis, since kinetic energy sources are of interest, the term ‘load’ is strictly applied to weight-bearing forces on the joints.

Staying active despite having all of these symptoms might be challenging for people with OA. A method for encouraging people with OA to maintain an active lifestyle, even when they are feeling pain or stiffness, is to monitor their everyday physical activities (Consolvo et al., 2006). Consolvo et al. (2006) showed in their study, that by giving people with OA a device that monitors their physical activity, their overall physical activity increases.

In contrast to the beneficial effects of moderate activity, risk factors for OA might be excessive mechanical load in sports (Litwic et al., 2013). Lievense et al. (2003) conclude that there is moderate evidence for a positive association between hip OA and sporting activities, such as running, soccer, athletic activities, and ballet dancing. Blagojevic et al. (2010) found out in there systematic literature review that cohort studies generally suggested an increased risk of knee OA in those who exercise more regularly or intensely. However, there are newer studies by Williams (2013) looking into the effects of running and walking on OA and hip replacement risk. They found out that, for recreational runners who even substantially exceed current guideline activity levels, and participate in multiple marathons annually, running does not appear to increase OA and hip replacement risk. Also, Krampla et al. (2001) did not find a significant correlation between OA and marathon running. They explored MRI of the knee in marathon runners before and after the competition. Krampla et al. (2001) concluded that in healthy individuals no negative long-term-effects were experienced. However, pre-existing high-grade lesions of the menisci might be a predisposing risk for osteoarthritis, triggered by the stress of long-distance running (Krampla et al., 2001). The effect of running and walking is not very clear after this paragraph, due to the disagreements
in the literature. However, the majority of the literature indicates that weight-bearing impact loading through joints (without the effect of motion) during physical activity appears as a risk factor in the progression of OA and will be in the focus of this thesis.

A common surrogate in biomechanics for measuring impact loading on the lower limb joints is load rate that is measured with a force plate (Daoud et al., 2012; Milne et al., 2005; Milner et al., 2006, 2007). Load rate is defined as the rate of change of load with respect to time. Hence, this thesis will use load rate as a surrogate for impact loading on the lower limb joints. The load rate on the lower limb joints will be estimated with wearable accelerometers (smartphones and smartwatches) during physical activity. The benefit of using wearables is that they are cost-effective and easy to use for everyday life, especially since wearables are widespread tools used by people of all ages (Ofcom, 2015). By the end of this thesis, a smartphone app will be developed, so-called OApp™, which can monitor estimated load rate on the lower limb joints during physical activity and gives the option to the user to score their pain on a visual analogue scale.

The aim of the thesis is to validate and test the wearable accelerometers to estimate load rate on the lower limb joints. The research in this thesis should lead to further research to develop a clinical app to support people with OA to adhere to an active lifestyle by seeing how their load rate during physical activities might be associated with their OA-related joint pain. Further, it could help clinicians to understand the relationship between load rate and OA-related pain better and to research this relationship while testing pain ameliorating drugs or treatments.

1.1 Structure of the thesis

This thesis continues with a literature review in Chapter 2, which concludes with the research’s rationale, aims and objectives.

To achieve the aims (Section 2.6) of this thesis, four studies had to be accomplished, where each study built on the results of the previous one. The titles of the four studies are:

1. Study 1: Validity of Load Rate Estimates using Accelerometers during Physical Activity on an Anti-gravity Treadmill (Chapter 4).

2. Study 2: Classification and Quantification of Load Rate Estimates of Everyday Activities using the Accelerometer of Wearables (Chapter 5).

3. Study 3: Monitoring Load Rate Estimates in Daily Life using the Accelerometer of Wearables (Chapter 6).
Chapter 1 Introduction

4. Study 4: Single Case Study to Relate Load Rate Estimates and Osteoarthritis-related Pain using the Accelerometer of Wearables (Chapter 8).

The general methods in the studies are described in Chapter 3, such as study designs, equipment, participants, data collection and data analysis between the studies are described. Differences in the methodology are discussed in detail in each study chapter. Before starting with Study 1, a pilot study (Section 4.3) was accomplished to refine the feasibility of the methods in Study 1. During the pilot study, a participant wore a smartphone between his shoulder blades, which recorded acceleration data while he was walking, jogging and running on an anti-gravity treadmill. The anti-gravity treadmill was used to reduce the weight of the participant during running to test different load conditions on the lower limbs. The acceleration data were then transformed into load rate data, which were correlated with the adjusted gravity on the anti-gravity treadmill.

Following the pilot study, Study 1 (Chapter 4) involved attaching wearables to 12 healthy participants and acceleration data were recorded while participants performed locomotive activities on an anti-gravity treadmill. The acceleration data were then transformed into load rate data with a load rate magnitude algorithm, which was used to develop a linear mixed regression model to explore whether wearables are a suitable tool for estimating load rate.

During Study 2 (Chapter 5), 11 healthy participants equipped with wearables performed locomotive activities at different velocities on various terrains (asphalt, grass, and gravel). Load rate data on the lower limb joints were estimated from the different activities during an outdoor circuit to create a three-layered algorithm consisting of load rate extraction with the load rate algorithm, including a classifier and categoriser, thereby splitting the load rate into low, moderate, and intensive load rate.

Study 3 (Chapter 6) was, in contrast to Study 1 and Study 2, a prospective study and had aimed to assess the ability of a developed smartphone app to record continuously load rate data. The app was able to transform the acceleration data to mean load rate data directly on board on the phone. The app which, was developed during Study 3, was a prototype and was extended in Study 4 to the $OApp^{TM}$.

Before commencing Study 4, a patient and public involvement (PPI) representative event took place where people with lower limb OA were asked for their opinions about the design and structure of the $OApp^{TM}$, which provided important information for Study 4. The difference between the app in Study 3 and $OApp^{TM}$ was, that it additionally gave the user the option to score their pain on the user-interface of $OApp^{TM}$. Hence, $OApp^{TM}$ was able to record load rate and gave the option to record pain on the user-interface.

In Study 4 the app from Study 3 was extended with the option to record OA related pain on a visual analogue scale from 0-10. Study 4 (Chapter 8) then ascertained whether the equipment (load rate magnitude algorithm, smartphone and smartwatch) were able to
find a relationship between load rate and OA-related pain. Study 4 consisted of a case study of a female with hip OA who was scheduled for an intra-articular corticosteroids injections (CSI). She carried wearables during the study for 14 days (seven days pre and seven days post-injection), on which the OApp™ was installed. OApp™ estimated her daily load rate and the participant was asked to score her OA-related pain on the OApp™.

Hyperlinks are used within this thesis, which are marked when pointing on them with the mouse on the pdf file.
Chapter 2

Background and literature review

The background discusses important concepts in this thesis and the literature review analyses the existing literature in order to better prepare for and understand the aim of this thesis. The background covers the important characteristics of cartilage in this thesis and the literature review first delves into the physical activity monitoring methods which are already used in clinical setting. Furthermore, the literature regarding load monitoring and how this thesis approaches it were examined.

After deciding which method was to be used in this thesis, the literature regarding signal processing was analysed in order to help the reader understand how the physical activity monitoring method was to be applied. The literature review for each study, on the other hand, is conducted in each of the study chapters.

2.1 Background

The interest in the mechanics of human movement began with studies conducted by ancient Greeks, where the topic was approached from a more philosophical standpoint rather than from experimental evidence (Beasley, 1982). Movement studies with a more scientific approach began in the 17th Century (ibid). Giovanni Alfonso Borelli is often called the father of biomechanics (ibid). In his book, De Motu Animalium, he explained muscular movement and other body functions according to the laws of statics and dynamics (Borelli, 1680). He utilised muscle analysis and mathematical illustrations of movement, such as running and jumping.

Experiments with detailed measurements and analyses of body movements considering the forces involved in those movements started in earnest in the 19th Century, whereas the 20th Century saw a large increase in research in kinematics and kinetics and used sophisticated equipment (Beasley, 1982), such as instrumented knee implants (Bergmann et al., 2014), force plates (Ehrg et al., 2011) or inertial sensors (Kobsar et al., 2017).
Due to the mechanical properties of cartilage, which will be discussed in the following section, the vertical impact loading on the lower limbs are the interest of this thesis. The forces transmitted through the joints is called load. Every part of the joint, such as cartilage, bone, muscle, ligaments and nerves, take part in load transmission. In mechanics, impact loading is a high force or shock applied over a short period of time when two or more bodies collide. To identify impact loading during activities on the lower limb joints, the change of load over time (i.e. the load rate) has to be estimated. The load rate will be discussed in more detail in Section 3.6.1.

2.1.1 The Mechanical properties of cartilage as a material

Articular cartilage is a biological tissue composed of interacting solid matrices and their interstitial waters, where the solid matrix is composed mainly of collagen and cartilage molecules, such as proteoglycan (PG) and chondrocytes (Oei et al., 2014). Articular cartilage generally has three layers: superficial, middle and deep (see Figure 2.1).

![Figure 2.1: The overall structure of articular cartilage (Oei et al., 2014) consists of three layers: superficial, middle and deep. The layers include proteoglycan (PG), chondrocytes, and collagen fibrils. Left: a normal histology with an intact articular surface, a high concentration of PG in all layers, and tightly ordered collagen fibrils. Middle: degradation of PG in the superficial and middle layers which leads to either the disorganisation or reorientation of the superficial layer’s collagen network and water influx. Right: an additional loss of PG and, in turn, a fibrillation of the cartilage’s surface which thereby leads to advanced OA.](image)

The collagen fibrils of the deep layer of the cartilage matrix run perpendicularly to the articular surface. Moreover, in order to become parallel to the superficial layer, the fibrils curve in the middle layer. For simplicity, articular cartilage can be assumed to be a viscoelastic material (Özkaya et al., 2017). This assumption is theoretical and simplified, however, sufficient for the purpose of understanding what impact loading on the joints means. Viscoelastic materials show viscose and elastic behaviours, depending
on how fast the load is applied to the material (Özkaya et al., 2017). In Figure 2.2, one can see that the theoretical behaviour of viscoelastic materials depends on stress $\sigma$ and strain $\epsilon$ at different loading levels.

Figure 2.2: The stress ($\sigma$) and strain ($\epsilon$) curve for a viscoelastic material at different loading levels (Luo et al., 2014). The hysteresis between point OB shows a continuous line for the loading and a dotted line for the unloading process. Between points O and B, the material deforms during the loading process and goes back to its original shape at point O. The area between the continuous line and the dotted line is the absorbed energy. Passing point B initiates the cracking of the material. The hysteresis between point OM shows that, when the material passes the initial cracking point, the material will not go back to its original shape after unloading. It will reach point N rather than point O. The crack initiation means that the material is then in the stable crack growth phase until it reaches point C, at which point it transforms into an unstable crack growth. From this point, the material fails. The ratio of stress ($\sigma$) and strain ($\epsilon$) can vary with time in terms of the load which is applied to the viscoelastic material. The faster the load is applied to the viscoelastic material, the faster point B is reached.

In general, viscoelastic materials deform to a certain point with slow applied loading and deform back to its initial shape after unloading (Chawla and Meyers, 1999), which develops a hysteresis in the stress ($\sigma$) and strain ($\epsilon$) curve. The area in the hysteresis loop is the energy lost per unit volume in the entire deformation cycle (Chawla and Meyers, 1999) and can be seen as the energy that was absorbed by the system (Figure 2.2). However, in viscoelastic materials strain is time-dependent (Özkaya et al., 2017) and, with an increasing strain rate, the stress increases; hence, the crack initiation increases as well (Figure 2.3(a)). If the loads are applied quickly, the viscoelastic materials behave stiff like elastic materials and do not have a sufficient time to react to the load. This, in turn, may lead the material not to deform, thereby causing it to break. Figure 2.3(a) and Figure 2.3(b) show the time dependency of the strain. However, it should not be
forgotten, that biological tissues such as cartilage and bone have different coefficients for their viscoelasticity, meaning that they have a different strain-stress relationship, hence, different crack initiation points (Özkaya et al., 2017).

However, more detailed in vivo research has to be done to explore the behavior of cartilage in responds to loading regarding the progression of OA. Sutter et al. (2015) directly measured local tibiofemoral cartilage strains in response to a dynamic hopping activity in normal healthy knees and found a significant difference in the cartilage thickness after the hopping. Lad et al. (2016) measured local tibiofemoral cartilage strains in response to walking in normal healthy knees and found a significant difference in the cartilage thickness after walking. Both methods are ways to explore the cartilage after loading and might help to understand the material behavior of articular cartilage. Both methods used in Sutter et al. (2015) and Lad et al. (2016) need to be used in a epidemiological study, to explore the behavior of cartilage in responds to loading regarding the progression of OA.

From the current knowledge it can be assumed, out of simplicity reasons, that a lower load rate has less impact on the cartilage than a high load rate due to the viscoelastic behaviour of the cartilage. Applying the previous knowledge about viscoelastic material to articular cartilage, it can be concluded that over a threshold of load rate, the cartilage might not have sufficient time for deformation to occur. When no sufficient time is given for the cartilage, it becomes stiff (Jurvelin et al., 1997) and, in turn, brittle. Thus, the cracking initiation point of the cartilage will be at an earlier stage. Advanced OA is related with net loss and damage of collagen fibrils and loss of PGs (Oei et al., 2014) (Figure 2.1), which lead to decrease of the structure of cartilage. It can be concluded that applied load, depending on time (load rate), goes into nano-fracturing and into the thinning of the three layers of cartilage.
2.1.2 Load rate during locomotion measured with the gold standard equipment

In sports science, repetitive loading is known to be a key complement in the pathophysiology of stress fractures (Milner et al., 2007). Schaffler et al. (1989) showed that repeated loading at higher load rates, such as that which occur during running, are more damaging for fracturing than repeated loading at lower load rates. Tibial stress fractures are related to vertical load rates (Milner et al., 2006). Davis et al. (2004) showed on a large group of runners that higher positive vertical load rates were found mostly in people with tibial stress fractures in comparison to controls. In Figure 2.4, a gait cycle is showing the force over time when the foot makes the first contact with the ground during running.

Figure 2.4: (A) shows a gait pattern with an impact peak at the moment the foot begins to have contact with the ground. (B) shows a gait pattern with a foot strike that does not cause an impact peak. (Daoud et al., 2012)

Milner et al. (2006) hypothesised that high load rates contribute to some extent to injuries. This generates high stresses and strains in parts of the joint, such as the cartilage and bone, because of their viscoelastic properties, which can, over repeated cycles, consequently lead to damage (Milner et al., 2006; Daoud et al., 2012). The studies mentioned in this paragraph are focusing on tibial stress fractures, which can heal and might not effect the post fracture running gait, which would be different for OA or cartilage damage.

Gill and O’Connor (2003), Pamukoff et al. (2016) and Addison and Lieberman (2015) investigated in vertical ground reaction force rates. Gill and O’Connor (2003) investigated gait patterns of normal subjects to establish the gait determinants responsible for producing large impulses at heelstrike. Pamukoff et al. (2016) found that while walking at a standardized speed, obese subjects displayed greater instantaneous vertical loading
rates. Hence, they suggest that obesity may contribute to knee osteoarthritis, as greater loading rates are related to cartilage degeneration (Pamukoff et al., 2016). Addison and Lieberman (2015) investigated how variations in footwear heel stiffness influenced several aspects of walking and heel strike running impact peaks, which is known to be a variable for various repetitive stress injuries. They found that both vertical impulse and effective mass increase when wearing less stiff footwear. Further, they found out that there is a significant inverse relationship between impact loading rate and vertical impulse. Which means that less stiff footwear might be related to lower impact loading rates. Mündermann et al. (2005) found that patients with more severe knee OA had greater first peak knee adduction moments, meaning that they had higher load rate values than their matched control subjects. This, however, has to be carefully interpreted, because the higher load rate in people with OA can be a secondary effect of their disease and not from their initial gait pattern. Øiestad et al. (2015) found that daily walking is not associated with the progression of OA. The finding of Øiestad et al. (2015) confirmed the assumption that activities with lower load rates might be less associated with the progression of OA than activities with higher load rates. It should be also mentioned that Timmins et al. (2017) did not find a relationship between running and OA in their literature review. They found that moderate to low quality evidence suggests no association with OA.

The literature is not clear on this topic, however, a majority of the literature mentioned above indicate that high load rate values can be used as an indicator for impact loading on the lower limb joints and might be connected with the progression of OA. However, with these load rate estimates it is not possible to identify if different joints are more affected by the impact loading or gain more detailed information about impact loading. Further, it has to be mentioned that ground reaction force load rates are not necessarily the accurate predictor of knee joint load rate. This might be since ground reaction forces can be modulated through muscle contraction and limb segment motion in order to dampen joint contact force load rate (Sutter et al., 2015; Lad et al., 2016).

### 2.2 Literature review: clinical use of physical activity monitoring

A physical active lifestyle is recommended to stay healthy and prevent diseases and for people with a variety of diseases (Haskell et al., 2007; Valderrabano and Steiger, 2010). In some academic studies (Murphy et al., 2008; Adesola and Ololade, 2013), physical activity is seen as a variable, which needs to be quantified and measured. Multiple subjective and objective measurements exist, which are analysed in the following section. The following literature review had the purpose to give the reader a brief overview of validated methods to monitor physical activity over the time.
2.2.1 Method of literature review

To select the literature required, a list of keyword was used for the search engine. The key words are related to physical activity monitoring in clinical use. These terms were found during preliminary study in the literature used in this work. The following keywords were used:

- Physical activity
- Monitoring
- Recording
- Clinical Use

Studies were included if they: (1) were peer-reviewed studies in the English language, (2) in clinical use (3) measuring physical activity, (4) recognized valid measurement tools. The valid measurements were determined as validated against gold standard equipment. Studies were excluded if they were not recognised valid measurements for physical activity and if the method was repeated.

The following databases were searched within the time frame of 2000-2016: The PubMed, CINAHL, Health and Physical Psychosocial Instruments data bases, and the Centre for Disease Control Web site. The time span 2000-2016 was used to obtain the changes over time for physical activity monitoring.

Due to not being the major focus of this thesis, the results were ordered by relevance and only the top 100 were considered in this literature review. A three phase screening strategy was used to identify relevant articles. Firstly, potential relevant studies were identified by their title (100 studies). Secondly, these potential studies were filtered by their abstracts (remaining 25 studies). Finally, the full text of remaining articles was reviewed against the selection criteria (9 studies included). The results of this literature review was presented in the sections 2.2.2 and 2.2.3.

2.2.2 Subjective measurements

Self-report methods are a common, and probably the cheapest, way to monitor activity in clinics; nevertheless, their measurement validity has some limitations (Sallis and Saelens, 2000). Early methods of recording activity involved using paper-based recall questionnaires and diaries, where patients were asked to write down their activities and the joint pain that they had experienced whilst doing those activities from their memory (Roos and Lohmander, 2003).
Stone et al. (2003) reported that compliance rates in paper-based diaries cause difficulties related to data quality. According to Stone et al. (2003) a person’s willingness and recall accuracy may introduce errors in the data. Due to the dependence on self-motivation, completing paper-based diaries have a high level of unreliability (Stone et al., 2003). A study using instrumented diaries for pain entries demonstrated that paper diary compliance, which was reported to be 90%, was actually 11% (ibid). On 32% of all study days the paper diary binder was not opened, yet reported compliance for these days exceeded 90%. It showed that the participants filled multiple pain entries after the suggested time, because they forgot to enter their pain level and thought they can add afterwards the pain level. An alternative form of the self-report method is electronic diaries which, even though they are more expensive, they are also equipped with alarms and reminders to increase the compliance of the user.

Dinger et al. (2004) showed a correlation between Physical Activity Scale for the Elderly (PASE) and ActiGraph accelerometer of 0.91 for older individuals. However, a more recent study (Dyrstad et al., 2014) used a paper-based questionnaire (International Physical Activity Questionnaire, IPAQ) to monitor physical activity and to compare their outcomes with the ActiGraph accelerometer. They found large variations and poor agreement between the self-reported and accelerometer-measured physical activities (ibid). According to Dyrstad et al. (2014) recall questionnaires cause some issues with regards to the accuracy of the measurement of physical activity. Depending on the questionnaires, users have to recall their physical activity either over the past 24 hours, the past week, or even for longer periods. The recall over a longer period of time may include memory errors for the user. However, PASE seems like an accurate method to monitor physical activity for elderly people according to Dinger et al. (2004).

Jacobs Jr et al. (1993) showed that light to moderate physical activities tended to not be reported. According to Jacobs Jr et al. (1993) questionnaires have the tendency to target specific populations, such as the elderly or adolescents, and consist of questions which reflect more specific group patterns. The outcome of the literature review of subjective measurements is listed in Table 2.1 and Table 2.2.
Table 2.1: Outcome and quality of relevant studies for measuring physical activity in clinical use (Subjective measurements, part 1)

<table>
<thead>
<tr>
<th>Author and title</th>
<th>Objective</th>
<th>Method</th>
<th>Subject Information</th>
<th>Conclusion</th>
<th>Strength</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consolvo et al. (2006), Design Requirements for Technologies that Encourage Physical activity</td>
<td>To assess the pedometer’s being a suitable method for encouraging people to become more physically active.</td>
<td>Three weeks, in situ, of daily step counting as a measure of physical activity. Included sharing with friends and step goals.</td>
<td>13 healthy subjects</td>
<td>Pedometers or other accelerometers can encourage people to become more physically active.</td>
<td>Mixed qualitative and quantitative, exploitative, observational design, interviews, social support</td>
<td>The sample size of the pilot study; the pedometer under-represented overall physical activity; effort can be different on inclines; step goals were one-size-fits-all; lack of deeper understanding of motivational behavioural change and barriers</td>
</tr>
<tr>
<td>Sallis and Saelens (2000), Assessment of Physical Activity by Self-Report: Status, Limitations, and Future Directions</td>
<td>To summarise the findings regarding test-retest reliability, criterion-related validity, and content validity for physical activity self-reports which were developed in the 1990s.</td>
<td>Literature review of self-report measurement of physical activity, objective measurements: accelerometers, direct observation, and heart rate monitors</td>
<td>Literature review</td>
<td>Most studies showed self-reports do not provide accurate estimates of the absolute amount of physician activity</td>
<td>Questionnaires are for one specific group</td>
<td>Little information was provided regarding the study’s method.</td>
</tr>
<tr>
<td>Author and title</td>
<td>Objective</td>
<td>Method</td>
<td>Subject Information</td>
<td>Conclusion</td>
<td>Strength</td>
<td>Weaknesses</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>--------</td>
<td>----------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>Stone et al. (2003), Patient compliance with paper and electronic diaries</td>
<td>To assess the compliance of paper-based diaries</td>
<td>Paper diaries or electronic diaries with time-stamped entries were kept in order to track diary use for 24 days</td>
<td>84 subjects who reported experiencing pain for 3 or 4 days per week</td>
<td>Compliance was compared between paper and electric diaries; the compliance of paper diaries was, in reality, only 11%, but was reported as being 98%; compliance even decreased over three weeks’ time; the method is interesting seeing as the sensor was located at the binder.</td>
<td>Comparison of groups between paper and electronic diaries in order to determine whether compliance depends on group behaviour.</td>
<td>Electronic diaries had features which supported compliance (alarm clock, carrying device); study was not designed to tell which feature was responsible for the compliance.</td>
</tr>
<tr>
<td>Jacobs Jr et al. (1993), A simultaneous evaluation of 10 commonly used physical activity questionnaires</td>
<td>To assess the reliability of 10 commonly used physical activity questionnaires</td>
<td>Ten common physical activity questionnaires were evaluated for reliability and validity; 4 week physical activity histories; 2 day accelerometer readings.</td>
<td>78 men and women age 20-59, test-retest reliability</td>
<td>One month reliability was high in all questionnaires; no questionnaire correlated with the accelerometer data.</td>
<td>That is compared the questionnaires with two accelerometers.</td>
<td>Most of the questionnaires regarded high intensity activities.</td>
</tr>
<tr>
<td>Dyrstad et al. (2014) Comparison of self-reported versus accelerometer-measured physical activity</td>
<td>To compare physical activity and sedentary time from IPAQ with ActiGraph</td>
<td>7 days of wearing an ActiGraph and answering a questionnaire at the end of the seven days.</td>
<td>1751 adults</td>
<td>The participants reported through IPAQ more vigorous physical activity and less sedentary time; sedentary time was compared to the accelerometer data.</td>
<td>Large population sample size from a wide age range and the random inclusion of rural and urban populations.</td>
<td>A large number of participants did not answer all of the IPAQ questions. The sensor and questionnaire was sent via mail and does not necessarily represent the same 7 days.</td>
</tr>
</tbody>
</table>
2.2.3 Objective measurement

Forms of objective measurements are direct observation, pedometers, or accelerometers. Direct observation was one of the very first forms of objective measurements and provides valid and reliable information (Friedewald, 1985). But direct observation is time consuming and the duration of such observation is usually very short.

The next step of monitoring activities was the implementation of body-worn pedometers which measure the numbers of steps taken and which have been demonstrated to be a useful measurement of steps taken (Silva et al., 2002). The limitations of pedometers as an activity monitoring tool are that they are not able to record load on the joints or the intensity of the activity. Moreover, the accuracy of the pedometer decreases in people with a slow gait, something which is common in older adults or obese individuals (Martin et al., 2012).

Another validated method is the use of body-worn tri-axial accelerometers for physical activity monitoring (Item-Glatthorn et al., 2012). The validation of accelerometer use for physical activity monitoring was discussed in Section 4.1.1. Body-worn accelerometers have micro electro-mechanical systems (MEMS) sensors integrated within them which measure raw accelerations. These can measure raw accelerometer data, which can be used for identifying more complex activities than the pedometer. Bieber et al. (2013) explored the ability of the Actiwatch Spectrum from Philips (Philips, 2018) and its industrial application. Neugebauer et al. (2014) and Meyer et al. (2015) accessed the validity of the ActiGraph (ActiGraph, 2017) against the gold standard equipment (the force plate) and found a correlations between the peak acceleration of the ActiGraph and the force plate peak ground reaction forces. Grant et al. (2006) validated with the optical motion capture professional system ActivPal (ActivPal, 2018) and found that ActivPal activity monitor is a valid and reliable measure of posture and motion during everyday physical activities. Other forms of physical activity monitoring are motion capture systems with multiple connected accelerometers, such as Xsens (Xsens, 2018). Fernández-Baena et al. (2012) validated Xsens system (Xsens, 2018) against an optical motion capture professional system and obtained a range of disparity that guaranties enough precision for most of the clinical rehabilitation treatments prescribed nowadays for patients. Kobsar et al. (2017) used a four-accelerometer system and found out that three sensors were best able to identify responders. However, they suggest that a simplified two sensor array at the back and thigh may be the most ideal configuration to provide clinicians with an efficient way to use to optimize treatment.

All these accelerometers however are an additional investment for the user, and which means they have to carry an additional device. The more convenient smartphone and smartwatch accelerometer seem like a cheaper and easier method to monitor physical activity. Nishiguchi et al. (2012) validated smartphone based accelerometers. They found out that all the gait parameter results obtained by the smartphone showed statistical
significance and that the smartphone showed considerable correlations with the same parameter results obtained by the tri-axial accelerometer ($r = 0.99 - 0.82; p < 0.01$). Their study indicates that the smartphone with a gait analysis application used in their study has the capacity to quantify gait parameters with a degree of accuracy that is comparable to that of the tri-axial accelerometer (Nishiguchi et al., 2012). The outcome of the literature review of objective measurements is listed in Table 2.5 and Table 2.6.

There has been very little research done on the estimation of impact loading on the joints with accelerometers, which is the focus of the present thesis. Previous work in the estimation of impact loading on the joints with accelerometers was discussed in Section 2.3. To find an indicator for impact loading, one has to delve into a better understanding of impact loading on the joints and how to indicate this via sensors. The key might be to use different features in comparison to the ones which are used currently in activity monitoring.
Table 2.3: Outcome and quality of relevant studies for measuring physical activity in clinical use (Objective measurements, part 1)

<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Study Design</th>
<th>Literature Review</th>
<th>Findings</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bassett Jr and John (2010), Use of pedometers and accelerometers in clinical populations: validity and reliability issues</td>
<td>To introduce clinicians to various activity monitors that are commercially available</td>
<td>Literature review of the scientific literature from the past 20 years; utilised the most common types of physical activity monitoring; was devised by establishing validity and reliability.</td>
<td>Pedometers and accelerometers are useful for tracking ambulatory physical activity in clinical populations</td>
<td>8 pedometers are listed in detail, including their benefits and disadvantages.</td>
</tr>
<tr>
<td>Silva et al. (2002), Average patient walking activity approaches 2 million cycles per year</td>
<td>Evaluated the accuracy of accelerometers which were used for assessing walking activities.</td>
<td>Participants wore the 2D accelerometer called the Step Activity Monitor (SAM), as well as a pedometer for 4 days representing usual walking activity.</td>
<td>33 subjects with at least 2 years status-post successful hip replacement and walking activity was not limited</td>
<td>Pedometer had an under-recording of 34% compared to SAM; measurements of the SAM and the pedometer were highly correlated.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cost efficiency; the comparison of two objective measurements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The pedometer had limitations, including not being reliable for recording the walking activity of obese and elderly people.</td>
</tr>
<tr>
<td>Author and title</td>
<td>Objective</td>
<td>Method</td>
<td>Subject Information</td>
<td>Conclusion</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>--------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Martin et al. (2012), Pedometer accuracy in slow walking older adults</td>
<td>To determine pedometer accuracy during slow overground walking in older adults</td>
<td>Participants wore five different brands of pedometer while walking 3 velocities between 0.3 and 0.9 ( \frac{m}{s} ), one self selected cadence over 80 meters</td>
<td>18 Participants</td>
<td>Pedometer accuracy decreases with slower walking speeds across all devices; the use of pedometers for monitoring step counts in healthy older adults with slower gait velocities is problematic.</td>
</tr>
<tr>
<td>Item-Glatthorn et al. (2012), Validity of Intelligent device for Energy expenditure and activity accelerometer System for Quantitative Gait Analysis in Patients With Hip OA</td>
<td>To evaluate the concurrent validity of an accelerometer-based system with a criterion instrument in orthopedic patients</td>
<td>Validity study; participants walked at normal and fast velocities, evaluating the concurrent validity of an accelerometer-based system (Intelligent Device for Energy Expenditure and Activity) with a criterion instrument (Gaitrite).</td>
<td>26 men with unilateral hip OA</td>
<td>Quantitative gait analysis with IDEEA was satisfactory</td>
</tr>
</tbody>
</table>
2.2.4 Commercially available devices used for activity monitoring

Commonly used accelerometers in research are ActiGraph (ActiGraph, 2017), ActivPal (ActivPal, 2018) or GENEa (Activinsights, 2017), which are validated against the golden standard equipment (Meyer et al., 2015; Neugebauer et al., 2014). Even if their accuracy and functionality is sufficient for physical activity monitoring, a major disadvantage of these commercially available accelerometers is that they would be an additional device for the users. Smartphones conveniently include accelerometers, which is a device commonly used in public. Another advantage of using a smartphone is that apps can be developed and used by anybody on the specific operating system (e.g. Android or iOS). The operating systems such as Android or iOS, offer to develop apps, which have access to sensor data. A good example is the OApp, which was developed as part of this research work, which can be easily installed on any Android phone.

Commercially available activity monitoring tools such as the Activity app on Apple Watch (AppleWatch, 2018) or the Fitbit Charge 2 by Fitbit (Fitbit, 2018) might be useful tools to estimate overall activity (El-Amrawy and Nounou, 2015; Bunn et al., 2018), based on accelerometers, heart rate monitors and GPS. However, different than the aim of this thesis, the loading on the joints are not estimated. The Activity app on Apple Watch counts three quantities (AppleWatch, 2018): 1. Burned calories based on acceleration and heart rate data. 2. Exercise minutes, which is identified if the user has a higher heart rate than the heart rate for a brisk walk. 3. Hours user was standing for at least one minute, which is identified with the accelerometer. Similar to the Activity app on the Apple Watch, the Fitbit Charge 2 by Fitbit counts the steps, calories burned, measures the heart rate and counts the minutes of exercises (Fitbit, 2018). The Activity app on Apple Watch and Fitbit Charge 2 might help users to maintain a healthier lifestyle by motivating them to move (El-Amrawy and Nounou, 2015; Bunn et al., 2018), but no further information is given regarding their joint loading.

2.3 Literature review: estimation of loading on the joints with accelerometers

Some attempts have been made to measure the load on the joints during physical activity using accelerometers. This section is a brief literature review about monitoring loading, to understand what kind of research is done in this field. Due to being a novel approach the results for this literature review were spares and were varying amongst each other in their approaches. That is why further reviews were done in each study corresponding to its specific topic (Sections 4.1.1, 5.1.1, 6.1.1, and 8.1.1). The literature review strategy of the following chapters (Sections 4.1.1, 5.1.1, 6.1.1, and 8.1.1) was based on this literature review with additional papers found via Google Scholar for specific queries, such as “validation of accelerometer to estimate ground reaction forces”.

2.3.1 Method of literature review

To select the literature required, a list of keyword was used for the search engine. The keywords are related to loading on the joints during physical activity. The following keywords were used:

- Physical activity intensity
- Load intensity
- Peak load
- Load rate
- Accelerometer

Studies were included if they: (1) were peer-reviewed studies in the English language, (2) used accelerometer-based activity monitoring, and (3) tried to measure the loading on the joints during physical activity. Studies were excluded if they were not using acceleration to recognise the loading or were not used for the lower limbs.

The following databases were used for the research, focusing on the time frame between 2005-2016: Engineering Village, Medline, Web of Science, Scopus, IEEE Xplore, SciFinder Scholar. 2005 was chosen because the use of accelerometer for physical activity monitoring was a relatively new method. In addition, some papers were included which were referenced by papers read previously and thought to be important, as they contained results that affected this thesis. The source types were academic journals.

A three phase screening strategy was used to identify relevant articles. Firstly, potential relevant studies were identified by their title (47 studies). Secondly, these potential studies were filtered by their abstracts (remaining 14 studies). Finally, the full text of remaining articles was reviewed against the selection criteria (6 studies included). The results of this literature review are presented in the following section.

2.3.2 Estimations of loading on the joints with accelerometers

One approach for monitoring impact loading with accelerometers was taken by Kelley et al. (2014) who developed a method for assessing the potential bone-loading intensity of different locomotion activities. They developed a simplified load intensity algorithm for physical activity, which is the product of load (the product of body mass and measured acceleration) and load frequency. This was normalised with respect to body weight. The results of the study were that, the higher the velocity of walking or running, the higher the loading intensity \( p < .05, \) Cohen’s \( d = 1.15 \). Even though the load intensity
algorithm had a high correlation, the method was not validated against a gold standard equipment such as a force plate. Further, due to the viscoelastic behaviour of biological tissue it is hypothesised that biological tissue is more prone to loads with higher load rate (Özkaya et al., 2017) rather than the load intensity algorithm that Kelley et al. (2014) developed.

