LUND UNIVERSITY
School of Economics and Management

Physical Exercise and Psychological Well-Being

A Prognosis for Cost-effectiveness of Corporate Health Care

Christoffer Knutsson & Magnus Berg

NEKN06 - First Year Master Thesis

Supervisor: Jan Bietenbeck

May, 2016

Abstract

The purpose of this paper is to predict the cost-effectiveness of a preventive health care project based on physical exercise that will take place at *Medicon Village*, Lund, in the fall of 2016. Data from the *Health Survey for England* is used to develop an empirical model with individual fixed effects that estimates the relationship between physical exercise and psychological well-being. The frequency of different kinds of physical exercise is used to explain several indicators of psychological well-being. The empirical model is then matched with survey responses collected from *Medicon Village* employees in order to realistically predict the impact of the project.

The results show that the empirical model fits the collected survey data well. Prediction intervals are calculated by inserting survey respondents' activity levels into the empirical model. This results in estimated levels of psychological well-being. 97.9% of these prediction intervals accurately cover the respondents' real reported value of given indicators of psychological well-being.

The cost-effectiveness of the preventive health care project will differ for every participating individual, due to different base line health statuses and sporting backgrounds, but also for every measure of psychological well-being. Physical exercise also does not impact all the indicators of psychological well-being equally. The model developed in this paper has the ability to take this into account and can predict the cost-effectiveness of the project for any given individual. Three examples of cost-effectiveness estimates based on our survey responses are presented in the results section.

Table of Contents

1) Introduction

Being healthy is important for several reasons. It is related to happiness, self-esteem, confidence, and the ability to cope with difficulties. In the corporate world it is also related to something more tangible, namely money. Large amounts of money are lost every year due to sick days as well as reduced productivity due to presenteeism – being at work but not being able to perform optimally due to illness. An article by Pauly, M.V., et al., (2008), provides estimates of this. The physical health of a company's employees is directly related to the performance of the company. Given this relationship between the health of employees and the financial well-being of the company, it is possible for companies to reap potentially significant rewards from investing in health care for their employees. Indeed, many companies are investing in corporate health care and the positive effects on employees health is well documented as shown by Huber, M., et al., (2015). In Sweden it is even customary that employees get health care contributions besides their salary in order to minimize sick days and presenteeism among the employees.

The purpose of this paper is to predict the cost-effectiveness of one particular exercise based preventive health care project that takes place at *Medicon Village* – a corporate complex consisting of various companies in the life science industry – in Lund. The choice of applying the model to *Medicon Village* is not random. Two companies from Lund, *BetaHälsan* and *Caramba Syd*, specialized in physiotherapy and corporate health care, are currently working together with *Gerdahallen*, a major training center in Lund, in order to establish a preventative health project that going to be implemented at *Medicon Village* in the fall of 2016. The goal of this project is to improve the physical and mental health of workers at *Medicon Village* companies. The participants' health will be assessed before and after participating in recommended training programs and lectures. The project necessitates an investment from the employer's side, and as many of the companies at *Medicon Village* are fairly small, with around 4 employees and small turnovers, it can potentially be a high stakes project for the companies. It is therefore of very high interest to the companies that substantial results from such a project are realized.

This paper's analysis holds two essential parts. For the first part, actual data is gathered from *Medicon Village* employees, who are potential participants in the preventive health care project. A survey, assessing the psychological well-being and physical activity of the respondent, is constructed and handed out to the employees outside *Medicon Village*'s lunch restaurant. The survey is answered online so that potential participants in the *Medicon Village* project also have the option to answer the survey at some other point in time, at home or at work. Five variables measuring psychological well-being are included in the survey. All of these are measured on a scale of one to five, where five represents the most positive option. The respondents are asked to think about the past week when answering the questions regarding their well-being. The five means that respondent has always felt useful, relaxed, etc. during the last week, and the one means the respondent has never felt useful, relaxed, etc. during the last week. The variables included are the following:

- Useful – How often has the respondent felt useful?
- Relaxed – How often has the respondent felt useful?
- Energy – How often has the respondent had energy left after work?
- Clear Thinking – How often has the respondent been able to concentrate and think clearly?
- Confidence – How often has the respondent felt confident and good about him-/herself?

The physical activity is measured by two variables, where one represents light physical activity and one measures more intense activities. The respondents are once again asked to think back to the last week when answering how often they walked for 10 or more consecutive minutes and how many times they exercised vigorously. Vigorously, in this context, means that the person has a clearly elevated rate of breathing and sweats noticeably. These two variables are measured on a scale zero to seven, where each category represents the number of days that the person performed each activity in the last week. The data was gathered by us, the authors, from the end of April 2016 through mid-May 2016.

The second part of the analysis is an empirical analysis based on existing data from England. A data panel holding answers from the *Health Survey for England* is being used to construct a regression model that can predict the outcomes of the *Medicon Village* project. The regression model regress several variables measuring psychological well-being on physical activity. The model provides estimates of the effect physical activity has on psychological well-being. More details about the English health survey will be brought up later in this paper. In order to determine whether the theoretical model can be used to predict the outcomes of the *Medicon Village* project, the physical activity levels of our survey respondents are inserted in the model. This produces a predicted value of the dependent variable, measuring psychological well-being. If the respondents' true values of psychological well-being fall within the theoretical model's prediction intervals, the model is considered a good fit.

In the final step of the analysis, the survey data gathered at *Medicon Village* is inserted in the regression model, creating predictions for what the future results of the preventative health care project (referred to as PHCP for the remainder of the paper) could be. These predictions are compared to the participation cost the companies have to pay to take part in the health care project, and thus the project's cost-effectiveness is estimated.

Before the methodology and results of this analysis are presented, previous literature on topics related to corporate health care is reviewed. After the previous literature section, the methodology is described in detail. This section explains the development of the regression model and how it will be connected to the survey data gathered at *Medicon Village.* In the section following the method segment the results are presented and explained. The paper is then finalized by a discussion regarding potential problems and the conclusions drawn from the entire analysis.

This paper can be helpful for companies investing in corporate PHCP's similar to the one analyzed here, as it provides an external assessment of the potential cost-effectiveness of such a project. The prognosis for the participants' health development is based on existing panel data, which gives further credibility to the results. The participants' projected health development is compared to the fee that companies would pay to participate, and this gives an estimate of the expected return of the project, in terms of health improvement per invested SEK, which is of central interest to companies thinking about investing in exercise based corporate health care.

2) Previous Literature

The number of papers written on the subject of corporate health care, and the benefits of being healthy in general, is staggering. Some of these are used as a theoretical foundation for the analysis in this paper, and these articles and their results are outlined in this section. Papers are used for primarily two purposes in this paper: establishing a connection between physical exercise and psychological well-being, and developing an understanding of the work related costs associated with poor health.

2.1) Physical Exercise and its Effect on Psychological Well-Being

In this subsection, the articles establishing the relationship between psychological well-being and physical exercise will be reviewed. In short we can conclude the following from these articles: There appears to be a positive relationship between physical exercise and psychological well-being. None of the papers manage to estimate a causal relationship between these variables though. Using panel data to investigate this topic one can possibly allow the estimation of the causal link between physical exercise and psychological well-being.

The first paper being reviewed, written by Bredahl et al. (2008) investigates the effect of ‘exercise on prescription’. The participants in the study answer a questionnaire regarding their psychological health before and after completing a 16 months long exercise program coupled with motivational lectures. Participants that drop out during the program still get to answer the questionnaire at the end. The authors find that many participants do feel better while regularly exercising, but that those who dropped out of the study early did not change their lifestyle in most cases. Another similar study performed by Coulsen et al. (2008) investigates the performance of workers who are allowed to exercise during office hours. The results show that 89 percent of the participants report a better work performance during days when they exercise compared to days when they do not. Also, respondents’ mood improves as the work day progresses on exercise days.

A recently published paper by Lindegård et al., (2015), studies the effect of physical exercise on stress-related exhaustion. The study follows 69 individuals who have been diagnosed with clinical burnout, depression, and anxiety. At the beginning of the study the patients’ health is assessed and personal training programs are given to the participants. The results are recorded after an 18 month period and show that all participants have improved significantly after the test period. Furthermore, the results also indicate that groups of compliers, those who actually follow their training regimen, improve significantly more than the non-compliers. Relating this result to our paper, it is of great importance that the participants in the planned preventive health care project actually follow their training programs if positive results are to be realized.

The study by Lindegård et al., (2015) also highlights some of the mental effects that can explain why physical exercise is positively related to mental health. The authors explain that aerobic training is positively related to reductions in perceived stress and improved executive function. Furthermore, both cardiovascular and resistance based training are positively related to improved sense of self accomplishment and decreased emotional exhaustion.

Gerber et al., (2014), have written a paper that concerns the effect of physical activity on work related stress. The authors use an approach where physical activity is used as a variable for categorizing the material. Questionnaire answers from 2660 health care and social insurance workers in Västra Götaland, Sweden, are analyzed to determine whether or not physically active people experience less problems with stress. The answers from the questionnaires are used to divide the data material into six different categories based on stress levels. The authors conclude that the division into precisely six categories is the best way to divide the data by looking at BIC, sample size adjusted BIC and log-likelihood measurements. The six categories are the following: Highly burdened, Stressed, Resilient to high stress, Moderately stressed, Resilient to moderate stress, Low stress and healthy. The results show that the prevalence of physically inactive people is lower in the Low stress and healthy category (5% inactive people) than the Stressed category (23% inactive people) and the Highly burdened category (35% inactive people). The prevalence of inactive people in the other categories is also higher than in the Low stress and healthy category, but these differences are not as striking. The Chi^2-value for the difference in physical activity levels between the groups is 175.02, with a p-value < 0,001.

As mentioned earlier in this subsection, none of these articles establish a causal link between physical exercise and psychological well-being. This greatly depends on the fact that most studies on the topic use analytical methods based on descriptive statistics.

2.2) Work Related Costs and Productivity Loss

This subsection deals with the other articles underlying our analysis, the ones regarding losses in productivity and money associated with poor health among the employees. We quickly summarize what one finds in these articles by saying: The estimated costs and productivity losses due to poor health vary greatly. This is not strange as there are no standardized methods of measuring these losses. It is also clear that costs and losses, as well as gains, attributed to changes in a staff's health, depend greatly on specific context.

A study from Finland, performed by Tuomi et al., (2004), makes a broad investigation of how work-place well-being is related to a number of different variables among workers in the metal industry and retail trade. Among these variables, the authors find a highly significant – p-value < 0.01 – positive correlation between increased physical exercise and work ability. The work ability is measured as an index, consisting of seven items related to self-rated work performance and focus. The amount of physical exercise the respondent undertakes is measured on a five category scale, ranging from 0 through 4, where 0 is not exercising at all and 4 is exercising vigorously at least four times per week. The magnitude of the correlation coefficient is 0.07, meaning that moving from one category of physical activity to the next increases work ability by seven percent.

In yet another paper, by Proper et al., (2004), the authors conduct a cost benefit analysis to determine how effective worksite exercise counselling is. 299 participants were divided into an intervention group of 131 individuals and a control group of 168 individuals. The participants are followed for nine months and the intervention costs are compared to the indirect monetary gains from reduced sick leave. The participants' health status is determined in the beginning of the study and the intervention group is offered individual specific plans for exercise regimes and diet. The intervention cost per employee is €430 and after the intervention period, the cost of sick leave is €635lower per individual for the intervention group than the control group. The standard deviation is, however, very large in both the intervention and control groups and the confidence interval for the difference between the groups covers zero. According the authors, the lack of precision is due to a few individuals creating un-proportionate costs as a result of their poor health.

The authors also make a cost-effectiveness analysis to determine how physical health is affected by the intervention. The intervention group show better cardio-respiratory fitness. Their sub-maximal heart rate is approximately 5 beats per minute lower compared to the control group. The confidence interval does, however, cover zero this time as well, so there might not be an effect. When the authors look at the amount of upper-extremity syndromes, such as neck pain or aching elbows, there is a significant effect in favor of the intervention group. Intervention group participants are approximately 10% less likely to experience these kinds of problems.

Uegaki, et al., (2010) investigate the topic of evaluating occupational health from an employer's perspective. This paper is a review of 34 papers published in either Dutch or English. The authors attempt to establish whether it's possible to compare estimated costs and benefits from occupational health investments. The conclusion is that one cannot do this. The methods used to estimate the costs and benefits vary widely and many of them are not comparable. As one can also suspect, there is no "right answer" in the sense that one method beats the others. In the papers being reviewed, the method for estimating costs and benefits is contextual. No standardized measures are used. This is relevant for us as the prognosis we make in this papers needs to reflect this uncertainty.

An article by Hemp (2004) published in the *Harvard Business Review* provides an indication of how much presenteeism reduces productivity. The magnitude of the productivity loss obviously depends on the type of illness, but the article provides estimates of the productivity loss due some illnesses that can be thwarted by physical exercise. The estimated productivity loss due to migraine is 4.9%. The same number for depression is 7.6% and 5.5% is the estimated productivity loss due to lower back pain. Arthritis is associated with a 5.9% productivity loss. These different illnesses are not always equally prevalent, but if one is making the slightly simplistic assumption that they are, the average productivity loss is 5.975%.

In Sweden the organization *Försäkringskassan* publishes the official statistical reports regarding the work related costs of ill health. There is even an online instrument developed for calculating how much one day of sick leave costs the employer. Using this instrument one can estimate the hourly cost of an employee who is absent from work due to sickness. In Sweden, the first day is a "waiting day", meaning that the sick person does not get any reimbursement on this day. In between day 2 and day 14 the sick employee gets the highest reimbursement, which corresponds to at least 80 percent of the income and benefits lost. An example of this is given below:

Table 1 – Costs of Absenteeism in Sweden

Employee's yearly salary (SEK)	Employer's hourly cost (SEK)
300,000	225
400,000	301
500,000	375

Source: Försäkringskassan, 2016.

According to *Försäkringskassan's* statistics for 2015, people are absent from work roughly 28 days per year in Scania. This measure is calculated as the amount of days that *Försäkringskassan* pays reimbursements due to ill-health divided by the number of registered people in the region.

The group of individuals that is analyzed in our paper consists of almost exclusively academically educated individuals working in a research intensive field. A study by Baecke et al. (1982) concludes that highly educated people are less physically active in their daily work. This is not surprising and implies that we should expect to see measurable and significant results from the health intervention program we are evaluating. The study by Baecke et al. (1982) also estimates the reliability of their questionnaire based test-retest design and concludes that the reliability ranges from approximately 75 to 90 percent, depending on the category of questions. The participants who choose to take part in the study are given the questionnaires, and within seven days of recording the answers, physical measurements of the participants are taken. These are then compared to the answers on the participant's questionnaire to check if they answered truthfully and accurately. The fact that the reliability is as high as it is indicates that using a test-retest design is appropriate when analyzing questions regarding people's physical habits and health.

Puig-Ribera et al., (2015) investigate the relationship between time spent sitting and physical activity in relation to on-the-job productivity. 557 Spanish university employees completed international, standardized questionnaires regarding levels of physical activity and well-being at work. The answering participants are divided into three groups depending on how physically active they are. Study participants in the different groups report that they experience work capacity limitations with regard to: i) scheduling demands ii) performing mental interpersonal tasks and iii) delivering output:

Table 2 – Experienced Work Capacity Limitations

	i)	ii)	iii)
Low Activity	22.60%	24.42%	28.16%
Moderate Activity	15.86%	20.16%	23.73%
Highly Active	14.67%	17.12%	21.24%

i) denotes scheduling demands, ii) performing mental interpersonal tasks, and iii) delivering output.

From table 2 one can clearly see that inactive people tend to experience greater challenges at work than active people. For instance, inactive study participants report that they experience 54 percent, $(22.60 - 14.67)/14.67$, more capacity limitations with regard to scheduling demands and 43 percent, $(24.42 - 17.12)/17.12$, more capacity limitations with regards to performing mental interpersonal tasks. Interestingly though, the study participants do not appear to differ as much when it comes to estimates of lost work productivity due to presenteeism and absenteeism. In the study, the estimate of lost productivity for the highly active group is 4.36% and the same number for the inactive group is 5.99%. The moderately active is estimated to lose 4.95% work productivity. This is not very surprising as it is difficult to precisely measure productivity in this type of work. This situation will likely occur in our analysis as well, as many of the participants in the PHCP have research jobs and jobs that are of academic nature.

Presenteeism is a difficult topic in the scientific community. The concept is fairly new and the methodological map for investigating the issue has not been completely drawn yet. Absenteeism, on the other hand, has been an area of interest for a longer time and one can find fairly precise estimates of the costs associated with this. An article dealing with the difficult issue of presenteeism is written by Pauly, M.V., et al., (2008). This paper uses an empirical approach to estimate the costs of presenteeism. The authors ask managers of different types of firms to assess the problems presenteeism would cause in their firms. The managers answer on a one to four scale, where one means that the productivity loss caused by presenteeism is completely negligible to a specific worker's team, and the four means that it would lead to a complete team shut down in case workers are not as productive as they are when they are healthy. The authors then run ordered probit regressions with the managers' answers as dependent variable and time sensitivity, availability of perfect substitutes, and team production as explanatory variables. The authors find that presenteeism's cost per day, measured as percent of the sick workers salary, depends greatly on the tasks that this person performs. The presenteeism cost of engineers is estimated to be the highest at 75%. The lowest costs are the ones for auto service technicians and hotel maids, estimated to be 12.5%.

This subsection, as well as the first previous literature section, is concluded by the realization that the costs and productivity losses due to poor employee health are extremely difficult to estimate. It is clear that poor employee health constitutes a problem, but the magnitude of the problem depends greatly on the context in which it occurs. At *Medicon Village*, the companies are quite small on average, with workers co-operating in tight teams that are engaged in the very advanced life science field. This suggests that problems due to presenteeism, as well as absenteeism, would be substantial.

3) Method

The method section of this paper is divided into four subsections. The first explains the work that is done to get the *Medicon Village* data through our own survey and the second subsection explains the data obtained from the *Health Survey from England*. The third subsection deals with the regression analysis, based on the *Health Survey for England,* and how this is constructed to emulate the PHCP. This ultimately leads to the last section presenting the model used for predicting the cost-effectiveness of the PHCP.

3.1) Data Gathering at Medicon Village

Our study uses a test-retest method, with one survey being sent out to all the companies at *Medicon Village* before the start of the PHCP. The same survey will be sent to the companies at *Medicon Village* next year, as the PHCP is completed. As mentioned in the introduction, the survey is constructed by us and includes five variables measuring psychological well-being and two variables measuring physical activity. In addition, the survey includes some questions regarding demographics, such as the age, sex, initials, and job title of the respondent. The reason why these are included is partly because they are used as control variables – age and gender – and partly because they allow us to identify the respondents – via initials and job title – when the second survey is handed out next year, when the evaluation of the PHCP takes place.

The first survey goes out to companies that will take part, but also to those companies that will not take part in the project, as these will help establish a control group for the treatment. This first survey wave cannot be used to evaluate the potential effects of the project on its own, but the data collected constitutes a cross-sectional measurement that will be used as the project is evaluated. By doing this, one can determine the treatment effect of the PHCP.

During week 16 and 17, 2016, we situated ourselves in the lobby outside the lunch restaurant at *Medicon Village* and collected survey responses from the employees as they had their lunch breaks. The survey was constructed online, using *Survey Monkey's* survey design tools. A laptop was brought for the respondents to fill out their answers on. The screen was turned away the entire time and we, the authors, are completely unaware of what the respondents answer. The fact that the survey uses an online format means that busy lunch guests who do not have time to complete the survey right there and then can continue to do so at a later point in time. We also informed the lunch guests passing by where they could find the survey, were they interested in filling it out on their own somewhere else. Answers from roughly 50 employees were collected.

The purpose of this paper is to develop a cost-effectiveness estimate for the *Medicon Village* project that can be shared with the involved parties. For this purpose, the cross-sectional data gathered by us is used to assess the accuracy of the empirical model outlined in the next subsection. If one is going to be able to make such a comparison, the variables in the two data sets need to match. When constructing the survey being handed out to the *Medicon Village* employees, we took great care to make sure that the questions on our survey match the variables in the English data panel. The accuracy of the empirical model is investigated by using it to predict the current psychological well-being of our survey respondents. If our respondents' answers fall within the empirical model's prediction intervals, the model specification is considered successful. In the next subsection, the data obtained from the *Health Survey from England* is explained.