Other features for identifying impact loading can be found in elite sports research. Jarming et al. (2015) tried to estimate jump frequency in volleyball, which has a high loading on the joints, by using peak vertical acceleration (PVA) and peak resultant acceleration (PRA). They hypothesised that jump movements present a greater PVA or PRA than non-jump movements. This hypothesis could not be proven, because different movements, such as side-to-side shuffle steps, shuttle runs, and floor dives, did not show significant differences to jump movements. They recommended that future researchers not use either PVA nor PRA measured by an accelerometer for estimating jump frequency as a verified system.

Gabbett et al. (2010) used the acceleration rate magnitude estimated with the minimaxX (Catapult (2017), Catapult Innovations, Melbourne, Australia) for identifying collisions and horizontal impact forces on professional rugby league players during pre-season and in-season skills training sessions. The results showed a strong correlation ($r = 0.96$, $p < 0.01$) between collisions recorded via the acceleration rate magnitude and those recognised from video recordings. The algorithm was suitable for detecting collision and horizontal impact forces.

Barrett et al. (2014) used the same acceleration rate magnitude measurement as Gabbett et al. (2010) for team-sport players performing different velocities on a treadmill, obtaining moderate to high test-retest. They demonstrated with measures of exercise intensity for a single individual ($r = 0.92$ to $0.98$); nevertheless, the results also showed that one cannot compare exercise intensities between individuals using that method. This can be due to hidden or unconsidered variables, such as anatomy, muscle strength and gait of individuals.

Considering the above mentioned studies, the acceleration rate magnitude appears to be a suitable feature for detecting impact loading. The acceleration rate magnitude does not consider the individual weight of participants, which might cause the poor accuracy between participants. This thesis therefore included the weight of participants: estimating the load rate with accelerometer as the product of acceleration rate magnitude and weight. The validation of acceleration rate magnitude and load rate estimates against the gold standard equipment is discussed in Chapter 4.

The load rate is a valid measurement for impact loading on the joints, which was shown by Daoud et al. (2012), Milner et al. (2006) and Milner et al. (2007) against the gold standard equipment. These studies measured load rate in biomechanical laboratories with a force plate as a surrogate for impact loading on the lower limb joints. This thesis
concentrated on estimating the load rate of the lower limbs in everyday life using the accelerometer of wearables (smartphone and smartwatch) for the purpose of achieving a more accessible tool outside the biomechanical laboratory.
Table 2.5: The outcome and quality of studies measuring impact loading (part 1).

<table>
<thead>
<tr>
<th>Author and title</th>
<th>Objective</th>
<th>Method</th>
<th>Subject Information</th>
<th>Conclusion</th>
<th>Strength</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannini et al. (2013), Activity recognition using single accelerometer places at the wrist ankle</td>
<td>To obtain an algorithm to process wrist and ankle raw data and classify behavior into four broad activity classes: ambulation, cycling, sedentary and other</td>
<td>Accelerometers were worn on the wrist and ankle; 26 daily activities were performed at the 2s, 4s, 12.8s windows; Feature: signal magnitude vector and frequency-domain features; four activity levels: ambulation, cycling, sedentary and other; 90 Hz, SVM</td>
<td>33 healthy subjects</td>
<td>Classification algorithm with 13 features shows good classification into four activity intensity categories</td>
<td>Used leave-on-subject-out cross-validation. Classification of broad intensity categorises</td>
<td>Was not able to identify within categories; used frequency domain to classify activity intensity.</td>
</tr>
<tr>
<td>Zhang et al. (2012), Physical Activity Classification using the GENEA Wrist Worn Accelerometer</td>
<td>To develop methods to classify physical activities into walking, running, household or sedentary activities based on raw acceleration data; compare wrist-worn with waist-worn</td>
<td>10-12 semi-structured activities in the laboratory and outdoors. Accelerometer: waist, left wrist, right wrist, 80Hz. Classify activities into sedentary, household, walking and running.</td>
<td>60 healthy subjects</td>
<td>Algorithms suitable for use with wrist-worn accelerometers for detecting certain types of physical activity.</td>
<td>Classification of broad intensity categorises; compared waist and wrist; compared multiple classifiers</td>
<td>A 10-fold cross validation was conducted; it is due to that validation technique that the accuracy is so high.</td>
</tr>
<tr>
<td>Author and title (2014),</td>
<td>Objective</td>
<td>Method</td>
<td>Subject Information</td>
<td>Conclusion</td>
<td>Strength</td>
<td>Weaknesses</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>--------</td>
<td>---------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>Kelley et al.</td>
<td>To develop a new method for assessing potential bone-loading intensities.</td>
<td>8 self-assessed activities; load intensity was measured from a loading magnitude and frequency activity of 50Hz.</td>
<td>30 healthy subjects</td>
<td>Developed a new method for measuring the loading intensity of physical activity on bone by using an accelerometer.</td>
<td>Bone behaviour was first researched; based on the finding of that research, an algorithm was developed.</td>
<td>There was no evaluation of another valid system; trials were self-assessed; a method for measuring intensity was self-developed.</td>
</tr>
<tr>
<td>Jarning et al. (2015), Application of a triaxial accelerometer to estimate jump frequency in volleyball</td>
<td>To explore the possibility of using peak vertical acceleration or peak resultant acceleration</td>
<td>Explore the possibility of using peak vertical acceleration and peak resultant acceleration to estimate jump frequency; acceleration and video data was also collected at 100Hz.</td>
<td>20 male elite volleyball player</td>
<td>Neither peak vertical acceleration nor peak resultant acceleration differed significantly between jumping movements and non-jumping movements; it was not an applicable method.</td>
<td>Elite athletes; high sampling frequency (100Hz)</td>
<td>Data analysis simply ANOVA; choice of PVA and PRA weak for jumps; did not include any signal processing procedures.</td>
</tr>
<tr>
<td>Barrett et al. (2014), PlayerLoad™ reliability, convergent validity, and influence of Unit Position During Treadmill Running</td>
<td>To establish the test-retest reliability and convergent validity of PlayerLoad™</td>
<td>Performed two standardised incremental treadmill running tests, each three times each; the test was separated by 7 days and performed at the same time every day according to an accelerometer and heart rate monitor.</td>
<td>44 team-sports players</td>
<td>PlayerLoad™ had a moderate to high test-retest reliability and demonstrated convergent validity with measures of exercise intensity</td>
<td>Test-retest reliability</td>
<td>The system was not validated by another force measuring system such as a force plate.</td>
</tr>
</tbody>
</table>
Chapter 2 Background and literature review

2.4 Literature review: signal processing within physical activity monitoring

This thesis utilises MEMS accelerometer within a wearable device, which is an objective measurement method to measure acceleration. Wearables are defined in this thesis as smartphones and smartwatches and their MEMS accelerometer are used as easily commercially available and cost effective accelerometers sensors. Smartphones and smartwatches were used because they are more accessible for the broader population than other commercially available accelerometers such as ActiGraph (ActiGraph, 2017) or ActivPal (ActivPal, 2018). The overall aim is to develop a tool, that can easily be used by the broader OA population rather than to make them to buy another device. To estimate load rate, the raw acceleration data has to be processed into meaningful features.

2.4.1 Method of literature review

To select the literature required, a list of keywords were used for the search engine. The key words are related to signal processing within physical activity monitoring. The following keywords were used:

- Physical activity
- Exercise
- Accelerometer
- Smartphone
- Machine learning
- Classification

Studies were included if they: (1) were peer-reviewed studies in the English language, (2) used accelerometer-based activity monitoring, (3) covered daily activities or exercises, (4) had a clear method to classify acceleration data, (5) had a clear validation method, and (6) used supervised and/or unsupervised learning. Studies were excluded if they were limited to upper extremity, head or wheelchair activity, also, if they focused on sleep monitoring.

The following databases were searched within the time frame of 2005-2016: Engineering Village, Medline, Web of Science, Scopus, IEEE Xplore, SciFinder Scholar. The references of primary studies identified were scanned to identify further relevant citations.
A three phase screening strategy was used to identify relevant articles. Firstly, potentially relevant studies were identified by their title (519 studies). Secondly, they were further filtered by their abstracts (62 studies). Finally, the full texts of remaining articles were reviewed against the selection criteria (17 studies). The results of this literature review was presented in the following section.

2.4.2 Overall sequences of signal processing

The signal processing includes choice of activity, data acquisition, pre-processing, and further data processing, such as regression or classification. Figure 2.5 illustrates the stages in signal processing.

![Figure 2.5: Pre-processing sequences of the raw data.](image)

The choice of the monitored activity (yellow) will influence the further activity monitoring steps and needs to be clear at the beginning of the analysis. Data acquisition (blue) is the second step in activity monitoring, which includes the sampling frequency of the data collection. The next step is the data pre-processing (pink), which consists of choosing the data windowing and feature extraction. Data windowing divides the data into packages for the purpose of transforming them into features, whereas feature extraction transforms raw data into useful quantities. These features were used in further data processing (green) and supported the division of the data into different classes or helped develop statistical models.

Data acquisition and the pre-processing of the data can affect the accuracy of the classifier, as well as the quality-of-fit of the model. That is why they need to be carefully chosen. The steps in Figure 2.5 and the data analyse methods depend on which activities and what aspect of the activity should be analysed. To obtain a high accuracy, one has to understand the different steps and their effects on the end result. This section helps to understand the choices of signal processing in the following studies of this thesis demonstrated in Chapters 4, 5, 6, and 8. A tabular review of all of the mentioned studies in this section are listed in Appendix A.
2.4.3 Data acquisition

The choice of data sampling frequency will influence classification accuracy. A critical step in classifying data is its acquisition. If the aim is to detect a specific activity, rather than the load of the activity on the joints, the choice of the sampling frequency can depend on the actual frequency of the activity that should be classified. Both Gyllensten and Bonomi (2011) and Khan et al. (2014) use 20Hz in order to classify activities with varying frequencies, thereby leading to a lack of accuracy in the classification. Deng et al. (2014), on the other hand, chose to use 50Hz, thereby achieving a higher accuracy. In terms of differences in accuracy, however, it is difficult to compare the studies that have been done in this field because the accuracy is affected by more than one factor of the data processing.

Bersch et al. (2014) and Zhang et al. (2012), for example, analyse the effects of a range of sampling frequencies. Zhang et al. (2012) (5Hz-80Hz) tried to determine whether reduced sampling still provides acceptable classification under the same conditions and physical activities. They found that sampling frequencies over 10Hz were not associated with higher significant accuracies in everyday activities. Furthermore, Bersch et al. (2014) found that sampling frequencies over 10Hz achieve nearly the same classification accuracies and sampling frequencies, with frequencies above 20Hz resulting in only minor gains. Nevertheless, Zhang et al. (2012) confirms that high sampling frequencies may become more important when discriminating different activity types that have similar frequency levels, such as walking and climbing stairs. Therefore, a higher sampling frequency is required if the load of activities are being determined.

2.4.4 Data windowing

As mentioned before, the choice of window size is affected by other parts of the pre-processing, such as sampling frequency, classification, and model method. The range of the window sizes of the aforementioned studies were from 0.5s to 30s. Bersch et al. (2014) and Sama et al. (2011) compared different window sizes with each other. In their study, Bersch et al. (2014) showed that the optimum window size was between 9s-13s. Furthermore, they showed that higher frequencies achieve the best accuracy for shorter window sizes. Sama et al. (2011), on the other hand, compared the accuracy of different window sizes with the classifier support vector machine and found a maximum in the range between 8s and 11s.

2.4.5 Features extraction

The next step for the pre-processing of the data is that of feature extraction. Three different types of features exist (Long et al., 2009).
Chapter 2 Background and literature review

• Time-domain: Time-domain features are based on physical quantity time. Raw accelerometer data is always time-based and, hence, can be transformed into time-domain features. For the load rate of activities, the time-domain features are useful (Chung et al., 2008). Examples are mean, variance, correlation, and maximum value of the signal.

• Frequency-domain: Frequency-domain features are based on the frequency of the signal (Long et al., 2009). Usually, the Fourier transformation will be applied to the raw accelerometer in order to transform the data from time-based to frequency-based (Long et al., 2009). To monitor activities with the same intensity but with different periodicities, it is useful to use frequency-domain features (Zhang et al., 2012). Every activity has a specific cycle frequency and, consequently, is identifiable with the frequency. Examples are the energy of the signal, the entropy of the signal, and the components of the Fourier transformation.

• Spatial-domain: Spatial-domain features are based on the position and orientation of the device (Long et al., 2009). In order to identify directions and postures, the spatial domain is appropriate. An example of this feature domain is the orientation variation. This feature domain, however, will not be discussed further in this thesis.

2.5 Research rationale

The growing body of work in the field of physical activity monitoring via wearables (Khan et al., 2014; Gabbett et al., 2010; Bersch et al., 2014) classified simple activities like running, walking, resting, and climbing up and down stairs and used acceleration features to identify these classes. The previously mentioned studies and health devices focused on identifying activities rather than extracting biomechanical meaningful information about the particular activities, such as impact loading in the joints. Contrary to this, many studies focus on measuring the load rate on the joints with medical devices, such as instrumented knee implants (Bergmann et al., 2014) or in biomechanical laboratories with force plates (Milner et al., 2006; Daoud et al., 2012; Mündermann et al., 2005), which can yield detailed information about the load rate transmitted through the joints. These detailed measurements, however, have sophisticated laboratory requirements which cannot be used by patients in their everyday lives. As a result, research is needed to find a tool that can estimate impact loading on the lower limb joints outside of the laboratory environment.

The wearables’ micro electro-mechanical systems (MEMS) sensors are easily commercially available inert sensors, which is the motivation to use wearables to develop an app for the purpose of estimating the load rate on the lower limb joints. Since many people in the UK own a smartphone or other kind of wearable (Ofcom, 2015), everyone who has a wearable can use the app. Thus, people are not forced to buy other expensive
devices, such as ActivPal (ActivPal, 2018), ActiGraph (ActiGraph, 2017), or Actiwatch (Philips, 2018), just to monitor the estimated impact loading. Hence, the smartphone is a convenient and cost efficient tool. Another motivation to develop an app is that the Android operation system (OS) is a developer-friendly platform which makes it easy for researchers to develop the requisite tools. This thesis, however, does not concentrate on the development of the app; rather, its focus is that of analysing and interpreting the data in a meaningful way.

Knowing that OA can have multiple known and unknown factors for progression, this thesis only concentrates on the weight-bearing loading on the joints. This thesis is divided into four studies, with each one focusing on validating the system, testing the algorithm for classification of activities with different loading, testing its ability to record continuously, and testing it in an clinical atmosphere on one person with OA as preparation for a clinical trial. All four studies set the technical bases for the use in an clinical trial.

Furthermore, these four studies can help future researches develop a wearable app that monitors load rate and OA pain and, hence, support people with OA to balance their joint loading while having an active lifestyle. This thesis is the beginning of research in estimating load rate on the lower limb joints during physical activity via wearables and should serve as the initiation of an app that will, in the future, be a form of intervention for people with OA or will serve to help researchers understand the progression of OA better. With this future app, people with OA should be able to monitor their load rate and, therefore, understand how load rate on their lower limbs affects their OA-related pain.

2.6 Overall aim

The aim of this thesis was to develop, validate, and test the algorithm which estimates load rate on the lower limb joints with the accelerometers of wearables. The algorithm was used at the end of this work in the form of a smartphone app, called OApp™. The OApp™ monitors the load rate on the lower limb joints and has the option for its users to score the intensity of their OA-related pain.

In order to develop, validate, and test the algorithm, the following listed aims should be achieved:

1. To assess the validation of load rate estimated with wearables against the gold standard equipment, the force plate, during locomotive activities (walking, jogging, running) on an anti-gravity treadmill.
2. To develop a classifier which distinguishes between activities outside of the laboratory environment while also estimating the load rate associated with these activities.

3. To assess the ability of the technologies (i.e., the smartphone and smartwatch with the developed app) to estimate load rate on the lower limbs during everyday physical activities during a time period of seven days.

4. To assess the ability of the non-interventional app, called OApp™, to monitor load rate and OA-related pain in preparation for use as a tool in clinical studies to explore the relationship between load rate and OA-related pain.

2.7 Summary

In the background section (2.1) the mechanical properties of cartilage and the motivation to monitor load rate were explained. Cartilage mainly has viscoelastic material characteristics, which is prone to high load rate. The literature review was separated into three subsections: Load Rate Monitoring, Clinical Use of Physical Activity Monitoring, and Signal Processing. In Section 2.2 the chronological use of physical activity monitoring methods in clinics were described, were the main idea came to use accelerometry for physical activity monitoring in this thesis. Further, there is no evidence in the literature, to the knowledge of the author, of monitoring load rate with accelerometers, which would help people with OA to identify impact loading on their lower limb joints during everyday physical activity. Here is where the motivation for this thesis is derived. Further, in Section 2.4 the literature review for signal processing was conducted, which let the reader understand the decisions of the author to process the signal that was measured and analysed. In Section 2.5 the rationale for the thesis was described by explaining the lack in research in load rate monitoring with accelerometry. Section 2.6 lists the aims of this thesis, which will be answered later in each study.
Chapter 3

General methodology

Since this thesis is on the life science interface, which here is a collaboration between computational engineering and health science, it is important for the readers with different backgrounds to understand the different terms and methods used. This chapter helps to understand the general methodology of the thesis and further methods used in each study are explained in the corresponding chapters.

3.1 Study designs

To obtain a reliable observation in the following studies, the method had to be carefully chosen depending on the aim. Three major study types exist: quantitative, qualitative, and mixed method (Marczyk et al., 2005). Quantitative methods include all types of design which measure observations (hence the name) and use mathematical or statistical tools for calculating a result which, in turn, leads to an interpretation of the result and, therefore, to a conclusion based on those calculations. Quantitative research can be driven by a hypothesis (see Chapter 4) which has to either be proved or disproved. Not all quantitative studies, though, are driven by hypothesis (see Chapter 6). Furthermore, quantitative research can be experimental or non-experimental (e.g., surveys; Creswell (2013)). Qualitative research, on the other hand, is interested in human behaviour and habits. These are measured by means of verbal data via interviews and/or surveys. Finally, as the name implies, the mixed methods approach is a design type which uses both qualitative and quantitative methods together. This thesis utilised quantitative, rather than qualitative, methods.

There are different research methods which fall under ‘quantitative research’: viz., case control, observational, cohort, longitudinal, and cross-sectional (Marczyk et al., 2005). Case control studies are used to compare treatments or explore the cause of diseases and utilise a group of people with a disease or other phenomena and compares the results.
with a control group (i.e. people without the disease). Medical records or interview data are used to link the disease or phenomena with statistical analysis. The difference to the observational design is that in observational studies the researcher cannot interfere in the experiment with treatment or medication. Moreover, cohort studies investigate a particular group with certain medical conditions or medications over a period of time. Longitudinal (also called prospective) studies, on the other hand, make observations over a period of days/weeks/months/years and allow researchers to explore long-term effects in a human population. Conversely, cross sectional studies make observations at a specific point in time.

Studies 1 and 2 (see Chapters 4 and 5) were quantitative, cross-sectional experimental studies. They were both conducted at one specific point in time and had an experimental character. Studies 3 and Study 4 (see Chapters 6 and 8) were quantitative, observational prospective studies, which obtained data from the participants for 7 days and two weeks, respectively.

3.2 Equipment

The main devices used throughout this thesis were the sensors of wearables (i.e. a smartphone and smartwatch). The equipment used in specific studies will be detailed in later chapters in their corresponding methodology sections.

3.2.1 Microelectro mechanical systems (MEMS)

In this thesis, the wearables’ micro electro-mechanical systems (MEMS) were used as a sensor for activity monitoring for the purpose of collecting acceleration data. MEMS are systems consisting of mechanical sensors. The most important MEMS devices are sensors using piezoresistive, capacitive, and vibration sensing mechanisms and are made up from components between 1 to 100 µm (Bao, 2005). The gross thickness of a MEMS ranges between 0.02mm and 1.0mm (Bao, 2005). The device electronics use integrated circuit, batch processing techniques and micro-mechanical components, which are fabricated from sophisticated forms of silicon or other substances using micro-machining processes (Bao, 2005).

3.2.2 The Sony® Xperia™ Z5 Compact and SmartWatch 3

The Sony® Xperia™ Z5 Compact (Sony, 2016b) (127 × 65 × 8.9mm, 138g) and SmartWatch 3 (Sony, 2016a) (51 × 36 × 10mm, 38g) were used in the studies to monitor the estimated load rate on the lower limb joints accumulated over time. The MEMS accelerometer of the devices were collecting acceleration data. The advantage of the Sony
devices is that they are waterproof. These smartphones and smartwatches are equipped with triaxial accelerometers to capture body movement in three orthogonal directions. The smartphone and smartwatch run on the Android operation system (OS).

### 3.3 Participants

Participants in the pilot study, Study 1, 2 and 3 were healthy people with the inclusion criteria:

- Aged 18 years and above
- Able to give written, informed consent
- Without any lower limb pathologies or any musculoskeletal, neurological, or systemic diseases or other physical disabilities which may have limited their mobility

The pilot study, Study 1 and 2 had further exclusion criteria:

- With respiratory diseases which might cause problems while running.

Participants in the pilot study, Study 1, 2 and 3 were recruited via posters on multiple noticeboards around the University of Southampton. Once a participant showed interest, the researchers sent an email to them with the participant information sheet and an invitation to the study.

The participant in Study 4 was a person with OA in her lower limbs (hip, knee, or ankle). The inclusion criteria were:

- Aged 18 years and above
- Able to give written, informed consent
- Diagnosed with OA in their lower limbs
- Scheduled for a intra-articular corticosteroids injections (CSI) for their joint with OA

The exclusion criteria for Study 4 were:

- Inability to walk without a walking aid
- Neurological or systemic illnesses or other causes of pain than OA
- Contra-indication to the CSI (infection, bleeding diathesis)
3.4 Ethical approval

The pilot study, Study 1, 2 and 3 were approved by the Faculty of Health Science and Engineering and Environment Ethics Committee at the University of Southampton (no.17086 and no.30034, see Appendix B). Study 4 was approved by the Heath Research Authority of the National Heath Service in UK (REC: 16/NI/0228, IRAS ID: 211265, Appendix B). Study 4 was on the National Institute for Health Research portfolio (CPMS ID: 34674). All studies were conducted in compliance with the Helsinki Declaration (June 1964).

3.5 Data collection

For all studies in this thesis the acceleration data of the smartphones and smartwatches were sampled at 50Hz. Since Zhang et al. (2012) found that sampling frequencies over 10Hz were not associated with higher significant accuracies and Bersch et al. (2014) found that sampling frequencies over 10Hz achieve nearly the same classification accuracies and sampling frequencies, with frequencies above 20Hz resulting in only minor gains, it was decided to choose the 50Hz. 50Hz is higher than the suggestions from Zhang et al. (2012) and Bersch et al. (2014), which would just lead to more accuracy. In the pilot study, Study 1 and Study 2 the data were then processed on a computer. For Study 3 and 4 the acceleration data was collected and semi-processed on board of the smartphone and further processed off-line on the computer.

3.6 Data analysis

In each study the acceleration data was transformed into a load rate value, which was discussed in this following section. The following data analysis for each study was different and was discussed in the methodology sections of each separate study. The data were processed using MATLAB (Version R2016b or R2017a, The Math Works, Natick, MA).

3.6.1 Load rate

The load on the joints is derived from Newton’s second law and can be written as

$$ F_L = m \ a, $$

(3.1)
where $F_L$ is the load, which is affected by gravitational forces, $a$ is the acceleration, and $m$ is the body mass. The mass $m$ includes all the mass of the participant, including the mass below the joints. In the studies conducted in this thesis, $m$ is constant due to the minimal change of weight of the participants. As mentioned in Section 2.1 it is assumed that higher load rates on the joints are more damaging than lower load rates. Even if the theory of assuming that cartilage is a viscoelastic material is very simplified, this simplistic theory is sufficient enough for the purpose of this thesis, which is to identify the most impacting forces on the lower limb joints.

The time derivation of acceleration is the physical quantity **jerk**:

$$j = \frac{da}{dt} = \frac{d^2v}{dt^2} = \frac{d^3x}{dt^3} \quad (3.2)$$

where $j$ is the jerk, $v$ is the velocity, $x$ is the position, and $t$ is the time. The **load rate function** with constant $m$ is:

$$\dot{F}_L = \frac{dF_L}{dt} = m \frac{da}{dt} = m \cdot j, \quad (3.3)$$

where $\dot{F}_L$ is the load rate, including the quantity jerk multiplied by $m$. For the data analysis, the differential calculus 3.3 can be simplified with a numerical analysis in order to calculate the load rate magnitude in three spatial directions from the wearable accelerometer’s sensor data:

$$\left| \frac{\Delta \dot{F}_L}{\Delta t} \right| = m \sqrt{\left( \frac{a_x.t_2 - a_x.t_1}{t_2 - t_1} \right)^2 + \left( \frac{a_y.t_2 - a_y.t_1}{t_2 - t_1} \right)^2 + \left( \frac{a_z.t_2 - a_z.t_1}{t_2 - t_1} \right)^2} \quad (3.4)$$

where $a_x$ is the acceleration in $x$ direction, $a_y$ is the acceleration in $y$ direction, $a_z$ is the acceleration in $z$ direction, $n$ number of time stamps. With units of kg m s$^{-3}$= N s$^{-1}$, the load rate magnitude was used as the load rate ($m$=meter, $s$=seconds, N=newton, kg=kilogram). The advantage of the Equation 3.4 is that it considers all three spatial directions, because the direction of the wearable is not always known. Each user might position the phone differently on the body, which means that the direction of the smartphone is not always the same.

### 3.6.2 Concept of uncertainty

Due to noise, the finite size of the data, and the chosen model, which results in that the physical measurements entails a specific degree of uncertainty (Bishop, 2006). Uncertainty can be divided into measurement and model uncertainty (ibid). If $Y$ is the
observed physical measurement and \( X \) are predictors, a general relationship between them both can be written as

\[
Y = f(X) + \epsilon, \tag{3.5}
\]

where \( f \) is an unknown function of \( X \) and \( \epsilon \) is the random error term, which is independent from \( X \) and has a mean of zero (Friedman et al., 2001). In this thesis, the estimation of \( f \) mainly serves for inference. The prediction for Equation 3.7 can be written as

\[
\hat{Y} = \hat{f}(X), \tag{3.6}
\]

where \( \hat{f} \) is the estimation for \( f \), \( \hat{Y} \) is the prediction for \( Y \) and the error term averages to zero (Friedman et al., 2001).

The model uncertainty, which can also be called reducible error (Friedman et al., 2001), can be reduced by potentially improving the accuracy of the estimated model \( \hat{f} \) by using the most appropriate statistical methods for estimating \( f \) (Friedman et al., 2001). This means that, by applying the appropriate statistics, the reducible error can be minimised. Hence, a sufficient estimate of the model uncertainty can be calculated with the methods which were used throughout this thesis, such as validation and resampling methods.

However, even if the estimation \( \hat{Y} \) is exactly the same as the model \( f(X) \) in the form of \( \hat{Y} = f(x) \), the prediction would still have some error (Friedman et al., 2001). This is due to the fact that \( Y \) is also a function of \( \epsilon \). The remaining error is known as the irreducible error, which can also be called the measurement uncertainty (Friedman et al., 2001). This approach can be explained in the following equation:

\[
E(Y - \hat{Y})^2 = E[f(X) + \epsilon - \hat{f}(X)]^2 = (f(X) - \hat{f}(X))^2 + Var(\epsilon), \tag{3.7}
\]

where \( E(Y - \hat{Y})^2 \) is the expected value and \( Var(\epsilon) \) is the variance related with the error term \( \epsilon \). The irreducible error can be due to manufacture variation or technical unreliability, which do not fall under the influence range of the researchers. The aim of this thesis, though, is to minimise the reducible error by using the most appropriate statistics and to provide a prediction for the uncertainty in the model.

Validation and resampling methods were useful in this thesis for the purpose of obtaining a better indication of the uncertainties of the results obtained by each study (Efron and Tibshirani, 1994; Hastie et al., 2001). Resampling means that the training set of the data is repeatedly drawn to refit a model of interest in terms of each sample in order to obtain
additional information about the fitted model. Bootstrapping and cross-validation are commonly used validation and resampling methods, which are explained below.

### 3.6.3 Cross-validation

Building models with a small sample size can lead to biases in favour of the sample, which might not be representative for the entire population. Consequently, the model becomes a wrong estimation and, therefore, useless. Cross-validation helps to validate a model with a small sample size (Stone, 1974). Generally, in order to validate a model, the data set is split into a training set and a test set for the purpose of testing the model with independent new observations (Hastie et al., 2001). The training set helps to build the model and the test set acts as a new observation and is used to validate the model by testing the model’s possible errors in terms of independent data. Splitting a small data set into a test set and a training set means that the test set’s valuable observations are wasted insofar as they are only used for validation. In data sets with big sample sizes, the model would be affected minimally by the loss of some observations for the purpose of validation. For a data set with a small sample size, on the other hand, a few observations can inertly change the model and its accuracy.

Cross-validation repeatedly splits the data set into a training and a test set. At each iteration, the split is performed randomly, and the model is trained with the new training set, then validated with the new test set. The results of every validation iteration are averaged and used as the final validation result. Different cross-validation methods exist: leave-one-out cross-validation (LOOCV), k-fold cross-validation (k-fold CV) and leave-one-subject-out cross-validation (LOSOCV). The LOOCV randomly excludes one point in the data set and uses this point as a test set and the rest of the data as a training set. This is repeated n times and the results of all of the iterations are averaged. Figure 3.1 shows the schematic structure of LOOCV.

An advantage of the LOOCV is that almost all the data set is used for creating the model, which means that the training set contains \(n - 1\) observations. A disadvantage of this method is that it is computationally expensive because \(n\) models have to be built.

The k-fold CV method is a less computationally expensive method, where the original sample is separated into \(k\) equally-sized subsamples. Of the \(k\) subsamples, a single subsample is used as the validation data for testing the model, and the remaining \(k - 1\) subsamples are used as training data. The process is then repeated \(k\) times, with each of the \(k\) subsamples being left out. Figure 3.2 shows the schematic structure of a 5-fold CV.

An advantage of the k-fold CV over the LOOCV is that the k-fold CV is less computationally expensive due to its just building \(k\) models, with \(k\) usually equaling either 5
Chapter 3 General methodology

Figure 3.1: The data set is repeatedly separated into training set (marine blue) and the test set that contains only one observation (light blue). The test result is then estimated by averaging the $n$ results. This figure was inspired by Friedman et al. (2001).

Figure 3.2: The data set is randomly separated into $k = 5$ non-overlapping groups. Each group will be a testing set (light blue), and the remaining points are the training set (marine blue). The test result is then estimated by averaging the five results. This figure was inspired by Friedman et al. (2001).

or 10. This method, however, does have the disadvantage of being less likely to give an accurate estimate of the target metric.

The LOSOCV method is similar to the k-fold CV and separates for each validation iteration all of the data points of one participant. In contrast to the k-fold CV, the test set is not randomly separated; rather, the data set of one single participant is taken out. The LOSOCV method is useful when different participants are involved because it evaluates the model with each participant’s hidden variables being excluded. The LOSOCV shows how the model performs on participants not included in the training set, thereby remaining closer to reality.
3.6.4 Bootstrapping

The bootstrapping method is a resampling method that can be used to quantify the uncertainty of an estimated model by calculating quality-of-fit variables and confidence intervals (Efron and Tibshirani, 1994; Hastie et al., 2001). Similar to cross-validation, it is beneficial to use the bootstrapping method for small sample sizes where the uncertainty of a model is higher. The bootstrapping method creates multiple bootstrapping sample sets with the same sample length as the origin data by drawing, with replacement, random observations from the original data set. The idea behind the bootstrapping method is that it is assumed that the original data set represents the population from which it was drawn. Resampling the original data set is as if many samples from the population were being drawn. The sampling is performed with replacement, which means that the observations can repeatedly occur in the bootstrapping sample set. Assuming $S(Z)$ is any parameter computed from the original data set $Z$, then the mean is followed,

$$\hat{S} = \frac{1}{B} \sum_{b=1}^{B} S(Z^b)$$

where $B$ is the number of the bootstrapping sample sets. $\hat{S}$ serves as an estimate from the original data set. Figure 3.3 illustrates a bootstrapping approach on a small sample with $n = 3$ observations.

The bootstrap distribution will be nearly normal if the number of bootstrap sample sets are large (Hesterberg et al., 2005). The normal bootstrap distribution is due to the central limit theorem, which says that the sample distribution is approximately normal if the number of samples is large (Hesterberg et al., 2005). The normal bootstrap distribution and further characteristics of the bootstrapping method create multiple advantages:

- The normality of the bootstrap distribution statistical allows statistical tests such as t-test or ANOVA, which require normal distributed data sets, to be used to find significance.
- Due to the normality of the bootstrap distribution it is not required that sample sizes are large.
- Confidence intervals can be calculated to estimate the uncertainty in the model and to check the stability of the results.

A disadvantage of the bootstrapping method is that it can be computationally expensive due to it producing a large number of bootstrapping sample sets which serve as data sets. Further, it is important to note that the bootstrapping method does not replace
Chapter 3 General methodology

or add to the original data. The bootstrapping method is used as a way to estimate the variance in a result based on the original data (Hesterberg et al., 2005).

3.7 Summary

Section 3.1 explained the difference of the study designs between the studies. The pilot study, Study 1 and 2 were cross sectional experimental and Study 3 and 4 were exploratory prospective designs. In Section 3.2 first the actual sensor system (MEMS) of wearables were described followed by the wearables that were used throughout the thesis. Specific equipment used in the different studies are described later in the corresponding chapters. The participants of the first four studies were healthy people and in Study 4 the participants was one person with OA. In Section 3.6 the parts of the data analysis are described, which are used throughout the thesis. The main equation (Equation 3.4, load rate magnitude) was derived from Newton’s second law (Section 3.6.1) and is used
in this thesis as surrogate measurement for impact loading. Resampling methods to help
obtain a better indication of the uncertainties of the results of this thesis are described in
Section 3.6.3 and Section 3.6.4. Further data analysis was discussed in specific chapters.
Chapter 4

Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

4.1 Introduction

The first step towards using wearables (smartwatch, smartphone) as a suitable tool for load rate estimations, is to validate the estimates. The load rate data estimated from acceleration data collected by the MEMS accelerometer sensors of wearables are compared to load rate data of the gold standard equipment in the biomechanical laboratory, the force plate. This chapter starts by analysing previous validation studies of the accelerometer to estimate ground reaction forces or other impact loadings on the joints. It follows with a pilot study (Section 4.3), which was conducted before Study 1 to refine the feasibility of Study 1 and follows than with the Study 1.

4.1.1 Previous work in the literature

Commercially available acceleration sensors are commonly used for physical activity monitoring during everyday life (Meyer et al., 2015; Neugebauer et al., 2014). Meyer et al. (2015) assessed the validity of ActiGraph (ActiGraph, 2017) and GENEA (Activinsights, 2017) accelerometers using force plates during different daily physical activities in children. The peak acceleration on the accelerometers and the peak ground reaction force on the force plate were averaged. ANOVA and correlation analyses were conducted to determine the validity of these accelerometers using ground reaction force. Both studies showed a high correlation between the estimated peak ground reaction forces of the
accelerometer and the measured peak ground reaction forces of the force plate. Neugebauer et al. (2014) developed a method for estimating peak vertical and braking ground reaction forces with accelerometers which they then validated against a force plate. The errors that were obtained are for peak vertical ground reaction forces (8.3%) and braking ground reaction forces (17.8%). Errors for the estimated peak ground reaction forces of the accelerometer were 8.3% (in an individual of 70kg, an error of 8.3% would be around 5.81kg which equates to around 56N). However, looking at their height $R^2$ value of 94% it can be said that their estimation of the ground reaction force is an indication of the ground reaction force and not an estimation of the real ground reaction force. More research is suggested before being used in interventions.

Both studies validated the use of accelerometers as tools for estimating peak ground reaction forces on force plates, where both studies’ results had high correlation coefficients values. However, the focus of their validation was the peak ground reaction forces and not the load rate which was used in the studies to estimate impact loading during locomotive activities (Davis et al., 2004; Milner et al., 2006). Even if it seems as if the peak ground reaction force has some relation to impact loading, the present study will adopt the theory that load rates might be a better indicator for impact loading on the lower limb joints (Davis et al., 2004; Milner et al., 2006). While load rate estimates are considered here, there may be other indicators for joint damage such as biomechanics, age, strength, gender, or predisposing conditions (Litwic et al., 2013), which are not included in this thesis.

Other features used for identifying impact loading can be found in elite sports research. Gabbett et al. (2010) used the acceleration rate magnitude with the MinimaxX (Catapult, 2017) for identifying collisions and horizontal impact forces on professional rugby league players during pre-season and in-season skills training sessions. The results showed a high correlation ($r = 0.96, p < 0.01$) between collisions recorded via the acceleration rate magnitude and those recognised from video recordings. The algorithm was suitable for detecting collisions and horizontal impact forces. Hollville et al. (2016) validated the accelerometers using MinimaxX against a force plate by calculating the mean acceleration rate magnitude of the accelerometer and force plate specific to a team sport activity performed on the force plate. The correlation between the accelerometer data and the force plate data were between 0.74 and 0.93. Gabbett et al. (2010) and Hollville et al. (2016) utilised acceleration rate magnitude, which appears to be a suitable method for capturing impact loadings on the lower limb joints.

Wundersitz et al. (2013) assessed the validity of a MinimaxX accelerometer worn on the upper body for estimating peak forces during running and change-of-direction tasks. Peak vertical acceleration and acceleration magnitude values [m s$^{-2}$] were converted to force values [N] via Newton’s second law of motion (i.e. multiplying by the participant’s body mass) and were compared against the peak ground reaction force from the
force plate. Resultant accelerometer measures showed no to strong significant correlations ($r = 0.00 - 0.76$). They showed that accelerometers worn on the upper body could provide a relative measure of peak impact force experienced during running and two change-of-direction tasks (45° and 90°). This approach involved inserting the participant’s body mass in the equation, which includes one of the hidden variables that Gabbett et al. (2010) and Hollville et al. (2016) did not use. Since the accelerometer was attached to the upper body of the individuals, the actual accelerometer measurements came from the upper body where a lighter force was applied. This can be discussed as not being an accurate way of measuring load. Nevertheless, as an estimation, it had a high correlation to the ground reaction force and, hence, might be seen as a valid method for estimating ground reaction force with accelerometers.

Hollville et al. (2016) and Wundersitz et al. (2013) validated two different acceleration values against the force plate data: the mean acceleration rate (jerk) magnitude and the peak force (peak acceleration multiplied by the participant’s body mass). The approach in the present study is a combination of both quantities and means that the accelerometer rate magnitude was multiplied by the participants’ body mass to obtain an estimation of the load rate. Load rate was used during this thesis because in sports science it serves to detect impact loading on the lower limb joints during locomotive activities on force plates (Davis et al., 2004; Milner et al., 2006). The outcome of this literature review was summarised in Table 4.1 and Table 4.2.
# Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

## Objective

To assess the validity of accelerometers using force plates (i.e. ground reaction forces) during performance of different tasks of daily activity.

## Method

- **ActiGraph and GENEA** was worn at the hip. Participants completed activities on the force plate: walking, jogging, running, landings from the boxes of different height, rope skipping, dancing.

## Subject Information

13 healthy children

## Conclusion

Data from both accelerometers correlated with ground reaction force ($r=0.90$ and $0.89)$.

## Strength

- Seven repeats of trials

## Weaknesses

- No information how to maintain speed. Just use of Pearson's correlation coefficient.
- Short sampling period and no habituation of running due to 10 meters of space.

---

## Author and title

Meyer et al. (2015), Validation of two accelerometers to determine mechanical loading of physical activities in children

## Objective

To develop a statistically based model to estimate vertical and peak braking ground reaction forces.

## Method

- ActiGraph was worn on the right hip and participants completed six walking and six running trials.
- Accelerometer data multiplied with body weight of participants.

## Subject Information

39 healthy participants

## Conclusion

Average absolute percentage differences were 8.3% and 17.8%.

## Strength

- Six repeats of trials and consideration of body mass in calculations to estimate ground reaction forces.

## Weaknesses

- Braking ground reaction force might not be very meaningful full next to the peak load. More interest to the derivation from the braking ground reaction force is of complex generalized regression model, which led to low error values.
- Short sampling period no habituation of running due to 10 meters of space.

---

### Table 4.1: Outcome and quality of relevant studies validating acceleration sensors against the golden standard equipment.
<table>
<thead>
<tr>
<th>Author and title</th>
<th>Objective</th>
<th>Method</th>
<th>Subject Information</th>
<th>Conclusion</th>
<th>Strength</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hollville et al. (2016), MinimaxX player load as an index of centre of mass displacements? A validation study.</td>
<td>To assess the acceleration rate magnitude computed by the MinimaxX accelerometers by comparing it to the acceleration rate magnitude computed by a gold standard method based on in series force platforms.</td>
<td>Participants were instrumented with two MinimaxX during specific team sport displacements performed on the force plates.</td>
<td>Fourteen participants</td>
<td>Pearson correlation coefficients were ranged from 0.74 to 0.93 while the coefficients of variation varied from 6.9 to 16.4%. The acceleration rate magnitude parameter computed with MinimaxX accelerometers seems able to characterize the physical demands in team sports</td>
<td>Accelerometer rate magnitude</td>
<td>No repeated measurements, which lead to a small sample size. No habituation of running due to 15 meters of space.</td>
</tr>
<tr>
<td>Wundersitz et al. (2013), Validity of an upper-body-mounted accelerometer to measure peak vertical and resultant force during running and change-of-direction tasks</td>
<td>To assess the validity of a tri-axial accelerometer worn on the upper body to estimate peak forces during running and change-of-direction tasks.</td>
<td>Participants completed four different running and change-of-direction tasks (0°, 45°, 90°, and 180°). Peak craniaudal and resultant acceleration was converted to force and compared against peak force plate ground reaction force.</td>
<td>Seventeen participants</td>
<td>Resultant accelerometer measures showed no to strong significant correlations (r =0.00-0.76). Accelerometers worn on the upper body, can provide a relative measure of peak impact force experienced during running and two change-of-direction tasks.</td>
<td>Multiplied with weight to obtain force</td>
<td>Short sampling period. No habituation of running due to 10 meters of space.</td>
</tr>
</tbody>
</table>

Table 4.2: Outcome and quality of relevant studies validating acceleration sensors against the golden standard equipment.
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

4.1.2 Rationale

Although previous studies using accelerometers for the purpose of estimating ground reaction forces or accelerometer rates showed good correlations to force plate data (Meyer et al., 2015; Neugebauer et al., 2014; Hollville et al., 2016), validation studies assessing the relationship between load rate estimated with wearables and force plates are still necessary. If further research is conducted using the load rate magnitude algorithm, the algorithm and the devices needs to be validated against the gold standard equipment, the force plate. Study 1 serves as a validation study for using commercially available acceleration sensors, such as wearables, as suitable tools for estimating load rate on the lower limbs.

4.2 Aim and objectives

To assess the validation of load rate estimated with wearables against the gold standard equipment, the force plate, during locomotive activities (walking, jogging, running) on an anti-gravity treadmill.

4.2.1 Objectives

1. To explore the correlation between load rate estimated from the wearables against the load rate estimated from the force plate during locomotive activities (walk, jog, run) on an anti-gravity treadmill with different bodyweight percentages (110%, 100%, 90%, 80%, 60%, 30%).

2. To support the results from Objective 1 through a deeper exploration of the relationship between load rate estimated from the wearables against the load rate estimated from the force plate by developing a linear regression model.

3. To compare the wearables data and force plated data within different speed conditions.

4. To compare wearables on different body part positions (between shoulder blades, right hip, left wrist and right wrist).

4.2.2 Hypothesis

There is a positive correlation between load rate data from the force plate and the load rate data from the wearables.
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

4.3 Pilot single case study to refine feasibility for monitoring load rate with wearables

4.3.1 Introduction of pilot study

The feasibility of the data collection of Study 1 was tested in this pilot study. To measure how participants perform locomotive activities with different weight-bearing loading conditions, they were asked to walk and jog on an anti-gravity treadmill. Since the authors had never conducted studies with anti-gravity treadmill before, the feasibility of collecting data had to be tested. Furthermore, the relation between the estimated load rate and the body-weights of the participants had to be explored.

4.3.2 Aim of pilot study

This pilot study was aimed to explore the ability of smartphones to estimate load rate through the lower limb joints with an accelerometer while participants walked or ran with different percentages of body weight and varying velocities on an anti-gravity treadmill.

4.3.3 Methods of pilot study

An adult male wore a vest with a smartphone placed between his shoulder blades. The smartphone collected acceleration data during one-minute periods while the participant walked and jogged (5 km h\(^{-1}\) and 8 km h\(^{-1}\) and body weight percentages (20\%, 50\%, 80\%, 90\%, 100\%, in random order) on an anti-gravity treadmill. The anti-gravity treadmill Alter-G is able to lift the user up, hence, reducing the weight-bearing loading on the lower limb joints. The weight of the participants was adjusted with a bodyweight percentage during the study. The data from the middle 20 seconds of each test period were used in the analysis. The load rate magnitude (Equation 3.4) was calculated at a sampling rate of 50 Hz from the raw acceleration data collected by the smartphone during the study. Pearson’s correlation coefficient was calculated between the load rate estimated with the smartphone and the adjusted bodyweight percentage on the anti-gravity treadmill to reveal the relationship between these two variables.

4.3.4 Results of pilot study

The vertical load rate magnitude data were highly correlated with percentage body-weight (for 5 km h\(^{-1}\) \(r = 0.96\), for 8 km h\(^{-1}\) \(r = 0.99\), for 12 km h\(^{-1}\) \(r = 0.99\)).
4.3.5 Discussion and conclusion of pilot study

All three speed conditions showed high correlations between load rate magnitude estimated with the smartphone and the bodyweight adjusted during the trail. The correlation between the measured vertical load rate and the adjusted bodyweight percentage shows that it is feasible to use the smartphone for Study 1. Furthermore, the results illustrated that the study’s design is feasible. Feasibility, in this context, means that the study’s procedure was physically achievable for both the participant and the researcher and that the smartphone would be able to detect differences in load rate regarding different body percentages. The results of this pilot study led the researchers to define Hypothesis 4.2.2. The pilot study was presented as a poster at the Physical Activity and Osteoarthritis conference at Loughborough University on the 17 December 2015.

It is feasible to collect data using a smartphone in a training facility. The findings showed that the calculated load rate from the acceleration data via a smartphone correlated with bodyweight percentage. The smartphone can therefore potentially be used to distinguish between different load conditions. Hence, Study 1 is feasible.

4.4 Methods

Recordings were made of acceleration values with the MEMS sensors of wearables. These were to be worn by participants while performing locomotive activities on an anti-gravity treadmill. The choice to use the anti-gravity treadmill during the present study was justified because the system was able to vary the load that the participants experienced through their joints. This enabled the collection more data points for varying speeds. Another advantage of using the anti-gravity treadmill was that it has an integrated force plate.

4.4.1 Study design

The study design was cross-sectional and experimental as explained in Section 3.1. The study compares load rate estimated with the accelerometer data of wearables to the load rate data estimated with the force plate during locomotive activities at different body weight percentages and speeds on an anti-gravity treadmill.

4.4.2 Participants and recruitment

Twelve healthy adults (female n=4, male n=8; aged 26 ± 3 years; height: 175 ± 15 cm; body mass: 71 ± 9 kg; means± standard deviation) participated in the study. Participants were recruited via posters on multiple noticeboards around the University
of Southampton. Once a participant showed interest, the researchers sent an email to them with the participant information sheet and an invitation to the study. Based on the screening, which excluded those with lower limb pathologies or any musculoskeletal, neurological, or systemic diseases or other physical disabilities which may have limited their mobility, 12 of 18 volunteers accepted the invitation. Data collection took place at Southampton Football Club’s training facilities. The sample of convenience of 12 participants was chosen due to limited time and access to the facility. Each participant completed 18 different trials, making a total of 216 data points.

4.4.3 Ethical approval

The study was approved by the Faculty of Health Science Ethics Committee at the University of Southampton (no.17086, see Appendix B). Each session took approximately 45 minutes, including briefing, obtaining informed written consent (Participant information sheet and Consent form: Appendix E), and the data collection was done while the participants were walking and running on an anti-gravity treadmill.

4.4.4 Equipment

Study 1 utilised two smartphones, two smartwatches, an anti-gravity treadmill, and a data acquisition device.

**Smartphones and smartwatches** All participants were asked to put on an elastic sports vest holding Smartphone 1: between the shoulder blades (SP 1) which was positioned in such a way as to have it located between their shoulder blades. This location aligns with elite sports practice (Barrett et al., 2014; Gabbett et al., 2010) where athletes wear accelerometers between their shoulder blades. Smartwatch 1: right wrist (SW 1), on the other hand, was placed on the right wrist. To assess the effect of the position of the wearables, data from another set of wearables were collected from 6 participants from the original 12 participants: SW 2 was attached to the lateral upper right leg with a cohesive tape, and SP 2 was placed on the left wrist. Only the data of the 6 participants were available due to technical limitations. The lateral upper right leg was chosen to represent the usual position on the body of the smartphone: the hip pocket. The brand of the smartphones used was Sony® Xperia® Z Compact, and the brand of the smartwatches was Moto 360 from Motorola®. The acceleration data was stored on the phone and was transmitted after the study on the laptop, where the data analysis was conducted.

**Alter-G Anti-Gravity Treadmill** The anti-gravity treadmill M320 from Alter-G® (Figure 4.1) enables the body weight of the user to be reduced per percentage body
weight; that means it reduces the ground reaction force. The anti-gravity treadmill comes with customised neoprene compression shorts that ensure an airtight seal in the enclosure. The body weight is adjustable between 20% and 100%. A force plate is located beneath the treadmill. It measures the body weight so the computer can determine how much positive-pressure is required to create a lifting force to lower the body weight while walking or running on the treadmill. Additionally, it can be used as a force plate to validate the wearable data.

![Figure 4.1: Alter-G® anti-gravity treadmill (Photo: Willem Eerland)](image)

**Data Acquisition Device** The National Instruments® Multifunction M series for USB-6211™ (National-Instruments, 2016b) is a high-performance data acquisition device that has analogue inputs, and is easily portable via USB ports on laptops. The ground of the anti-gravity treadmill consists of four load cells which serve as a force plate. The ground reaction force signals on the four load cells were collected with a sampling frequency of 128Hz with four analogue inputs on the NI DAQ USB™ device which were connected to four output pins at the pressure control board of the treadmill to collect voltage signals from 0-5V. The ported signal was collected with the LabVIEW™ (National-Instruments, 2016a) software with the help of the data acquisition assistant. 128Hz was the sampling frequency of the data force plate which was set by the Alter-G Anti-Gravity Treadmill. Before the data collection, the force plate was calibrated with 25 weights between 0 and 90kg. The weights which were used were weighed on a digital milligram scale and then placed in the middle of the force plate. The voltage signal for each weight was used for building a linear function ($R^2 = 0.99971$), which transformed voltage signal into a mechanical signal. The force plate data was stored directly on a laptop.
4.4.5 Procedure

Before the participants performed the locomotive activities on the treadmill they were asked to carry 10% of their body weight in a weight vest to allow >100% gravity to be tested. Each session consisted of 3 active phases followed by two resting periods. The three active phases were walking (5 km h\(^{-1}\)), jogging (8 km h\(^{-1}\)) and running (12 km h\(^{-1}\)) at six body weight percentages (30%, 60%, 80%, 90%, 100%, 110%, the order of which was randomised), giving a total of 18 trials per participant (Figure 4.2). These weight adjustments were chosen to obtain a broad range of loading conditions, considering the limited time available at the facility and the need to minimise fatigue. The upper percentage range was of more interest, hence, more percentage values in the higher range.

The speed conditions 5 km h\(^{-1}\), 8 km h\(^{-1}\) and 12 km h\(^{-1}\), were chosen to obtain a broad locomotive range from walking, to jogging and then to running. Each trial lasted 90 seconds, with the smartphone, smartwatch, and force plate data being collected simultaneously. The first 20 seconds of recording served as a period of habituation and were discarded before the data were processed. The next 60 seconds were used for data processing, while the last 10 seconds of each trial were discarded to avoid recording possible behaviour changes associated with the trial ending. Using the mean of the anti-gravity treadmill and wearables values would help to align the values.

4.4.6 Data analysis

The data acquisition, data preprocessing, feature extraction, and validation methods which were utilised in all three studies are explained in Section 3.6. Further data analysis methods which were explicitly adhered to in Study 1 are explained in this subsection.

The data were processed using MATLAB (Version R2016b, The Math Works, Natick, MA). The raw acceleration data was transformed into a mean load rate magnitude (Equation 3.4). No filtering was conducted since the values of interest were peak events and a filter might smoothen the signal which might be of interested (Meyer et al., 2015). An example of the calculated load rate magnitude of four gait cycles of Participant 4.
during running are depicted in Figure 4.3. To calculated the mean load rate magnitude, the load rate magnitude was averaged over the time.

![Figure 4.3: Four gait cycles of load rate magnitude of Participant 4 during running on the anti-gravity treadmill.](image)

To achieve Objective 1, Pearson’s correlation coefficient (Mukaka, 2012) was calculated from the whole data set with 186 data points (18 data points for each of the 12 participants with some data points excluded as explained in Section 4.5) of the load rate data from the wearables and the load rate data from the force plate for the purpose of testing the hypothesis of Study 1. Due to uncertainties, which might occur due to weak study design (e.g. small sample size or not considerations of confounders), it might be useful to support the results in the first objective, with further analysis of the relationship between wearables data and force plate data.

For this reason, Objective 2 followed, which involved a much more thorough analysis to explore the relationship between wearable data and force plate data. Three linear methods were developed: Model 1 with just fixed effects, Model 2 with fixed effects and a random intercept and Model 3 fixed effects and random slope and intercept. Model 1 was developed to obtain a more general model. Model 2 and Model 3 were developed to obtain participant specific models. To increase the uncertainty in the models, bootstrapping and cross-validation methods were used. The purpose of building three models was to explore the relationship between wearables data and force plate data and to obtain the simplest model with the highest quality-of-fit. The quality-of-fit parameters, which were obtained from the bootstrapping vectors for each model, were compared in a one-way analysis of variance (ANOVA) (Fisher, 1919). The ANOVA determined if there was a significant difference between the mean of the quality-of-fit variable ($R^2$ and root mean squared error ratio (RMSER), explained below) of the models followed by a pairwise comparison to see which model was the best.
Objective 3 was achieved by developing Model 1, Model 2 and Model 3 for each speed (5km h$^{-1}$, 8km h$^{-1}$, 12km h$^{-1}$). Including the speed condition had the purpose to make the results comparable to the results of Neugebauer et al. (2014).

Objective 4 was achieved by building Model 1, Model 2 and Model 3 for the four different devices. ANOVA was used to determine significant differences between the means of the $R^2$ values of the four devices followed by a pairwise comparison to analyse the differences between the devices.

**Linear mixed regression model** A linear model analysis was used to explore the relationship between the mean load rate magnitude estimate, hereafter referred to as load rate, from the wearables and the force plate. The linear mixed regression model is a statistical model that contains fixed and random effects and is typically used for repeated measurement in the same statistical unit (Henderson et al., 1959). The idea behind the linear mixed regression model is that the data show a linear trend across subjects while the intercept and slope can be different (Pinheiro, 2005).

In general, the linear mixed regression model, is a combination of two approaches: (1) estimation of the overall trend and (2) separate model for each subject to account for between-subject differences. The first approach assumes the populations’ average model, with response variable $y$ defined as

$$ y = \alpha + \mathbf{x} \beta + \epsilon $$

(4.1)

where $\mathbf{x}$ is an multi-dimensional vector of the observed responses and $\alpha$ and $\beta$ are vectors of the fixed effects that denote the intercept and slope of the predictor variables. The within-subject errors $\epsilon$ is assumed to be distributed as $\mathcal{N}(0, \sigma^2)$.

The second approach uses subject-specific coefficients $a_i$ and $b_i$ to consider differences between subjects:

$$ y_{m,i} = a_i + \mathbf{x}_i b_i + \epsilon_i $$

(4.2)

where $y_{m,i}$ corresponds to data for observation $m$ and group $i$, $a_i$ and $b_i$ vectors represent each participant’s random intercept and slope deviations from the corresponding predictor variable parameter. These approaches would lead to incorrect standard errors and neglect of trends observed across subjects (Pinheiro, 2005). The linear mixed regression model includes both the average model and the subject-specific model as
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

\[ y_{m,i} = \alpha + x_i \beta + a_i + b_i x_{m,i} + \epsilon_i . \]  
(4.3)

The \( a_i \) and \( b_i \) are assumed to be independent with distribution \( \mathcal{N}(0, \Sigma) \). The parameter \( \Sigma \) indicates the variance in the population distribution and, therefore, the degree of heterogeneity in the subjects. The within-subject errors \( \epsilon_i \) are assumed to be to follow a normal distribution \( \mathcal{N}(0, \Lambda_i) \), independent of \( a_i \) and \( b_i \). Hence, the normality of the distribution of the residual was tested with the Kolmogorov-Smirnov test.

**Model Formalisation** In this study, a linear mixed regression model was chosen due to the existence of hidden variables which were not measured while collecting the data, such as anatomy, muscle strength, and the style of gait of the individuals. In the case of the present study, the load rate data from the force plate was the response variable, the data from the wearables, the predictor variables, and the participants was the grouping variable. The data were used to build three different linear regression models to test the hypothesis in 4.2.2: Model 1 consists of only fixed effects (M1) and is detailed as follows:

\[ y_{M1}^{M1} = \alpha_{\text{Wear}} + \beta_{\text{Wear}} x_{m,i}, \]  
(4.4)

which only considers the population’s average behaviour and ignores the between-subject variation in ambulatory activities. Model 2: linear mixed model with random intercept (M2), with fixed effects and a random intercept, is as follows:

\[ y_{M2}^{M2} = \alpha_{\text{Wear}} + \beta_{\text{Wear}} x_{m,i} + a_i , \]  
(4.5)

which assumes that the between-subject variation just depends on the random intercept. Model 3: linear mixed model with random intercept and slope (M3), with fixed effects and a random slope and intercept, is as follows:

\[ y_{M3}^{M3} = \alpha_{\text{Wear}} + \beta_{\text{Wear}} x_{m,i} + a_i + b_i x_{m,i} , \]  
(4.6)

which considers the population’s average behaviour and the between-subject variation in ambulatory activities. The load rate magnitude (Equation 3.4) from the force plate (the response variable) is \( y_{m,i} \) with observation \( m \) and participant \( i \). \( \alpha_{\text{Wear}} \) and \( \beta_{\text{Wear}} \) are the intercept and slope of the estimated load rate of the wearables (fixed effect predictor variables), and \( a_i \) and \( b_i \) are the intercept and slope of each participant (random effect predictor variables).
To obtain a better indication of uncertainty in the models, the bootstrapping method was used (Section 3.6.4). One thousand bootstrap vectors were created and cross-validated (Section 3.6.3). For every vector three models were built, Model 1, Model 2 and Model 3, and the quality-of-fit variable (see next paragraph) was calculated and then averaged, with confidence intervals calculated based on the 1,000 bootstrap samples.

**Measuring the Quality of Fit**  
For each of the 1000 bootstrap vectors, and each model (Model 1, Model 2 and Model 3), the $R^2$ value and root mean squared error ratio (RMSER) were calculated and then averaged. The RMSER is a way of measuring how well the prediction of the models match the observations with relation to the mean observed value. The RMSER estimation with the bootstrap method was as follows:

$$RMSER = \frac{1}{B} \sum_{b=1}^{B} \left( \frac{1}{n} \sum_{m=1}^{n} \left( \frac{\hat{Y}_m - Y_m}{\bar{Y}} \right)^2 \right)$$  \hspace{1cm} (4.7)$$

where $\hat{Y}_m$ is a prediction, $Y_m$ is an observation corresponding to the input of the function which generated the predictions, and $B$ is the number of bootstrapped samples. $RMSER$ serves as an estimate from the original data set, where the model out of Model 1, Model 2 and Model 3 with the smallest $RMSER$ was of interest. The root mean squared error is not a ratio, but normalising it with the mean observation value of the training set makes it proportional (i.e. between 0 and 1) and easier to interpret. In comparison to the mean absolute error, the RMSER has the advantage that it squares the errors before they are averaged, which gives a relatively high weight to a large error.

The $R^2$ helps to quantify the extent to which the model fits the data. The $R^2$ is the ratio between the *residual sum of squares* (RSS) and the *total sum of squares* (TSS), which makes it proportional and takes values between 0 and 1. The estimated $R^2$ from the original data set was defined as

$$\hat{R}^2 = \frac{1}{B} \sum_{b=1}^{B} \left( 1 - \frac{\sum_{m=1}^{n} (Y_m - \hat{Y}_m)^2}{\sum_{m=1}^{n} (Y_i - \bar{Y})^2} \right)_b$$

$$= \frac{1}{B} \sum_{b=1}^{B} \left( 1 - \frac{RSS}{TSS} \right)_b = \frac{1}{B} \sum_{b=1}^{B} R^2_b$$  \hspace{1cm} (4.8)$$

where $R^2$ is the value of $B$ bootstrap sample sets. Disadvantages of both parameters are that they can lead to overfitting which, in turn, leads to the wrong choice of model (Hastie et al., 2001).
Both $R^2$ and RMSER were considered in the present study in order to consolidate our results and make them more comparable to previous work such as the work of Neugebauer et al. (2014). Algorithm 4.1 shows the pseudocode containing the bootstrapping and cross-validation used to obtain $R^2$ and RMSER.

```
devices := ['device_1','device_2','device_3','device_4']

for each device in devices 1 to 4
    boot_sets := bootstrap(data,count=1000)
    for each boot_set in boot_sets 1 to 1000
        for each cross validation 1 to 1000
            ex := rand(1,length(boot_set)) % random index between 1 and
            the length of the original data set
            ex_point = boot_set[ex] % apoint index to value in bootstrap
            data set
            train_set = boot_set[0:ex-1] + boot_set[ex+1:length(boot_set)]
            % exclude randomly chosen data point

            model1 := linear_reg(train_set)
            model2 := linear_reg(train_set,rand_intercept=True)
            model3 := linear_reg(train_set,rand_intercept=True,
                         rand_slope=True)

            model1_R2 := model1_R2 + calc_R2(model1)
            model2_R2 := model2_R2 + calc_R2(model2)
            model3_R2 := model3_R2 + calc_R2(model3)

            model1_MSE := model1_MSE + calc_MSE(model1)
            model2_MSE := model2_MSE + calc_MSE(model2)
            model3_MSE := model3_MSE + calc_MSE(model3)
        
        mean_model1_R2 := mean_model1_R2 + calc_mean(model1_R2)
        mean_model2_R2 := mean_model2_R2 + calc_mean(model2_R2)
        mean_model3_R2 := mean_model3_R2 + calc_mean(model3_R2)

        mean_model1_MSE := mean_model1_MSE + calc_mean(model1_MSE)
        mean_model2_MSE := mean_model2_MSE + calc_mean(model2_MSE)
        mean_model3_MSE := mean_model3_MSE + calc_mean(model3_MSE)
    
    Mean_model1_R2 := Mean_model1_R2 + calc_mean(mean_model1_R2)
    Mean_model2_R2 := Mean_model2_R2 + calc_mean(mean_model2_R2)
```
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

Mean_model3_R2 := Mean_model3_R2 + calc_mean(mean_model3_R2)

Mean_model1_MSE := Mean_model1_MSE + calc_mean(mean_model1_MSE)

Mean_model2_MSE := Mean_model2_MSE + calc_mean(mean_model2_MSE)

Mean_model3_MSE := Mean_model3_MSE + calc_mean(mean_model3_MSE)

end

Algorithm 4.1: Pseudocode for the data analysis of Study 1

The one-way ANOVA was used to determine if there was a significant difference between the mean of the bootstrapped $R^2$ and RMSER values of Model 1, Model 2 and Model 3 followed by a pairwise comparison with the Bonferroni correction (Bonferroni, 1936). The $R^2$ and RMSER values of each model were normally distributed ($p > 0.15$). One-way ANOVA with the Bonferroni correction was used to compare the $R^2$ values of the four different devices. The Bonferroni correction was used to include the effect of comparing multiple groups. Hence, the desired $p$-value has to be divided by the number of comparisons being conducted ($\alpha/n$, where $n$ is the number of comparisons being conducted). For the comparison of the models, a value of $\alpha = 0.05/4 = 0.0125$ was used for significance. For comparing the devices with each other, the value $\alpha = 0.05/6 = 0.0083$ was used for significance.

4.5 Results

Participants 1 to 11 completed all percentage bodyweight trials at the three speeds mentioned above. Participant 12, however, was only able to complete the 5km h$^{-1}$ and 8km h$^{-1}$ trials, and was not able to complete the trial for the 12km h$^{-1}$ speed due to time restrictions. Furthermore, the complete data from participant 1 and the 5km h$^{-1}$ data from participant 2 were not usable due to technical issues. Therefore, a total of 186 trials were analysed.

Pearson’s correlation coefficient revealed high correlation between the load rate data from the wearables and the load rate data from the force plate with $r_{\text{smarphone}} = 0.7810$ ($p < 0.0001$) and $r_{\text{smarwatch}} = 0.7765$ ($p < 0.0001$). The use of Pearson’s correlation was justified by the fact that the mean load rate data could be assumed to be normally distributed because Kolmogorov-Smirnov test failed to reject the null hypothesis that the load rate data comes from a standard normal distribution ($p = 0.26, \alpha = 0.05$). The linear relationship between load rates from the wearables and the load rates from the force plate can also be seen in Figures 4.4 and 4.5. The left plots show all data points of SP 1 and SW 1 and the right plots include the regression.

The residuals were normally distributed (Kolmogorov-Smirnov, for all $p > 0.15$). The residuals of models Model 1, Model 2 and Model 3 are shown respectively in Figures
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

Figure 4.4: Left: Whole dataset of all participants for Smartphone 1; right: same data set with a linear regression line.

Figure 4.5: Left: Whole dataset of all participants for Smartwatch 1; right: same data set with a linear regression line.

Figure 4.6: Left: Residuals vs the fitted values. Right: fitted response vs. observed response in Model 1.
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

4.6(a), 4.7(a) and 4.8(a). Figures 4.6(b), 4.7(b) and 4.8(b) show the fitted response values to the observed response values of those same models.

For both the $R^2$ and RMSER values, 95% confidence intervals were calculated using MATLAB (Table 4.3). The $R^2$ values of the three models for SP 1 are $R^2_{M1} = 0.62 \pm 0.11$, $R^2_{M2} = 0.72 \pm 0.13$, $R^2_{M3} = 0.77 \pm 0.09$. A linear relationship exists for all models between wearables and the force plate (Table 4.3).

The one-way ANOVA showed that the three models had a significant difference as can be seen in Figure 4.9(b) for the $R^2$ values ($p < 0.0001$) and in Figure 4.10(b) for the RMSER values ($p < 0.0001$). The pairwise comparison showed which model is the best fit (see Table 4.4).
Table 4.3: The $R^2_{\text{Model}}$ and RMSER values for all participants using all of the smartphone and smartwatch data which was collected.

<table>
<thead>
<tr>
<th>Device</th>
<th>M1 lower CI</th>
<th>M1 upper CI</th>
<th>M2 lower CI</th>
<th>M2 upper CI</th>
<th>M3 lower CI</th>
<th>M3 upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1 - $R^2$</td>
<td>0.6166</td>
<td>0.4984</td>
<td>0.7279</td>
<td>0.7192</td>
<td>0.5889</td>
<td>0.8211</td>
</tr>
<tr>
<td>SP1 - RMSER</td>
<td>0.0145</td>
<td>0.0123</td>
<td>0.0166</td>
<td>0.0124</td>
<td>0.0099</td>
<td>0.0150</td>
</tr>
<tr>
<td>SW1 - $R^2$</td>
<td>0.6078</td>
<td>0.5040</td>
<td>0.7064</td>
<td>0.6658</td>
<td>0.5240</td>
<td>0.7745</td>
</tr>
<tr>
<td>SW1 - RMSER</td>
<td>0.0140</td>
<td>0.0121</td>
<td>0.0158</td>
<td>0.0129</td>
<td>0.0107</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

Figure 4.9: The results from the one-way ANOVA test, comparing the 1.000 $R^2$ from Model 1, Model 2 and Model 3.

Figure 4.10: The results from the one-way ANOVA test, comparing the 1.000 RMSER from Model 1, Model 2 and Model 3.

The performances of the models for the three different speed conditions are for Model 3 from SP 1 $R^2_{5km/h} = 0.93 \pm 0.07$, $R^2_{8km/h} = 0.88 \pm 0.11$, and $R^2_{12km/h} = 0.92 \pm 0.06$, (see
Table 4.4: Pairwise comparison with the Bonferroni correction. All differences in the models were significant ($p < 0.0001, \alpha = 0.05/4 = 0.0125$).

<table>
<thead>
<tr>
<th>Models</th>
<th>M1-Upper CV</th>
<th>M1-Lower CV</th>
<th>M2-Upper CV</th>
<th>M2-Lower CV</th>
<th>M3-Upper CV</th>
<th>M3-Lower CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP-$R^2$</td>
<td>0.103</td>
<td>0.097</td>
<td>0.108</td>
<td>0.149</td>
<td>0.143</td>
<td>0.155</td>
</tr>
<tr>
<td>SP-RMSER</td>
<td>-0.0021</td>
<td>-0.0022</td>
<td>-0.0020</td>
<td>-0.0032</td>
<td>-0.0033</td>
<td>-0.0031</td>
</tr>
<tr>
<td>SW-$R^2$</td>
<td>0.058</td>
<td>0.052</td>
<td>0.064</td>
<td>0.118</td>
<td>0.112</td>
<td>0.124</td>
</tr>
<tr>
<td>SW-RMSER</td>
<td>-0.0011</td>
<td>-0.0012</td>
<td>-0.0010</td>
<td>-0.0023</td>
<td>-0.0024</td>
<td>-0.0022</td>
</tr>
</tbody>
</table>

Table 4.5: The $R^2_{speed}$ values for each speed for the smartphone between the shoulder blades.