3.2) Health Survey for England Data

In order to establish a basis for the cost-effectiveness prediction one first needs to create a model that can be used for this purpose. Data from the *Health Survey for England* is being used to construct such a model in this paper. The survey started in 1991 and is an ongoing annual survey with a focus on monitoring any potential trends in the state of health of the nation. The entire data set contains little over 250,000 observations. A majority of these are however dropped, either because of incomplete information or because the respondents belong to an unrepresentative age bracket. Roughly speaking, around 12,000 observations are used when running the outlined models. An exception is the model estimated using subsets of the data. In this case, one subset holds a little more than 2,000 observations and the other a little more than 9,000. For the purpose of our analysis the data is restricted to the years 2009-2014. This is done due to the fact that the questions in the original health survey are changed every few years, but during this time period the questions are consistent. The questions being used include demographic identifiers, indicators of psychological wellbeing, as well as measures of physical activity. The variables indicating psychological wellbeing are determined on a 1-5 scale, where the number 5 represents the top criteria. Physical activity is measured on a scale 0-7, indicating the number of the days the individual has been active during the week. The questions in the questionnaire being sent to *Medicon Village* match the questions used in the English survey and a copy of the questionnaire is found in Appendix A. Using the same questions ensures that the answers collected from *Medicon Village* are comparable to the answers of the English survey. This comparability is a requirement when making a prognosis for the project. The descriptive statistics of the data from both the *Medicon Village* respondents and the *Health Survey for England* is presented in Appendix B.

The initial step when analyzing the data is determining what type of analysis it is suitable for. Regression analysis is used to determine the influence of physical activity on psychological well-being. The *Health Survey for England* only collects data on the ordinal level, with the scales described in the previous paragraph, and as the survey we construct matches the English one, our questionnaire only contains questions that are ordinal as well. This means that the data, in its basic form, is not very similar to conventional datasets. Despite the lack of interpretability, panel data regression models are used to estimate the effects of physical exercise on psychological well-being. Treating the ordinal variables as if though they were continuous and running the regression provides enough information to draw meaningful conclusions. This is illustrated by a brief example: Imagine that one runs a regression with self-reported ability to concentrate as dependent variable and number of walks per week as independent variable. Let's say that the ability to concentrate is reported on a 5 degree scale and that the $\hat{\beta}$-coefficient is 0.35. Although this does not say that the ability to concentrate increases by an exact amount, it does say that higher ratings of ability to concentrate are more frequent for people who take walks more often. Given that the regression model is correctly specified, a result like this implies that physical exercise in the form of walks has a positive effect on people's ability to concentrate. The usage of linear models to analyze categorical data has also been assessed by for instance C. N. Norris, et al., (2006). They use categorical data from a clinical setting and compare linear, logistic and ordinal models. Their results indicate that the linear model was indeed a good fit.

It might seem a little unintuitive to fit a linear regression model to ordinal data, where no exact differences between categories can be observed, but it is a thoroughly tested method that performs fairly well. It also has the advantage that it is easy to administer and carry out. Another paper on the subject of regression with ordinal variables by Torra, et al, (2006) observes that this method usually works well when the questions have five or more categories:

Before specifying any regression model a correlation matrix is constructed to check the correlations between the variables. The correlation matrix is presented in table 3. This allows one to check for worryingly high correlations between the predictors, which might cause collinearity issues.

Table 3 - Spearman Rank Correlations

	Walks	Vigorous	Age	Gender
Walks	1			
Vigorous	0.20***	1		
Age	-0.06***	-0.08***	1	
Gender	0.01	0.18***	0.03***	1

The rank correlations are estimated using 12,114 complete observations and the significance levels are the conventional 0.05 ("*"), 0.01 ("**"), and 0.001 ("***").

Based on this first-step check, one can conclude that there are no apparent multicollinearity problems. The correlations are significant, but that depends on the large sample size. Having a sample as large as the one used in this paper leads to very small standard errors. This implies that one must look more carefully at the magnitude of the correlations than their significance levels, and as one can see, none of these are worryingly large. We therefore conclude that there should not be any major collinearity problems.

3.3) Regression Analysis Based on the Health Survey for England

The first step of the regression analysis is dropping some of the observations. In order for the theoretical prognosis model to match the practical scenario we are facing, it makes sense to only include respondents in the age bracket 16 through 65. After restricting the data to include only observations of the right age, the regression analysis is started by off by running a pooled OLS on the entire remainder of the dataset. This model is presented in the results section, table 4. The regression equation for this model has the following specification:

$$y_i = \beta_0 + \boldsymbol{\beta_1} Walks_i + \boldsymbol{\beta_2} Vigorous_i + \varepsilon_i \qquad (1)$$

The dependent variable, y_i, represents any one of the different variables that indicate mental well-being. All of these are used as dependent variables separately, with the same set of explanatory variables. In all of the regression equations in this paper, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$, written in bold font, are vectors of beta coefficients. They hold one coefficient for each category of exercise frequency. The equation presented above is a slightly naïve model specification in the sense that it will not measure the causal relationship between physical exercise and psychological well-being. There are most likely a few control variables that should be included. This simplistic model does, however, serve a purpose. One would expect the coefficients in the model to change quite dramatically when more sophisticated model specifications are made. If they do not, it would suggest that our method of analysis is missing something essential.

Efforts to get closer to a causal relationship are made by including controls for age and gender in the model. Age is clearly connected to one's health status, so this variable has an obvious place in the model. The gender of respondents should also be included, as men and women have different exercise habits. A study by Craft, et al., (2014), finds that women and men exert different levels of effort when exercising. The authors also find that women tend to have physical appearance as a training motive more often than men. There has however been an equalizing trend over the last few years, and given that the data set used in this analysis is modern, one would suspect that there might not be any differences of noticeable magnitude between the exercise habits of men and women. The model including these control variables has the following specification:

$$y_i = \beta_0 + \boldsymbol{\beta_1} Walks_i + \boldsymbol{\beta_2} Vigorous_i + \beta_3 Age_i + \alpha_i + \varepsilon_i \tag{2}$$

The α_i indicates the gender of the respondent and the results from this model are presented in table 5.

The data set used for this analysis has a panel component which can be exploited to get even further to estimating the causal relationship between physical exercise and psychological well-being. By including individual fixed effects in the model, one can remove the individual differences driving both psychological well-being and levels of physical activity. These fixed effects could for instance include different sporting backgrounds, genetics, socio-economic background, eating habits while growing up, etc. There are numerous examples of background factors that can drive both psychological well-being and physical exercise habits, and as it is extremely hard to measure these accurately and including them as variables in a regression model, excluding them by running a fixed effects regression is a viable option. The fixed effects model is specified in the following way:

$$y_{it} = \beta_0 + \boldsymbol{\beta_1} Walks_{it} + \boldsymbol{\beta_2} Vigorous_{it} + \beta_3 Age_{it} + \omega_{it} \tag{3}$$

The α_i, indicating the gender of the respondent, is removed from the model as this is, most often, one the fixed effects described above. The error term, ω_{it}, is the conventional panel data error, $\omega_{it} = u_i + \varepsilon_{it}$, where u_i represents individual fixed effects and the ε_{it} represents the residual type error term, assumed to be independent of the rest of the model and identically distributed over all time periods, *t*.

As panel data is used it can be of interest to see if the coefficients and significance levels change dramatically when the models are estimated with random effects instead. There are proponents of random effects estimation, even though it is often considered an odd estimation method. Random effects estimation is more efficient, but it is also carrying an inbuilt bias. It's built around the very strong assumption that the individual error terms are completely independent of all the other covariates, $E[u_i|x'_{it}] = 0$, where u and x'_{it} denote the error term and a vector of covariates respectively. A paper by Clark & Linzer (2015) compares fixed and random effects to try and determine which should or should not be used. They do this by fitting both types of models to over 2,000 samples, generated by Monte Carlo simulation. One often finds the recommendation to use a Hausman test to try and determine whether fixed or random effects should be used. This is not always a good idea, though, as explained by the authors:

"Scholars are sometimes advised to use a Hausman (1978) specification test to detect violations of the random-effects modeling assumption that the explanatory variables are orthogonal to the unit effects. A "significant" test result is taken as evidence of a correlation between x and αj, implying that the random-effects model should be rejected in favor of the fixed-effects model. However in most applications, the true correlation between the covariates and unit effects is not exactly zero. Therefore, if the Hausman test fails to reject the null hypothesis of orthogonality, it is most likely not because the true correlation is zero—and, hence, that the random-effects estimator is unbiased. Rather, it is likely that the test has insufficient statistical power to reliably distinguish a small correlation from zero correlation. When using the random-effects model, there will still be bias (if perhaps negligible) in estimates of β, even if the Hausman test does not find a significant result. Of course, in many cases, a biased (random-effects) estimator can be preferable to an unbiased (fixed-effects) estimator if the former provides sufficient variance reduction over the latter. The Hausman test does not help evaluate this trade-off." (Clark & Linzer, 2015, p. 402-403)

The Monte Carlo simulations performed by Clark and Linzer indicate that the random effects method is more efficient when a short panel is used, given that the bias is not too large. A short panel is used in this paper, and the Husman test is used for the five dependent variables as it is interesting to see what estimation method is suggested. The test results are presented in Appendix BC. The model above is estimated with both fixed and random effects so that the results can be compared. These results are presented alongside each other in table 6. Although the random effects estimates are more efficient than the fixed effects estimates the fixed effects model is still preferred. The preference for the fixed effects model is based in that the random effects estimation uses feasible generalized leased squares estimation, which means the underlying individual differences that push the results in a certain direction are not accounted for. As mentioned earlier in this section, this implies that unless one is able to identify all potentially relevant control variables and include these in the model specification, it is extremely difficult to get close to estimating a causal relationship between variables. Running a fixed effects regression with individual fixed effects instead somewhat remedies this problem, as the fixed effects automatically remove some of the underlying individual factors that drive the results without them needing to be precisely identified.

As the model specification has been determined, the data set is divided into two subsets, with self-reported general health status, in period one, as divisor. This is done in order to make the data match the practical situation that the participants in the PHCP will face. The participants in the project are screened and divided into different groups, depending on their health status. The relatively unhealthy group will take part in light exercise classes, such as taking walks. The relatively healthy group will be recommended different exercise classes, such as HIIT, short for High Intensity Interval Training. As the groups will exercise in different ways and have fundamentally different health statuses, it is not likely that they will respond equally to all types of exercise. Therefore the English data is divided into two subsets, one relative healthy and one relatively unhealthy, to create a context that is better suited for making the prognosis. The fixed effects regression including age and gender controls will be fitted to both subsets. In the unhealthy subset, only walks are included in the regression, as this is a good proxy for the light exercise that the unhealthy group in the PHCP will undertake. The same model, but with vigorous activity added, is fitted to the healthy subset. This proxies the situation the healthy participants in the PHCP will face, as these participants are free to choose to participate in the lighter exercise classes as well. The fixed effects models estimated for each subset provide the coefficient estimates that are used when making the prognosis for the different groups. The results from these models are found in table 7.

3.4) Matching Data and Making Predictions

As the results of the different regression models have been checked, the most realistic model, the version with the two subsets, is compared to the data set containing our survey respondents' actual answers. When one uses regression analysis to make a prediction one inserts values that the explanatory variables could take into the model equation. Given that the specification of the model is correct, this is a consistent way to estimate a value of some dependent variable. What one is usually interested in, though, is the certainty of the prediction. In order to assess this, one needs to create an interval for the predicted outcomes. Verbeek (2012) explains how a prediction interval for a linear regression model is constructed in some basic cases. For a more thorough review of this, see Appendix D.

In practice, STATA is used for all calculations in this paper, and when the prediction interval is calculated one simply specifies what is to be predicted when writing the program commands. The point of including Appendix D on prediction intervals is the following: From the expressions in the appendix it is clear that a prediction interval is much broader than a confidence interval. The difference between the two types of intervals lies in the inclusion of the estimated error term's variance. When one wants to predict one individual value, an error term has to be included. On average, the error terms are zero, so when constructing normal confidence intervals, this source of variation can be ignored. The concept of prediction intervals, and how they differ from confidence intervals, is explained in a pedagogical way by Faraway (2002).

The prediction intervals are calculated at the same time as the regression models are estimated. This is done because the standard errors of all the models' estimated parts are reported by default, which means that the final step of this analysis, i.e. validating the chosen model, becomes very straight forward. The responses from our survey are inserted into the regression model of choice to produce a prediction and its corresponding interval. Checking whether or not our respondents' real answers actually fall within this prediction interval will disclose if the theoretical model can be used to predict the outcomes of the PHCP at *Medicon Village*. How often the real, reported values of the dependent variables fall within the model's prediction intervals is presented in the results section.

As mentioned in the previous literature section, any prognosis made should reflect the uncertainty discussed by Uegaki et al. (2010). So far a theoretical foundation for a cost-effectiveness prognosis has been laid, but no direct conclusions can be drawn from this. Using a regression model such as the one in this paper to make predictions regarding the cost-effectiveness of a corporate healthcare projects is merely one possible method. Clearly, the predictions will vary widely depending on the values inserted in the model, reflecting the uncertainty mentioned in the Uegaki paper. The predicted cost-effectiveness will differ for every individual participant. This might seem an unattractive feature of this prediction method, as no simple answer is given to the question regarding the entire project's cost-effectiveness, but it can also mean that the regression model used to predict the cost-effectiveness is a flexible method that can be adjusted to fit many different situations. If the administrator of a health care project has some background information on the potential participants, this information can be used to predict an outcome for every single participant. In this particular case both the participation cost and underlying health variables of the potential participants are known, which means that one can quite easily estimate the health improvement per SEK invested. Whether it is a good method or not can really only be assessed as the PHCP is evaluated in approximately one year's time. The cost-effectiveness of a few potential participants is presented in table 8, and these estimates depict different scenarios that could be realized when the project is carried out. A consideration that obviously needs to be kept in mind is that the compliance of participants can be very poor, or very good, which would greatly off-set any prediction. If they go to every exercise class, or none, the cost-effectiveness of the project is most likely not covered by the estimates. As Lindegård et al., (2015), find in their study, compliers benefit significantly more from taking part in exercise based health care projects than non-compliers.

4) Results

In this section the results from the models will be presented. All the relevant output from the estimated models is presented in tables 4 through 7. In all of the tables, the variables measuring physical activity are factor variables that use the alternative 0 as reference. If one takes “Walks” as an example, the coefficients on the row numbered 1 show how much the indicators of psychological well-being increase if the respondent walks for 10 consecutive minutes one day per week, as compared to walking for 10 consecutive minutes zero days per week.

In table 4 below, we see the results from the pooled OLS model, without controls. It is clear that almost everything is significant at the 0.001 level. As mentioned in the method section, this is not a very realistic description of the real relationship between physical activity and psychological well-being. The estimates are expected to change quite a bit when control variables and panel components are introduced. This model does not need much attention, but the sign of the coefficients indicate that there is a positive relationship between physical activity and psychological relationship. One can however observe two noticeable facts about the coefficients: The magnitude of the coefficients make intuitive sense. Psychological well-being and physical exercise are linked to each other, but there are also other factors affecting a person’s psychological well-being, which suggests that increasing the amount of physical exercise should not have a massive impact on the dependent variables in the model. One can also see that more is better, generally speaking, as the coefficients are larger for frequent physical activity. From this model’s output it is very difficult to determine whether intense or light exercise is best for improving psychological well-being though. The more sophisticated models will provide clearer and truer estimates of this relationship.

Table 4 – Pooled OLS

	Usefulness	Relaxed	Energy	Clear Thinking	Confidence in Self
Walks					
1	0.25***	0.08	0.15**	0.09*	0.10*
2	0.26***	0.17***	0.21***	0.17***	0.18***
3	0.33***	0.21***	0.27***	0.19***	0.22***
4	0.28***	0.19***	0.29***	0.16***	0.19***
5	0.37***	0.22***	0.34***	0.22***	0.26***
6	0.33***	0.20***	0.29***	0.25***	0.25***
7	0.37***	0.23***	0.36***	0.24***	0.26***
Vigorous					
1	0.14***	0.07*	0.17***	0.08**	0.14***
2	0.16***	0.15***	0.31***	0.13***	0.19***
3	0.23***	0.18***	0.37***	0.15***	0.22***
4	0.20***	0.23***	0.45***	0.20***	0.29***
5	0.15***	0.21***	0.32***	0.15***	0.21***
6	0.18***	0.28***	0.45***	0.15***	0.30***
7	0.23***	0.16***	0.38***	0.17***	0.27***

In all of the tables, the conventional significance levels are used, and the asterisks have the following meaning: "*" means significant at the 0.05 level, "**" means that the estimate is significant at the 0.01 level, and "***" means the estimate is significant at the 0.001 level. The table presents the marginal effects from the pooled OLS models featuring only indicators of physical exercise as explanatory variables. The coefficients should be interpreted the following way: The zero-category is always the baseline category. A positive coefficient shows how large of an improvement the person would make if s/he were to change from exercising zero times per week to exercising more frequently.

The results from the OLS model including the control variables are presented in table 5 below. The significance levels change slightly as the control variables are introduced – most of them grow by 0.02, and some of them are at most reduced by 0.04. As the changes are so small, the coefficients are still of a magnitude that makes sense and most coefficients are still significant at the 0.001 level (***), which does not seem to accurately reflect a real world relationship. This model depicts a scenario much closer to reality than the very basic pooled OLS specification, but the panel component of the data must be included if the model should estimate anything close to a causal relationship, which is needed when making the cost-effectiveness prediction for the PHCP. As a final note regarding this model specification, one can see that it is still not possible to determine which of the exercise types is more efficient.

Table 5 – Pooled OLS Including Control Variables

	Usefulness	Relaxed	Energy	Clear Thinking	Confidence in Self
Walks					
1	0.26***	0.10*	0.15**	0.10*	0.11*
2	0.26***	0.19***	0.22***	0.18***	0.19***
3	0.33***	0.23***	0.28***	0.20***	0.23***
4	0.28***	0.21***	0.30***	0.18***	0.21***
5	0.38***	0.25***	0.34***	0.24***	0.28***
6	0.34***	0.23***	0.29***	0.27***	0.26***
7	0.37***	0.26***	0.37***	0.26***	0.28***
Vigorous					
1	0.16***	0.07*	0.15***	0.08**	0.13***
2	0.18***	0.15***	0.29***	0.14***	0.18***
3	0.25***	0.18***	0.35***	0.16***	0.21***
4	0.23***	0.22***	0.41***	0.20***	0.27***
5	0.18***	0.19***	0.28***	0.15***	0.18***
6	0.22***	0.25***	0.40***	0.15***	0.26***
7	0.25***	0.14***	0.35***	0.16***	0.24***
Age	0.004***	0.005***	-0.003***	0.005***	0.001*
Gender	-0.06***	0.11***	0.13***	0.03*	0.12***

The table presents the marginal effects from the pooled OLS models featuring indicators of physical exercise as well as the respondents' age and gender as explanatory variables. The coefficients should be interpreted the following way: The zero-category is always the baseline category. A positive coefficient shows how large of an improvement the person would make if s/he were to change from exercising zero times per week to exercising more frequently.

The next, logically following step in the analysis is to include the individual fixed effects. The results from the model specification including these are presented in table 6. In this table we also include the random effects (RE) estimates of the model. These are said to be more efficient – which they clearly are – but also biased. They are used as a type of intuitive robustness check in the sense that the fixed effects (FE) estimates are compared to the random effects estimates, the latter most probably being incorrect. The FE and RE estimates should be of the same sign, but as the random effects estimates do not exclude the individual time invariant differences from the estimation, the coefficients estimated using random effects are most probably driven by some underlying factors – e.g. genetics, amount of physical activity and sports played in early years, socio-economic background etc. – that most likely affect a person's psychological well-being, as well as how they respond to physical activity. This means that one would expect the RE estimates to be flagged as highly significant more often than the FE estimates. One would also want to see that some of the estimated coefficients differ quite a bit, as the random effects are very likely to be biased in this case. If the coefficients would be very similar both in terms of magnitude and significance, this would cause suspicion regarding the accuracy of the fixed effects estimates.