<table>
<thead>
<tr>
<th>Speed</th>
<th>M1 lower CI</th>
<th>M1 upper CI</th>
<th>M2 lower CI</th>
<th>M2 upper CI</th>
<th>M3 lower CI</th>
<th>M3 upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>5km h$^{-1}$</td>
<td>0.5450</td>
<td>0.3470</td>
<td>0.7181</td>
<td>0.8878</td>
<td>0.8878</td>
<td>0.9350</td>
</tr>
<tr>
<td>8km h$^{-1}$</td>
<td>0.4584</td>
<td>0.2811</td>
<td>0.6317</td>
<td>0.8057</td>
<td>0.6812</td>
<td>0.8993</td>
</tr>
<tr>
<td>12km h$^{-1}$</td>
<td>0.3397</td>
<td>0.1109</td>
<td>0.5830</td>
<td>0.6275</td>
<td>0.3361</td>
<td>0.8546</td>
</tr>
</tbody>
</table>

Table 4.5).

When compared to one another, the $R^2$ values of the four devices showed the following results (all for Model 3): $R^2_{SP1} = 0.83 \pm 0.08$, $R^2_{SP2} = 0.83 \pm 0.10$, $R^2_{SW1} = 0.81 \pm 0.09$, $R^2_{SW2} = 0.83 \pm 0.10$ (Table 4.6). The $R_2$ values of each model and device were normally distributed (Kolmogorov-Smirnov, $p_{5km/h} = 0.89$, $p_{8km/h} = 0.28$, $p_{12km/h} = 0.81$). A one-way ANOVA with pairwise comparison with the Bonferroni correction led to results in Table 4.7. The data for the present study are available on Github (Nazirizadeh et al., 2017).

4.6 Discussion

The present findings show $R^2$-values between 0.34 – 0.93 for force plate and wearable estimates of load rate data while the participants performed locomotive activities on an anti-gravity treadmill. In this section, the different models (Model 1, Model 2, Model
Chapter 4 Study 1: Validity of load rate estimates using accelerometers during physical activity on an anti-gravity treadmill

Table 4.6: The $R^2_{\text{device}}$ values are compared with the different positions which those devices were placed on the body from 6 of the participants

<table>
<thead>
<tr>
<th>Device</th>
<th>M1 lower CI</th>
<th>M1 upper CI</th>
<th>M2 lower CI</th>
<th>M2 upper CI</th>
<th>M3 lower CI</th>
<th>M3 upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1- $R^2$</td>
<td>0.6729 0.5402 0.7932</td>
<td>0.7958 0.6893 0.8791</td>
<td>0.8336 0.7474 0.9044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW1- $R^2$</td>
<td>0.6624 0.5506 0.7642</td>
<td>0.7407 0.6192 0.8286</td>
<td>0.8108 0.7104 0.8934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP2- $R^2$</td>
<td>0.7620 0.6401 0.8597</td>
<td>0.7999 0.6736 0.8898</td>
<td>0.8286 0.7358 0.9064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW2- $R^2$</td>
<td>0.7092 0.6012 0.8083</td>
<td>0.7870 0.6801 0.8706</td>
<td>0.8279 0.7223 0.9054</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: Pairwise comparison of four devices with the Bonferroni correction ($\alpha = 0.05/6 = 0.0083$).

<table>
<thead>
<tr>
<th>Devices</th>
<th>Mean difference of $R^2$</th>
<th>upper CI</th>
<th>lower CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1-SW1</td>
<td>0.0227***</td>
<td>0.0175</td>
<td>0.0279</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>SP1-SP2</td>
<td>0.0049</td>
<td>-0.0002</td>
<td>0.0102</td>
<td>0.0714</td>
</tr>
<tr>
<td>SP1-SW2</td>
<td>0.0057**</td>
<td>0.0005</td>
<td>0.0109</td>
<td>0.0249</td>
</tr>
<tr>
<td>SW1-SP2</td>
<td>-0.0177***</td>
<td>-0.0230</td>
<td>-0.0125</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>SW1-SW2</td>
<td>-0.0170***</td>
<td>-0.0223</td>
<td>-0.0118</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>SP2-SW2</td>
<td>0.0007</td>
<td>-0.0045</td>
<td>0.0059</td>
<td>1</td>
</tr>
</tbody>
</table>

*** $p << 0.0001$, ** $p < 0.05$, * $p < 0.1$

3), the models with different speed conditions (5km h$^{-1}$, 8km h$^{-1}$, 12km h$^{-1}$), and the difference between the wearables on different body parts were discussed.

A linear relationship was recognised as being present between the SP 1 and SW 1 data and the force plate data for Model 3 (Figures 4.4 and 4.5). Differences are visible between the three different speed conditions when looking at the right plots. In Figures 4.4 and 4.5, on the right, the linear regression lines for walking have a higher slope than the slopes of the jogging or running data. Looking in detail at the slope of the walking data, the wearables seem to underestimate the load rate in comparison to the force plate. This indicates that wearables might slightly underestimate the load rate for low intensity activities. Underestimation means that the predictive model, which was used would estimate a lower value for the estimated load rate and. For jogging and running, however, it seems that the wearables mostly overestimated the load rate data
in comparison to the force plate data. Indicating that wearables might overestimate the load rate for moderate to vigorous activities.

**Correlation between wearables and force plate (Objective 1)** The high correlation found between load rate data from the wearables and the force plate can be described as a “high positive correlation” according to Hinkle et al. (2003), from their ‘Rule of Thumbs for Interpreting the Size of a Correlation Coefficient’. To support these results and gain further knowledge of the uncertainty in the validation of wearables as suitable tools to estimate load rate on the lower limbs, the linear mixed regression model analysis was conducted.

**Linear regression model (Objective 2)** The plots of the residuals of Model 1, Model 2 and Model 3 support the residuals’ normality of distribution (Figures 4.6(a), 4.7(a) and 4.8(a)). The normality of the data allowed the linear mixed regression model to be utilised in this study, as the condition of using the linear mixed regression model is to have a normally distributed error term. The data points in Figure 4.6(b) are more widely distributed than in Figures 4.7(b) and 4.8(b). This shows that Model 1 had the lowest $R^2$ and the highest RMSER. This was confirmed in the data analysis (Table 4.3).

To assess the validation of load rate estimated with wearables against the force plate during locomotive activities two linear mixed regression models and a linear regression model were developed. A one-way ANOVA showed that all models were significantly different from each other ($p < 0.0001$), which can be seen in Figure 4.9(b) and Figure 4.10(b). The pairwise comparison helped to identify the best model (Table 4.4). The difference between Model 1 and Model 3 were the highest with $\Delta R^2_{M3,M1} = 0.149$, $\Delta RMSER_{M3,M1} = -0.0032$. Hence, knowing that Model 3 had the highest $R^2$ and
lowest RMSER values would lead to the decision that Model 3 ($R^2_{M3} = 0.77 \pm 0.09$) is the best performing model. Model 3 included, in comparison to Model 1, random slope and intercept effects, which takes into account unknown participant-specific characteristics, such as muscle structure, skeletal structure, or participant height, all of which are hidden variables for the model. To examine a simpler model, the random slope of Model 3 was excluded: i.e., Model 2 with a fixed effect and a random intercept, which led to a lower $R^2_{M2} = 0.72 \pm 0.13$. Therefore, Model 2 implies that different participants did, indeed, have hidden variables which, in turn, influenced the slope and intercept of the function. Nevertheless, the improvement of Model 3 over Model 2 was small, with $\Delta R^2_{M3,M2} = 0.046$.

It was essential to consider Model 1 ($R^2_{M1} = 0.62$), with just fixed effects, to be able to develop a general model, which was the baseline model. Adding random slope and intercept effects creates a more accurate model but with the disadvantage of being a less generalisable model. Neugebauer et al. (2014) also created linear mixed models for their analysis, which were in comparison to the models in the present study more complex. They considered these variables as predictor variables: acceleration data, the mass of participants, type of activity (walk=0, run=1), and interaction between acceleration data and type of activity. This complex model yielded a small absolute error value of 8.3%, where the type of activity had the most significance in the model. This led to the decision to conduct further analysis considering the speed condition (5 km h$^{-1}$, 8 km h$^{-1}$ and 12 km h$^{-1}$) to be able to compare the model from Neugebauer et al. (2014) with Model 3 in Study 1.

**Different speed conditions (Objective 3)** When comparing the different speed conditions recorded with SP 1, it can be seen that the $R^2$ values do not vary substantially ($R^2_{5km/h} = 0.93 \pm 0.07$; $R^2_{8km/h} = 0.88 \pm 0.11$; $R^2_{12km/h} = 0.92 \pm 0.06$, Table 4.5). The $R^2$ value for the 5 km h$^{-1}$, however, was the highest. This implies that the model was suited for monitoring people using wearables at varying speeds: e.g. covering the range from people with a slower gait to people with faster gaits. Knowing the speed of the locomotive activity increases the $R^2$ substantially and yields similar results to those of Neugebauer et al. (2014). However, Model 3 is less complex and has just one prediction variable (load rate estimated by wearables) and one grouping variable (participant), which leads to a direct relation between load rate estimated by wearables and load rate estimated by force plates.

**Wearables attached to different body parts (Objective 4)** When comparing the wearables (SP 1, SW 1, SP 2, SW 2) attached to different body parts, there were differences noticed (here Model 3 results were compared). All devices had very similar $R^2$ values ($R^2_{SP1} = 0.83$; $R^2_{SP2} = 0.83$; $R^2_{SW1} = 0.81$; $R^2_{SW2} = 0.83$, see Table 4.6). However, Table 4.7 shows that SW 1 differed from the other three devices ($\alpha = 0.05/6 = 0.0083$).
SW 1 was on the right wrist, which most often deviated from a consistent motion, such as stroking one’s hair, looking at the smartwatch, or gesticulating. These results imply that the suitability of wearables as a surrogate for ground reaction load is largely independent of location on the body. Rowlands and Stiles (2012) had similar results and found out that wrist-worn monitors show a similar relationship with GRF as hip-worn monitors. However, the authors suggest using the non-dominant wrist to avoid confounders as future working. Also, it was concluded that the positions at the upper back between the shoulder blades and at the right hip are not significantly different and hence both positions are recommended for future work. In Study 3 and Study 4 the use of the smartwatches to estimate load rate on the lower limbs will not be included.

**Comparative analysis**

As shown, low-intensity activities were underestimated, and that moderate to vigorous locomotive activities were overestimated, by the wearables. This is similar to the findings of Meyer et al. (2015), where all data obtained by accelerometers systematically overestimated peak acceleration in comparison to the force plate. The findings of Meyer et al. (2015), however, do not match with the discovery of Martin et al. (2012) that slow gaits are under-represented using accelerometers and pedometers. Study 1 used three models (one linear regression and two linear mixed models).

The comparison between the three models may help other researchers understand the generalisability of the methods used in Study 1. Neugebauer et al. (2014) used a complex generalised regression model, which included acceleration, weight, type of activity and the interaction between the type of activity and acceleration. The generalisation of their model is difficult due to its complexity. The models used in Study 1 are kept as simple as possible. Hence, the load rates estimated with the wearables and force plate are directly related to the models. Another finding was that knowing the speed of the activity increased the quality-of-fit. Considering the speed led to similar results to Neugebauer et al. (2014), which included the type of locomotion (walk or run). However, the inclusion of the speed variable in the model makes the model less general, hence, less useful for the monitoring of everyday living.

A major limitation of the four previous validation studies (Meyer et al., 2015; Neugebauer et al., 2014; Hollville et al., 2016; Wundersitz et al., 2013) is that all force plates were placed in the middle of the laboratory, thereby giving the participants between 10-15 meters to perform the activities. Except for Hollville et al. (2016), who used six force plates, all other studies used one force plate in the middle of the room. One force plate means that, for each trial, just one data point was estimated for the force plate. Hollville et al. (2016) and Wundersitz et al. (2013) repeated their trials around six to seven times to obtain a better estimate of the uncertainty. The force plate integrated treadmill, on the other hand, generated 128 * 60 = 7680 data points (128 Hz sampling frequency; 60 s sampling time); hence, a better estimate of the uncertainty was made. Furthermore,
with the treadmill, a period of habituation for 20s of walking, jogging or running was possible during each trial, which would not have been possible if the participants just had 10-15m in which to do the activities.

Additionally, to obtain a better estimate of the uncertainties in the models in Study 1, the bootstrapping and cross-validation methods were used. Another limitation, which Meyer et al. (2015) and Neugebauer et al. (2014) had, is that by identifying the peak acceleration the algorithm becomes more complex. The advantage in using load rate estimates is that the algorithm does not have to identify peaks, which may lead to errors when analysing noisy signals.

4.7 Limitations and strengths of the study

One of the weaknesses of Study 1 was the limited number of participants. A larger number of participants would have been desirable, but due to time and feasibility restrictions on the facility, the number was kept to 12 participants. More problematic than the sample size was the homogeneity of the participants, who were all from a specific group of people (e.g., 22 to 26 years of age, from European areas, higher educated, athletic, but with different gender). Thus, this sample does not represent the general population.

Confounders such as poor biomechanics, which might lead to OA or in general to joint damage were not considered. The lower limb has been treated as a singular object, rather than a complex chain of joints, whose interactions might vary based on age, strength, gender or predisposing conditions. Further research is suggested considering these confounders (see Section 9.2). Also the trainers (shoes) were not standardised as this could be a confounder as they have different absorption properties.

Another weakness of the present study was the limited number of wearables attached to the participants. To compare wearables and their positions, having the participants wear more wearables would have given a better understanding of the position of the wearables on the participants’ bodies and how they affect the load rate data. Another weakness of Study 1 were some technical issues which made some of the data not usable. Moreover, the $R^2$ and RMSER values had space for improvement. A higher quality-of-fit might be achieved with a higher sampling rate from the devices.

Study 1, however, had multiple strengths, such as the comparison of three models and the direct relationship between wearable data and force plate data in the models to obtain better generalisability. The main strengths of the study were the use of an anti-gravity treadmill with an integrated force plate which allowed participants’ joint load to be varied. The different loading conditions allowed the model to consider a broad spectrum of loading on the joints, which can range from the loading who people with degenerative joints diseases might have to the loading who athletes might have. The
integrated force plate made it possible to obtain much more data points from the force plate in comparison to the previously mentioned validation studies, in which participants walked across the plate embedded on the floor. This, together with the data analysis with the bootstrapping and cross-validation methods, increased the understanding of the uncertainty in the models.

4.8 Conclusion

Wearable technology, commonly used in the activity tracking of everyday life activities, appears to provide an acceptable level of accuracy for estimating load rate on the lower limbs while locomotive activities. The acceptability was assumed here because the correlation found between load rate data from the wearables and the force plate can be described as a “high positive correlation” from the guidelines of Hinkle et al. (2003).

There was no significant difference observed between the wearables worn on the hip and between the shoulder blades. However, there was a significant difference between the smartwatch on the right hand and the smartwatch on the left hand. The models’ $R^2$ increased when considering the different speed conditions. The model’s performance was similarly high with varying speeds. Therefore, it is suitable for a range of activities, from everyday to the athletic.

These results support further research in using wearables to estimate load rate, which may lead to a progressive development in healthcare and the self-management of arthritis and exercise. People would be able to monitor the load rate of their lower limbs during everyday life and exercises without being dependent on expensive and non-mobile tools. Wearables with load rate estimation may provide an easy, objective, and cost-effective method for people to measure their activity regarding the decrease of the load rate through their joints while having a physical active lifestyle.

The validity of wearables estimating load rate using a force plate was demonstrated in Study 1. Study 2 (next chapter), on the other hand, tested the load rate estimations using wearables during different activities conducted outside the laboratory environment.
Chapter 5

Study 2: Classification and quantification of load rate estimates of everyday activities using wearable accelerometers

5.1 Introduction

After assessing the validity of smartphones and smartwatches to estimate load rate in a laboratory environment during locomotive activities in Study 1 (Chapter 4), the load rate algorithm had to be tested outside the laboratory environment.

Physical activity monitoring with inertial sensors is a growing field in research with applications in elite sports (Gabbett et al., 2010; Barrett et al., 2014) and clinics (Item-Glatthorn et al., 2012; Silva et al., 2002). Commercially available inertial sensors make it possible for one to count steps, measure distances, and record the time taken to complete a physical activity, all of which might have a positive effect on the behaviour of people regarding their physical activity (Consolvo et al., 2006). As mentioned in Section 2.2.3, elite sports research has led to the development of methods for monitoring horizontal impact events with inertial sensors during contact sports (Gabbett et al., 2010). Also mentioned in Section 2.2.3, there are studies that measure acceleration data with inertial sensors and then classify multiple activities with machine learning methods (Zhang et al., 2012; Mannini et al., 2013). The above-mentioned research areas seem promising for improving people’s physical performance. Nevertheless, in no studies has the estimation of loading on joints during physical activity been addressed. Estimating the loading on the joints during physical activity in everyday life with commercially
available inertial sensors could potentially benefit various populations, such as people with a high risk of developing degenerative joint diseases or people with arthritis.

5.1.1 Previous work in the literature

Some attempts were made to measure physical activity intensities using accelerometers. Mannini et al. (2013) and Zhang et al. (2012) performed studies which classified activities into four general intensity categories: sedentary (lying down, sitting, internet searching, reading, typing, writing, standing still), cycling, ambulation (natural walking, treadmill walking, carrying a box, going up and down stairs), and other activities. The mean and standard deviation of the signal magnitude vector and frequency-domain features (such as dominant frequency, the power of the dominant frequency, second dominant frequency, the power of second dominant frequency, total power, and a dominant frequency between 0.6-2.5 Hz) were used in these studies. Zhang et al. (2012) used split mode validation (2/3) and 10-fold cross-validation with which they achieved accuracies of 99% and 96% for waist-worn and wrist-worn accelerometers, respectively. Mannini et al. (2013) achieved accuracies of 95% and 85.2% with a leave-one-subject-out cross-validation for ankle-worn and wrist-worn accelerometers, respectively. Mannini et al. (2013) mentioned that frequency-domain features gave the most accurate results, with additional features having minimal effects. In general, the intensity of the activities chosen depended on the frequency of those activities because they involved the repetitive cyclical motions of various parts of the body. This might have been the reason why Mannini et al. (2013) mentioned that frequency-domain features gave a better accuracy and were sufficient for distinguishing between activity intensities.

The main weakness in the work of Mannini et al. (2013) was that, although the classifiers with frequency-domain features were able to distinguish between the four categories, they were not able to distinguish within the categories themselves. The categories, that were distinguished were broad (ambulation, sedentary, cycling, and other activities), hence, the classification accuracy is high.

Classification Two approaches exist for human activity classification: supervised learning (with labelled data) and unsupervised learning (without labelled data). Since labelled data is available in this work, a supervised learning model was used (Bishop, 2006).

Existing studies use classifiers such as hidden Markov model (HMM) (Trabelsi et al., 2013), logistic regression (Zhang et al., 2012), decision tree (Chung et al., 2008; Zhang et al., 2012), support vector machine (SVM) (Mannini et al., 2013; Zhang et al., 2012; Khan et al., 2014), naïve Bayes (Bersch et al., 2014), K-nearest neighbour (KNN) (Bersch
et al., 2014; Albert et al., 2012), neural network (Gyllensten and Bonomi, 2011), multilayer perceptron (Bayat et al., 2014), and random forest (Bersch et al., 2014; Bayat et al., 2014). How well the classifier distinguishes the test set of the data correctly can be quantified with the accuracy. To obtain this value, the test data is classified with the classifier and both the true and predicted values are compared. The percentage of correct predictions is then calculated to determine the accuracy.

Out of the literature mentioned above the SVM method had the first (five out of eight) or second highest (three out of eight) accuracy in the studies mentioned in Tables A.1, A.2, and A.3 and showed a high accuracy in general. It is difficult to compare these studies due to their different aims and designs. There is, however, a noticeable trend showing that the SVM’s accuracies are among the highest in all of the studies mentioned in Tables A.1, A.2, and A.3.

Validation Multiplication validation techniques exist for classification, such as k-fold cross-validation (k-fold CV) (Zhang et al., 2012), leave-one-out cross-validation (LOOCV) (Cawley and Talbot, 2003), and leave-one-subject-out cross-validation (LOSOCV) (Mannini et al., 2013; Esterman et al., 2010; Gyllensten and Bonomi, 2011). The metric that was used in the abovementioned studies was the accuracy. The accuracy was also used in this study to cross validate data. LOOCV is computationally expensive, but Zhang et al. (2012), who used k-fold CV, obtained a higher accuracy in comparison with Mannini et al. (2013), who used LOSOCV, even though their study design was similar. In comparison to the k-fold CV and LOOCV, where random data points are separated from the dataset to serve as test data, LOSOCV separates a participant’s entire feature set to use it as test data, thereby allowing for one to assess a model’s ability to generalise over other participants. The accuracy estimation, however, will decrease when LOSOCV is used since the data obtained from participants will not be related to the data used in the model. Hence, no dependent data is involved in the model. Nonetheless, the estimation is more accurate if one uses the model on different participants.

Activities to classify Epidemiological studies concentrating on activities with high load rate might show a more detailed relationship between to the progression of OA and high load rate. Spector et al. (1996), for instance, found that long-term weight-bearing sports, such as middle- and long-distance running, are related to OA at the hip and knee. Moreover, Cooper et al. (1994) discovered that repeated knee bending is a risk factor for knee OA and that risk may be higher in jobs which entail both knee bending and mechanical loading. The risk of knee OA was significantly increased in participants whose main job included climbing more than ten flights of stairs per day (OR 2.7, 95% CI 1.2 – 6.1). However, it should be also mentioned that Timmins et al. (2017) did not find a relationship between running and OA in their literature review. They found out that moderate to low quality evidence suggests no association with OA diagnosis. However,
Chapter 5 Study 2: Classification and quantification of load rate estimates of everyday activities using wearable accelerometers

Lievense et al. (2003) conclude that there is moderate evidence for a positive association between hip OA and sporting activities, such as running, soccer, athletic activities, and ballet dancing. Blagojevic et al. (2010) found out in their systematic literature review that cohort studies, generally suggested an increased risk of knee OA in those who exercise more regularly or intensely. The literature in this field contradicts, however, the researchers are interested in exploring the load rate estimates from smartphones and smartwatches outside the laboratory environment.

Furthermore, in Tessutti et al. (2012) recreational runners were asked to run twice for 40 meters on asphalt, concrete, natural grass and rubber with in-shoe pressure patterns. They explored whether running on grass may reduce the stress on the musculoskeletal system compared with the stress when running on harder surfaces, such as asphalt and concrete. To the knowledge of the authors, there has been no literature for estimating loading on the joints with accelerometers during locomotive activities and no classifier which examines the effects of different activities on different terrains. These results led to the decision in Study 2 to examine if load rate as a feature can classify different terrains outside the laboratory and, ergo, identify more activities with higher impact loading.

5.1.2 Rationale

Load rate is a standard feature in biomechanical science used to estimate impact loading on the lower limbs (e.g. gait analysis during running). However, there are no classification studies to the knowledge of the author, that have used load rate as a feature for the classification of activities. By doing so, the researcher can classify activities and quantify load rate, which is an indicator for impact loading. The present study aimed to estimate load rate with wearables to classify activities instead of using the raw acceleration data (Zhang et al., 2012; Mannini et al., 2013). After assessing the validation in Study 1 the load rate magnitude algorithm has to be tested outside the laboratory environment. One option to test the load rate magnitude for physical activity monitoring is to use it for activity classification. Hence, the testing in Study 2.

5.2 Aim and objectives

5.2.1 Aim

The aim of Study 2 was to develop a classifier which distinguishes between activities outside of the laboratory environment while also estimating the load rate associated with these activities.
5.2.2 Objectives

1. To explore and compare the accuracies of an activity classifier which uses load rate or acceleration features.

2. To explore and compare the accuracies of a terrain classifier which uses load rate or acceleration features.

3. To compare the ability of the smartwatch to the smartphone in Objectives 1 and 2.

4. To compare the accuracies between LOSOCV and 10-fold cross-validation (CV) to make the results comparable to other studies.

5. To create a proof of concept with a three-layered algorithm (Layer 1: feature extraction, Layer 2: classification, Layer 3: categorisation).

5.2.3 Hypothesis

The accuracy of a physical activity classifier is not significantly different when using load rate features compared to acceleration features.

5.3 Methods

During this study acceleration data of multiple activities (walking 5km h\(^{-1}\), jogging 8km h\(^{-1}\), running 12km h\(^{-1}\), climbing up and down stairs) on different terrains (asphalt, grass, and gravel) with a smartphone and a smartwatch were collected. From these data a three-layered algorithm was developed: 1) load rate estimation from accelerometer data from the wearable device sensors, 2) activity classification, and 3) load rate categorisation. The algorithm enabled joint loading to be estimated and indicated which activities related to which levels of load.

5.3.1 Study design

The study design was cross-sectional and experimental.

5.3.2 Participants and recruitment

Eleven healthy adults (female: n=4; male: n=7; age: 26.3 ± 3.3 years; height: 175.16 ± 14.84 cm; body mass: 71.22±8.52 kg; mean and standard deviation) participated in this study. All participants were active in their normal lives on a regular basis. Participants
were recruited via posters on multiple noticeboards around the University of Southampton. Once a participant showed interest, the researchers sent an email to them with the participant information sheet and an invitation to the study. Based on the screening, which excluded those with lower limb pathologies or any musculoskeletal, neurological, or systemic diseases or other physical disabilities that limited their mobility, 11 of 18 volunteers accepted the invitation.

5.3.3 Ethical approval

The study was approved by the University of Southampton’s Faculty of Health Sciences’ Ethics Committee (no. 17086, see Appendix B).

5.3.4 Equipment

Recordings were made of acceleration values from the sensors on a smartphone (Sony® Xperia Z1 Compact™, 137 g) and a smartwatch (Motorola Moto 360™, 49g) worn by participants while completing different activities in outdoor parkland. The smartphone was inserted into the right pocket and then attached with a band, as this position is a place to carry the phone in normal life. The non-dominant wrist was chosen because the previous study of the authors validated that the smartwatch on the non-dominant wrist was a suitable tool to estimate load rate during locomotive activities in an experimental environment Section 4.6. Further, a GPS watch was given to the participants to wear on their dominant hand, to estimate and maintain speed. The weight of the equipment did not influence participants’ activities or loading (see Figure 5.1).

5.3.5 Procedure

Each session in this study took, on average, 45 minutes and included briefing the participants, obtaining their written, informed, consent (consent and participant information sheet: Appendix E), and performing the different activities in the park. After the briefing, participants were asked to attach the equipment to their bodies.

The participants walked (5km h\(^{-1}\)), jogged (8km h\(^{-1}\)) and ran (12km h\(^{-1}\)) in a straight line on three different terrains (asphalt, grass, gravel) for 60 seconds. Moreover, they climbed up and down stairs for approximately 5 seconds. They did this for only 5 seconds because it was difficult to find a set of stairs that would have taken longer than approximately 5 seconds without interruption. Every participant completed 13 trials. The wearables recorded acceleration with a sampling frequency of 50 Hz. Table 5.1 shows each trial conducted.
Chapter 5 Study 2: Classification and quantification of load rate estimates of everyday activities using wearable accelerometers

5.3.6 Data analysis

After the data were collected, they were processed with MATLAB (Version R2016b, The Math Works, Natick, MA). The first 10 seconds of recording of the walking, jogging, and running served as a period of habituation and were discarded before the data was processed. The next 40 seconds were used for data processing, while the last 10 seconds of each trial were discarded to avoid recording possible behaviour changes associated with the trial ending. The data of the stairs climbing was discarded before 1 second and discarded after 4 seconds. The raw data was transformed into the load rate magnitude (load rate) and acceleration magnitude (acceleration).

Both the acceleration and load rate feature sets were each transformed to time-domain features: mean, maximum value, minimum value, variance, standard deviation, root mean square (RMS); and frequency-domain features: mean frequency, dominant frequency, entropy and energy (10 × acceleration, plus 10 × load rate = 20 features in total, see Table 5.2). The dominant frequency was found with the ‘find peak’ function.
in Matlab. The energy and entropy (Reiss et al., 2010) of the signal were obtained by a Fast Fourier transformation (Welch, 1967).

Table 5.2: Feature matrix

<table>
<thead>
<tr>
<th>Domain</th>
<th>Load rate based features</th>
<th>Acceleration based features</th>
</tr>
</thead>
<tbody>
<tr>
<td>time-domain features</td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>maximum value</td>
<td>maximum value</td>
</tr>
<tr>
<td></td>
<td>minimum value</td>
<td>minimum value</td>
</tr>
<tr>
<td></td>
<td>variance</td>
<td>variance</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>standard deviation</td>
</tr>
<tr>
<td></td>
<td>root mean square (RMS)</td>
<td>root mean square (RMS)</td>
</tr>
<tr>
<td>frequency-domain features</td>
<td>mean frequency</td>
<td>mean frequency</td>
</tr>
<tr>
<td></td>
<td>dominant frequency</td>
<td>dominant frequency</td>
</tr>
<tr>
<td></td>
<td>entropy</td>
<td>entropy</td>
</tr>
<tr>
<td></td>
<td>energy</td>
<td>energy</td>
</tr>
</tbody>
</table>

**Theory of classification - support vector machine** As mentioned in Section 5.1.1, to identify activities Study 2 uses support vector machine (SVM). The SVM is a supervised learning classifier which assumes that the set of data points are IID (Bishop, 2006). To classify an IID data set in a binary system, a separating line has to be set. Generally, there are infinite options in a 2D space for setting a separation line between the classes. Multiple examples of separation lines can be seen in Figure 5.2:

![Figure 5.2: Different options for separating lines exists for a linearly separable set of 2D-data. This figure was inspired by Smola and Schölkopf (1998).](image)

The SVM classifies a set of data points so that the margin between the classes has the broadest data-free space. This is called the Large Margin Classifier (Smola and Schölkopf, 1998). The SVMs classify linear separable data and attempt to minimise the generalisation error by maximising the margin between data points (support vectors) on both sides of the separating hyperplanes. In Figure 5.3, one may observe that the method only uses marginal points to set up an optimal hyperplane.
SVM is a linear classifier, but can easily be transformed into a nonlinear classifier by mapping the input data into a high-dimensional feature space. Thus, as can be seen in Figure 5.4, a set of data which, itself, is nonlinear in a two-dimensional space, can become linear in a multi-dimensional space.

By choosing the largest margin, the smallest error will be achieved. This will lead to a good generalisation performance. This principle is called structural risk minimisation (Schölkopf and Smola, 2001).

Consider a linearly separable dataset consisting of $n$ observations:
\[(x_k, y_k) \forall k = 1, 2, ..., n \quad (5.1)\]

\[x_k \in \mathbb{R} \quad \text{and} \quad y_k \in \{-1, 1\} \quad (5.2)\]

Previously, a separating line in a 2D space has been discussed for the purpose of explaining the SVM theory. Henceforth, a multi-dimensional space will be considered with the aim of finding a separating plane which corresponds to a decision function:

\[g(x) = \text{sgn}(w^T x + b) \quad (5.3)\]

‘sgn’ stands for the sign function, which can be written in real numbers as:

\[\text{sgn}(x) := \begin{cases} 
-1 & \text{if } x < 0 \\
0 & \text{if } x = 0 \\
1 & \text{if } x > 0.
\end{cases} \quad (5.4)\]

The separating plane in Equation 5.3 is defined by the orthogonal vector \(w\) and a threshold \(b\) such that:

\[y_k \left( \frac{w^T x_k}{||w||} - b \right) \geq M \quad , \quad (5.5)\]

where \(M\) is the margin. The optimal hyperplane is unique among the separating hyperplanes since it differs regarding its maximum margin of separation between any of the training points and the hyperplane. In Equation 5.5, the relative distances between the data points, the hyperplane, and their mathematical relationship can be observed.

The hyperplane can be written as \(H = \left( x \left( \frac{w^T x_k}{||w||} \right) - b \right)\); consequently, the distance between a point with the hyperplane is \(d_k = y_k \left( \frac{w^T x_k}{||w||} \right)\). Moreover, the distance between the origin and the data \(x_k\) as \(d_{ok} = \frac{w^T x_k}{||w||}\) will be obtained. \(w\) and \(b\) will be found to maximise \(M\) subject to the constraints:

\[y_k \left( \frac{w^T x_k}{||w||} - b \right) \geq M, \quad \text{for all } \quad k = 1, 2, ..., n \quad (5.6)\]

Furthermore, dividing through \(M\), and defining \(w' = w / (M ||w||)\) and \(b' = b/M\) the following constraints will be obtained. In order to construct the optimal hyperplane, the following expression must be minimised subject to the following constraints
\[ \min \tau(w) = \frac{||w'||^2}{2} \quad , \quad (5.7) \]

subject to \( y_k \left( w^T x_k - b' \right) \geq 1 \) for all \( k = 1, 2, ..., n \) \quad . \quad (5.8)

The function \( \tau \) in Equation 5.7 is called the objective function, while Equation 5.8 is an inequality constraint. Together, they form a constrained optimisation problem. These kinds of problem are solved by introducing Lagrange multipliers \( \alpha \geq 0 \) and a Lagrangian shown below:

\[ L(w, b, \alpha_k) = \frac{1}{2} ||w||^2 - \sum_{k=1}^{n} \alpha_k (y_k (w^T x_k - b) - 1) \quad . \quad (5.9) \]

The Lagrangian \( L \) has to be minimised with respect to the primal variables \( w \) and \( b \) and maximised with respect to the dual variables \( \alpha_k \). This means that the saddle point has to be found. According to the Karush-Kuhn-Tucker (KKT) conditions, the derivation of \( L \) generates the following saddle points:

\[ \frac{\partial}{\partial b} L(w, b, \alpha_k) = 0 \quad \text{and} \quad \frac{\partial}{\partial w} = 0 \quad . \quad (5.10) \]

This leads to:

\[ \sum_{k=1}^{n} \alpha_k y_k = 0 \quad , \quad (5.11) \]

and consequently:
\[ \mathbf{w} = \sum_{k=1}^{n} \alpha_k \mathbf{y}_k \mathbf{x}_k = 0 \quad . \] 

(5.12)

When \( \alpha_k > 0 \) the constraints are exactly met so that \( y_k \left( \mathbf{x}_k^T \mathbf{w}' - b' \right) = 1 \) is obtained. These data points are known as support vectors (SVs)(see Figure 5.6). The SVs lie on the margin and the remaining training examples are irrelevant for the classifier.

![Figure 5.6: The support vectors are lying on the margin with the KKT conditions. This figure was inspired by Smola and Schölkopf (1998).](image)

By substituting 5.11 and 5.12 into the Lagrangian 5.9, the primal variables \( \mathbf{w} \) and \( b \) are eliminated and the dual optimisation problem is found. The problem 5.13 is usually solved in practice:

\[
\begin{align*}
\max_{\alpha} W(\alpha) &= \sum_{k=1}^{n} \frac{1}{2} \sum_{k,l=1}^{n} \alpha_k \alpha_l y_k y_l \mathbf{x}_k^T \mathbf{x}_l , \\
\text{subject to} \quad &\alpha_k \geq 0 \quad \text{for all} \quad k = 1, 2, ..., n \quad \text{and} \quad \sum_{k=1}^{n} \alpha_k y_k .
\end{align*}
\]

(5.13)

Equation 5.13 can be solved with quadratic programming algorithms. Using 5.12, the hyperplane decision function 5.3 can be written as:

\[
g(\mathbf{x}) = \text{sgn} \left( \sum_{k=1}^{n} y_k \alpha_k \mathbf{x}_k^T \mathbf{x} + b \right) .
\]

(5.14)

The high accuracies obtained with SVM, which were discussed in Section 5.1.1, was a key reason why SVM was chosen for Study 2.