Table 6 – Comparing Fixed and Random Effects

	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
	Usefulness		Relaxed		Energy		Clear Thinking		Confidence in Self	
Walks										
1	0.17	0.26 ***	0.20	0.09 *	0.06	0.14 **	0.24	0.10 *	0.15	0.10 *
2	0.28 *	0.27 ***	0.26 *	0.18 ***	0.35 **	0.21 ***	0.26 *	0.17 ***	0.25 *	0.18 ***
3	0.27 *	0.34 ***	0.35 **	0.22 ***	0.41 ***	0.27 ***	0.38 ***	0.20 ***	0.27 *	0.22 ***
4	0.17	0.29 ***	0.07	0.20 ***	0.18	0.28 ***	0.25 *	0.18 ***	0.10	0.20 ***
5	0.25 *	0.38 ***	0.16	0.24 ***	0.31 **	0.32 ***	0.28 **	0.24 ***	0.27 *	0.26 ***
6	0.25	0.35 ***	0.32 **	0.22 ***	0.30 *	0.28 ***	0.41 ***	0.27 ***	0.20	0.25 ***
7	0.36 ***	0.38 ***	0.24 **	0.25 ***	0.35 ***	0.35 ***	0.40 ***	0.25 ***	0.28 **	0.26 ***
Vigorous										
1	0.16	0.15 ***	0.008	0.07 **	0.23 **	0.16 ***	0.10	0.09 ***	0.16	0.14 ***
2	0.33 ***	0.17 ***	0.18 *	0.16 ***	0.37 ***	0.31 ***	0.27 ***	0.14 ***	0.36 ***	0.19 ***
3	0.32 ***	0.24 ***	0.27 ***	0.19 ***	0.56 ***	0.37 ***	0.21 **	0.16 ***	0.30 ***	0.23 ***
4	0.30 **	0.22 ***	0.34 ***	0.24 ***	0.55 ***	0.44 ***	0.25 **	0.21 ***	0.41 ***	0.29 ***
5	0.14	0.16 ***	0.29 **	0.22 ***	0.52 ***	0.32 ***	0.20 *	0.16 ***	0.25 **	0.21 ***
6	0.45 **	0.20 ***	0.41 **	0.29 ***	0.66 ***	0.45 ***	0.38 **	0.17 ***	0.48 **	0.31 ***
7	0.21	0.23 ***	0.15	0.16 ***	0.48 ***	0.37 ***	0.27 *	0.17 ***	0.36 **	0.27 ***
Age	0.002	0.004 ***	0.005 *	0.005 ***	-0.001	-0.002 ***	0.005 **	0.005 ***	-0.001	0.001 *

The table presents the marginal effects from both the fixed effects and random effects models, featuring indicators of physical exercise and respondents' age as explanatory variables. The coefficients should be interpreted the following way: The zero-category is always the baseline category. A positive coefficient shows how large of an improvement the person would make if s/he were to change from exercising zero times per week to exercising more frequently.

Looking at the coefficients in table 6, the anticipated differences between the two methods of parameter estimation are observed. The random effects estimates are often significant at the 0.001 level (***), whereas the fixed effects estimates are much less likely to be that significant. Also, the estimates almost constantly differ. One of the few occasions where they actually match fairly well is found when looking at the effect of vigorous training on the reported feeling of usefulness. Training vigorously five days per week produces estimated parameters of 0.18 for FE and 0.17 for random effects, the latter being significant at the 0.001 level and the former not being significant at all. Still looking at the same variables, but seven days per week instead of five, the FE estimate is 0.24, significant at the 0.05 level, and the RE estimate is 0.25, significant at the 0.001 level. Results such as these indicate that using fixed effects is a more conservative and less naïve approach than running random effects models on the data. The results here tell us that one is most likely better off using the fixed effects models. The magnitudes of the coefficients are still reasonable, regardless of using random or fixed effects. If this was not the case for the fixed effects models, we would find ourselves being in trouble. This is however not the case, which is consistent with the theoretical foundation we base our model selection on, so we proceed with the fixed effects method.

It is still very difficult to determine if one of the exercise regimes can be considered better than the other though. Training more often does seem to produce better results than just training a few times per week. All of the categories are compared to the option that the respondents do not exercise at all, i.e. the zero-category. There are some exceptions though. For instance walking for 10 consecutive minutes three times per week seems to have a much greater impact of perceived energy levels than taking walks more often. We see no logical explanation for this phenomenon and attribute it to uneven distribution among answers in the data. Some categories are simply a lot more common than others and consequently produce better estimates. In the previous literature section a similar problem, observed by Proper et al., (2004), is mentioned. In this paper, the resulting cost-benefit estimates of corporate physical exercise programs are quite prominently skewed because of a few individuals being of very poor health in the beginning of the study. In our dataset it is quite rare that people are physically active more than 4 times per week. This means that the top categories, category 6 and 7, are a lot less frequent. If we are “unlucky” and get some answers from individuals who are indeed exercising a lot, but are also leading otherwise stressful lives, these individuals can impact the estimated parameters and distort them in the way observed in the paper by Proper et al.

So far it seems that a fixed effects model including control variables for respondents’ age and gender produce the most realistic estimates. In order to improve the similarity between the theoretical results underlying the prognosis and the practical situation we are facing the data is divided into subsets. In the PHCP, *BetaHälsan* will screen the participants into healthy and unhealthy groups, and these will be recommended different training programmes. This health screening is simulated by creating a cut-off in the English respondents' reported general health status. An answer greater than or equal to 4 means that a respondent is put in the high health group (HH). The rest of the respondents, answering 3 or lower, are put in the relatively low health group (LH). The results from the models being run on these subsets are presented below in table 7 and there are a few things that should be considered before drawing any conclusions from this table. The LH subset only contains 2,422 observations, which is possibly too few to produce accurate estimates. This explains why there are so few significant estimates. The HH subset holds 9,595 observations, which means that the standard errors are much smaller in this subset. Comparing the magnitudes of the estimates in the different subsets, one can still conclude that the magnitudes of the coefficients seem realistic. This indicates that these subset models potentially reflect the true relationship between physical exercise and psychological well-being, as they are the most case specific and realistic models we can construct. It is still not possible to clearly say that one type of exercise is better than the other. There are some indications that the healthy subset benefits more from vigorous exercise. This can be seen in the models using *relaxed* and *energy* as dependent variables.

Table 7 – Subsets

	LH	HH	LH	HH	LH	HH	LH	HH	LH	HH
	Usefulness		Relaxed		Energy		Clear Thinking		Confidence in Self	
Walks										
1	0.05	0.25	0.49	0.17	0.38	0.05	-0.08	0.32 *	0.29	0.14
2	0.29	0.30 *	0.24	0.28 *	0.12	0.46 ***	0.42	0.22	0.15	0.29 *
3	0.05	0.37 **	0.48	0.35 **	0.48	0.46 ***	0.44	0.39 **	0.06	0.34 **
4	0.17	0.21	0.08	0.13	0.31	0.23	0.24	0.30 *	0.13	0.12
5	0.05	0.33 **	0.13	0.21	0.40	0.34 **	0.09	0.36 ***	0.07	0.33 **
6	0.24	0.30 *	0.08	0.45 **	0.35	0.36 *	0.81 **	0.34 **	0.31	0.21
7	0.47 *	0.38 ***	0.36	0.26 **	0.63 **	0.35 ***	0.46 *	0.43 ***	0.38	0.30 **
Vigorous										
1		0.16		0.01		0.21 *		0.04		0.14
2		0.29 ***		0.17 *		0.32 ***		0.25 **		0.30 ***
3		0.27 **		0.29 **		0.53 ***		0.16 *		0.27 **
4		0.18		0.26 *		0.48 ***		0.08		0.33 **
5		0.11		0.34 ***		0.58 ***		0.21 *		0.29 **
6		0.44 **		0.57 ***		0.75 ***		0.33 *		0.42 *
7		0.15		0.19		0.36 **		0.19		0.24
Age	-0.01	0.004	0.01	0.007 **	-0.01	0.002	0.03 *	0.01 ***	0.02	0.002

The table presents the marginal effects from the fixed effects models being run on sample subsets. The models feature indicators of physical exercise and respondents' age as explanatory variables. The coefficients should be interpreted the following way: The zero-category is always the baseline category. A positive coefficient shows how large of an improvement the person would make if s/he were to change from exercising zero times per week to exercising more frequently.

It is unfortunate that only a few significant estimates are observed in the relatively unhealthy subset, given that this is the most realistic scenario we can construct using the data available to us. Regardless of the significance levels of the estimates, this is the model we will use to construct the prognosis. The practical implication of the non-significant results in the LH subset is that the interval for the prognosis will be broader for the light type of exercise than the vigorous exercise. This can be a potential problem, but the model still produces estimates of realistic magnitude, using a method that carries little bias and gets close to estimating a causal relationship. The model specification used on the subsets definitely seems the best one, regardless of statistical significance.

The models used for making the predictions are the following:

Relatively healthy subset:

$$y_{it} = \beta_0 + \boldsymbol{\beta_1} Walks_{it} + \boldsymbol{\beta_2} Vigorous_{it} + \beta_3 Age_{it} + \alpha_i + \omega_{it} \tag{4}$$

Relatively unhealthy subset:

$$y_{it} = \beta_0 + \boldsymbol{\beta_1} Walks_{it} + \beta_3 Age_{it} + \alpha_i + \omega_{it} \tag{5}$$

The subscripts have the same meaning as described in the method section above. The total number of predictions made with these models is 235. Out of these, 230 predictions are successful in the sense that our survey respondents' real values fall within the models prediction intervals. This translates to a success rate of 97.9%. A complete list of the prediction intervals is found in Appendix E.

As discussed earlier, an accurate estimate of the entire project's cost-effectiveness is very difficult to produce, but in table 8 below one finds some suggestions on what the cost-effectiveness could be for different individuals. The individuals included in the table below are drawn from our surveys actual respondents and are thus real potential participants in the PHCP. This should realistically reflect the situation the companies are facing. It should however be mentioned that these calculations are made under certain assumptions. It is assumed that the participants will comply with their training programs and that the suggested exercise sessions will not function as substitutes for other training. The cost of participation, 2,500 SEK per person, is an approximate cost derived by the companies planning the project. It is based on the assumption that there will be roughly 50 participants, taking part in three exercise sessions per week. The calculations one finds in table 8 are made for one active person, one moderately active person, and one fairly inactive person. The low activity individual walks three times per week and does not exercise vigorously at all. This person belongs to the relatively unhealthy subset, and is assumed to add three walking sessions per week. The moderately active person walks four times per week and exercises vigorously once per week. This individual belongs to the healthy subset and gets to choose from both walking sessions and more intense training. He/she is assumed to add one walk per week and two sessions of vigorous training. The active person walks six times per week and exercises vigorously three times a week. This person is assumed to add three sessions of vigorous training per week.

The calculations in table 8 are made in the following way: The base level is the present, reported score for each of the dependent variables. The predicted outcome level is derived with the parameter estimates in the regression model including the subsets. As a person increases the training frequency, he or she moves from one category of the explanatory variables to some other category. The effect of this change is measured as the difference between the estimated beta coefficients, $\hat{\beta}$ of the new category minus $\hat{\beta}$ of the old category, and this difference is added to the base level. The cost-effectiveness is measured in percentage improvement per invested 1,000 SEK. The percentage improvement is calculated in the following way:

$$Percentage\ Improvement = \left(\frac{Predicted\ Outcome\ Level}{Base\ Level} - 1\right) \times 100 \qquad (6)$$

This is then divided by the cost, 2,500 SEK.