**Activity classification and validation** For Objective 1, 2 and 3 four classifiers were built: a load rate and an acceleration based classifier for activities and a load rate
and an acceleration based classifier for terrains. Preliminary testing was undertaken with Matlab’s “Classification Learner” app with the complete acceleration and load rate based dataset of all participants. Based on the features in 5.2, the following supervised learning methods were compared with the “Classification Learner” app using a 10-fold CV: discriminant method (linear and complex), decision trees (complex, medium, simple), support vector machines (kernels: linear, quadratic, cubic, fine Gaussian, medium Gaussian, coarse Gaussian), and K-nearest neighbours (distance measure: fine, medium, coarse, cosine, cubic, weighted).

The app has an advantage such that, for an initial analysis, it offers an easy and quick comparison between multiple classification methods. After using the app, a further, more accurate analysis was accomplished. The app showed the preliminary results that, for a 10-fold CV, the SVM classifier had the highest accuracy among the methods listed above (activity classification accuracy: 92.58 ± 0.7%, terrain classification accuracy: 78.71 ± 1.9%). Among the SVM classifiers, the highest accuracy was obtained with the Gaussian kernel. Therefore, it was used for further analysis. As mentioned before, in addition to performing best regarding classification accuracy, a vital advantage of the SVM method is that it only stores and classifies using data points at the margin, thereby significantly reducing computational intensity.

Further preliminary testing was made with Matlab’s “Classification Learner” app to recognize which four features are the most affecting features amongst the features listed in Tables 5.2. The four features with the highest accuracies (see 5.3) were taken and different feature combinations were compared with the aim of obtaining the highest accuracy in a minimal computing time. The number of the four features was a number of convenience to reduce computational time by while not losing accuracy.

Table 5.3: Four features with the highest accuracy for the smartphone

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Four features with highest accuracy</th>
<th>Accuracy of acceleration based feature set</th>
<th>Accuracy load rate based feature set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity classifier</td>
<td>Mean, RMS, energy, entropy</td>
<td>58.9%, 64.1%, 72.2%, 69.4%</td>
<td>73.1%, 71.3%, 72.0%, 71.6%</td>
</tr>
<tr>
<td>Terrain classifier</td>
<td>Maximum value, standard deviation, mean frequency, dominant frequency</td>
<td>41.2%, 41.0%, 48.7%, 48.3%</td>
<td>41.0%, 40.8%, 48.9%, 48.6%</td>
</tr>
</tbody>
</table>

Further preliminary testing of the effect of data windowing sizes with the complete load rate and acceleration based feature set was assessed with a Gaussian kernel SVM classifier and LOSOCV selected. The load rate and acceleration data were windowed, and the effect of different windowing lengths (3s, 4s, 5s, 6s, 7s, 8s) were compared. The
preliminary test showed that the window sizes 7s for the activity and 3s for the terrain
classifier had the best accuracy/time ratio. Further analyses were conducted with 7s for
the activity classifier and 3s for the terrain classifier as window sizes.

The main results were based on 100 LOSOCV bootstrap samples with a Gaussian kernel
SVM classifier. Confidence intervals were calculated with one-sigma distances from the
mean. To determine whether there is a significant difference between acceleration and
load rate based classification, a t-test was conducted ($\alpha = 0.05$).

To achieve Objective 4 the accuracy of LOSOCV was compared to the 10-fold CV. While
more computationally intensive than 10-fold CV, LOSOCV indicates the models’ ability
to generalise across participants. However, comparing the 10-fold CV was chosen to
make it comparable to other studies.

**Activity Categorisation**  In order to achieve Objective 5, the load rate data were par-
titioned into three load rate categories after the classification. The categories are: low,
moderate, intensive. The load rate data during walking in the current study were used
to determine load rate category thresholds. The low/moderate and moderate/intensive
thresholds were chosen by the mean of the walking load rate of the current study plus
and minus the standard deviation (lower threshold = $x - \sigma$, upper threshold = $x + \sigma$).
These bounds are imagined as suggestions for specific people groups with degenerative
joint diseases. These thresholds are not guidelines and just serve as an example how
categories can be applied to the algorithm. Further research has to be done to obtain
threshold to apply on a classifier specially made for groups of people with different joint
loading needs.

The three-layered algorithm with load rate generator, SVM and load rate categorisation
was implemented for each participant (see structure in Figure 5.7)

```plaintext
1 devices := ['device_1','device_2']
2 raw_acceleration:= raw_data(participants = 1:11) % acceleration data of
  whole participants
3
4 for each device in devices
5  acceleration_set:=calc_acceleration(raw_acceleration)
6  loadrate_set:=calc_loadrate(raw_acceleration)
7
8 feature_acceleration:=calc_feature(acceleration_set)
9 feature_loadrate:=calc_feature(loadrate_set)
10 for i= 1:100
11  ex := rand(participant)
12  test_acceleration = feature_acceleration[ex]
```
Figure 5.7: The three-layered load rate algorithm consists of a feature generator, support vector machine classifier, and categorisation. L= low (blue); M= moderate (green); I=intensive (red).

Algorithm 5.1: Pseudocode for the data analysis of Study 2

```plaintext
test_loadrate = feature_loadrate[ex]
train_acceleration = feature_acceleration[0:ex-1] +
feature_acceleration[ex+1:length(feature_acceleration)]
train_loadrate = feature_loadrate[0:ex-1] + feature_loadrate[ex+1:
length(feature_loadrate)]

acceleration_classifier:=SVM(train_acceleration)
loadrate_classifier:=SVM(train_loadrate)

predicted_class_acceleration:=acceleration_classifier(test_acceleration)
predicted_class_loadrate:=loadrate_classifier(test_loadrate)

accuracy_acceleration:= accuracy_acceleration + calc_accuracy(
predicted_class_acceleration, true_class_acceleration)
accuracy_loadrate:= accuracy_loadrate + calc_accuracy(
predicted_class_loadrate, true_class_loadrate)

end
mean_accuracy_acceleration := mean_accuracy_acceleration + calc_mean(
accuracy_acceleration)
mean_accuracy_loadrate := mean_accuracy_loadrate + calc_mean(
accuracy_loadrate)
end
```
5.4 Results

Participants 1 to 11 completed all activity trials. Therefore, a total of 143 trials were analysed.

Including all features in the classifier of activities yields accuracies with LOSOCV of 62.7%(38.2 – 76.4) for smartphone data and 60.9%(53.9 – 75.5) for the smartwatch. A higher accuracy was obtained by only using the four features with the highest importance. These are the acceleration based features (mean acceleration, RMS acceleration, energy acceleration, entropy acceleration), which yield an accuracy of 80% for the smartphone data based classifier and 80% for the smartwatch data based classifier (Table 5.4).

Table 5.4: Activity accuracy with LOSOCV. Both feature sets had no significantly different means (p > 0.05).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Accuracy acceleration</th>
<th>Accuracy load rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean, RMS, energy, entropy (smartphone)</td>
<td>78.55% (68.8-89.6)</td>
<td>80.3% (70.0-89.6)</td>
</tr>
<tr>
<td>Mean, RMS, energy, entropy (smartwatch)</td>
<td>76.49% (64.1-90.1)</td>
<td>80.1% (63.5-91.4)</td>
</tr>
</tbody>
</table>

Classifying different terrains (asphalt, grass, gravel) within different activities showed a poor classification accuracy with the LOSOCV of 49.2% (30.5 – 66.7) for the smartphone data and 48.2% (29.7 – 68.2) for the smartwatch data if all above-mentioned features are used. Combining the four load rate features with the highest importance (maximum value, standard deviation, mean frequency, dominant frequency) yields accuracy, with the LOSOCV, of 57% for the smartphone and 59% for the smartwatch. Acceleration based features showed accuracy, with the LOSOCV, of 57% for the smartphone and 59% for the smartwatch (table 5.5).

Table 5.5: Terrain accuracy with LOSO cross validation. No significant difference between both feature sets (p > 0.05).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Accuracy acceleration</th>
<th>Accuracy load rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum value, standard deviation, mean frequency, dominant frequency (smartphone)</td>
<td>57.33% (45.5-73.7)</td>
<td>56.5% (44.9-67.3)</td>
</tr>
<tr>
<td>Maximum value, standard deviation, mean frequency, dominant frequency (smartwatch)</td>
<td>59.3% (44.9-74.1)</td>
<td>59.4% (44.8-71.8)</td>
</tr>
</tbody>
</table>

As mentioned before, the same activity and terrain classifier with the same acceleration and load rate based features were validated with the 10-fold CV methods. Including all features in the classifier of activities, yields accuracies with the 10-fold CV of 97%(96.8 – 97.4) for smartphone data. Reducing the feature number to the four features (mean acceleration, RMS acceleration, energy acceleration, entropy acceleration)
with the highest importance (Table 5.3) decreases the overall accuracy, which contrasts the results obtained by the LOSOCV, where the reduction of features increases the overall accuracy. This yielded an accuracy of 88% with acceleration based features and 87% for the load rate based features with the smartphone data based classifier (Table 5.6).

Classifying different terrains (asphalt, grass, gravel path) within different activities shows a much higher accuracy with the 10-fold CV of 86.4% (85.0 – 88.0) for with the smartphone data if all above-mentioned features in the acceleration and load rate based features are used. Combining the four load rate based features with the highest accuracies (maximum value, standard deviation, mean frequency, dominant frequency) yields the accuracy with the load rate based features with the 10-fold CV of 65% for the smartphone. Acceleration based features give the accuracy of 64% for the smartphone (Table 5.6).

Table 5.6: Activity and terrain accuracies with 10-fold cross validation. No significant difference between both feature sets ($p > 0.05$).

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Feature</th>
<th>Accuracy acceleration</th>
<th>Accuracy load rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity classifier</td>
<td>Mean, RMS, energy, entropy</td>
<td>88.3% (87.4- 89.3)</td>
<td>86.7% (85.7- 87.6)</td>
</tr>
<tr>
<td>Terrain classifier</td>
<td>Maximum value, standard deviation, mean frequency, dominant frequency</td>
<td>63.8% (60.9- 66.6)</td>
<td>65.5% (63.3- 67.7)</td>
</tr>
</tbody>
</table>

After the classification, the data was partitioned into three different categories based on the load rate: low, moderate and intensive (see section 5.3.6). The low/moderate threshold was chosen as $x_{low} = \mu_{walk} - \sigma_{walk} = 2.8 \times 10^3$ and the moderate/intensive threshold as $x_{up} = \mu_{walk} + \sigma_{walk} = 5.6 \times 10^3$, where $\mu_{walk}$ is the mean and $\sigma_{walk}$ is the standard deviation of all walking load rate data (5km h$^{-1}$ data of all participants). The proof of concept of the three-layered algorithm is depicted in Figures 5.8 and 5.9. These two figures show how the true and the predicted values of the activity and terrain classifier were partitioned into low, moderate and intensive.
Chapter 5 Study 2: Classification and quantification of load rate estimates of everyday activities using wearable accelerometers

Figure 5.8: True (left) and predicted (right) proportions of time spent walking (top), jogging (middle) and running (bottom) for all participants. The load rate was divided into three categories: low (blue), moderate (green), and intensive (yellow).
Figure 5.9: True (left) and predicted (right) proportions of time spent on asphalt (top), grass (middle) and gravel (bottom) for all participants. The load rate was divided into three categories: low (blue), moderate (green), and intensive (yellow).
5.5 Discussion

The results show that load rate based features had an accuracy of 80% (leave-one-subject-out cross-validation (LOSOCV)) for the activity classifier and an accuracy of 57% (LOSOCV) for the terrain classifier. Different studies had difficulties to distinguish classes with similar intensity (Zhang et al., 2012; Mannini et al., 2013), however, the classifier in the present study was able to distinguish between classes with similar intensity, which is a novel finding.

Activity classifier (Objective 1) There was no significant difference in the accuracy of the classifier when using load rate based features in comparison to acceleration based features (Table 5.4). With this result, hypothesis 5.2.3 was confirmed. It can be said that the load rate based features are as good as the conventionally used acceleration based features. The biomechanically meaningful load rate based features can be used for physical activity classification if loading features on the lower limbs are needed.

Even if acceptable accuracies (Zhang et al., 2012; Mannini et al., 2013) were achieved with the activity classification, the confidence intervals were large. The large confidence intervals are likely due to the challenge of classifying activities with similar intensity (Zhang et al., 2012; Mannini et al., 2013). For example, running fast on asphalt or running upstairs are in the same intensity category and, hence, difficult to distinguish with the accelerometers used. Another possible explanation for the large confidence intervals is the difficulty of maintaining a specific speed by the participants. The GPS, which helped the participants to maintain a specific speed condition, was not accurate when participants were surrounded by high objects, such as trees.

Terrain classifier (Objective 2) The low to moderate terrain classification accuracy can be due to the relatively low sampling frequency of the wearables (50Hz). The importance of the features in Table 5.3 show that the high accuracy came from the frequency-domain features for the terrain classification, which illustrates that the joint loading features were less influential than expected.

Comparison between smartphone and smartwatch (Objective 3) The classification accuracy obtained from models trained on smartwatch data were similar to ones trained on smartphone data (Table 5.4 and Table 5.5). The similar accuracy was because the arm swing activities during locomotive activities within an experimental environment were related to the intensity of an activity. However, it is emphasised that the validity of using a smartwatch for physical activity monitoring based on estimated load rate is questionable. The similar accuracies are due to other factors, such as stronger arm swing with higher activity intensity. The activities that were used during this study
mainly involved the lower limbs, and the similar results are due to a relation between lower limbs activity and arm swing. Further, the data collection of this study was conducted outdoors and activities were not conducted under everyday real-life conditions. More research has to be done outside under everyday real-life conditions to obtain the correlation between smartphone and smartwatch data.

Comparison between LOSOCV and 10-fold CV (Objective 4) The 10-fold CV accuracy results were 6-10% higher than the results from the LOSOCV (see example in 5.6; load rate activity classification with LOSOCV: 80% (70%-90%); load rate activity classification with 10-fold CV: 87% (86%-88%)). Further, the confidence intervals of the 10-fold CV results are between 1 and 2%. However, for the LOSOCV values the confidence intervals differed between 10 and 15%. These different results show how difficult it is to compare activity classification studies in the literature directly. The LOSOCV is an approach to make the classifier more general by utilising the independent data of each participant for validation.

Three-layered algorithm (Objective 5) In their study, Crowell et al. (2010) showed that participants were able to reduce impact loading during running with the use of real-time visual feedback. This consisted of a screen showing them their load rate, which was monitored by a force plate, while they ran on a treadmill. Participants used this information to modify their running style, leading to reduced load rate. In a similar way, wearable based data can be used to help people to reduce joint loading to the ‘moderate’ band. This was a proof of concept and the thresholds of the categories need to be validated meaningfully. The thresholds should be tailored to the needs of a specific group of people, such as people with a degenerative joint disease, athletes with high impact on the joints and other people with a high risk of arthritis.

Comparative analysis The work of Gyllensten and Bonomi (2011) had slightly lower accuracies in comparison to the present study (80%) for their outdoor testing (raw acceleration based features, SVM, LOSOCV: 75.6% ± 10.4%). However, their activity types (ambulation, cycling, sedentary activities and other activities) were at different intensity levels, which made it easier to achieve a higher accuracy than the present study. The activity classes in the present study were all ambulatory. The difficulty in physical activity classification is to distinguish between activities within the same intensity level. Mannini et al. (2013) tried to distinguish between activity intensities but failed. However, the present study concentrated on classifying activities within the same intensity level. Activities such as walking fast upstairs and jogging are in the same intensity category, making them difficult to distinguish. However, they could be classified in the present study with an accuracy of 80%, which is a novel finding.
Mannini et al. (2013) obtained 95% accuracy for their ankle based accelerometer data and 84.7% for wrist-based accelerometer with SVM and LOSOCV. The higher accuracies were because of the use of an ankle-based accelerometer, which measures larger accelerations than a waist-based accelerometer used in this study. However, an ankle-based accelerometer would require another device for physical activity classification. To avoid buying more devices for physical activity monitoring, it was more appropriate to use commercially available inertial sensors (wearables), which are already used in the everyday life of the general population in England. The differences in wrist-based classification was around 4% higher for Mannini et al. (2013) compared to the current study.

Their data was collected in a laboratory-based environment, which made it easier to maintain experimental settings. In the current study, the data was collected outside the laboratory environment. The maintenance of speed was difficult and only based on a GPS watch carried by the participants. To maintain the speeds, there was an adaptation phase where participants needed to find the correct speed. Even though the first and last 10 seconds of data collection were cut off, variations in the speeds of participants were not accounted for. The classes with ‘5 km h\(^{-1}\)’, ‘8 km h\(^{-1}\)’, and ‘12 km h\(^{-1}\)’ be better labelled ‘walking’, ‘jogging’ and ‘running’.

The study design of Zhang et al. (2012) was very similar to Mannini et al. (2013). However, Zhang et al. (2012) used 10-fold CV and achieved for the activity intensity classifier accuracy of 99% for waist worn accelerometer and 97% for wrist-worn accelerometer. The different accuracies found by Zhang et al. (2012) were due to the different activity classes they used (walking, running, household or sedentary). The walking category (including stairs) represented moderate intensity, the running category represented the vigorous intensity and the sedentary category represented inactive phases. These classes are broader than the classes chosen in the current study (walking, jogging, running, climbing up stairs fast, climbing up stairs slowly, climbing downstairs fast, climbing downstairs slowly). Further, accuracies were higher because they used the 10-fold CV. This study uses LOSOSV, which is a more representative test of the classification performance in real life scenarios. Comparing the waist-based data, the results of this study, with the 10-fold CV, achieved higher accuracies in comparison to the validation with LOSOCV. Due to different classes and different cross-validation methods, direct comparisons are difficult to make for the results of the current study and Zhang et al. (2012) study.

The results of the present study promote the use of a combination of time-domain features and frequency-domain features, which is in disagreement with the suggestions of Mannini et al. (2013) and Zhang et al. (2012) due to their suggestions, that frequency-domain features are sufficient for physical activity monitoring. It can be seen in Table 5.3 that for the activity classifier, the mean load rate has the highest single accuracy. However, the mean load rate, which was a time-domain feature, was as important as the
frequency-domain features energy and entropy. The feature importance values in Table 5.3 indicate that time-domain and frequency-domain features are similarly important for the classification of activities chosen by this study. For the terrain classification, the frequency-domain features mean and dominate frequency, were the most important features, followed by time-domain features with 7% less accuracies. However, in combining the time-domain and frequency-domain features, the terrain classifier was able to classify with a LOSOCV up to 57% accuracy.

5.6 Limitations and strengths of the study

This study has a number of weaknesses, including a relatively small and homogeneous sample of adults. However, around 3000 data points were calculated from all participants. The equipment showed limitations, such as poor GPS signal in some spots and, therefore, influenced the running speed of the participants. The low/moderate/intensive thresholds in this proof of concept study have not been validated, and a longitudinal study is needed to make meaningful decision relating to such thresholds. Joint damage is not purely due to weight bearing loading but also due to poor movements during walking (Litwic et al., 2013). The biomechanical analysis was not included in the analyses, which is a limitation of the method. The author suggests further research, including the biomechanical aspects.

However, the present study also has several strengths, including the use of the bootstrapping method to help obtain a better estimate of uncertainty. A major strength was that the classes chosen were close in their intensity level, rather broad classes, which were used in previous studies. Similar intensity levels are for example running and running up stairs. Therefore, a more precise classification was achievable. Further, this study used a more biomechanically meaningful feature, the load rate. This study developed a terrain classifier. However, the accuracies were low and more research is needed in this field.

5.7 Conclusions

Wearable technology, commonly used in activity tracking of everyday life activities, appears to provide an acceptable level of accuracy for classifying activities and categorising joint loads on the lower limbs. The level of accuracy is called acceptable, because similar studies mentioned before have comparable results (Gyllensten and Bonomi, 2011; Mannini et al., 2013). The present findings show that support vector machines have an acceptable level of accuracy in classifying activities within the same intensities. This result can lead to physical activity monitoring for activities within the same intensity but different joint loading, which is a novel finding.
The hypothesis tested was confirmed: Load rate based features have similar accuracies to acceleration based features, and can, therefore, be used as features to classify activities. These results may lead to further research in physical activity monitoring in reference to joint loading. Low to moderate accuracies were found in classifying terrains. Further research is needed in terrain classification and in validating thresholds for categorisation of joint loading.
Chapter 6

Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

6.1 Introduction

Previous chapters in this thesis have assessed the validation and tested the estimation of load rate on the lower limbs via wearables and have illustrated that wearables might be suitable tools for estimating load rate on the lower limb joints. Studies 1 and 2 were experimental cross-sectional studies. The next step was that of testing the algorithm and wearables, used in the previous studies, outside the experimental environment. In order to create a useful tool which would incorporate the knowledge of the previous studies, a prospective study was required.

6.1.1 Previous work in the literature

Multiple prospective studies exist which monitor physical activity with an accelerometer-based unit. Gretebeck and Montoye (1992) determined how many days participants should be monitored in order to provide an estimate of habitual physical activity. Accelerometer data were measured from 30 participants who were monitored during their walking hours for seven continuous days. A repeated measures ANOVA showed no significant difference between working days. For weekend days significant differences were noticeable. They concluded that, for a seven-day period, less than 5% significance was observed and weekend days need to be included.

Coleman and Epstein (1998) used generalisability theory to estimate the number of monitoring days necessary to estimate reliable activity measured by the TriTracR3D for
seven days. They concluded that three to four days of monitoring would result in a reliable estimate of physical activity based on the TriTracR3D. Generalisability theory is a statistical theory for evaluating the reliability of behavioural measurements. Their results, however, contradict the result of Gretebeck and Montoye (1992). Coleman and Epstein (1998) assume that the activities of participants are equal each day; hence, they did not consider the physical activity behaviour changes which occurred during weekends.

Matthews et al. (2002) used variance partitioning techniques to examine the number of days needed to reliably estimate various accelerometer parameters, with physical activity being assessed for up to 21 consecutive days. They found that only three to four days of monitoring was required to achieve an 80% reliability and seven days were required to assess patterns of inactivity.

Monitoring load rate continuously over seven days might use a significant amount of internal storage on a smartphone. Sampling methods can be used so as not to continuously record the load rate but which still supply a sufficient estimate of the load rate over those seven days. Monte Carlo methods are a broad class of algorithms based on repeated random sampling to obtain a numerical result estimation (Hammersley, 2013). The simple Monte Carlo approach is that of observing random numbers and inferring the desired solution from the behavior of the random numbers (Hammersley, 2013). The desired solution will have a specific uncertainty, due just using some random observations and to uncertainty in the measurements (see Section 3.6.2). One way of reducing uncertainty in the results is by simply collecting more observations. This, however, is costly, both in terms of time and storage. There is a relationship between finding errors in the results and the number of observations made (Hammersley, 2013). Two of the main advantages of the Monte Carlo method is its generality and unbiased nature due to the way it samples randomly (Hammersley, 2013).

Wilson and Atkeson (2005) trained learning motion models for activity monitoring with four different binary sensors (motion detections, break-beam sensors, pressure mats, and contact switches). The parameters of motion models were either learned off-line via continuous measurements or onboard via a Markov chain Monte Carlo sampling method. The off-line methods had the highest accuracy, followed by the on board leading method.

Wei et al. (2011) developed a two-layered hidden Markov model (HMM) algorithm for long-term and continuous daily activity monitoring using a wearable body sensor network (using three sensors). The first layer processed sensory data locally at each device to reduce data transmissions. The second layer classified the activity sequence from the result of the local processing. The algorithm was time-efficient only in that the result of the decoding procedure in each device needed to be transmitted rather than the raw acceleration data. Wilson and Atkeson (2005) and Wei et al. (2011) used the Monte Carlo method for problems regarding physical activity monitoring via sensors, such as
data sampling or handling missing data for the purpose of obtaining a sufficient accuracy. The design of this study was in some parts inspired by the work of Wilson and Atkeson (2005) and Wei et al. (2011). However, both studies’ aims and approaches were different, hence, the studies are not comparable.

6.1.2 Rationale

The abovementioned studies monitored physical activity. None of them, however, considered impact loading on the lower limb joints during physical activity for a prospective study. In this study, an app was developed which monitored load rate as a surrogate for impact loading on the lower limb joints. The ability of this developed app to monitor in a continuous manner and its ability to process the acceleration data to mean jerk magnitude data onboard was explored.

Since it is unclear in the abovementioned literature how many days are necessary to record data in order to obtain reliable estimations for physical activity, it was necessary to analyse this phenomenon. Furthermore, even though the performance of the smartphone vis-à-vis the smartwatch was explored for daily activities, no clear conclusions were able to be drawn. Study 1 showed that the $R^2$ values were similar for smartphones and smartwatches for locomotive activities on a treadmill. In daily life, though, these might not correlate well and, ergo, might be not reliable for monitoring impact loading.

6.2 Aim and objectives

6.2.1 Aim

The aim of this study was to assess the ability of the smartphone and smartwatch, with the developed app, to estimate load rate on the lower limbs during everyday physical activities during a time period of seven days.

6.2.2 Objectives

1. To test the ability of the app to monitor load rate continuously
2. To explore how the smartphone and smartwatch performed in comparison to another
3. To test the error in the accuracy in order to obtain a reliable estimate for the load rate.
6.3 Methods

This study used the accelerometer sensors of wearables (a smartphone and a smart-watch) to monitor the load rate of the daily, real-life activities of healthy participants throughout a span of seven days. In contrast to the previous studies in this thesis and similar to Wilson and Atkeson (2005) and Wei et al. (2011), parts of the data processing in Study 3 were conducted onboard on the smartphone while other parts were conducted off-line using MATLAB. To save the storage space and battery life of the wearables, the app developed for Study 3 used a Markov chain Monte Carlo method for the purpose of sampling data for a period of seven days, which means that it randomly sampled 5 seconds of data from every minute of the day. The samples are ordered in sequential chains because they are distributed approximately between 0-55 seconds apart from each other throughout the day. The current observation is dependent on previous observations; thus, the assumption being independent and identically distribution (IID) is not sufficient. Hence, the data can be seen as a Markov chain. Later, the way of obtaining a sufficient accuracy with the Markov chain Monte Carlo method used in this study was described (Section 6.3.7).

6.3.1 Study design

The current study had a quantitative, exploratory, prospective design.

6.3.2 Participants and recruitment

Ten healthy adults (female n=6, male n=4; aged 27.2 ± 3.6 years; height: 172.6 ± 9.6cm; body mass: 73 ± 14.7kg; means ± standard deviation) participated in the study. Participants were recruited via posters on multiple noticeboards around the University of Southampton. Once a participant showed interest by contacting the lead researcher, they were sent an email with the participant information sheet and an invitation to the study. Based on the screening, which excluded those who had neurological or systemic illnesses or other physical disabilities which may have limited their mobility, 10 of the 12 volunteers accepted the invitation. The sample of convenience of 10 participants was chosen due to limited time. Each participant made, on average, eight thousand mean jerk magnitude data points.

6.3.3 Ethical approval

The study was approved by the University of Southampton’s Faculty of Engineering and Environment Ethics Committee (no.30213, see Appendix B).
6.3.4 Equipment

Recordings were made of mean jerk magnitude values from the wearables (see Section 3.2.2). Participants were given a smartphone (Sony® Xperia™ Z5 Compact (Sony, 2016b)) and a smartwatch (SmartWatch 3 (Sony, 2016a)), which both had the developed app installed on them, to carry around with them for seven days during their everyday lives. The smartphones and smartwatches changed during the studies, because the availability of the previously used devices were restricted. However, the validation was still valid because the devices have the same processor (Qualcomm Snapdragon), which included the same MEMS sensors. Participants were also asked to complete an activity diary (see Appendix D) throughout the seven days for the purpose of helping the researchers interpret the load rate data from the app. A paper-based activity diary asked for the time and duration of moderate to vigorous physical activities. Furthermore, the participants had the option of mentioning in their diaries whether the wearables had run out of battery or if they had forgotten to carry them around with them during any part of the days.

6.3.5 Procedure

After a short briefing about the app, the wearables, and the study, and after having received the participants’ written consent (participant information sheet and consent form: Appendix E), the participants were given a smartphone to be worn close to the hip, as well as a smartwatch (to be worn on their non-dominant wrist). They had been given the choice of either putting the smartphone in their trouser pocket or using the smartphone-belt provided by the researchers. The hip, as a location at which to place the smartphone, was chosen because the hip might be a potential location for carrying the smartphone in everyday life and because this location at the body was used in Study 1.

Jerk magnitude data were collected over the whole seven days for each participant. Participants were asked to carry the smartphone and smartwatch with them during the active phases of their day over the seven days of data collection and to write down moderate to vigorous activities into the diary. During their inactive phases, the participants were allowed to take off the wearables. The participants were not asked to do any specific exercises. Rather, they were just expected to conduct their everyday routines, including, if applicable, the exercises that they do for themselves. Seven days later, they were invited to return the smartphone, smartwatch, and paper-based activity diary to the researchers.
6.3.6 Data collection

Monte Carlo Method  The acceleration data of the smartphones and smartwatches were sampled at 50Hz. Due to storage limitations, however, a sampling method was chosen which did not continuously measure jerk magnitude data but still would have given a sufficient approximation for the overall load rate. For this reason, a probabilistic model was needed, which is an approximate inference method based on numerical sampling, also known as the Monte Carlo method (Bishop, 2006). If a general function \( f(x) \) with reference to a probability distribution \( p(x) \) is considered, the expectation of contentious variables can be written as

\[
E[f] = \int f(x)p(x)dx. \tag{6.1}
\]

In the case of discrete variables, the integral were replaced by summation. A set of samples \( x_i \) (where \( i = 1, \ldots, n \)), which are drawn independently from the distribution \( p(x) \), provides the general idea of the sampling methods. ‘Drawn independently’ means that they are sampled randomly without any sequential dependency (Hammersley, 2013). Drawing randomly from a data set will prevent a bias during the estimation process. The expectation (Equation 6.1) can be approximated by a finite sum as

\[
\hat{f} = \frac{1}{n} \sum_{i=1}^{n} f(x_i). \tag{6.2}
\]

The accuracy of the expected value is dependent on the sample size (Bishop, 2006). Looking at Equation 6.1, if the expectation is small in regions where \( p(x) \) is large, or if the opposite is the case, then expectation will be dominated by the wrong probability and a larger sample size will be required in order to obtain sufficient accuracy. By addressing Objective 3 the error with the sample size of seven days data collection was calculated, which is an indicator if seven days of data collection were sufficient enough.

Markov Chain  The independent and identically distribution (IID) assumption is made over data sets whose observations \( x_{(i)} \) (where \( i = 1, \ldots, n \)) are independent and identically distributed, thereby corresponding to Figure 6.1.

![Figure 6.1: An example of data, which is independent from each other and is identically distributed. This figure was inspired by Bishop (2006)](image-url)
Chapter 6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

The elements in the illustration are assumed to be independent and not linked to each other. Sequential data sets whose observations are linked to each other do not fulfill the IID assumption. If a current observation \( \{x_i\} \) is related to a previous observation \( x_{i-1} \) in the data set, the data set is called a \textit{first-order Markov chain} (see Figure 6.2).

![Figure 6.2: A first-order Markov chain of observation \( \{x_i\} \) in which the distribution \( p(x_i|x_{i-1}) \) of a particular observation \( x_i \) is dependent on the previous observation \( x_{i-1} \). This figure was inspired by Bishop (2006)](image)

In this study raw acceleration data was randomly collected for five seconds of each minute of each day for seven days. This method represents a random resampling method with the restriction of a sequence. Hence, the method used here can be seen as a Markov chain Monte Carlo method.

\textbf{Mean jerk magnitude extraction} A \textit{mean jerk magnitude} was calculated from the raw acceleration data over each five seconds onboard the wearable as follows:

\[
| \vec{j} | = \left| \frac{\Delta a}{\Delta t} \right| = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left( \frac{\Delta a_x}{\Delta t} \right)^2 + \left( \frac{\Delta a_y}{\Delta t} \right)^2 + \left( \frac{\Delta a_z}{\Delta t} \right)^2},
\]  

(6.3)

where \( \Delta a_x = \) is the difference in acceleration in the \( x \) direction between \( t = i \) and \( t = i + 1 \), \( \Delta a_y = \) is the difference in acceleration in the \( y \) direction between \( t = i \) and \( t = i + 1 \), \( \Delta a_z = \) is the difference in acceleration in the \( z \) direction between \( t = i \) and \( t = i + 1 \), and \( n = \) the number of time steps within the five second data sample.

\textbf{6.3.7 Data analysis}

The rest of the data analysis was conducted off-line with MATLAB (Version R2017a, The Math Works, Natick, MA) off-line. The mean jerk magnitude was multiplied with each participants mass, which resulted in the mean load rate magnitude. The mean load rate magnitude was then used for further data analysis.

For Objective 1 mentioned in Section 6.2, a moving average of 10 minute windows were calculated and plotted over the seven days for both the smartphone and smartwatch. The moving average was chosen to smooth out the fluctuation in the data and to recognise trends (Olaniyi et al., 2011). 10 minute windows are a relative short window for recognising trends but since it was an exploratory study a relative short window size was chosen. The diaries helped to interpret the data collected with the wearables.
In order to accomplish Objective 2, the Pearson’s correlation coefficients between the mean load rate magnitude data from the smartphone and smartwatch for each minute in the day was calculated. For the correlation test $60 \times 24 \times 7 \approx 10080$ (minute $\times$ hours a day $\times$ days) data points were used for each participant. The smartphone data was correlated for three cases: for all data, and if smartphone data was $> 200 \text{ kg m s}^{-3}$ and $> 300 \text{ kg m s}^{-3}$. This threshold was introduced, to find the correlation when participants were physically active with their lower limbs. This was followed with an ANOVA analysis with the Bonferroni correction Bonferroni (1936) to compare the three methods (significance level $\alpha = 0.05/3 \approx 0.01$).

For Objective 3, the Law of Large Numbers was used (Hsu and Robbins, 1947). The theorem states that, if the number of trials of a random process increases, the difference between the expected value and the actual value will become zero (ibid). As the sample size $n \rightarrow \infty$, the sample mean converges to the expected value, $x \rightarrow \mu$. To explore how many days are needed from a statistical point-of-view in order to obtain sufficient accuracy, the accumulated mean load rate magnitude was calculated to estimate an expected convergence. This convergence can show how many days of data collection was needed to obtain accurate results for the load rate. To estimate the required number of data points in Study 3, the confidence intervals have to be considered. This can be written as

$$\hat{x} - z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \hat{x} + z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}},$$

where $\alpha = 0.05$ and $z_{0.025} = 1.96$ for a confidence interval of 95%. If this equation is rearranged and, if one introduces an error $\varepsilon = |\hat{x} - \mu|$, an estimate for the number of sample can be approximated, as

$$n = \left( \frac{z_{\alpha/2} \sigma}{\varepsilon} \right)^2,$$

where $\sigma$ is the unknown population standard deviation which can be estimated by the sample standard deviation $\hat{s}$. With Equation 6.5, the approximate duration of the trial can be estimated in order to obtain an appropriate accuracy (Driels and Shin, 2004).

### 6.4 Results

All 10 participants successfully completed seven days of data collection. Participant 2’s smartphone’s load rate data were not recorded after the third day. However, the participants smartwatch load rate data were recorded continuously until the end of the seventh day. No data were measured for the seventh day from the smartwatch of
Participant 8. These two incidences might have been due to technical issues at the manufacturers level. Participant 4 did not submit a diary, hence, his data could not be interpreted. The moving average of load rate magnitude with 10 minutes window sizes for each participant can be seen in Figures 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, and 6.12.

![Figure 6.3: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 1.](image)

![Figure 6.4: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 2.](image)
Chapter 6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

For the correlation between smartphone and smartwatch data (Figure 6.13) linear interpolation was used if time stamps were missing for the smartphone to match the smartwatch (Meijering, 2002). The Pearson’s correlation coefficients between the mean load rate magnitude data from the smartphone and the smartwatch were calculated using all the data and with the threshold $200 \text{ kg m s}^{-3}$ and $300 \text{ m s}^{-3}$ for the smartphone data (Table 6.1). The mean correlation coefficient for all the data and participant was: $\bar{r}_{\text{all}} = 0.55$ and for the data excluding $< 200 \text{ kg m s}^{-3} \bar{r}_{\text{ex}200} = 0.57$ and $< 300 \text{ kg m s}^{-3} \bar{r}_{\text{ex}300} = 0.51$. 

Figure 6.5: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 3.

Figure 6.6: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 4.
Chapter 6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

Figure 6.7: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 5.

Due to the small $r$ observations size of 9 (10 participants excluded Participant 2) a non-paramedic ANOVA method was chosen (Kruskal-Wallis ANOVA). The Kruskal-Wallis ANOVA analysis showed that the mean $r$ values were not significantly different from each other ($p_1 = 1; p_2 = 0.52; p_3 = 0.11$).
Chapter 6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

Figure 6.9: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 7.

Figure 6.10: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 8.
Chapter 6 Study 3: Monitoring load rate estimates in daily life using wearable accelerometers

Figure 6.11: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 9.

Figure 6.12: The moving average load rate magnitude with 10 minute windows throughout a seven-day data collection period depicting Participant 10.
Figure 6.13: Mean load rate compared between smartphone and smartwatch on a loglog scale.
Table 6.1: Pearson’s correlation coefficients between smartphone and smartwatch. All values were significant ($p << 0.001$).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Correlation coefficient all data</th>
<th>Correlation coefficient $\frac{\Delta F}{\Delta t}$ smartphone $&gt; 200$</th>
<th>Correlation coefficient $\frac{\Delta F}{\Delta t}$ smartphone $&gt; 300$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.70</td>
<td>0.54</td>
</tr>
<tr>
<td>3</td>
<td>0.64</td>
<td>0.61</td>
<td>0.66</td>
</tr>
<tr>
<td>4</td>
<td>0.55</td>
<td>0.70</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>0.49</td>
<td>0.57</td>
<td>0.53</td>
</tr>
<tr>
<td>6</td>
<td>0.59</td>
<td>0.70</td>
<td>0.73</td>
</tr>
<tr>
<td>7</td>
<td>0.75</td>
<td>0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>8</td>
<td>0.61</td>
<td>0.64</td>
<td>0.54</td>
</tr>
<tr>
<td>9</td>
<td>0.47</td>
<td>0.58</td>
<td>0.55</td>
</tr>
<tr>
<td>10</td>
<td>0.73</td>
<td>0.62</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Figure 6.14: A comparison of the accumulated mean load rate magnitude with increasing sample size and convergence which was measured during the seven days of data collection for all participants.

The accumulated mean load rate magnitude with increasing sample size is depicted in Figure 6.14 which, in turn, converges over the values specific to each participant. The errors, which were estimated at a level of 95% confidence for seven days data collection, were all acceptable low ($< 3\%$) (Gretebeck and Montoye, 1992; Matthews et al., 2002) (Table 6.2).

The mean error of an accumulated mean load rate magnitude for the 10 participants was calculated as $\hat{e}_{\text{smartphone}} = 2.66\%$ and $\hat{e}_{\text{smartwatch}} = 1.99\%$ over seven days.
Table 6.2: Error estimation for seven days

<table>
<thead>
<tr>
<th>Participant</th>
<th>Estimated error smartphone</th>
<th>Estimated error smartwatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.51%</td>
<td>2.48%</td>
</tr>
<tr>
<td>2</td>
<td>2.49%</td>
<td>1.77%</td>
</tr>
<tr>
<td>3</td>
<td>2.51%</td>
<td>1.56%</td>
</tr>
<tr>
<td>4</td>
<td>2.66%</td>
<td>1.48%</td>
</tr>
<tr>
<td>5</td>
<td>2.41%</td>
<td>1.86%</td>
</tr>
<tr>
<td>6</td>
<td>2.96%</td>
<td>2.25%</td>
</tr>
<tr>
<td>7</td>
<td>2.08%</td>
<td>2.60%</td>
</tr>
<tr>
<td>8</td>
<td>3.08%</td>
<td>2.30%</td>
</tr>
<tr>
<td>9</td>
<td>2.93%</td>
<td>1.94%</td>
</tr>
<tr>
<td>10</td>
<td>2.98%</td>
<td>1.59%</td>
</tr>
</tbody>
</table>

6.5 Discussion

The results show that the app developed for this study was able to continuously monitor load rate estimates of the lower limb joints (6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, and 6.12).

Ability of the app (Objective 1) In general the diary data related to the monitored load rate data. Participant 1 was an active person, who was cycling regularly. In Figure 6.3 can be seen that 30 minutes cycling has a lower peak value than walking for the smartphone. However, for the smartwatches a very high peak can be seen when participant 1 was cycling. During the seven days Participant 1 cycled five times for 30 minutes and out of these five times, the smartwatch shows very high values for three of the times. The other two times the smartwatch battery was empty. The same phenomena can be seen in Participant 7 and 8 (Figures 6.9 and 6.10).

For longer cycling periods a relative low mean load rate was estimated for the smartphone, but a very high for the smartwatch. It can be said, that while cycling participants had low load rate estimations for the smartphone attached close to their waist, but high load rate estimations for the smartwatch which was attached to their non-dominant wrist. Cycling is an example, where the limitations of the smartwatch using the algorithm can be seen. Even though the jerk measurements might represent the jerk values on the wrist while cycling, the assumption to estimate the load rate on the lower limbs can no longer be made. The good correlation between smartphones and smartwatches in Study 1 do not agree with the findings for cycling in Study 3. The good correlation between smartphones and smartwatches in Study 1 might be due to the homogeneous and locomotive activities which were tested. However, the high peak for the smartwatches during cycling could lead to further research for cyclists or other occupations, who are
loading the upper body limbs, such as construction workers using chain saws or engine tamper.

Walking had low load rate estimations for both smartphone and smartwatch (Figures 6.8, 6.10 and 6.10) in comparison to running. The load rate estimation between smartphone and smartwatch appeared proportional during walking and running. This is in line with the results in Study 1, where a high correlation was found between smartphone and smartwatch. Other active was Yoga (Figure 6.8), where Participant 6 had very low load rate estimations. Further, some parts of the data showed linear behaviour, which in some points was reported as times when the battery of the wearables was empty. However, it was not always reported by participants when the wearables ran out of battery, which makes it difficult to understand if the app was the reason for not having data or if the smartphone was off. Participant 4 did not submit any diary, but mentioned that the smartwatch switched off due to low battery more quickly than the smartphone, which made it less practical to use.

**Comparison between smartphone and smartwatch (Objective 2)** A relationship between the data from the smartphone and the smartwatch is visible in Figure 6.13. In each plot of Figure 6.13 there was an accumulation of data with low load rate data from the smartphone. This indicated that for low smartphone data the smartphone and smartwatch data might not correlate, hence, the introduction of threshold values for the smartphone. The relationship between the smartphone and the smartwatch was revealed with Pearson’s correlation coefficient (see Table 6.1). The coefficients were calculated for the total data set, for the data set with \( \Delta f / \Delta t \text{ smartphone} > 200 \text{ kg m s}^{-3} \) and \( \Delta f / \Delta t \text{ smartphone} > 300 \text{ kg m s}^{-3} \). The threshold of \( \Delta f / \Delta t \text{ smartphone} > 200 \text{ kg m s}^{-3} \) was chosen visually with the aid of Figure 6.13. The mean correlation between the smartphone and the smartwatch is relatively low \( \bar{r}_{alt} = 0.55, \bar{r}_{ex200} = 0.57 \) and \( \bar{r}_{ex200} = 0.51 \). Introducing the threshold of \( \Delta f / \Delta t \text{ smartphone} > 200 \text{ kg m s}^{-3} \) increased the correlation about 2%. However, if the threshold was increased for more intensive activities with \( \Delta f / \Delta t \text{ smartphone} > 300 \text{ kg m s}^{-3} \) the mean correlation dropped from the about 4%.

The ANOVA showed that there were no significant differences in the mean of the three \( r \) values. The relative low correlation between smartphone and smartwatch shows again the limitation of the use of the smartwatch as a tool to estimate load rate on the lower limbs. Hence, the authors suggest that wrist worn accelerometers are not a suitable tool to estimate load rate on the lower limbs. However, the author encourages other researchers to explore the ability of wrist worn accelerometer to be use as a tool to estimate loading on the upper limbs.

**Error in the resampling method (Objective 3)** The convergences, which was visible in Figure 6.14, showed that for each participant the estimate of their load rate converged to a sufficient value. The average of all of the participants’ error estimation of an
accumulated mean load rate magnitude for the 10 participants was $\hat{e}_{\text{smartphone}} = 2.66\%$ and $\hat{e}_{\text{smartwatch}} = 1.99\%$ over seven days. The error, which was calculated in Table 6.2, was the difference between the estimated value (the converged value after 7 days) and the true value. This supports the estimations of Gretebeck and Montoye (1992), who suggested to monitor physical activity for a period of seven days. Gretebeck and Montoye (1992) concluded, the different behavior patterns of people between weekdays and weekends would effect the results. That is why the authors would suggest at least 7 days of data collection including weekdays and weekends.

**Comparative analysis**  
For the purpose of obtaining an appropriate degree of accuracy, at least seven days of data sampling are advised by the researcher whenever the Markov chain Monte Carlo sampling method with 5-second window sizes were randomly sampled for each minute throughout the day. The method of semi-processing the data onboard and post processing the rest off-line was successful. Similarly to Wei et al. (2011) and Wilson and Atkeson (2005) the Monte Carlo method supplied data with sufficient accuracy.

Previous studies mentioned above, mostly suggested seven days of data collection, including weekends. Based on the findings in this study an error of $\hat{e}_{\text{smartphone}} = 2.66\%$ was found for seven days data collection. This error is lower than the error thresholds (5%-20%) used in the previous studies, hence, a error term of 2.66% was considered as sufficient. Differences in the suggestions for data collection in the previous studies are noticeable. Whilst Gretebeck and Montoye (1992) and Matthews et al. (2002) suggested seven days of data collection Coleman and Epstein (1998) suggested three to four days. These differences can be due to different participant groups. The present study, Gretebeck and Montoye (1992) and Matthews et al. (2002) used healthy participants, which had occupations and different health levels. These three studies suggested seven days of data collection. It can be assumed that these healthy participants had varying activity levels spread during the seven day period. However, Coleman and Epstein (1998) just considered in-active college men, who’s physical activity might have not changed during the week. Hence, three to four days were sufficient in the study of Coleman and Epstein (1998). With these findings it can be said that for active people at least seven days of data collection might be representative. Further analysis has to be done with in-active participants to compare the results.

One of the findings in Study 1 and Study 2 indicated that smartphones and smartwatches (on the non-dominant wrist) were a suitable tool to estimate load rate on the lower limb joints. However, the difference between the smartphone and smartwatch estimations in this study might indicate the unreliability of the smartwatch for serving as tool for estimating the load rate on the lower limbs.
6.6 Limitations and strengths of the study

Study 3 has a number of weaknesses, including a relatively small and homogeneous sample of young adults. The participants were recruited from the University of Southampton; hence, they all shared a similar age and occupation. Furthermore, the duration of Study 3 was limited to seven days. Moreover, the smartphone was not fixed on the hip the whole time and the participants could have worn the smartphone at different positions on the body close to their hip, thereby possibly making it difficult to compare the different participants. However, this can be seen as an advantage as well, because it would make the app more generalizable. Even if the diary data were to help to interpret the load rate data, and some sort of uncertainty was expected in the diary data, it is emphasised that paper based diaries have limitations, as mentioned in the literature review (Section 2.2.2). By recalling their activities during the day participants might not have reported their activities accurately.

The study’s strengths were that the Markov chain Monte Carlo method was used, hence the ability of recording mean load rate magnitude could be shown. Time and storage can be saved in future studies using the same method. Another strength was that the study was a prospective study, in comparison to previously conducted cross-sectional studies. The researchers were able to observe how the app was working in a practical setting, which was a major strength looking at the usability of the app and the devices.

6.7 Conclusions

The wearable app appears to provide continuous load rate data. The present findings show that smartwatches are not an accurate method for monitoring impact loading on the lower limb joints for the participant group who took part in Study 3. This finding differs with the previous findings because in Study 1 and 2 only locomotive activities were considered such as walking and running. However, in Study 3 multiple different activities were conducted by the participants, where the upper limb did not move. Furthermore, a period of at least 7 days of data collection are recommended for an accuracy with an error of $\hat{e}_{\text{smartphone}} = 2.66\%$. In comparison to other studies mentioned above, the error of the load rate estimates is lower and hence it can be assumed that a reliable estimate was achieved after 7 days.

These positive results led to Study 4 being conducted, where the app was used in a clinical situation in a case study.
Chapter 7

Participant and public consultation: user interface for \textit{OApp}$^{\text{TM}}$ for people with lower limb joint osteoarthritis

7.1 Introduction

Conducting a clinical trial involves a lot of administrative and organisational effort. Hence, clinical trials should be thoroughly designed. An effective way of improving the design of clinical trials is to obtain consultation from the cohort focused during the trial (Brett et al., 2014). Patient and public involvement consultation sessions have the purpose to give people with a specific disease or problematic, which have a lay background, an active role in health science research (Brett et al., 2014). The idea behind the consultations session is to design research which is more closely to the needs of the specific people group and can lead to research of greater quality (Brett et al., 2014). In a patient and public involvement (PPI) consultation session people with a specific disease or problematic are invited to serve as representatives for their cohort and are asked for their consultation in the design of the research. According to Brett et al. (2014) positive impacts enhanced the quality and appropriateness of research were identified if PPI was conducted.

Gooberman-Hill et al. (2013) used PPI to enhance the design of a future clinical trial with people with OA. Different designs of tools potentially used during the future clinical trial were shown to the PPI representatives and their design features were discussed. The present PPI consultation session was inspired by the PPI consultation
session Gooberman-Hill et al. (2013). Guidance on how to conduct the PPI in accordance with research governance guidelines were obtained from the INVOLVE website (INVOLVE, 2017). For the purpose of exploring strategies to improve the adherence of participants using the OApp™ during Study 4, a PPI consultation session was arranged.

7.1.1 Rationale

Multiple ways exist for increasing and optimising the adherence of patients to record their OA-related pain. One option might be to use, in addition to the smartphone, a smartwatch which provides the option of recording OA-related pain into the smartwatch interface. Nevertheless, there has been no previous evidence about whether this would have any impact on utility or rates of adherence. There is, therefore, a need to discover whether people with OA would prefer a smartwatch or a smartphone when using the OApp™. Furthermore, the user interface of the OApp™ is an important component for exploring the PPI representatives’ perceptions about the app’s usefulness and applicability. All of these components could potentially have an impact on the future adherence and usage of the OApp™.

7.2 Purpose

The objective of the PPI group session was to find out whether people with lower limb OA preferred to input their OA-related pain into a smartwatch or smartphone. The researcher’s aim was to find out what kind of user interface the PPI representatives would prefer in the OApp™, how often they wanted to be reminded to input their OA-related pain, and whether they had any other comments or feedback that would help enhance the design of the OApp™.

The purpose of this consultation was to explore and understand the perspectives of people who would be using the OApp™ so that the user interface could be designed to incorporate the needs of people with OA. It is important to note that the outcome from this session were not intended to lead to research results.

7.3 Consultation Session

This PPI representative group was a practical group session. Different design options were introduced to the PPI group. A think aloud technique was used whereby people were encouraged to comment freely on the design, providing their opinions about what worked well for them and why. This was of interest to the researcher because it helped
to design OApp™ tailored to the people with lower limb OA who were going to be using it in Study 4.

The PPI representative group were people with self-reported lower limb joint pain. There were no limits for excluding representatives from being included in these group. People did not need a formal clinical diagnosis of OA and people with self-reported joint pain were also included. Potential representatives from the University of Southampton’s PPI representative database and the University’s Health Sciences research participant database were informed about the session by registered health and community PPI members. Ten PPI representatives were invited to take part in the session. Ten was a sample of convenience due to the time limitations given for preparing Study 4. Of these ten representatives, two women, both over retirement age, agreed to take part in the session. Another representative, which planned to come to the session, withdraw from participating few hours before the session. The aim was to obtain around eight representatives (same as Gooberman-Hill et al. (2013)), however, due to difficulties to find representatives and the time restriction, it was decided to continue with the session.

These two PPI representatives were invited to the Faculty of Health Sciences at the University of Southampton. It was here where the complete session took place. The OApp™ and Study 4 were introduced to the PPI representatives. Following that, the PPI representatives were allowed to explore the OApp™. Some screenshots of the user interface of OApp™ with different pain levels are depicted in Figures 7.1 and 7.2.

After the exploratory phase of the session, the PPI representatives were asked to answer questions posed by the research team (see Appendix C). Questions were about the user-interface design of OApp™ and the functionality of OApp™. The researchers were interested in, which user-interface design would make more sense to the representatives and how representatives would react if they had to use a smartphone and smartwatch during a study.

An open discussion took place in which questions were asked in order to obtain general feedback. The notes from the discussions are listed in Appendix C. Participants were asked at the end of the session to answer on paper the same questions discussed during the session and to compare their different answers both during and after the session.

The answers to the questions were used to improve the OApp™’s user interface. Thus, the answers of the PPI representatives determined whether the smartwatch would be included in the OApp™ and how often they would been ask to enter the pain score.

7.4 Outcome

The first PPI representative preferred to include the smartwatch in the study. She thought it would be enough to be reminded to score her pain three times a day, because
she did not experience much pain. She also said that she preferred to be reminded to score her pain. As for the user interface of the OApp™, she preferred a linear pain scale. The second PPI representative, on the other hand, preferred to have the option of scoring her pain, either on a smartphone or a smartwatch, depending on whether she was at home or outside. For the frequency of the pain score reminder, she preferred 3 times a day on a normal basis, but for clinical trials, she would have done it up to every second hour during the day. For the user interface of the OApp™, she was happy with both options, but when answering the question in the questionnaire, she chose the linear pain scale. She thought that having a smartwatch on one’s wrist which would remind someone to score their pain “would be a very good idea”.

Both PPI representatives felt positive about the study. Indeed, in the additional comments that they made, they said that it “Sounds like good research to help people with pain” and that it would be “good for assessing and managing pain levels”.

Both PPI representatives were hesitant about using wearable technologies at the beginning but, after exploring the technologies, and a bit further, they gained confidence using both the OApp™ and the wearable technologies seemed to enjoy each.
7.5 Limitations

It has to be acknowledged that one limitation was of only including two representatives, who were both older females and had differing views. Even if a higher percentage of people within the OA community are elderly women (Litwic et al., 2013), they were not representative of the whole OA community. As mentioned above, it was desirable to find more volunteering representatives, however, it was not feasible within the time available. And as mentioned before, one volunteer also withdraw from taking part few hours before the session.

7.6 Conclusion

The comments from the PPI representatives demonstrated that there were variances in preferences and that options to score pain either on the smartphone or the smartwatch were both valued. Due to the positive reaction to the smartwatch participants have the option to score their pain either on the phone or smartwatch. The preferred frequency of the pain score reminder varied between the participants. Hence, the researchers chose...
to take an intermediate answer between the two answers: i.e. five times a day. The OApp™ adopted a linear pain scale between 0 to 10 (0 = no pain, 10 = unimaginable) for scoring pain levels in Study 4.
Chapter 8

Study 4: Single case study to relate load rate estimates and osteoarthritis-related pain using wearables

8.1 Introduction

The main life-impairing symptoms of OA is pain (Litwic et al., 2013). As mentioned in the introductory chapter, there is currently no cure for OA despite the existence of various symptom-reducing drugs, such as intra-articular corticosteroids injections (CSI) (Arden et al., 2008) or joint replacement surgeries. Besides symptom-reducing treatments, Valderrabano and Steiger (2010) believe that moderate exercise delays and alleviates the symptoms of the disorder and prevents muscle atrophy and stiffness. The relationship between moderate physical activity and OA-related pain, however, was not confirmed in Valderrabano and Steiger’s work. Very little is known about the relationship between pain and physical activity in people with OA, which will be discussed in Section 8.1.1. Indeed, even less is known about the relationship between pain and load rate on the lower limbs during physical activity and this topic was addressed by the study in this chapter.

8.1.1 Previous work

The definition of pain used in this thesis for Study 4 was the one developed by the International Association for Study of Pain (IASP) (IASP, 2016). According to the ISAP, pain is an unpleasant sensory and emotional experience associated with actual or
potential tissue damage, which is described in such terms (ibid). An essential aspect of this definition is that it is always subjective. Also, as Lasch (2000) noted, the perception of pain is affected by experiences related to injuries incurred early in life and varies between people with different cultures and genders (Lasch, 2000).

According to Petursdottir et al. (2010), pain is a barrier to being active and influences the exercise behaviours in people with OA. The inactivity can lead to muscle atrophy, which is a common finding in people with OA (Valderrabano and Steiger, 2010). Valderrabano and Steiger (2010) emphasised that muscle weakness is a predisposing factor for OA. Functional ability and quality of life, such as physical function, sleep, moods, and other factors, can be significantly improved by reducing chronic pain in OA (Gerstle et al., 2001). According to Valderrabano and Steiger (2010), physical activity and muscles play an essential role in the joint adaptations and degeneration processes of OA. Muscles influence joint biomechanics as they produce movement, absorb loads, and provide dynamic joint stability. The stronger the muscles are, the more load might be absorbed and, consequently, the less pain the patient may experience (ibid). Less pain might lead to more physical activity and less muscle atrophy. This simplified theory requires more research to examine whether there is a relationship between pain and physical activity in people with OA, as well as to examine whether even more research is needed.

Adesola and Ololade (2013) examined the relationship between physical activity and pain, using a paper-based recall questionnaire (International Physical Activity Questionnaire) to measure the activity and a visual analogue scale to measure the pain. They found a significant association between occupation and pain intensity, but physical activity was not associated with pain intensity. They also discovered that an increased risk of knee pain is apparent in occupations which are likely to involve knee bending and possible heavy lifting.

Murphy et al. (2008) examined the relationship between pain and physical activity, but also fatigue. In contrast to Adesola and Ololade (2013), however, they found that pain and fatigue were negatively related to physical activity. They used an accelerometer and a device which recorded pain on a visual analogue scale. They focused on people who were still able to work and who reported having less pain, whereas Murphy et al. (2008) considered people with more advanced OA. Furthermore, Adesola and Ololade (2013) used subjective methods (i.e. a questionnaire and a visual analogue scale) exclusively to measure pain and activity in a cross-sectional way. Conversely, Murphy et al. (2008) used one objective method (i.e. continuous acceleration) for physical activity and one subjective measurement (viz. a visual analogue scale on a device, six data samples per day) for the pain over a period of five days. The conditions of these studies, however, are different and, therefore, not directly comparable. The combination in the work of Murphy et al. (2008) of objective and subjective measurements seems like a more reliable method for discovering a relationship between OA-related pain and physical activity. It is difficult, however, to compare and evaluate the conflicts in the literature.
since the relationships between physical activities, OA-related pain, and muscle atrophy are complex and need to be compared in similar situations.

Fransen et al. (2007) conducted an interventional randomised controlled trial where people with lower limb OA conducted either hydrotherapy or Tai Chi classes for 12 weeks, recording their pain and physical function. Both groups had a significant improvement in pain and physical functioning. Interestingly, the hydrotherapy group had more significant relief of joint pain. This result indicates that physical activity and pain are related and that physical activities with less weight-bearing loads on the joints may be more beneficial for decreasing the pain in the joints. Unfortunately, the study did not use any device or tool to record those physical activities. Hence, there is no direct comparison between physical activity and pain.

The literature mentioned above indicates that there is a relationship between physical activity and OA-related pain in the lower limb joints. Adesola and Ololade (2013) and Murphy et al. (2008) aimed to examine the relationship between physical activity and OA-related pain directly. None of the above-mentioned studies, however, considered load rate estimates as quantitative measurements. Fransen et al. (2007) compared two different exercises with different weight-bearing loads on the joints and found that there was a difference in the improvement of pain between those two different exercises. The literature, therefore, poses the following two questions: viz. whether there is a relationship between physical activity and OA-related pain and, furthermore, whether there are reliable methods for measuring these two factors. The outcome of the literature review on pain and physical activity monitoring is listed in Table 8.1.
<table>
<thead>
<tr>
<th>Author and title</th>
<th>Objective</th>
<th>Method</th>
<th>Subject Information</th>
<th>Conclusion</th>
<th>Strength</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adesola and Ololade (2013), Sociodemographic data, physical activity and pain intensity in patients with knee OA</td>
<td>To assess the association between sociodemographic data, physical activity and pain intensity in people with knee OA</td>
<td>Cross sectional survey research design, visual analogue scale (pain), IPAQ (physical activity level last 7 days)</td>
<td>100 subjects with knee OA</td>
<td>Occupation is associated with pain intensity but physical activity does not affect pain.</td>
<td>Large sample size, considering other demographic factors</td>
<td>Only one sample between pain and physical activity is subjective; recall of 7 days for questionnaire; baseline of pain participants not clear.</td>
</tr>
<tr>
<td>Murphy et al. (2008), The Impact of Momentary Pain and Fatigue on Physical Activity in Woman with OA</td>
<td>To examine the daily life patterns of pain and fatigue symptoms and objective physical activity in women with lower extremity OA.</td>
<td>Observational study, two laboratory visits and 5-day home data collection period (continues physical activity and six times symptom input)</td>
<td>60 women</td>
<td>Significant positive relationship between pain and physical activity, increasing momentary pain was associated with increased physical activity.</td>
<td>Control group, prospective study, pain progression and physical activity were compared over 5 days</td>
<td>No change of one of the variables, peak value of acceleration surrogate for physical activity, 32Hz, activity count is peak acceleration over 15 seconds</td>
</tr>
<tr>
<td>Fransen et al. (2007), Physical activity for Osteoarthritis management a randomized controlled clinical trial evaluating hydrotherapy or Tai Chi classes</td>
<td>To determine whether Tai Chi or hydrotherapy classes for individuals with hip or knee OA result in measurable clinical benefits</td>
<td>Randomised controlled trial; interventional; 12 weeks of hydrotherapy, Tai Chi, or a waiting list control group. Assessed 12 and 24 weeks after randomisation and included pain and physical function (WOMAC).</td>
<td>152 subjects with knee or hip OA</td>
<td>Hydrotherapy and Tai Chi can provide large and sustained improvement in physical function and pain. Hydrotherapy appeared to provide greater relief on joint pain.</td>
<td>Randomised, controlled, follow-up.</td>
<td>Physical intervention means that the study was not blinded.</td>
</tr>
</tbody>
</table>

Table 8.1: Outcome and quality of relevant studies to measure pain and physical activity
8.2 Rationale

In previous studies, a relationship between physical activity and the OA-related pain was noticed (Murphy et al., 2008). The work of Murphy et al. (2008) seemed reliable, however, more research is needed to determine the relationship between OA-related pain and joint loading. The purpose of this study was to develop and test a tool, which records these two variables, so that the relationship can be explored in further research.

In comparison to the work of Murphy et al. (2008) an additional factor was included in the present study: viz., the symptom-reducing intra-articular corticosteroids injections (CSI) (Arden et al., 2008). Arden et al. (2008) showed that using CSI leads to significant short-term pain relief. CSIs showed clinically important improvement in range of motion (Koester et al., 2007), which leads to improved function of the joint and the style of walking of the participant. The CSI enables a change in the pain variable, allowing to see if the $OApp^{TM}$ was able to recognise differences between pre- and post-injection.

8.3 Aim and objectives

8.3.1 Aim

The aim of Study 4 was to assess the ability of the non-interventional app, called $OApp^{TM}$, to monitor load rate and OA-related pain in preparation for use as a tool in clinical studies to explore the relationship between load rate and OA-related pain.

8.3.2 Objectives

1. To test the ability of the $OApp^{TM}$ to monitor load rate and OA-related pain.

2. To explore whether differences are noticeable in the pain and load rate relationship between the pre- and post-injection phases.

3. To compare pain and load rate data, which was recorded by the smartwatch and smartphone, with validated paper-based questionnaires the Hip and Groin Outcome Score (HAGOS) and the International Physical Activity Questionnaire (IPAQ).

4. To explore how the smartphone and smartwatch perform in comparison to one another in one person with OA.

5. To explore how data collected from the smartphone/smartwatch and OA-related pain are related.

6. To test the error in the data’s accuracy to obtain a reliable estimate of the load rate.
Chapter 8 Study 4: Single case study to relate load rate estimates and osteoarthritis-related pain using wearables

8.4 Methods

Study 4 assessed the ability of OApp™ with the purpose of providing proof of concept. Within Study 4, the non-interventional app, OApp™, was developed with the purpose of monitoring load rate through the lower limbs during physical activity via wearables and of giving the option to score OA-related pain using the touch screen of either a smartphone or a smartwatch. The patient and public involvement (PPI) consultation session was in preparation for Study 4 and helped to develop OApp™ (Chapter 7). People who had primary OA in their lower limb joints and had intra-articular corticosteroids injections (CSI) as their clinical routine care were invited to take part in the study. The study took 14 days to conduct, with seven days pre- and seven days post-CI. The CSI itself was not, however, part of the research in Study 4.

8.4.1 Study design

The current study had a quantitative, exploratory, proof-of-concept prospective design. One person with lower limb OA was given the OApp™ installed on both a smartphone and a smartwatch to monitor her load rate during physical activity and to score her current OA-related pain over a period of 14 days (seven days pre- and seven days post-injection).

8.4.2 Participant and recruitment

One adult (female; age: 65 years; height: 165 cm; body mass: 81 kg) with primary OA in her right hip who had a CSI scheduled seven days after the start of the study participated in the study. She was referred by her consultant to be scheduled for a CSI at the Nuffield Orthopaedic Centre Oxford. She was identified via a CSI list and then contacted via post and followed up, and was recruited though a phone call from a research nurse. Based on a screening process, which excluded those who needed supporting facilities (for instance, a wheelchair or crutches), who had neurological, systemic illnesses, who had other causes of pain apart from OA, or who had other physical disabilities, which may have limited their mobility, out of the five volunteers, she was the only one who accepted the invitation.

8.5 Ethical approval

This study was conducted in compliance with the Helsinki Declaration and was approved by the Health Research Authority of the National Health Service in the UK (REC: 16/NI/0228, IRAS ID: 211265). The study was registered with the National Institute for Health Research UK.
8.5.1 Equipment

8.5.1.1 Wearables

The OApp™, which was installed on both the smartphone and the smartwatch, was given to the participant for a total duration of 14 days (7 days pre-injection and seven days post-injection). Since the OApp™ is an advanced version of the app used in Study 3, the app records acceleration values from the sensors in the smartphone and the smartwatch and processes the acceleration data into a mean jerk magnitude onboard the smartphone. Further data analysis was done off-line using MATLAB.

The app randomly collected five seconds of jerk magnitude data every minute. The participant was asked to carry the smartphone close to her right hip, either on a smartphone belt (which was provided by the researcher) or in her pocket. Further, the participant was asked to enter either on the smartphone or smartwatch screen her OA-related pain score five times a day on an analogue scale, with scores between 0 and 10 (0: no pain; 10: unimaginable). The smartwatch reminded the participant to score her current OA-related pain (morning, before noon, afternoon, early evening, late evening). The data were stored in a database file on the smartphone and were extracted once the devices were returned to the researcher.

8.5.1.2 Questionnaires

The participant was asked to complete two questionnaires: the IPAQ (Craig et al., 2003) for physical activity and the pain part of the Hip and Groin Outcome Score (HAGOS) (Thorborg et al., 2011) for hip pain, after each seven days of the 14-day period. The participant was asked to assign which joints were painful on a manikin pre-, intra-, and post-study. These two questionnaires were chosen because their use did not include any additional costs and because they were international validated tools. The manikin was chosen because the researchers were interested if other joints were painful, hence, to identify confounders in the interpretation.

8.5.2 Procedure

In Study 3, a period of seven days of load rate monitoring was recommended for an accuracy with an error of $\hat{e}_{\text{smartphone}} = 2.66\%$. Since pre- and post-injection phases are assumed to show differences, each phase’s (both pre- and post-) duration was seven days.

The study involved three visits: Visit 1 (Introduction), Visit 2 (Injection), and Visit 3 (Return-wearables). Once the participant agreed to take part in the study, Visit 1 was
arranged, which occurred seven days before she was scheduled for the injection. At Visit 1, the participant signed a consent form instructing that she agreed to take part in the study, thereby giving her informed consent (participant information sheet and consent form: Appendix E). Her demographic details, including her height and weight and the joints in which she experienced pain, were recorded and the OApp™ was introduced to her in detail. The participant was asked to carry the wearables with her at all times when she was active. After seven days, at Visit 2, the participant was asked to fill the IPAQ and HAGOS questionnaires and to indicate which joints she was experiencing pain in on the manikin. The time at which she was administered the injection was recorded. Finally, seven days after she received her injection, at Visit 3, the participant was asked to fill the IPAQ and HAGOS questionnaires again and to indicate which joints she had experienced pain in on the manikin. The structure of Study 4’s procedure is depicted in Figure 8.1.

![Diagram](structure_of_study_4.png)

**Figure 8.1: Structure of the procedure of Study 4.**

### 8.5.3 Data collection

Data collection was conducted as in Study 3 (see Section 6.3.6). Raw acceleration data were randomly collected for five seconds (with a sampling frequency of 50 Hz) of each minute of each day for seven days. A mean jerk magnitude was calculated from the raw acceleration data over each five seconds onboard the wearables.

### 8.5.4 Data analysis

The rest of the data analysis was conducted off-line with MATLAB (Version R2017a, The Math Works, Natick, MA). The difference between the data analysis of Study 3 and Study 4 is that the variable pain was included in the analysis next to the variable load rate. A test of normal distribution had been conducted in Study 3, hence, a parametric analyses were employed.