Table 8 – Cost-effectiveness

	Base Level	Cost (SEK/person)	Predicted Outcome Level	Cost-effectiveness (in percentage improvement/1000SEK)
		Low Activity Individual		
Useful	3	2,500	3.19	2.53
Relaxed	2	2,500	1.60	-8
Energy	2	2,500	1.87	-2.6
Clear Thinking	3	2,500	3.37	4.93
Confidence	3	2,500	3.25	3.33
		Moderately Active Individual		
Useful	4	2,500	4.23	2.30
Relaxed	3	2,500	3.36	4.80
Energy	3	2,500	3.43	5.73
Clear Thinking	4	2,500	4.18	1.80
Confidence	3	2,500	3.34	4.53
		High Activity Individual		
Useful	4	2,500	4.17	1.7
Relaxed	3	2,500	3.28	3.73
Energy	3	2,500	3.22	2.93
Clear Thinking	4	2,500	4.17	1.7
Confidence	4	2,500	4.15	1.5

The estimated cost-effectiveness for three of our survey respondents is found in the table. The numbers in the right column are percentages and show the improvement in psychological well-being per invested 1,000 SEK.

As mentioned, the cost-effectiveness is estimated under the assumption that the participants actually follow the training programs they are provided. If it turns out they do not, it is easy to adjust the cost-effectiveness estimates in retrospect. This is one of the merits of the regression framework used to make the estimates. If somebody does not exercise as frequently, or more frequently, than that person is recommended and expected to, one can simply insert new values of exercise frequency in the model and get a new prediction. In table 8 one also observes that some of the estimates are negative, implying that investing in a project similar to the PHCP could lead to people being worse off in terms of psychological well-being. This is perhaps not a very realistic scenario, and it will be brought up in the discussion section.

5) Discussion

The discussion segment of this paper will raise a few points that need to be criticized in chronological order. The discussion is started off by critiquing the English data used for building the theoretical model. After that some methodological concerns are raised, and finally the discussion is rounded off by examining the results and conclusions.

5.1) Data

As always, the sample taken from the *Health Survey for England* may not represent the population very well. When the interviews are conducted over the phone during office hours, mostly old and young people answer, which we see when we drop these from the dataset. A vast majority of the individuals are discarded. Those of working age who answer the home/private phone during office hours might not be representative of the working majority.

One could also suspect that the answer scale to the survey's questions is not accurate enough. Nearly everyone answers 4 or 5, more or less meaning that they describe their health state as very good. People's tendency to always answer the most positive alternatives means that any estimated beta coefficients might underestimate the true relationship between physical exercise and mental well-being. Realistically speaking, somebody in a population has to feel the worst, somebody has to feel the best, and most people will be somewhere in between. It is not realistic that everybody feels more or less equally good. Of course it is impossible for everybody to know how they feel in comparison to others, so one is simply not very likely to get a truthful measure of people's well-being by asking questions in this way.

Another problem we have encountered is connected to the collection of survey responses. The survey response rate was surprisingly low. As far as we can see this depends on the fact that this paper and the actual PHCP ran according to different schedules. By the time we had to collect the survey responses, only a few companies at *Medicon Village* were involved in the project, implying that many of the potential respondents had little knowledge of what the survey actually concerned. If the survey would have been handed out at a later stage, it is likely that more people would see the value of participating and the response rate would most likely have been substantially better.

5.2) Method

The first concern that is raised has to do with the treatment of the variables in the regression models. All the variables are factor variables when collected, but not all of them are treated as such when used in the regression models. The dependent variables, namely the indicators of mental well-being, are treated as continuous variables, whereas the explanatory variables measuring level of physical activity are kept in their factor form. The analysis is carried out in this way as the companies who are meant to take part of the results are interested in knowing how different scenarios of exercise frequency compare to each other. This could for instance be the following: Say that somebody walks five times per week and exercises vigorously once per week. If that person would change his or her exercise pattern to three walks per week and three vigorous work outs per week, what would the effect of making such a change be? Also, the relationship between exercise frequency and mental well-being is not a linear one, so treating the explanatory variables as continuous variables might not be a good way to represent the data. Doing that would produce one linear marginal effect for the variable, not taking the nonlinearity of the relationship into account. When a person reaches a certain amount of weekly exercise, the coefficients seize to grow. There appears to be diminishing returns from training.

Furthermore, one could question the variables used, or rather the ones that are not used. The models estimated in this paper only include two control variables, age and gender. One could easily imagine that other controls should also be included. One such example is the respondent's education. This variable is actually included in the data set used for estimating the models. The primary reason for excluding it from the analysis is that it greatly reduces the number of observations used in the estimations. Very few individuals answer both this question and the questions regarding physical activity. This means that one is left with very few complete observations that can be used to estimate the model. One can look at the relatively unhealthy subset to get an idea of the problem: In many of the possible categories of the explanatory variables, only around 20 highly educated individuals are included. Before excluding the education variable, its correlation to each of the dependent variables is checked and found to be very weak in magnitude. Ideally, one would probably want to include more control variables in the model, education being one of them, but considering how small the data set becomes as this is included, and how weakly it correlates with the dependent variables, it simply does not seem worth it.

5.3) Results

The goal of this paper is to estimate the cost-effectiveness of a physical exercise based health care project. The final result of the analysis is arguably not an estimate, but a model which can be used to create estimates. This might be disappointing to some readers, but given the fact that the cost-effectiveness of such a project will be different for every individual, it does not really make sense to give one estimate that should be representative of every individual. One could of course take the averages of the responses to our survey and insert these in the regression model. This would produce a cost-effectiveness estimate for some hypothetical "average" person, but that estimate would also be wrong in more or less every case. It makes more sense to report the model and some possible scenarios, so that the person reading the paper can get an idea of what might be expected for different individuals.

The realized results of the project will be greatly affected by participants' compliance with their training programs, but even if they do comply with the programs they are given it could be extremely difficult to tell what the effects of the PHCP actually are. If the participants simply substitute training that they would do in their free time anyway for the recommended exercise sessions in the PHCP, one would probably not be able to say that the project improved this particular participant's health status. Unless the project actually makes somebody exercise more than they already do, it will most probably be ineffective. So far, we have not come up with a good idea to solve this problem. It seems impossible to tell whether a person will substitute the training he or she already does for the training provided in the PHCP. Ultimately this could lead to the estimated cost-effectiveness of the project being completely wrong. Because of this, we need to state that the results are based on the assumption that the PHCP's training sessions will not substitute other training. We have a suspicion that this might prove to be a quite unrealistic assumption.

Another concern regarding the results in table 8 is presented. This table holds the final cost-effectiveness estimates and is meant to illustrate a few possible scenarios. Some of the results in this table are negative, implying that investing in exercise based preventive health care can have a negative impact on somebody's health status. This seems counterintuitive and strange and is a result of the fact that some of the parameter in the model that the predictions are based on are negative. These parameters are not statistically significant, so one really cannot say whether or not the negative effect of physical exercise on health is real. This does not mean that they are necessarily completely wrong either. There are a few imaginable situations where more physical exercise can have a negative impact on mental health. If the boss of a company pays the participation fee for a similar project and tells the staff that he paid the fee and that they are all to participate, this might cause stress with some employees. Perhaps they feel more or less obligated to participate by some external force, even if they do not want to themselves. In the tables where the results are presented, one can also see that exercising more does not necessarily mean that one feels better. This could simply be because the training in itself becomes very time consuming as its frequency reaches a certain point. This could in turn lead to changing priorities and neglecting other things that could make the person happy. In short: It is possible, at least in theory, that exercising does not have a positive effect on mental well-being, at least not in the short run.

Now, basing the estimation on a model without many significant parameters can be considered a bad choice. We choose to do this anyway because it is of great importance that the model is constructed in a way that mirrors the practical situation as well possible. This paper evaluates a project that is being planned in real life and could, as mentioned early in the paper, be of potentially high stakes for some participating companies with small turnovers. Keeping this consideration in mind, we do not think that it is necessarily a bad thing that a model without significant parameters of great magnitude is used to make the predictions. We make rather careful estimates of the cost-effectiveness, and in a practical situation it is also plausible that the project does not have an impact on the participants health, due to reasons such as lack of compliance, just substituting exercise sessions that would have taken place anyway for exercising via work, or just the fact that people's psychological well-being is so greatly affected by factors that lie outside the control of any training plan or health care project.

6) Conclusion

In fulfilling its purpose, this paper uses regression analysis as a framework for estimating the cost-effectiveness of physical exercise based health care projects. Authors of earlier studies, e.g Uegaki et al., (2010), have found that it is virtually impossible to compare cost-effectiveness and cost-benefit analyses of such projects as there are no standardized ways of making the calculations. Using a regression model to make predictions of cost-effectiveness is one theoretically possible way of doing the calculations. It has the advantages of being fairly flexible, meaning that it is easy to get an estimated cost-effectiveness for each individual of interest. As always though, the accuracy of these estimates depends on the model specification being correct. Furthermore, the prediction intervals produced can be quite wide, which might mean that they are of little practical use.

In this analysis, mental well-being is regressed on frequency of physical exercise. The most realistic model we can come up with, given the panel data set and scenario we are working with, is a fixed effects model including respondents' age as controls variable. Using the fixed effects design hopefully removes the individual specific factors that affect both the mental well-being and exercise habits of the respondents. If one wants to get close to estimating a causal relationship, these aspects are necessarily removed from the model.

The most realistic model estimated in the paper is compared to cross sectional data gathered from a different time and place, namely at *Medicon Village* in Lund, Sweden, in the spring of 2016. In 97.9% of the cases the real reported values of the mental well-being indicators fall within the model's prediction intervals. This might lead one to draw the conclusion that the model is fairly well thought out, but whether the point estimates of cost-effectiveness are reasonably accurate or not can only be assessed by comparing these estimates to actual cost-effectiveness, which means that one needs to evaluate an actual project and compare the actual, realized results to the predictions. In short: only time can tell if the ideas underlying this regression based method for estimating cost-effectiveness are practically sound.