For Objective 1, the moving average of 10 minutes of the mean load rate magnitude of the smartphone and smartwatch data was calculated. These data, as well as the pain data, were plotted over time both pre- and post-injection.
For Objective 2, the mean of the smartphone, smartwatch, and pain data over seven days were calculated and compared pre- and post-injection. The paired sample t-test ($\alpha = 0.05$) was used to identify differences between pre- and post-injection in smartphone, smartwatch and pain data.

For Objective 3 HAGOS was compared to the pain that was self-recorded in the wearables and IPAQ was compared with the measured load rate from the wearables.

For Objective 4, the Pearson’s correlation coefficients between the load rate magnitude data from both the smartphone and smartwatch were calculated and averaged. The significance of the correlation was decided with a value of $\alpha = 0.05$.

For Objective 5, a non-parametric method was used, because the normality of the pain data was rejected according to the Kolmogorov-Smirnov test. The Spearman’s rank correlation coefficients between the mean load rate magnitude data from the smartphone and smartwatch and the pain data were calculated using all the mean load rate magnitudes between the two pain scores and the following pain score. The significance of the correlation was decided with a value of $\alpha = 0.05$.

Objective 6 was achieved by using the Law of Large Numbers (Hsu and Robbins, 1947), by calculating an accumulated mean load rate magnitude, and by estimating the duration of sample size to achieve the same error which was obtained during Study 3 for healthy participants.

### 8.6 Results

OApp™ monitored the estimated load rate continuously and the pain though out the 14 days. There were improvements in all measures pre- to post-injection on the wearables, as well as the self-report measures (HAGOS and IPAQ).

The moving average of the load rate magnitude with 10-minute window sizes for the participant both pre- and post-injection can be seen in Figure 8.2. OApp™ was able to monitor continuously the estimated load rate for the smartphone and smartwatch (see Figure 8.2).

To be able to note differences between pre- and post-injection, the mean of the collected load rate data with smartphone and smartwatch and the pain scores were listed in Table 8.2. The load rate data from smartphone, smartwatch, and pain data were all significantly different between pre- and post-injection. However, no minimal clinically important difference was calculated, which has to be done in further research. Furthermore, the results of the IPAQ questionnaire were listed in the Metabolic Equivalent of Task (MET) -minutes per week and the HAGOS results were from 0-4 (0: no pain; 4: extreme pain). The wearable data and the questionnaire data were presented in
one table to be able to compare differences in pre- and post-injection by the OApp™ and the self-reported recall questionnaires. Significant tests were not possible for the questionnaire data because only two data points were collected.

Table 8.2: Mean load rate and mean pain for each seven day period both pre- and post-injection.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pre injection</th>
<th>Post injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone load rate [kg m s(^{-3})]</td>
<td>714.92*</td>
<td>699.94*</td>
</tr>
<tr>
<td>Smartwatch load rate [kg m s(^{-3})]</td>
<td>1885.20*</td>
<td>1812.99*</td>
</tr>
<tr>
<td>IPAQ [MET-minutes/week]</td>
<td>17103</td>
<td>21984</td>
</tr>
<tr>
<td>Pain score [0-10]</td>
<td>6.21* (intensive-very intensive)</td>
<td>4.52* (distressing-very distressing)</td>
</tr>
<tr>
<td>HAGOS [0-4]</td>
<td>2.75 (moderate-sever)</td>
<td>0.625 (non-mild)</td>
</tr>
</tbody>
</table>

*Significant \((p < 0.0001)\) difference pre- and post injection.

The mean load rate magnitude collected by the smartphone and the smartwatch were depicted in Figure 8.3 to see if the relationship between smartphone and smartwatch was visible. Load rates from the smartphone and smartwatch were highly correlated \((> 0.8)\) both pre and post injection (Table 8.3).

The mean load rate magnitude collected by the smartphone or smartwatch and the pain score were depicted in Figure 8.4 to see if the relationship between load rate and pain
Figure 8.3: The mean load rate magnitude compared between the smartphone and smartwatch data on a log-log scale both pre- and post-injection.

Table 8.3: The Pearson’s correlation coefficients between the smartphone and smartwatch.

<table>
<thead>
<tr>
<th></th>
<th>Pre Injection</th>
<th>Post Injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient all data SP-SW</td>
<td>0.84*</td>
<td>0.81*</td>
</tr>
</tbody>
</table>

*Significant \((p < 0.0001)\) correlation.

was visible. The load rate magnitude was averaged between two pain scores, and linear interpolation was used if time stamps were missing for the smartphone or smartwatch to match the pain score timing (Meijering, 2002). Load rates from the smartphone or smartwatch and the pain score were low \((> 0.35)\) to no correlation \((< 0.35)\)(see Table 8.4).

Table 8.4: The Spearman’s correlation coefficients between smartphone/smartwatch and pain score.

<table>
<thead>
<tr>
<th></th>
<th>Correlation coefficient Smartphone - Pain</th>
<th>p-value</th>
<th>Correlation coefficient Smartwatch - Pain</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre Injection</td>
<td>0.38*</td>
<td>0.03</td>
<td>0.31</td>
<td>0.08</td>
</tr>
<tr>
<td>Post Injection</td>
<td>0.26</td>
<td>0.14</td>
<td>0.17</td>
<td>0.3</td>
</tr>
</tbody>
</table>

*Significant \((p < 0.05)\) correlated.
Chapter 8 Study 4: Single case study to relate load rate estimates and osteoarthritis-related pain using wearables

The mean of the load rate magnitude in relation to an increasing sample size were depicted in Figure 8.5.

An estimate of the load rate was achieved with an error of 2.66% for the smartphone after $t_{SP}^{2.66\%} = 80856064\text{ms}$ (i.e. approximately 0.94 days) and with an error of 2.64% for the smartwatch after $t_{SW}^{1.99\%} = 127622209\text{ms}$ (i.e. approximately 1.48 days). The error values were achieved after seven days of data collection in Study 3 and used as a comparison for Study 4.

8.7 Discussion

The results of the present study show that the load rate data and the pain data were monitored successfully throughout the 14 days data collection by the OApp™ (Figure 8.2). Significant differences were noticed between pre- and post-injection of the load rate data from the smartphone, smartwatch and the pain data (Table 8.2). However, the significant differences were minimal.

Ability of OApp™ (Objective 1) It can be seen in Figure 8.2 that the values for the smartwatch are higher than the ones from the smartphone for the participant with OA in her hip. The higher values of the smartwatch were not noticeable in Study 3,
Chapter 8 Study 4: Single case study to relate load rate estimates and osteoarthritis-related pain using wearables

Figure 8.5: The accumulated mean load rate magnitude with increasing sample size and convergence, which was measured during the 14 days of data collection both pre- and post-injection.

except when participants were cycling outside. The higher values of the smartwatch in Study 4 may be due to the different activity level of the participant in Study 4 to Study 3. The mean pain level of the participant pre-injection was 6.21 out of 10 (Table 8.2), which is equivalent to intensive to very intensive. Her high level of pain in the joints with OA and, hence, her limited physical activity with her lower limbs might explain why the smartphone data were much lower than that for the smartwatch data. The load rate levels recorded by the smartphone were noticeably lower than the ones for healthy participants in Study 3. The smartwatch data, however, do not seem lower than the ones from healthy participants, showing that she used her arms relatively normally. These findings show that estimating load rate on the lower limb joints with smartwatches has limitations and that for people with OA in their lower limbs, using a smartwatch is not a suitable tool for estimating load rate on the lower limb joints due to their activity level.

Differences pre- and post-injection (Objectives 2 and 3) Differences are noticeable between pre- and post-injection in Table 8.2. As previously mentioned, it was expected that the participant would improve their biomechanics during walking post-injection (Koester et al., 2007). This is reflected in the post-injection data, where a significant lower load rate was measured. The IPAQ questionnaire showed that the participant had more physical activity post-injection, which was assumed since she was feeling less pain.
It is clear that a difference is noticeable between the load rate measured by both the smartphone and smartwatch pre- and post-injection; nonetheless, more research is needed to explore the difference between the pre- and post-injection. The pain level recorded by both the wearables and the HAGOS questionnaire decreased post-injection. The decreasing pain level was expected since the literature existing confirms there was a reduction in pain after a CSI was administered (Arden et al., 2008).

**Correlation between smartphone and smartwatch (Objective 4)** The smartphone and smartwatch data correlated well with one another for the participant with lower limb OA (Table 8.3), which can be also seen in Figure 8.3. The post-injection correlation was lower, which was due to the different activity patterns of the participant. It is worth noticing, however, that the correlation for the participant with lower limb OA was higher than that for healthy participants. The significant difference in the values between the smartphone and the smartwatch (Table 8.2) indicates that the smartwatch should not be recommended for estimating load rate on the lower limb joints. More participants, however, would be needed to undergo this study to verify this assumption.

**Correlation between load rate and pain (Objective 5)** The highest and only significant correlation (low positive correlation according to Hinkle et al. (2003) with \( p = 0.03 \)) observed was between the smartphone and the pain pre-injection (Figure 8.4). The correlation decreased post-injection. More research is needed to understand and to relate OA-related pain and estimated load rate on the lower limb joints.

**Accuracy of the load rate data (Objective 6)** To obtain a sufficient estimation of the load rate, \( t_{SP}^{2.66\%} = 80856064\text{ms} \), approximately 0.94 days and \( t_{SW}^{99\%} = 127622209\text{ms} \), approximately 1.48 days were estimated. The value 2.66\% was used because it was the value obtained after 7 days and it was assumed to be a reliable estimate. In Figure 8.5 the converges of the load rate estimate is achieved much faster than to the figures in Study 3. In comparison to Study 3 (7 days were recommended), the time that was needed for a sufficient estimate was much shorter. The regular physical activity patterns of the participant in Study 4 in comparison to the participants in Study 3 were different. Her physical activity seems equally distributed over the seven days, both pre- and post-injection (Figure 8.5) vis-à-vis the physical activity patterns of the participants in Study 3 (Figure 6.14). The participant in Study 4 had no occupation and seemed regularly low to moderate physical active, however, the participants in Study 3 had mostly desk-based occupations and were regularly performing vigorous physical activities.

**Comparative analysis** The data collection method in the present study was similar to the data collection in the work of Murphy et al. (2008). The participants in the Murphy et al. (2008) study were people with lower limb OA reporting pain for at least
three months with a level of mild severity. The present participant had a pre-injection pain level of moderate-sever, which was higher than the inclusion criteria of Murphy et al. (2008). A different mobility and progression of OA might make the comparison impossible. However, a period of five days data collection, as Murphy et al. (2008) used, was short even if they found significant correlations between physical activity and pain. Further, Murphy et al. (2008) used a wrist based accelerometer with a sampling frequency of 32Hz and a pain scale between 0-4. In the current study, smartphone and smartwatches were highly correlated but in Study 3 the smartwatch was not seen as a suitable tool to monitor load rate nor physical activity. In the present study the smartwatch data were noticeably higher and it was concluded that it was not suitable to monitor lower limb load rate or physical activity level. However, comparisons are also difficult due to the different size in the number of the participants.

8.8 Limitations and strengths of the study

The main weakness of the present study is that just one participant was recruited. Due to ethical approval restrictions, the researchers had difficulties during the clinical recruitment stage. The researcher were not experienced in clinical study recruitment, hence, the recruitment process was not feasible and a new recruitment process needed new ethical approval. Hence, an amendment to the ethical approval delayed the recruitment process. Furthermore, the duration of the study was limited to a 14 day period, which might not reveal the true relationship between estimated load rate on the lower limbs and OA-related pain. The progression of OA and its symptoms, such as pain, have complicated relationships which might be influenced by further variables, such as poor biomechanics, how pain is experienced by each individual or genetics (Litwic et al., 2013). Nevertheless, Study 4 will be continued to reach adequate sample size of participants at the Nuffield Orthopaedic Centre in Oxford. Adequate sample size was defined as sample of convenience, such as $n = 20$ for an initial group study. Further, a limitation of the data analysis was that the load rate data had to be normalised with the mean acceleration value before comparison. This would relate the load rate data to the actual physical activity of the participant. This has to be done, because the physical activity of the participants can vary between the days. Further, no minimal clinically important difference was calculated, which has to be done in the continuing clinical study.

The strength of the present study was the introduction of the CSI into the procedure. This made it possible to compare the two phases and to show whether the relationship between pain and estimated load rate changed, i.e. to test the sensitivity of the measures used. Moreover, another strength of the study was that it adopted the suggestion made by Stone et al. (2003): viz., the participant was reminded to score her pain on the smartwatch. This, in turn, led to a high level of compliance since she did not have to take out her phone every time she needed to score her pain. Furthermore, the Monte
Carlo method enabled to save storage space, because the method did not record load rate continuously but sampled randomly.

8.9 Conclusions

OApp™ was able to continually estimate load rate with a sufficient degree of accuracy, with significant differences being found between pre- and post-injection. Research is warranted for exploring the relationship between load rate and OA with OApp™. The smartphone was recommended as a tool to estimate load rate on the lower limbs for people with lower limb OA. The smartwatch, on the other hand, was not recommended for estimating load rate on the lower limbs during the everyday lives of people with lower limb OA. This was due to the significant differences between the smartphone and smartwatch data which was discerned in the present study, even if high correlations were found between the smartphone and smartwatch. For people with an regular activities over a typical week, approximately one day of data collection may be sufficient. In general, further research is recommended to be conducted with an adequate number of participants and with an extended period of time for exploring whether OApp™ can capture a relationship between pain and estimated load rate on the lower limb joints.
Chapter 9

Conclusions and future research

This thesis developed, assessed the validation, and tested a load rate magnitude algorithm which was able to estimate load rate on the lower limb joints using the accelerometer sensors of wearables. The algorithm was used in the form of a smartphone app. At the end of this thesis, the smartphone app, the so-called OApp\textsuperscript{TM}, monitored estimated load rate on the lower limb joints and OA-related pain, which can be used in further research.

In Chapter 2, the viscoelastic characteristic of the biomechanical tissue, such as cartilage, was discussed (Section 2.1.1), which was that the viscoelastic materials’ strain - stress behaviours are time-dependent and, therefore, are prone to higher load rates during loading. In Section 2.1.2 the importance of using external load rate as an indicator for impact loading on the lower limb joints was empathised. This thesis included the weight of participants, estimating the load rate with the use of an accelerometer as the product of acceleration rate magnitude and the body mass of participants. The estimated load rate magnitude was used as a surrogate of impact loading on the lower limb joints during locomotive activities.

In Study 1 (Chapter 4), the validation of load rates estimated with wearables against the gold standard equipment (the force plate) was assessed. Before Study 1, a feasibility study was conducted which refined the methods used in Study 1 (Section 4.3). In Study 1, thanks to the anti-gravity treadmill, each participant delivered for each of the three speed conditions six different weight conditions, creating for each participant 18 sample points. For Studies 1 and 2, the bootstrap method and the cross-validation process were used for the purpose to obtain a better estimate of the uncertainty. A significant correlation was found between the load rate from the force plate and the wearables (smartphone: $R^2 = 0.77$; smartwatch: $R^2 = 0.73$). Wearables can estimate load rate, and the high correlation with the force plate data supports their use as a surrogate when assessing lower limb joint loading in people with osteoarthritis (Section 4.5).
In Study 2 (Chapter 5), a classifier which distinguished between activities outside of the laboratory environment was developed, and the load rate associated with these activities was estimated. A three-layered algorithm consists of 1) load rate estimation from accelerometer data from smart wearable device sensors; 2) activity classification; and 3) load rate categorisation.

Support vector machine classifiers were used to distinguish terrains (accuracy: 56±10%) and activities (accuracy: 80 ± 10%) with load rate-based features. Following classification, within-activity load was categorised into ‘low’, ‘moderate’ and ‘intensive’ loading (Section 5.4). Wearables’ accelerometer sensors, commonly used in the activity tracking of everyday life activities, appears to provide a high level of accuracy for classifying activities and categorising joint loads on the lower limbs during locomotive activities. The classifier in the present study was able to distinguish between classes with similar intensity. The terrain classifier had a low accuracy; further research, therefore, is needed.

In Study 3 (Chapter 8), the equipment’s (i.e. the smartphone and smartwatch with the developed app) ability to estimate load rate on the lower limbs during everyday physical activities. The outcome of Study 3 confirmed that smartphones worn on the hip can be used for estimating load rate during everyday life and that the load rate can be estimated directly from the smartphones in the form of an app. The findings in Study 3 showed that smartwatches were not a reliable method for monitoring impact loading on the lower limb joints for the participant group which took part in Study 3. Furthermore, a period of at least seven days of data collection was recommended for an accuracy with an error of $\hat{e}_{\text{smartphone}} = 2.66\%$ and $\hat{e}_{\text{smartwatch}} = 1.99\%$. The app is recommended to be used for locomotive activates, such as walking and running. The equipment shows limitations with activities including loading on the joints but with lower load rates. Activities with high loading on the joints and low load rates, such as weightlifting, are hypothesised to be activities which are not able to be detected.

In Study 4 (Chapter 8), the ability of a developed non-interventional app, called $OApp^\text{TM}$, was assessed to monitor OA-related pain and load rate using wearables, with the ultimate aim of providing proof of concept. Before Study 4, a patient and public involvement (PPI) consultation session (Chapter 7) was conducted with the purpose of exploring strategies for improving the adherence of participants using the app during Study 4. Considering the consultations of the PPI representatives, the app in Study 3 was further developed for Study 4 and renamed $OApp^\text{TM}$. $OApp^\text{TM}$ was able to estimate load rate with a sufficient accuracy. In the end, differences were discovered between pre- and post-injection OA-related pain. $OApp^\text{TM}$ is recommended to be used in further research an adequate number of participants and a more extended period to explore whether the $OApp^\text{TM}$ can capture a relationship between pain and estimated load rate on the lower limb joints. The limitation of Study 4, was that the $OApp^\text{TM}$ was only validated with a small sample size of one person with OA. Nevertheless, these pilot studies
set the bases for further research in the field of load rate monitoring with commercially available acceleration sensors for people with OA.

A limitation of this thesis is that estimating load rate with wearables might be misleading for the end user. It has to be made clear that the load rate measured with wearables is merely an estimation and can only be used as an indicator for load rate. The end user should know that higher estimated load rates are used as a surrogate for impact loading on the lower limb joints in this thesis and that there are more factors which lead to joint damage and OA which were not considered, such as the poor biomechanics of individuals or genetics.

9.1 Research contribution

This work’s contributions are listed below:

1. Assessed the validation of the smartphone and smartwatch to estimate load rate of the lower limbs against the golden standard equipment, the force plate, during locomotive activities on a treadmill.

2. Developed a classifier which distinguished between locomotive activities outside of the laboratory environment and estimated the load rate associated with these activities.

3. Developed a terrain classifier which was not able to distinguish between different terrains during locomotive activities. Further research is recommended.

4. Tested the equipment’s (wearables with the installed app and included algorithm) ability to estimate load rate on the lower limbs during everyday physical activities during a period of seven days outside the experimental environment.

5. Tested the correlation between the smartphone and smartwatch’s ability to monitor estimated load rate in everyday life and found that the smartwatch is not a suitable tool for estimating load rate on the lower limb joints in non-locomotive activities.

6. Discovered that a period of at least seven days of data collection is recommended for an accuracy with an error of $\hat{e}_{\text{smartphone}} = 2.66\%$ if five seconds are randomly sampled from each minute of the day in healthy people with unequally distributed physical activity patterns.

7. Explored the ability of the developed app, called $OApp^\text{TM}$, to relate OA-related pain with continuous load rate monitored via wearables and found that no such relationship was noticeable. Further research is needed.
8. Explored if the $OApp^\text{TM}$ was able to record noticeable differences pre- and post-injection in momentary OA-related pain and estimated load rate recorded with the wearables.

9. Found that a period of at least one day of data collection is recommended for an accuracy with an error of $\hat{e}_{\text{smartphone}} = 2.66\%$ if five seconds randomly are sampled from each minute of a day in people with lower limb OA with equally distributed physical activity patterns.

The following are how the contributions responded to all of the proposed research aims in each study (some contributions did not directly correspond to a research aim, but were a by-product of this work).

1. The aim of Study 1 was to assess the validity of load rate estimated with wearables against the gold standard equipment, the force plate. This contributions corresponds with contribution 1.

2. The aim of Study 2 was to develop a classifier which classifies activities outside of the laboratory environment while also estimating over the load rate associated with these activities. This contributions corresponds with contributions 2 and 3.

3. The aim of Study 3 was to assess our equipment’s (i.e. the smartphone and smartwatch with the developed app) ability at estimating load rate on the lower limbs during everyday physical activities during a period of seven days. This contributions corresponds with contributions 4, 5 and 6.

4. The aim of Study 4 was to assess the ability of the non-interventional app, called $OApp^\text{TM}$, to monitor load rate and OA-related pain in preparation for use as a tool in clinical studies to explore the relationship between load rate and OA-related pain. This contributions corresponds with contributions 7, 8 and 9.

The validity of the smartphone including the load rate magnitude algorithm was assessed against the gold standard equipment in the laboratory, was tested outside the laboratory environment, and its ability was assessed to record estimated load rate on the lower limbs outside the experimental environment in a prospective study. Furthermore, this thesis tried to assess the ability of wearables to relate OA-related pain and estimated load rate. Further research, however, is required.

The strength of this work was that the research was conducted on a life science interface. The interdisciplinary research connected research in engineering and data science with life science disciplines for the improvement of the health and well being of people with OA.
9.2 Future work

This thesis sets the basis for further research in the field of load rate monitoring with commercially available acceleration sensors for people with OA. This section suggests future work to complement the contributions of this work.

9.2.1 Extension of Study 4

OApp™ needs to be validated with an adequate number of people with OA. The ethical approval of Study 4 is continuing until the end of 2018, where more people with OA will be recruited from the Nuffield Orthopaedic Centre in Oxford for to validate OApp™.

9.2.2 Clinical and commercial use of OApp™

OApp™ has great potential to be used clinically as an interventional tool. Therefore, OApp™ has to be further developed into an interventional tool where users can see their monitored load rate and pain on the screen of the app. In that way, people with OA will be able to monitor their estimated load rate on their lower limbs and score their pain and, in turn, observe how their physical activity influences their OA-related pain. A longitudinal study validating the advanced version of OApp™ with an adequate number of people with OA is suggested to assess its ability to help people exercise safely with low load rate levels. A control group is also suggested which will not be able to see their load rate and pain on the app to test the differences between them and the focus group. Introducing a control group to the study might enable the validation of the OApp™ as a clinical and commercially available tool for estimating load rate through the lower limb joints and for relating OA-related pain.

9.2.3 Epidemiological studies

Further, it is suggested to use the load rate magnitude algorithm to conduct epidemiological studies on OA to discover the relationship between the progression of OA and weight-bearing joint loading with commercially available acceleration sensors. The researcher suggests that hip-worn accelerometers are given for several years to people with a high risk of developing OA on their lower limbs, such as athletes who participate in regular high load rate activities. During the study, it is suggested that imaging tests are conducted, such as X-ray or magnetic resonance imaging, to regular test the progression of OA on their lower limb joints. Moreover, motion capture analysis is suggested to exclude confounders, such as poor biomechanics, from the analysis. This epidemiological study might reveal the relationship between load rate on the lower limb joints and the progression of OA. Another suggestion would be that the load rate magnitude algorithm
can be used to conduct further research in the epidemiological studies, which were already carried out, such as Farr et al. (2008) study. The idea is that the accelerometer data collected by Farr et al. (2008) can be transformed to a mean load rate magnitude. The results can then be compared to those obtained in Farr et al. (2008) and can, thus, help researchers better understand the progression of OA.

9.2.4 Load rate-based classifier

The terrain classifier in Study 2 (Chapter 5) was the first step undertaken in this field. However, the terrain classification had low accuracy and more research is suggested. The low accuracy, of the terrain classifier was assumed to be due to the relatively low sampling frequency of the wearables since they were not able to capture the difference. The MEMS industry is growing fast and the sampling frequencies of wearables, or commercially available sensors in general, can be assumed to increase within the near future. A higher sampling frequency and more extended periods of data collection might help to increase their ability to classify terrains during locomotive activities. Furthermore, the activity classifier in Study 2 had a high accuracy, but there is room for improvement.

A high sampling frequency, more extended periods of data collection, and a significant number of participants might help to improve the accuracy of the classifier. If terrains can be distinguished with load rate-based classifiers, then footwear might be able to be distinguished as well. The researcher recommends further research in barefoot, minimalist and lightweight shod running. De Wit et al. (2000) and Bonacci et al. (2013) researched in this field. However, they were conducted in the laboratory with force plates. The researcher suggests researching the different impact characteristics of the shoes in-depth with the help of the load rate magnitude algorithm outside the laboratory environment.

9.3 Summary

The contributions made by this thesis have an impact on the future of estimated load rate monitoring on the lower limb joints with commercially available acceleration sensors. The results of this thesis can be used for further research, which might support people with OA to exercise safely with less weight-bearing joint loading or, furthermore, might help researchers to understand better the relation between load rate on the lower limb joints and OA-related pain.

In summary, this thesis has succeeded in producing novel contributions which have resulted in direct key contributions on the academic work, not only of the researcher but of other PhD students as well. Not only was the load rate magnitude algorithm, which was built into the OApp™, validated and tested, but research contributions have
been made which could facilitate and expedite future work in load rate monitoring using commercially available acceleration sensors.
Appendix A

Results of the literature review
signal processing within physical
activity monitoring
Table A.1: Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 1)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sensor Location</th>
<th>Sampling Frequency</th>
<th>Segmentation window</th>
<th>features</th>
<th>classification Method</th>
<th>Test subject information</th>
<th>Activity</th>
<th>Statistics</th>
<th>Accuracy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al (2012)</td>
<td>right wrist</td>
<td>5Hz, 10Hz, 20Hz, 40Hz, 80Hz</td>
<td>12.8 s</td>
<td>average, deviation (STD), dominant frequency, power of dominant frequency, total power, wavelet decomposition</td>
<td>logistic regression, decision tree, support vector machine (SVM)</td>
<td>60 healthy subjects (37F/23M, ages 40-60)</td>
<td>lying, standing, seated computer work, 4km walk, 5km walk, 6km walk, walking up and down stairs, free-living 6km walk, two household activities randomly selected from window washing, washing up, shelf stacking, and sweeping, running (8km, 10km, 12km)</td>
<td>Kappa statistics, mean absolute error, root mean square error</td>
<td>SVM: 97.1% ± 0.72%, logistic regression: 96.4% ± 0.91%, Bayesian network: 90.26% ± 1.78%</td>
<td>good comparison of different frequencies</td>
</tr>
<tr>
<td>Gyllensten and Bonomi (2011)</td>
<td>waist</td>
<td>20 Hz</td>
<td>6.4 s</td>
<td>mean, standard deviation, kurtosis, skewness range, cross-axis correlation, accelerometer angle, spectral energy, spectral entropy, peak frequencies, cross-spectral densities</td>
<td>decision tree, neural network (NN), SVM, majority voting</td>
<td>20 healthy subject (10F/10M)</td>
<td>lying, sitting, standing, dynamic transitions, walking, cycling, running</td>
<td>leave-one-subject out cross validation (LOSO CV), holdout validation</td>
<td>Laboratory data: Majority voting: 95.1% ± 4.2%, SVM: 95.4% ± 5.1%, NN: 91.4% ± 6.7%, decision tree: 92.2% ± 6.6% Daily-Life Data Majority voting: 75.7% ± 9.6%, SVM: 75.6% ± 10.4%, NN: 74.8% ± 9.7%, decision tree: 72.2% ± 10.3%</td>
<td>Laboratory data vs. daily-life: laboratory measures of accuracy are not reliable indicators for classification performance in daily life. While the models were stable under laboratory conditions, performance in free-living conditions was significantly lower.</td>
</tr>
<tr>
<td>Khan et al (2014)</td>
<td>trousers front pockets, trousers back pockets, jackets inner pocket</td>
<td>20 Hz</td>
<td>3.5 s</td>
<td>mean, standard deviation, correlation, signal magnitude area (SMA), coefficients of time series analysis: autoregressive (AR) analysis, moving average (MA) analysis</td>
<td>SVM</td>
<td>30 healthy subject (12F/18M, ages 26-35)</td>
<td>walking, walking on treadmill, running, running on treadmill, going upstairs, going downstairs, riding elevator up, riding elevator down, hopping, riding a bike, sitting/standing, watching TV, vacuuming, driving a car, riding a bus</td>
<td>offline recognition via 10-fold cross-validation, subject-independent offline recognition via LOSO, offline evaluation of different sensors and combination of sensors, subject-independent online recognition on smartphones using eight new subjects, comparison with previous HAR systems</td>
<td>offline subject-dependent test: 99.1%, offline subject-independent test: 94%, online subject-independent test: 24%</td>
<td>Sensor fusion with accelerometer, pressure sensor, microphone. System employed nonlinear discriminant approach (kernel discriminant analysis) with nonlinear classifier (SVM). Study comparison kernel, and use joint time-domain features.</td>
</tr>
<tr>
<td>Authors</td>
<td>Sensor Location</td>
<td>Sampling frequency</td>
<td>Segmentation window</td>
<td>Features</td>
<td>Classification Method</td>
<td>Test subject information</td>
<td>Activity</td>
<td>Statistics</td>
<td>Accuracy</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>---------------------------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Long et al. (2009)</td>
<td>waist</td>
<td>20 Hz</td>
<td>16 s</td>
<td>time, frequency, spatial domain</td>
<td>Bayesian classification</td>
<td>24 healthy subject (11F/13M age 26-55)</td>
<td>walking, running, cycling, driving, sport</td>
<td>LOSO CV, 10-fold CV</td>
<td>LOSO CV: walking=80.3%, running=92.9%, cycling=49.4%, driving=94.3%, sports=71.3%, 1-fold CV: walking=77.3%, running=95.5%, cycling=49.4%, driving=86.9%, sports=57.9%</td>
<td>Used FCA</td>
</tr>
<tr>
<td>Chung et al. (2008)</td>
<td>chest</td>
<td>50 Hz</td>
<td>3 s</td>
<td>Fourier components, power spectral density (PSD)</td>
<td>decision tree</td>
<td>1 subject (10 experiments each 5min)</td>
<td>rest, walking, running</td>
<td>frequency domain</td>
<td>81.25%</td>
<td>Frequency domain can be used to distinguish between activity, but not intensity. Study has a small amount of validation through participants, and no clear validation method.</td>
</tr>
<tr>
<td>Bersch et al. (2014)</td>
<td>ankle, thigh, wrist, hip, upper arm</td>
<td>10 Hz, 20 Hz, 30 Hz, 40 Hz, 50 Hz, 60 Hz</td>
<td>32 different sizes 0.5s-24s</td>
<td>root mean square (RMS), mean, SMA, signal vector magnitude, energy, entropy, FFTPeak, STD</td>
<td>Naive Bayes, SMO (based on SVM), KNN, KStar, MultiClass-Classifier, bagging, decision table, 348, and random forest</td>
<td>20 subjects (7F/13M, age 16-29)</td>
<td>walking, sitting, walking, carrying an item, standing still, lying down, climbing stairs</td>
<td>ANOVA, 10-fold CV</td>
<td>ANOVA to find out the correlation between, sampling frequency, segmentation window/ method, classification method.</td>
<td></td>
</tr>
<tr>
<td>Deng et al. (2014)</td>
<td>waist</td>
<td>50 Hz</td>
<td>2.56 s with 50% overlapping</td>
<td>mean, standard deviation, energy, mean-crossing rate, maximum value, minimum value, first quartile, second quartile, third quartile, PSD, FFT</td>
<td>kernel extreme learning machine + SVM</td>
<td>30 healthy subjects (age 19-48)</td>
<td>walking, walking upstairs, walking downstairs, sitting, standing, lying</td>
<td>Cross-person validation</td>
<td>RKELM=98.49%, Huang’s KELM=99.05%, ELM=96.45%, SVM=98.77%</td>
<td>More accurate method then SVM to record activities, but not suitable because expensive storage</td>
</tr>
<tr>
<td>Albert et al. (2012)</td>
<td>waist, back of the subject</td>
<td>20 Hz</td>
<td>10 s</td>
<td>moments, smoothed root mean square, extremum, histogram, Fourier components, mean acceleration magnitude, mean</td>
<td>SVM, sparse multinomial logistic regression (SMLR), Naive Bayes, decision tree, K-nearest neighbours (KNN)</td>
<td>15 healthy subject (4F/7M, age 22-50)</td>
<td>slip-backward, trip-forward, left/right lateral</td>
<td>10-fold cross validation; subject-wise cross validation</td>
<td>decision trees, KNN=94-98%, Naive Bayes=63-86%, SVM/ SMLR=99%</td>
<td>Example for movement with high acceleration</td>
</tr>
</tbody>
</table>

Table A.2: Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 2)
<table>
<thead>
<tr>
<th>Authors</th>
<th>Sensor Location</th>
<th>Sampling frequency</th>
<th>Segmentation window</th>
<th>Features</th>
<th>Classification Method</th>
<th>Test subject information</th>
<th>Activity Information</th>
<th>Statistics</th>
<th>Accuracy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayat et al. (2014)</td>
<td>hand, pocket</td>
<td>100 Hz</td>
<td>1.28 s with 50% overlapping</td>
<td>mean, MinMax, root mean square, average of peak frequency (APF), VarAPF, STD, correlation, MinMax</td>
<td>multilayer perceptron, random forest, logistic model trees (LMT), SVM, simple logistic, LogitBoost</td>
<td>4 healthy subject, 29-33</td>
<td>running, slow-walk, fast-walk, aerobic dancing, stairs-up, stairs-down</td>
<td>10-fold CV</td>
<td>in-hand: multilayer perceptron= 89.48%, SVM= 88.76%, random forest= 87.55%, LMT= 85.89%, simple logistic= 85.41%, logit boost= 82.54%, in-pocket: multilayer perceptron= 89.72%, SVM= 72.27%, random forest= 85.15%, LMT= 85.04%, simple logistic= 85.054%, logit boost= 82.24%</td>
<td>Compares different classifier to combination of classifier. Combining multiple good classifiers can improve accuracy, efficiency and robustness over single classifiers.</td>
</tr>
<tr>
<td>Gastin et al. (2014)</td>
<td>upper back between shoulder blades</td>
<td>100 Hz</td>
<td>player load</td>
<td>player load</td>
<td>decision tree</td>
<td>20 healthy subjects (elite male Australian football (AF) from one AF league team, ages 20-29)</td>
<td>low, medium and high tackle intensity</td>
<td>cross validation</td>
<td>78%</td>
<td>Low-intensity tackles were not as easily detected as medium and high-intensity tackles.</td>
</tr>
<tr>
<td>Sama et al. (2011)</td>
<td>waist</td>
<td>50 Hz</td>
<td>20 different sizes 1-30 s</td>
<td>module, orientation angles, vertical and forward components, energy expenditure</td>
<td>SVM</td>
<td>10 healthy subject (ages 25-50)</td>
<td>stay steady, walk, sit down, sit, stand up</td>
<td>10-fold CV</td>
<td>optimum vector length(7s)= 91.06%, stand up= 75.56%, sit down= 89.09%, steady 93.33%, walk= 95.56%</td>
<td>Comparison of feature vector length and kernels. Feature selection algorithm combined with kernel methods has provided has been used. PCA</td>
</tr>
</tbody>
</table>

Table A.3: Comparing different conditions of studies focusing on activity monitoring with acceleration measurements (part 3)
Appendix B

Documents of ethical approval
Submission Number: 17086
Submission Name: Monitoring mechanical loading of the body during physical activity via smartphones
This is an email to let you know your submission has been reviewed and approved by your supervisor.
It has now been sent to the Ethics committee for review.