The final results of this paper are the following: The point estimates of cost-effectiveness range from +1.70% to +5.73% improvements in psychological well-being per 1,000 SEK invested. The improvements differ depending on individual levels of physical activity and dependent variable used in the model. There are some odd results as well, showing negative cost-effectiveness estimates of -5% and -8%. In all likelihood the negative estimates are not realistic, but depend on low answer frequency for certain categories in the data. The estimates are made for three individuals who answer the survey handed out to *Medicon Village* and should not be considered universally true. They are merely suggestions of cost-effectiveness, but as the analysis leading up to them is thorough and conservative, the estimates serve their advisory purpose.

7) References

Baecke, J., et al., (1982). A Short Questionnaire for the Measurement of Habitual Physical Activity in Epidemiological Studies. *The American Journal of Clinical Nutrition*. Nov. 1982. 936-942.

Bredahl et al., (2008). Exercise on Prescription. Effect of attendance on participants' psychological factors in a Danish version of Exercise on Prescription: A Study Protocol. *BMC Health Services Research* 2008. 8:139

Clark, T. & Linzer, D., (2015). Should I Use Fixed or Random Effects? *Political Science Research and Methods*. 2015. Vol. 3. No 2. pp. 399-408.

Coulsen, J.C., et al., (2008). Exercising at work and self-reported work performance. *International Journal of Workplace Health Management*. Vol. 1 No. 3-2008.

Craft, B., et al., (2014), Gender Differences in Exercise Habits and Quality of Life Reports: Assessing the Moderating Effects of Reasons for Exercise, *International Journal of Liberal Arts and Social Science*, Vol. 2, No. 5, June 2014.

Faraway J. (2002), Practical Regression and Anova using R, p. 39.

Gerber, M., et al.. (2014). Physical Activity in Employees with Differing Occupational Stress and Mental Health Profiles: A Latent Profile Analysis. *Psychology of Sports and Exercise*. Aug. 8. 2014. pp. 649-658.

Gould, W., (2013), How can there be an intercept in the fixed-effects model estimated by xtreg, fe?, *Stata – Data Analysis and Statistical Software*, URL: http://www.stata.com/support/faqs/statistics/intercept-in-fixed-effects-model/, accessed: 2016-04-22.

Hemp, P., (2004). Presenteeism – At Work but out of It. *Harvard Business Review*. Oct. 2004.

Huber, M., et al., (2015), Workplace Health Promotion and Labour Market Performance of Employees, *Journal of Health Economics*, Vol. 43, September 2015, pp. 170-189.

Lindegård, A., et al., (2015). Changes in Mental Health in Compliers and Non-Compliers with Physical Activity Recommendations in Patients with Stress-Related Exhaustion. *BMC Psychiatry*. Nov. 4. 2015.

Norris, C.M., et al., (2006). Ordinal Regression Model and the Linear Regression Model Were Superior to the Logistic Regression Model. *Journal of Clinical Epidemiology*. 59. 2006. pp. 448-456.

Pauly, M.V., et al., (2008), Valuing Reductions in On-the-Job Illness: 'Presenteeism' from Managerial and Economic Perspectives, *Health Economics*, No. 17, pp. 469-485.

Proper, K.., et al.. (2004). Costs. Benefits and Effectiveness of Worksite Physical Activity Counseling from the Employer's Perspective. *Scandinavian Journal of Work. Environment & Health*. 2004;30(1):36-46.

Puig-Ribera, A., et al.. (2015). Self-Reported Sitting Time and Physical Activity: Interactive Associations with Mental Well-Being and Productivity in Office Employees. *BMC Public Health*. Jan. 31. 2015.

Torra, V., et al.. (2006). Regression for Ordinal Variables without Underlying Continuous Variables. *Information Sciences*.176. 2006. pp. 465-474.

Tuomi, K.., et al., (2004). Organizational Practices. Work Demands and the Well-Being of Employees: a Follow-Up Study in the Metal Industry and Retail Trade. *Occupational Medicine*. 2004. Vol 54. 115-121.

Uegaki, K.,. et al., (2010). Economic Evaluation of Occupational Health Interventions from a Company's Perspective: A Systematic Review of Methods to Estimate the Cost of Health-Related Productivity Loss. *Journal of Occupational Rehabilitation*. 2011. 21: 90-99.

Verbeek M., (2012), A Guide to Modern Econometrics, John Wiley & Sons Ltd., 4th edition, pp. 6-53, 377-378.

Försäkringskassan. 2016. Ohälsotalet efter Län och Åldersgrupp. https://www.forsakringskassan.se/wps/portal/statistik/statistik_och_analys2/Sjuk/ohalsomatt/!ut/p/a1/pVJbb4IwGP0tPuyRtIWCvDYgOIw4YN2kL0stoig3Q7ds_37FZAlOnVvWh-Zrck57LgUMLAGr-Vux4bJoal72Z2a94KnrIN-

Bs8XcI5Dcu6GJw0j3CVSAdAjwk8iFBNFYn8aW_5BYp3z_0ZtAQoKF4egOmgVffHhlqQdu8J8BA0zUspVbkIqmlutadus72ElloZPF_nTcvaq92fKyayou5XDuL2pFkYEUIbQyhGFpGEOh4QwLbcV5rmW5mQuBhOB6fmb8TNmN4I7ChxBInzwFweHcjqiOKf4OuJDtEfBDeKkSOb6uYgySP7oOfvEfit3hwIhqpW_jXYLl_2ppK0or2_jQ9rENDbPckNHoEyKW-vs!/?1dmy&urile=wcm%3apath%3a%2Fcontentse%2Fstatistik%2Fstatistik%2Fsjuk%2Fohalsomatt%2Fohtal_offstat1%2Fohtal_offstat1.

Försäkringskassan. 2016. Beräkna Kostnader för Sjukfrånvaro. https://gate1.forsakringskassan.se/agetjanster/public/sjukfranvarokalkyl/startSjukfranvarokalkyl.do

Data:

NatCen Social Research. University College London. Department of Epidemiology and Public Health. (2016). *Health Survey for England. 2009-2014*. [2016-03--30]. UK Data Service. SN: 7919. http://dx.doi.org/10.5255/UKDA-SN-7919-1.

Appendix A

Hälsoenkät

Enkäten utgör en del av ett Masterarbete i Hälsoekonomi på Lunds Universitet och tar max 5 minuter att besvara. Vi undersöker de potentiella effekterna en friskvårdssatsning på *Medicon Village* skulle kunna ha. Syftet med arbetet är att göra en prognos för hur kostnadseffektivt ett sådant friskvårdsprojekt skulle vara.

Ditt svar är viktig för att skapa en så god förutsättning för analys som möjligt. Även om ni inte kommer delta i ett sådant projekt så är Ert svar viktigt. Era uppgifter kommer i sådana fall ingå i en kontrollgrupp för undersökningen. Ni kommer få ta del av resultatet som sedan kan användas som underlag för framtida beslut.

Ditt deltagande är helt anonymt. Personuppgifter behandlas enligt PUL. Personuppgifter används enbart för att dela upp datamaterialet och kommer raderas innan resultaten publiceras.

Initialer:

Ålder:

Kön:

Har du universitetsutbildning?

- Ja
- Nej

Vad har du för yrkestitel?

VÄLBEFINNANDE

Hur bedömer du ditt allmänna hälsotillstånd?
- o Mycket bra
- o Bra
- o Någorlunda
- o Dåligt
- o Mycket dåligt

När Du besvarar följande frågor, fundera över hur du känt dig den senaste veckan.

1) Har du känt dig produktiv?
- o Aldrig
- o Sällan
- o Ibland
- o Ofta
- o Alltid

2) Har du känt dig avspänd eller stressad?
- o Alltid stressad
- o Oftast stressad och sällan avspänd
- o Båda känslor lika vanliga
- o Oftast avspänd och sällan stressad
- o Alltid avspänd

3) Har du haft energi över efter arbetsdagen?
- o Aldrig
- o Sällan
- o Ibland
- o Ofta
- o Alltid

4) Har du kunnat koncentrera dig ordentligt?
- o Aldrig
- o Sällan
- o Ibland
- o Ofta
- o Alltid

5) Har du känt dig tillfreds med dig själv?
- o Aldrig
- o Sällan
- o Ibland
- o Ofta
- o Alltid

6) Har du känt dig positiv?
- o Aldrig
- o Sällan
- o Ibland
- o Ofta
- o Alltid

MOTIONSVANOR

7) **Hur många dagar har du promenerat minst 10 sammanhängande minuter?**
- o 0
- o 1
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7

8) **Hur många dagar fått måttlig motion?** (Ev. promenader ej inräknade)
- o 0
- o 1
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7

9) **Hur många dagar har du tränat intensivt?** (Med intensiv träning menas att du svettats och blivit rejält andfådd)
- o 0
- o 1
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7

Appendix B

Descriptive statistics of the data from the *Medicon Village* respondents

Variable	Observations	Mean	Standard Deviation	Min	Max
General Health	50	4.15	0.65	2	5
Useful	50	4.02	0.64	3	5
Relaxed	50	3.30	0.90	1	5
Energy	50	3.62	0.82	2	5
Clear Thinking	50	4.00	0.68	3	5
Confidence	50	3.98	0.70	3	5
Walks	50	4.87	1.87	1	7
Vigorous	50	1.74	1.65	0	5

Descriptive statistics of the data from the *Health Survey for England.*

Variable	Observations	Mean	Standard Deviation	Min	Max
General Health	35,642	4.08	0.9	1	5
Useful	26,931	3.67	0.89	1	5
Relaxed	26,905	3.34	0.88	1	5
Energy	26,941	3.06	0.96	1	5
Clear Thinking	26,975	3.86	0.82	1	5
Confidence	26,978	3.58	0.9	1	5
Walks	12,298	4.87	2.42	0	7
Vigorous	12,27	1.93	2.19	0	7

Appendix C

Results from the five Hausman tests:

Dependent Variable of the Tested Model	P-Value
Useful	0.5122
Relaxed	0.5813
Energy	0.3126
Clear Thinking	0.5588
Confidence	0.6723

The test statistic is calculated by calculating the squared difference between the fixed and random effects' estimated coefficients, $\left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)^2$, and then dividing the difference between the respective coefficients variances, $v\left(\hat{\beta}_{FE}\right) - v(\hat{\beta}_{RE})$. The test statistic follows a Chi^2-distribution. A significant test values indicate that one should reject the random effects method. Clearly, these results indicate that the models should be estimated using random effects. For reasons explained in the method section, this estimation method is rejected anyway.