Comments
None
Click here to view your submission

ERGO: Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL
Subject: Your Ethics Amendment (Ethics ID:30213) has been reviewed and approved

Date: 29 August 2017 at 09:17

To: s.nazirizadeh@soton.ac.uk

Submission Number 30213:
This email is to confirm that the amendment request to your ethics form (Monitoring estimated load rate with smartphones and smartwatches (Amendment 1)) has been approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

Comments
None

Click here to view your submission
Coordinator: Susan Nazirizadeh

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL
Dear Prof Arden,

Study title: The relationship between lower limb joint pain and physical activity recorded via smartphone in people with osteoarthritis
IRAS project ID: 211265
REC reference: 16/NI/0228
Sponsor: University of Southampton

I am pleased to confirm that HRA Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. Please read Appendix B carefully, in particular the following sections:

- Participating NHS organisations in England – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities.
- Confirmation of capacity and capability - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.
It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices
The HRA Approval letter contains the following appendices:
- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval
The document “After Ethical Review – guidance for sponsors and investigators”, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:
- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:
- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the After Ethical Review document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope
HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback
The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application
procedure. If you wish to make your views known please email the HRA at hra.approval@nhs.net. Additionally, one of our staff would be happy to call and discuss your experience of HRA Approval.

**HRA Training**

We are pleased to welcome researchers and research management staff at our training days – see details at [http://www.hra.nhs.uk/hra-training/](http://www.hra.nhs.uk/hra-training/)

Your IRAS project ID is **211265**. Please quote this on all correspondence.

Yours sincerely

Simon Connolly  
Senior Assessor

Email: hra.approval@nhs.net

**Copy to:**  
Diana Galpin, University of Southampton  
Jennifer Peach, University Hospital Southampton NHS Foundation Trust R&D Department
## Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

<table>
<thead>
<tr>
<th>Document</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies of advertisement materials for research participants</td>
<td>1</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>[Advertisement Poster Oxford]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copies of advertisement materials for research participants</td>
<td>1</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>[Advertisement Poster Southampton]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)</td>
<td>1</td>
<td>12 August 2016</td>
</tr>
<tr>
<td>[Insurance confirmation Letter from sponsor]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRAS Application Form [IRAS_Form_05102016]</td>
<td></td>
<td>05 October 2016</td>
</tr>
<tr>
<td>Other [Correspondence from MHRA to confirm not a device study]</td>
<td></td>
<td>21 September 2016</td>
</tr>
<tr>
<td>Other [Risk Assessment]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other [Schedule of events]</td>
<td>1</td>
<td>02 August 2016</td>
</tr>
<tr>
<td>Other [mNCA]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant information sheet (PIS) [Participant Information Sheet Southampton, amended]</td>
<td>9</td>
<td>17 October 2016</td>
</tr>
<tr>
<td>Summary CV for Chief Investigator (CI) [Nigel Arden CV]</td>
<td>1</td>
<td>01 October 2015</td>
</tr>
<tr>
<td>Summary CV for student [Susan's CV]</td>
<td></td>
<td>25 August 2016</td>
</tr>
<tr>
<td>Summary CV for student [Jimmy Caroupapoulié CV]</td>
<td></td>
<td>21 September 2016</td>
</tr>
<tr>
<td>Summary CV for supervisor (student research) [Maria Stokes CV]</td>
<td></td>
<td>24 August 2016</td>
</tr>
<tr>
<td>Summary CV for supervisor (student research) [Alex Forrester CV]</td>
<td></td>
<td>25 August 2016</td>
</tr>
<tr>
<td>Validated questionnaire [Knee injury Osteoarthritis Outcome Score (KOOS)]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Validated questionnaire [Hip and Groin Outcome Score (HAGOS)]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Validated questionnaire [International Physical Activity Questionnaire (IPAQ)]</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the, participating NHS organisations, capacity and capability and Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: Diana Galpin
Tel: 02380595058
Email: rgoinfo@soton.ac.uk

HRA assessment criteria

<table>
<thead>
<tr>
<th>Section</th>
<th>HRA Assessment Criteria</th>
<th>Compliant with Standards</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>IRAS application completed correctly</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>2.1</td>
<td>Participant information/consent documents and consent process</td>
<td>Yes</td>
<td>Requested that IRAS reference appear on PIS and consent forms.</td>
</tr>
<tr>
<td>3.1</td>
<td>Protocol assessment</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>4.1</td>
<td>Allocation of responsibilities and rights are agreed and documented</td>
<td>Yes</td>
<td>HRA has been informed that sponsor intends for unmodified mNCA to be used with participating NHS organisations.</td>
</tr>
<tr>
<td>4.2</td>
<td>Insurance/indemnity arrangements assessed</td>
<td>Yes</td>
<td>Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this research study</td>
</tr>
<tr>
<td>Section</td>
<td>HRA Assessment Criteria</td>
<td>Compliant with Standards</td>
<td>Comments</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>4.3</td>
<td>Financial arrangements assessed</td>
<td>Yes</td>
<td>Study forms part of PhD. No funding to be provided to participating NHS organisations by sponsor.</td>
</tr>
<tr>
<td>5.1</td>
<td>Compliance with the Data Protection Act and data security issues assessed</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>5.2</td>
<td>CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed</td>
<td>Not Applicable</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>Compliance with any applicable laws or regulations</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>6.1</td>
<td>NHS Research Ethics Committee favourable opinion received for applicable studies</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>6.2</td>
<td>CTIMPS – Clinical Trials Authorisation (CTA) letter received</td>
<td>Not Applicable</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>Devices – MHRA notice of no objection received</td>
<td>Not Applicable</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>Other regulatory approvals and authorisations received</td>
<td>Not Applicable</td>
<td></td>
</tr>
</tbody>
</table>

**Participating NHS Organisations in England**

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

At participating NHS organisations participants will be recruited and the research activities described in submission will take place.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local LCRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for
participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

### Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.

<table>
<thead>
<tr>
<th>Participating NHS organisations in England will be expected to formally confirm their capacity and capability to host this research.</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Following issue of this letter, participating NHS organisations in England may now confirm to the sponsor their capacity and capability to host this research, when ready to do so. How capacity and capacity will be confirmed is detailed in the Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) section of this appendix.</td>
</tr>
<tr>
<td>• The Assessing, Arranging, and Confirming document on the HRA website provides further information for the sponsor and NHS organisations on assessing, arranging and confirming capacity and capability.</td>
</tr>
</tbody>
</table>

### Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).

<table>
<thead>
<tr>
<th>A principal investigator will be in place at each participating NHS organisation. Any study training required will be provided.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCP training is not a generic training expectation, in line with the HRA statement on training expectations.</td>
</tr>
</tbody>
</table>

### HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken.

| Where existing arrangements are not in place university researchers will require a letter of access to complete research activities within the NHS. It will need to be confirmed that appropriate DBS and occupational health checks have taken place. |

### Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.

| The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio. |

---

Page 7 of 7
11 October 2016

Prof Nigel K Arden
Nuffield Department of Orthopaedics
Windmill Road
University of Oxford
OX3 7LD

Dear Prof Arden

Study title: The relationship between lower limb joint pain and physical activity recorded via smartphone in people with osteoarthritis
REC reference: 16/NI/0228
IRAS project ID: 211265

The Proportionate Review Sub-committee of the HSC REC A reviewed the above application on 10 October 2016.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Mr Matthew Mills, RECA@hscni.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval.

Office for Research Ethics Committees (ORECNI)
Customer Care & Performance Directorate
Lissue Industrial Estate West
Rathdown Walk
Moira Road
Lisburn
BT28 2RF
Tel: 028 95361400
www.orecni.hscni.net
HSC REC A
Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion with additional conditions of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Please amend the participant information sheet as follows:

1. Replace ‘waste’ with ‘waist’.
2. Include ‘This study has been given a favourable opinion by Health and Social Care Research Ethics Committee A.’

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).


Where a NHS organisation’s role in the study is limited to identifying and referring potential participants to research sites (“participant identification centre”), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.
To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see “Conditions of the favourable opinion”).

Approved documents

The documents reviewed and approved were:

<table>
<thead>
<tr>
<th>Document</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies of advertisement materials for research participants [Poster Oxford]</td>
<td>1</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>Copies of advertisement materials for research participants [Poster Southampton]</td>
<td>1</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Insurance confirmation Letter from sponsor]</td>
<td>1</td>
<td>12 August 2016</td>
</tr>
<tr>
<td>IRAS Application Form [IRAS_Form_05102016]</td>
<td></td>
<td>05 October 2016</td>
</tr>
<tr>
<td>IRAS Checklist XML [Checklist_05102016]</td>
<td></td>
<td>05 October 2016</td>
</tr>
<tr>
<td>MHRA Notice of No Objection Letter (Medical Devices) and relevant correspondence [Responds email]</td>
<td></td>
<td>21 September 2016</td>
</tr>
<tr>
<td>Other [Risk Assessment]</td>
<td>1</td>
<td>09 June 2016</td>
</tr>
<tr>
<td>Other [mNCA form blank]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant information sheet (PIS) [Participant Information Sheet Oxford]</td>
<td>8</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>Other [Susan GCP ]</td>
<td></td>
<td>20 October 2015</td>
</tr>
<tr>
<td>Participant consent form [Informed Consent Form Oxford]</td>
<td>8</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>Participant consent form [Informed Consent Form Southampton ]</td>
<td>8</td>
<td>03 August 2016</td>
</tr>
<tr>
<td>Summary CV for Chief Investigator (CI) [Nigel Arden CV]</td>
<td>1</td>
<td>01 October 2015</td>
</tr>
<tr>
<td>Summary CV for student [Susan's CV]</td>
<td></td>
<td>25 August 2016</td>
</tr>
<tr>
<td>Summary CV for student [Jimmy Caroupapoli#233; CV]</td>
<td></td>
<td>21 September 2016</td>
</tr>
<tr>
<td>Summary CV for supervisor (student research) [Maria Stokes CV]</td>
<td></td>
<td>24 August 2016</td>
</tr>
<tr>
<td>Summary CV for supervisor (student research) [Alex Forrester CV]</td>
<td></td>
<td>25 August 2016</td>
</tr>
<tr>
<td>Validated questionnaire [Knee injury Osteoarthritis Outcome Score (KOOS)]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Validated questionnaire [Hip and Groin Outcome Score (HAGOS)]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Validated questionnaire [International Physical Activity Questionnaire (IPAQ)]</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Membership of the Proportionate Review Sub-Committee
The members of the Sub-Committee who took part in the review are listed on the attached sheet.

**Statement of compliance**

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

**After ethical review**

**Reporting requirements**

The attached document “After ethical review – guidance for researchers” gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

**User Feedback**

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: [http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/](http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/)

**HRA Training**

We are pleased to welcome researchers and R&D staff at our training days – see details at [http://www.hra.nhs.uk/hra-training/](http://www.hra.nhs.uk/hra-training/)

With the Committee’s best wishes for the success of this project.

**16/Ni/0228** Please quote this number on all correspondence

Yours sincerely

Ms Celia Diver-Hall
pp HSC REC A Alternate Vice-Chair (Chair of the meeting)

Email: RECA@hscni.net

**Enclosures:**

- List of names and professions of members who took part in the review
- “After ethical review – guidance for researchers” [SL-AR2]
Attendance at PRS Sub-Committee of the REC meeting on 10 October 2016

Committee Members:

<table>
<thead>
<tr>
<th>Name</th>
<th>Profession</th>
<th>Present</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms Margaret Brady</td>
<td>Deputy Chief Education Welfare Officer Operations</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Mrs Celia Diver-Hall</td>
<td>Macmillan Acute Oncology Nurse</td>
<td>Yes</td>
<td>Chair</td>
</tr>
<tr>
<td>Dr Orla Quigley</td>
<td>General Practitioner (Retired)</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Also in attendance:

<table>
<thead>
<tr>
<th>Name</th>
<th>Position (or reason for attending)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Siobhan McGrath</td>
<td>Head of the ORECNI</td>
</tr>
</tbody>
</table>
Appendix C

Questions and notes of the PPI session

C.1 Questions asked during PPI session

Imagine you are taking part in the 2-week study, which has just been described to you and discussed.

1. If you had to record your OA related pain regularly over a 2-week period, would you prefer to use a
   (a) Smartphone
   (b) Smartwatch
   Please explain your reason briefly.

2. The smartphone would record your physical activity automatically during each day of the 2 weeks. To record pain, do you think it might be easier and more convenient to use a smartwatch instead of a smartphone?
   (a) Yes
   (b) No
   Please explain your reason briefly

3. How often would you be prepared to enter your OA related pain level on the app?
   (a) Once a day
   (b) Three times a day
   (c) Five times a day
   (d) Ever second hour during the day
(e) Every hour during the day

Please explain your reason briefly

4. How would you prefer to record your pain?
   
   (a) On a scale that is a straight line from 1-10
   (b) Numbers in a circle (similar to a round clock face)

5. Please feel free to make any other comments.

C.2 Notes of the discussions

PPI representative 1

1. If you had to record your OA related pain regularly over a 2-week period, would you prefer to use a
   
   • Answer: “Smartwatch”

2. How often would you be prepared to enter your OA related pain level on the app?
   
   • Answer: “3 times / day would be enough for me, as I don’t get a lot of pain”

3. Would you lie to be reminded or record your pain on your own?
   
   • Answer: “be reminded”

4. How would you prefer to record your pain?
   
   • Answer: “On a linear scale from 1-10”

PPI representative 2

1. If you had to record your OA related pain regularly over a 2-week period, would you prefer to use a

   • Answer: “Probably the phone for normal things. Can I record it sometimes on one and sometimes on the other for recording activity? I wouldn’t have my phone in my pocket in the house, so would use the watch at home. Outside the house, I would carry the phone. Recording activity - Exercises in water, so would wear watch for hydro, rather than phone for that exercise. Recording pain - almost pain free due to hydro exercise. Depends on frequency. If it was 3 times a day, I would use the phone, if it was hourly, I would use the watch.”
2. How often would you be prepared to enter your OA related pain level on the app?
   - Answer: “I would choose three times a day.”

3. Would you like to be reminded or record your pain on your own?
   - Answer: “be reminded. A buzz ’that would be very good’ The only time I
     wouldn’t respond and complete the pain when I was driving and I may not
     remember to do it when I got home.”

4. How would you prefer to record your pain?
   - Answer: “Doesn’t really matter. Doesn’t make any difference to me. If I had
     to choose, I’d probably choose that one’ - pointer to linear scale.”
Appendix D

Diary for Study 3
Appendix 1: Diary

Study title: Monitoring estimated load rate with smartphones and smartwatches

Researcher name: Alex Forrester
Study reference: 30213

Participant's Name:

Please fill in the diary each day to show what physical activities you have undertaken, such as a run, sport, walk. Please state the time of day and how long each activity lasted (approximate times will be sufficient)

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Time</td>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E

Consent and participant information sheets
Participant Information Sheet

Study Title: Monitoring Mechanical Loading of the body during physical activity via Smartphones

Lead Researcher: Susan Nazirizadeh
Ethics number: 17086

Thank you for taking the time to read this information. Please read it carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

This study is part of the PhD project "Activity Monitoring via smartphones with reference to load on the lower limb joints". The project is lead by Susan Nazirizadeh, a PhD student in the Faculty of Health Sciences, with an engineering background. The PhD aims to develop a smartphone app, which estimates how much load goes through the joints to support people with osteoarthritis (OA) to have an active lifestyle.

Even if the reasons for the progression of OA are not entirely understood, an active lifestyle can control symptoms of OA such as pain and stiffness of joints. This project uses the smartphone to develop a clinical tool that can be used to support patients to have an active lifestyle. The smartphone is worn close to the body, for example in a pocket, and the sensors in the smartphone collect movement data of the user.

This study is one of three studies for the overall PhD project. The aim of this study is to collect data using a smartphone worn by a participant, while walking and running on an anti-gravity treadmill with different body weight percentages. An anti-gravity treadmill can lift the treadmill user up, so that the bodyweight of the user decreases and less load affects the joints. This is used for rehabilitation patients to reduce load on the joints. The movement data will help to estimate the load that goes through the joints in the smartphone app being developed.

Why have I been chosen?

You have expressed an interest in the study and are healthy and in the right age group. We are seeking healthy people, aged 18-60 years, because the participants have to walk and run at different speeds for 45min on a treadmill and then jog 30min outside.

Anyone with any of the following will not be able to take part in the study: lower limb pain, diagnosed with lower limb pathology, any musculoskeletal, neurological or systemic diseases, skin diseases, lower limb or spinal fractures and with cardiac issues and respiratory, eg. severe asthma.

What will happen to me if I take part?

The first part of the testing session will be the data collection for the anti-gravity treadmill study and will take place at the Southampton Football Club training facilities at Marchwood, where the treadmill is housed. The researcher team will provide a car to transport you to the facilities. But if you wish to go there by you one, the transportation cost will be reimbursed.

Friday, 17 November 2017, Version 2
The second part will take place on Southampton Common. These two sessions will be on two different days and the date and time of testing will be agreed between you and the researchers. The researcher will explain what you will be asked to do and you will have the opportunity to ask any questions before testing begins. If you are happy to take part in the study we will ask you to sign an informed consent form.

You will be asked to bring a pair of running shorts or tights and running shoes to the testing session, which will not restrict your movement. The testing will begin with providing personal details and your height and weight will be measured. You will be asked to wear neoprene compression shorts and a sports vest over your clothes, which will be provided by the researchers. There will be two smartphones, one smartwatch, one heart rate monitor and two additional motion sensors used for the first testing session. One smartphone will be placed in the sports vest, on your back between your shoulder blades and the second one will be placed at the right side of your hip. The smartwatch will be placed on your left wrist and the heart rate monitor will be placed under your shirt on your chest. Another pair of sensors will be strapped to your right leg. Additionally, you will be asked to wear a small rucksack with additional weight. After the preparation you will be asked to step into the anti-gravity treadmill and you will be sealed to the anti-gravity treadmill by zipping the shorts and treadmill opening together. After you are fixed into the machine the first trial will start.

Each trial will last two minutes and a button on the smartwatch will start and stop the recording. A researcher will be there to assist you for the whole session. Another researcher will be present to record the time and other information regarding the recorded data. The process will consist of three active phases (either walking or running) and two resting periods. The three phases will be walking 5 km/h, running 8 km/h and running 12 km/h. Every phase will consists of six trials and a resting period. The treadmill allows different levels of body weight to be exerted that are lighter than your weight usually is with the effect of gravity. Six different percentages of body weight will be used, in random order, during each of the three activity phases.

A second session will be carried out at least two days later and will take place on Southampton Common, close to the University. You will be asked again to wear shorts and running shoes, and will be provided with a sports vest to hold the smartphone. The same recording devices will be places in the same places on your body as on the first day. Before the session we will show you the path on a map. This time it will be a 30min jogging session, with Dr Alex Forrester accompanying you on a bike, to be sure that you are on the right path.

Are there any benefits in my taking part?

There are no direct health benefits to you for taking part in this study. The information obtained from this study will be used to develop a smartphone app, which shows OA patients the load on their joints caused by their physical activity.

Are there any risks involved?

There are minimal risks to taking part in this study and the sensors will not cause any pain or discomfort. There is a possibility that you may feel some delayed muscle pain after exercising if you are not used to running.
Will my participation be confidential?

All data collected will be kept confidential in compliance with the Data Protection Act and University of Southampton data protection policy. To keep data confidential all written data will be stored in a locked filing cabinet and all electronic data will be stored on password protected computers that only the researchers have access to. Your data will be anonymised by assigning a unique ID number to your data. No one, other than the researchers, will be able to link your name with the unique ID number. Data that will be published in journal articles, conferences or meetings will not report any names or unique ID numbers to maintain the anonymity of the data. All data will be kept for 10 years.

What happens if I change my mind?

Participation in the study is entirely voluntary and you may withdraw at any point during the study, without giving a reason or affecting your rights.

What happens if something goes wrong?

In the unlikely case that you fall/ or have any breathing problems or pain the trail will immediately be stopped. If you have a problem with your breathing an inhaler will be available for the first aid. But if the problems continue the emergency will be called to be sure that you get the right treatment.

If you have a concern or a complaint about this study you should contact Diana Galpin at the Research Governance Office (Address: University of Southampton, Building 37, Highfield, Southampton, SO17 1BJ; Tel: +44 (0)23 8059 5058; Email: rgoinfo@soton.ac.uk). If you remain unhappy and wish to complain formally Diana Galpin can provide you with details of the University of Southampton Complaints Procedure.

Where can I get more information?

If you require any further information or have any questions regarding taking part in this study please contact Susan Nazirizadeh on the details below:

Susan Nazirizadeh, Dipl.-Ing. Ph.D. Student
Faculty of Engineering and the Environment & Faculty of Health Sciences
Computational Engineering and Design Research Group &
Rehabilitation and Health Technologies Research Group
Building 176
Boldrewood Campus
University of Southampton
Southampton SO16 7QF
Tel. +44 (0) 23 8059 5194
Email: s.nazirizadeh@soton.ac.uk

Supervisors:
Dr Alex Forrester email: alexander.forrester@soton.ac.uk
Professor Maria Stokes email: m.stokes@soton.ac.uk
Patient Information Sheet

Monitoring estimated load rate with smartphones and smartwatches

Thank you for taking the time to read this information. Please read it carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

Invitation

You are invited to take part in our study, in which we will record your overall movement using a smartphone and a smartwatch. This will allow the analysis described below. Additionally, you are being asked to give consent for the results from this study to be published in medical journals and in conference presentations. Personal details and means of identification of individuals will remain confidential, and will not be made accessible to the public.

What is the research about?

This study is part of the PhD project “Activity Monitoring via smartphones with reference to load on the lower limb joints”. The project is led by Dr Alex Forrester at the University of Southampton, with a background in engineering. The aim of the study is to develop a smartphone app, which estimates impact loading on the lower limbs, in order to encourage people with osteoarthritis (OA) to maintain an active lifestyle.

Although, the reasons for the progression of OA are not entirely understood, an active lifestyle can improve symptoms of OA, such as pain and stiffness of joints. The current study might help to develop a clinical tool in a form of a smartphone app to encourage people with OA to maintain an active lifestyle. But before we can use the app, we would like to test it on healthy participants. You will be asked to wear a smartphone close to your body, for example in a pocket, so that the sensors in the smartphone can collect your movement data as efficiently as possible for one week.

The study aims to assess the ability of the smartphone and smartwatch to record physical activity and impact loading on your lower limb joints during seven days. You will not be asked to do any exercise outside your daily routine.

Why have I been chosen?

You have expressed an interest in the study and are healthy and in the right age group. We are seeking healthy people, aged 18 years and above, who are able to carry a smartphone and a smartwatch for one week.
Participants experiencing any of the following exclusion criteria will, unfortunately, not be able to take part in the study: needing a walking aid, such as wheelchair or crutches, and neurological or systemic illnesses.

**What will happen to me if I take part?**

The study will take one week in total for each participant with two visits, first visit and after 7 days the end visit. You will be invited to the University of Southampton Boldrewood campus for the visits. A research member will introduce you to the study and will show you how to use the smartphone and smartwatch, which includes an app that monitors your activity. A Smartphone and a smartwatch will be given to you. You will be asked to wear the smartphone close to your body around your waist. You will be given the choice to either wear it in your pocket or a smartphone-belt which will be provided by us. The smartphone and smartwatch will record together at the same time. The app will record your physical activity during the day. This means that you should carry the smartphone and smartwatch during the entire time that you are awake and moving around, within this week of data collection. We will not ask you for monitoring during sleep or resting periods. Also we will not ask for specific activities just your daily activity. You will also be given a one-page diary, where we ask you to write down your vigorous physical activities during the 7 days. The introductory session gives you the opportunity to familiarize yourself with the app and the study and to ask questions. The introduction session will take about 30 minutes. After this introduction you will be asked to take the smartphone and smartwatch with you for one week.

After the one trial week has concluded you will be asked to return the equipment and your travel expenses will be reimbursed.

**Are there any benefits in my taking part?**

There are no direct health benefits to you for taking part in this study. The information obtained from this study will be used to develop a smartphone app, which shows OA patients approximately how much strain is going through the joints during an activity.

**Are there any risks involved?**

There are no risks to taking part in this study and the sensors will not cause any pain or discomfort.

**Will my participation be confidential?**

Your privacy is very important to us and the study group will make every effort to protect it and make sure that any information that is released will not identify you. All data collected will be kept confidential as required by the Data Protection Act and University of...
Southampton data protection policy. To keep data confidential all written data will be stored in a locked filing cabinet and all electronic data will be stored on password protected computers that only the researchers have access to. Your data will be anonymised by assigning a unique ID number to your data. No one, other than the researchers, will be able to link your name with the unique ID number. Data that will be published in journal articles, conferences or meetings will not report any names or unique ID numbers to maintain the anonymity of the data. All data will be kept for 10 years.

What happens if I change my mind?

Participation in the study is entirely voluntary and you may withdraw at any point during the study, without giving a reason or affecting your rights.

What happens if something goes wrong?

Since we do not ask you to do any exercise, other than your normal daily activities, we do not anticipate any problems.

If you have a concern or a complaint about this study you should contact Susan Nazirizadeh (Address: University of Southampton, Building 176, Boldrewood Campus, Southampton, SO16 7QF; Tel: +44 (0)23 8059 8355; Email: s.nazirizadeh@soton.ac.uk).

Where can I get more information?

If you require any further information or have any questions regarding taking part in this study please contact Susan Nazirizadeh on the details below:

Susan Nazirizadeh, Dipl.-Ing.
Ph.D. Student
Faculty of Engineering and the Environment & Faculty of Health Sciences
Computational Engineering and Design Research Group &
Rehabilitation and Health Technologies Research Group
Building 176
Boldrewood Campus
University of Southampton
Southampton SO16 7QF
Tel. +44 (0) 23 8059 5194
Email: s.nazirizadeh@soton.ac.uk

Supervisors:
Dr Alex Forrester: email: alexander.forrester@soton.ac.uk
Professor Maria Stokes: email: m.stokes@soton.ac.uk
Professor Nigel Arden: email nigel.arden@ndorms.ox.ac.uk

Wednesday, 23 August 2017, Version 2
Patient Information Sheet

The relationship between lower limb joint pain and physical activity recorded via smartphone in people with osteoarthritis

Thank you for taking the time to read this information. Please read it carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

Invitation

You are invited to take part in our study, in which we will record your overall movement using a smartphone, and assess the pain in your joints using a smartwatch. This will allow the analysis described below. Additionally, you are being asked to give consent for the results from this study to be published in medical journals and in conference presentations. Personal details and means of identification of individuals will remain confidential, and will not be made accessible to the public.

What is the research about?

This study is part of the PhD project “Activity Monitoring via smartphones with reference to load on the lower limb joints”. The project is led by Susan Nazirizadeh, a PhD student in the Faculty of Health Sciences at the University of Southampton, with a background in engineering. The aim of the study is to develop a smartphone app, which estimates impact loading on the lower limbs, in order to encourage people with osteoarthritis (OA) to maintain an active lifestyle.

Although, the reasons for the progression of OA are not entirely understood, an active lifestyle can improve symptoms of OA, such as pain and stiffness of joints. The current study might help to develop a clinical tool in a form of a smartphone app to encourage people with OA to maintain an active lifestyle. For this purpose, you will be asked to wear a smartphone close to your body, for example in a pocket, so that the sensors in the smartphone can collect your movement data as efficiently as possible for two weeks. Additionally, you will be asked to score your OA related pain on a smartwatch.
The study aims to assess the pain that the patient is experiencing with a smartwatch whilst performing the physical activity recorded in the background by the smartphone. This study has been given a favourable opinion by Heath and Social Care Research Ethics Committee A.

Why have I been chosen?
You have expressed an interest in the study, you have OA in the lower limbs and you are in the right age group. We are seeking 20 participants aged 18 and over with lower limb OA affecting either their knee, hip, ankle or foot, who are receiving corticosteroid injections as part of their clinical practice.

Participants experiencing any of the following exclusion criteria will, unfortunately, not be able to take part in the study: difficulties walking and needing an aid, such as wheelchair or crutches, neurological or systemic illnesses, or causes of pain other than OA.

What will happen to me if I take part?

The study will take two weeks in total for each participant, consisting of the week before you have a corticosteroid injection in your lower limbs and the week after you obtained the injection. You will be invited to the Botnar Research Centre in the Nuffield Orthopaedic Centre for an introductory session one week before your injection. The researcher Susan Nazirizadeh (SN) will introduce you to the study and will show you how to use the smartphone app that will record your pain and physical activity. A Smartphone (Sony Xperia™) and a smartwatch will be given to you. You will be asked to wear the smartphone close to your body around your waist. You will be given the choice to either wear it in your pocket or a smartphone-belt which will be provided by us. The smartphone and smartwatch will record together at the same time. The app will record your physical activity during the day and you should score your OA related pain on the screen of the smartwatch whenever you experience pain in your lower limbs. The app will remind you five times a day (morning, before noon, afternoon, late afternoon, evening) to score your pain levels. This means that you should carry the smartphone and smartwatch during the entire time that you are awake and moving around, within these two weeks of data collection. We will not ask you for monitoring during sleep or resting periods. Also we will not ask for specific activities just your daily activity. The introductory session gives you the opportunity to familiarize yourself with the app and the study and to ask questions. The introduction session will take about 30 minutes. After this introduction you will be asked to take the smartphone and smartwatch with you for one week.

After one week you will be invited to return and obtain the corticosteroid injections from your rheumatologist, Professor Nigel Arden. After this session you will be asked to use the app for one further week. After each week of the trial period you will be asked to fill two
questionnaires: one related to your physical activity and one related to the OA related pain.

After the two trial weeks have concluded you will be asked to return the equipment and your travel expenses for the introduction session and the feedback session will be reimbursed.

Are there any benefits in my taking part?

There are no direct health benefits to you for taking part in this study. The information obtained from this study will be used to develop a smartphone app, which shows OA patients approximately how much strain is going through the joints during an activity.

Are there any risks involved?

There are no risks to taking part in this study and the sensors will not cause any pain or discomfort. Your clinical treatment, including the corticosteroid injection, is not part of the research and will be explained to you by your doctor.

Will my participation be confidential?

Your privacy is very important to us and the study group will make every effort to protect it and make sure that any information that is released will not identify you. All data collected will be kept confidential as required by the Data Protection Act and University of Southampton data protection policy. To keep data confidential all written data will be stored in a locked filing cabinet and all electronic data will be stored on password protected computers that only the researchers have access to. Your data will be anonymised by assigning a unique ID number to your data. No one, other than the researchers, will be able to link your name with the unique ID number. Data that will be published in journal articles, conferences or meetings will not report any names or unique ID numbers to maintain the anonymity of the data. All data will be kept for 10 years.

What happens if I change my mind?

Participation in the study is entirely voluntary and you may withdraw at any point during the study, without giving a reason or affecting your rights.

What happens if something goes wrong?

Since we do not ask you to do any exercise, other than your normal daily activities, we do not anticipate any problems.
If you have a concern or a complaint about this study you should contact Diana Galpin at the Research Governance Office (Address: University of Southampton, Building 37, Highfield, Southampton, SO17 1B; Tel: +44 (0)23 8059 5058; Email: rgoinfo@soton.ac.uk). If you remain unhappy and wish to complain formally Diana Galpin can provide you with details of the University of Southampton Complaints Procedure.

**Where can I get more information?**

If you require any further information or have any questions regarding taking part in this study please contact Susan Nazirizadeh on the details below:

Susan Nazirizadeh, Dipl.-Ing.
Ph.D. Student

Faculty of Engineering and the Environment & Faculty of Health Sciences
Computational Engineering and Design Research Group &
Rehabilitation and Health Technologies Research Group
Building 176
Boldrewood Campus
University of Southampton
Southampton SO16 7QF
Tel. +44 (0) 23 8059 5194
Email: s.nazirizadeh@soton.ac.uk

Supervisors:
Dr Alex Forrester: email: alexander.forrester@soton.ac.uk
Professor Maria Stokes: email: m.stokes@soton.ac.uk
Professor Nigel Arden: email nigel.arden@ndorms.ox.ac.uk
CONSENT FORM (Version2)

Study title: Monitoring Mechanical Loading of the body during physical activity via Smartphones

Researcher name: Susan Nazirizadeh (PhD student)
Study reference: 17086
Ethics reference: 17086

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (10th September /version 1 of participant information sheet) and have had the opportunity to ask questions about the study.

I agree to take part in this research project and agree for my data to be used for the purpose of this study

I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected

I am happy to be contacted regarding other unspecified research projects. I therefore consent to the University retaining my personal details on a database, kept separately from the research data detailed above. The 'validity' of my consent is conditional upon the University complying with the Data Protection Act and I understand that I can request my details be removed from this database at any time.

Data Protection
I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name).................................................................

Signature of participant.................................................................

Date.................................................................

Sunday, 15 November 2015, Version 2, Faculty of Health Science
CONSENT FORM

Study title: Monitoring estimated load rate with smartphones and smartwatches

Researcher name: Alex Forrester
Study reference: XXX

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (23rd August 2017 /version 2 of participant information sheet) and have had the opportunity to ask questions about the study.

I agree to take part in this research project and agree for my data to be used for the purpose of this study.

I understand my participation is voluntary and I may withdraw at any time without my treatment or legal rights being affected.

Data Protection
I understand that the information collected with respect to me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name)……………………………………………………
Signature of participant……………………………………………………………..
Date…………………………………………………………………………………

Name of researcher (print name)…………………………………………………
Signature of researcher …………………………………………………………..
Date…………………………………………………………………………………

I have read and understood the information sheet (23rd August 2017 /version 2 of participant information sheet) and have had the opportunity to ask questions about the study.

I agree to take part in this research project and agree for my data to be used for the purpose of this study.

I understand my participation is voluntary and I may withdraw at any time without my treatment or legal rights being affected.

Wednesday, 23 August 2017, Version 2
CONSENT FORM

Study title: The relationship between lower limb joint pain and physical activity recorded via smartphone in people with osteoarthritis

Researcher name: Susan Nazirizadeh (PhD student)
Study reference: 18024
IRAS reference: 211265

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (17th October 2016 /version 9 of participant information sheet) and have had the opportunity to ask questions about the study.

I agree to take part in this research project and agree for my data to be used for the purpose of this study.

I understand my participation is voluntary and I may withdraw at any time without my treatment or legal rights being affected.

Data Protection
I understand that the information collected with respect to me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name)……………………………………………………………

Signature of participant……………………………………………………………………..

Date…………………………………………………………………………………………..

Name of researcher (print name)…………………………………………………………….

Signature of researcher ………………………………………………………………………..

Date…………………………………………………………………………………………..

Monday, 17 October 2016, Version 9
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