Appendix D

Shortly reviewing how prediction intervals are constructed will help create an intuitive understanding of prediction intervals and explain how they relate to this analysis. The simplest version of such an interval is made when constructing the interval from a cross-sectional model with just one explanatory variable and looks the following way:

$$PI_{y_o} = \hat{y}_0 \pm 1.96 \times \sqrt{\widehat{\sigma^2} \times \left(1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_{\forall i}(x_i - \bar{x})^2}\right)}$$

Where $\widehat{\sigma^2}$ is the estimated variance of the error term, $\frac{1}{N-1}\sum_{\forall i} e_i^2$. The letter N is the total number of observations and the subscript 0 just indicates that one is predicting a value of the dependent variable, *y,* for some hypothetical person called individual 0. In making this prediction, one inserts a hypothetical value of the explanatory variable, x_o, in the regression equation. Formulated this way, with the 1.96 multiplier, the prediction interval is obviously based on the assumption that the underlying parameter estimate is normally distributed. Instead of 1.96, one could use the *t*-value, for instance, corresponding to some desired significance level and correct amount of degrees of freedom. Now, if there are more explanatory variables than one, the interval is most easily written using vector notation. It will look the following way:

$$PI_{y_0} = \hat{y}_0 \pm 1.96 \times \sqrt{\widehat{\sigma^2} \times (1 + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)}$$

The bold font indicates that the symbols in the expression refer to vectors of variables and the apostrophe indicates that the vector is transposed. The transposed vector holds all the explanatory variables for the hypothetical individual 0: $\boldsymbol{x}_0' = (x_{00}, x_{01}, x_{02}, \dots, x_{0k})$. The error component of the expression, $\widehat{\sigma^2}$, is still the estimated variance of the residuals:

$$\widehat{\sigma^2} = \frac{1}{N-K}\sum_{\forall i} e_i^2$$

The letter N still denotes the total number of observations in the sample, and K denotes the number of parameters estimated. These notations become a little bit messier when a time dimension is added, but the idea behind it is still pretty straight forward, as described by an example of the standard fixed effects case: When running a fixed effects regression, the predicted value of the hypothetical individual, called individual 0, is equal to the estimated parameters from the model plus some error term, which one does not know in advance. It can be written as $y_{0t} = \boldsymbol{x}'_{0t}\widehat{\boldsymbol{\beta}_{\boldsymbol{FE}}} + \omega_{0t}$ where the FE subscript indicates that this example concerns the fixed effects within estimate. The variance of this value, $V(y^*_{0t}) = V(\boldsymbol{x}'_{0t}\widehat{\boldsymbol{\beta}_{\boldsymbol{FE}}} + \omega_{0t})$, can be written in the following way:

$$\boldsymbol{x}'_{0t}V\left(\widehat{\boldsymbol{\beta}_{\boldsymbol{FE}}}\right)\boldsymbol{x}_{0t} + \sigma^2 \approx \hat{\sigma}^2(\boldsymbol{x}'_{0t}\left(\sum_{\forall t}\sum_{\forall i}\left((x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1}\right)\boldsymbol{x}_{0t} + 1)$$

In the above expression, $\hat{\sigma}^2 = \frac{1}{N(T-1)}\sum_{\forall t}\sum_{\forall i} e^2_{it}$, which simply means that the variance of the error term is estimated using the residuals. The prediction interval will hence be:

$$PI_{y_{0t}} = \hat{y}_{0t} \pm 1.96 \times \sqrt{\hat{\sigma}^2(\boldsymbol{x}'_{0t}\left(\sum_{\forall t}\sum_{\forall i}\left((x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1}\right)\boldsymbol{x}_{0t} + 1)}$$

As mentioned earlier in the paper, the most realistic regression model is used to predict the outcomes of our survey respondents. If the respondents' true, reported outcomes consistently fall within the model's prediction interval, we consider the model specification valid. A few things should be mentioned here: The *Medicon Village* survey only produces cross sectional data. This means that the $\boldsymbol{x}'_{0t}$ and $\boldsymbol{x}_{0t}$ vectors from the expression described in the method section only hold one time period, $\boldsymbol{t}$. The part of the expression describing the standard error of the residuals and model parameters are also slightly different looking, as these errors belong parameters estimated using the somewhat special fixed effects method which is default to Stata and described by Gould (2013).

Appendix E

LB	Actual	UB	LB	Actual	UB	LB	Actual	UB	LB	Actual	UB	LB	Actual	UB
	Usefulness			Relaxed			Energy			Clear Thinking			Confidence in Self	
2.38	4	5.34	1.80	3	4.87	1.14	4	4.36	2.76	4	5.52	1.94	5	4.98
2.40	5	5.38	1.96	4	5.04	1.49	4	4.79	2.80	4	5.56	2.27	4	5.37
2.49	4	5.51	2.03	3	5.26	1.54	.	5.02	2.84	4	5.70	2.18	3	5.38
2.13	4	5.00	1.99	3	5.21	1.71	3	5.15	2.50	4	5.28	2.11	4	5.15
2.20	4	5.18	1.92	4	5.09	1.62	5	5.04	2.62	4	5.44	2.20	5	5.28
2.23	4	5.15	1.84	4	4.88	1.29	4	4.49	2.69	4	5.41	2.08	5	5.10
	.			.			.			.			.	
2.15	4	5.06	1.74	4	4.78	1.29	3	4.49	2.52	3	5.24	2.08	4	5.10
2.42	5	5.40	1.79	2	4.91	1.25	3	4.55	2.72	3	5.54	2.04	3	5.16
2.19	3	5.08	1.57	2	4.57	1.08	3	4.26	2.54	3	5.32	1.84	3	4.80
2.45	4	5.44	1.88	2	5.05	1.37	3	4.71	2.76	4	5.62	2.17	3	5.33
2.34	4	5.33	2.09	1	5.27	1.62	3	5.04	2.91	3	5.73	2.20	3	5.28
2.41	4	5.39	1.91	4	5.07	1.26	5	4.62	2.64	5	5.34	1.99	4	5.07
2.25	4	5.12	2.00	4	5.20	1.58	3	4.96	2.72	4	5.54	2.25	4	5.33
2.32	3	5.24	1.78	4	6.02	1.24	4	5.34	2.46	4	6.46	2.06	4	5.32
2.36	4	5.32	2.03	3	5.14	1.62	4	4.98	2.72	4	5.44	2.14	4	5.18
2.27	5	5.26	1.81	4	4.99	1.43	5	4.85	2.59	5	5.41	2.03	5	5.11
2.35	5	5.23	2.13	4	5.33	1.58	5	4.96	2.93	4	5.75	2.25	5	5.33
2.39	4	5.44	2.01	4	5.19	1.70	4	5.16	2.73	4	5.53	2.31	4	5.47
2.47	4	5.40	1.83	5	4.89	1.14	4	4.26	2.80	4	5.60	2.05	4	5.11
2.35	3	4.67	1.61	2	5.68	1.11	2	4.31	2.36	3	5.48	1.90	3	4.92
	.			.			.			.			.	
2.31	4	5.25	1.81	4	4.93	1.16	3	4.42	2.72	5	5.54	1.98	4	5.06
2.26	4	5.17	1.68	4	4.72	1.10	4	4.30	2.56	4	5.28	1.91	5	4.93
2.24	4	5.12	1.63	4	4.62	1.08	4	4.26	2.63	4	5.41	1.84	4	4.80
2.43	4	5.43	1.90	4	4.97	1.41	4	4.69	2.81	5	5.59	2.17	5	5.25
2.47	3	5.47	1.86	4	6.35	1.12	3	5.44	2.65	5	7.09	1.90	3	4.98
2.44	3	5.48	2.05	3	5.26	1.66	3	5.10	2.65	3	5.39	2.35	4	5.51
2.42	5	5.43	1.89	4	4.96	1.41	4	4.69	2.80	5	5.58	2.17	4	5.25
2.37	4	5.33	1.86	4	4.89	1.38	2	4.62	2.75	3	5.47	2.14	4	5.20
2.30	4	5.23	1.75	3	4.73	1.36	3	4.58	2.72	4	5.50	2.07	3	5.07
2.40	5	5.33	1.67	4	4.66	1.17	5	4.39	2.74	5	5.52	1.90	5	4.90

2.40	3	5.39	1.97	4	5.15	1.43	3	4.85	2.85	4	5.67	2.03	3	5.11
2.21	4	5.20	1.93	3	5.11	1.62	3	5.04	2.65	4	5.47	2.20	4	5.28
2.36	3	5.33	1.84	2	5.00	1.26	2	4.62	2.53	3	5.23	1.99	3	5.07
2.21	5	5.17	1.79	2	4.86	1.33	4	4.55	2.60	5	5.36	2.11	4	5.15
2.37	3	5.29	1.65	3	5.89	1.05	3	5.15	2.38	3	6.38	1.89	3	5.15
2.44	4	5.38	1.78	4	4.82	1.28	4	4.56	2.80	4	5.62	2.03	4	5.07
2.37	4	5.30	2.03	2	5.11	1.46	3	4.64	2.80	5	5.56	2.09	4	5.09
2.32	4	5.28	1.98	3	5.09	1.62	4	4.98	2.64	4	5.36	2.14	5	5.18
2.47	4	5.49	1.82	5	4.94	1.34	4	4.66	2.71	4	5.51	2.13	4	5.25
2.54	5	5.59	2.00	3	5.18	1.51	4	4.97	2.85	4	5.65	2.14	5	5.30
2.44	5	5.44	1.90	3	5.05	1.46	4	4.86	2.64	4	5.42	2.00	4	5.06
2.20	3	5.11	1.80	3	4.84	1.29	5	4.49	2.62	3	5.34	2.08	4	5.10
2.29	4	5.25	1.57	3	4.60	1.19	3	4.43	2.41	4	5.13	1.97	4	5.03
2.53	4	5.53	1.83	2	4.90	1.22	2	4.50	2.83	3	5.61	2.00	3	5.08
2.29	4	5.15	1.80	3	4.95	1.27	4	4.61	2.64	4	5.42	1.95	4	4.99
2.19	5	5.73	1.87	3	5.27	1.47	4	5.26	2.59	4	5.79	2.09	4	5.55
1.52	4	5.52	1.24	3	4.93	0.72	4	4.50	1.79	5	5.70	1.08	4	5.04
	.			.			.			.			.	

Actual = The survey respondent's reported value

LB = Lower Bound of the prediction interval

UB = Upper Bound of the prediction interval