

**Original citation:**

Pink, Joshua, Petrou, Stavros, Williamson, Esther M., Williams, Mark A. and Lamb, S. E. (Sallie E.). (2016) Economic and health-related quality of life outcomes of whiplash associated disorders. *Spine*, 41 (17). pp. 1378-1386.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/79601>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

This is not the final version of record. <http://doi.org/10.1097/BRS.0000000000001512>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

**Economic and Health-Related Quality of Life Outcomes of Whiplash Associated  
Disorders**

Joshua Pink<sup>a</sup>, Stavros Petrou<sup>a</sup>, Esther Williamson<sup>b</sup>, Mark Williams<sup>b</sup>, Sarah E Lamb<sup>b</sup>

<sup>a</sup>Warwick Clinical Trials Unit,

Division of Health Sciences,

Warwick Medical School,

University of Warwick,

Coventry,

CV4 7AL

<sup>b</sup>Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences,

Medical Sciences Division,

University of Oxford,

Oxford,

OX3 7HE

Contact details:

[j.pink@warwick.ac.uk](mailto:j.pink@warwick.ac.uk)

+44(0)24 76151790

Acknowledgement: July 23, 2014

1st Revise: November 25, 2014

2nd Revise: January 19, 2016

Accept: January 30, 2016

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

No relevant financial activities outside the submitted work.

ACCEPTED

## Abstract

**Study design:** This study examines the links between severity of whiplash associated disorder and costs and health outcomes

**Objective:** The aim of this study was to estimate the economic costs and health state utilities associated with disability levels and recovery trajectories following acute whiplash injury.

**Summary of background data:** Data used were from the Managing Injuries of the Neck Trial, which collected information on 3,851 people over a 12 month period following acute whiplash injury.

**Methods:** Effects of whiplash associated disorder severity on economic costs (measured from a societal perspective and separately from a health and personal social services perspective) were estimated using two-part regression models, comprising probability of incurring a cost and the total cost, given one was incurred. Effects on health state utilities (measured using the EQ-5D and SF-6D) were estimated using ordinary least squares regression, and two-part models as for costs.

**Results:** There was a direct relationship between severity of disability following acute whiplash injury and economic costs. Between baseline and 4 months, average societal costs for those with no disability were £99.55 (UK£, 2009 prices), increasing to £668.53 for those with complete disability. Average societal costs for the whole sample were £234.15 over the first 4 months, decreasing to £127.51 between 8 and 12 months. Conversely, utility scores decreased with increased disability. The average EQ-5D utility score was 0.934 at 4 months for those with no disability, decreasing to 0.033 for those with complete disability. The average EQ-5D utility score for the whole sample increased from 0.587 immediately post-injury to 0.817 at 12 months. Relative costs and disutilities generated by the multivariate models are also presented by disability level and recovery trajectory.

Conclusions: These results provide estimates of the costs and health state utilities associated with disability levels and recovery trajectories following acute whiplash injury. They can be used to inform estimates of the cost-effectiveness of interventions targeting whiplash associated disorders.

**Key Words:** Whiplash, Outcome assessment, Health-related quality of life, Cost, Economic evaluation, Patient-reported outcome-measures

**Level of Evidence:** 3

ACCEPTED

## Introduction

An acute whiplash injury (AWI) is a common soft-tissue injury often caused by motor vehicle collisions but also trips or falls resulting in a sudden jolt to the neck. The most common complaint following AWI is neck pain but people can also experience headaches, arm pain, tingling and numbness. The prevalence of whiplash associated disorders (WAD) is high and increasing worldwide, particularly within developed countries<sup>1</sup>. The incidence of WAD in the United Kingdom (UK) is estimated to be around 400,000 per year<sup>2</sup>, with an increase of 25% in whiplash claims between 2002 and 2008<sup>2</sup>. Increases of a similar magnitude have also been reported in Germany<sup>3</sup> and the Netherlands<sup>4</sup>, whilst Canada<sup>5</sup>, Australia<sup>6,7</sup> and the USA<sup>8</sup> all have annual incidences of over 100 per 100,000 people.

Previous studies have documented the total societal burden of WAD, with an annual cost to the UK economy of over £3.1 billion in 2002, primarily made up of health service costs and productivity losses<sup>9</sup>. High societal costs have also been reported in the USA (\$9 billion<sup>10</sup>) and Europe (€10 billion<sup>1</sup>). There is also evidence on long-term health consequences, with 30-50% of people suffering reduced health-related quality of life (HRQoL) caused by chronic symptoms associated with WAD<sup>11</sup>. However, these studies present group level results, not considering individual patient characteristics and how they interact with costs and HRQoL<sup>12</sup>.

Evaluating the cost-effectiveness of possible interventions for WAD requires either prospectively collected patient-level data from a clinical trial/observational study, or the use of modelling techniques to simulate the costs and benefits of different interventions. These models often make use of estimates of patient-level costs and health utilities (preference-based outcomes rooted in economic theory<sup>13</sup>) from secondary data sources, stratified by clinical and socio-demographic characteristics, which can be applied where prospective data are not collected. Currently there are no such reference values available for WAD, a problem

which this study addresses using prospectively collected data from the Managing Injuries of the Neck Trial (MINT)<sup>14</sup>. A regression modelling approach was used to estimate economic costs and HRQoL outcomes associated with severity of WAD and different recovery trajectories.

ACCEPTED

## Materials and Methods

MINT ran during 2005-2009 in 15 NHS emergency departments (EDs) in the UK<sup>14</sup>. A cluster randomised controlled trial; it recruited 3,851 individuals with WAD of grades I-III<sup>15</sup>.

Centres were randomised to provide either active management (including *The Whiplash Book*<sup>16</sup>) or usual care. Then, patients whose symptoms persisted beyond 3 weeks and sought additional treatment were individually randomised to either a single or six physiotherapy sessions.

Participants were sent questionnaires within three days of ED attendance, including the Short-Form Health Survey version 1 (SF-12)<sup>17</sup> and the EQ-5D-3L<sup>18</sup>, which were used as baseline measurements. Postal questionnaire were sent at 4, 8 and 12 months post initial attendance, containing the SF-12 and EQ-5D, as well as the Neck Disability Index (NDI)<sup>19</sup>. Participants not responding within one week were sent a reminder letter, with those still not responding called over the telephone to obtain the core outcome measure set (NDI and EQ-5D).

Since this study's focus was the economic costs and HRQoL outcomes associated with different disability levels and recovery trajectories, rather than evaluations of the effectiveness or cost-effectiveness of the interventions in the trial, all MINT participants were included in these analyses, regardless of trial allocation.

### *Data collection instruments*

The NDI is a measure of neck pain related disability consisting of 10 questions measuring symptom severity and activity restriction, each scored from 0 (none) to 5 (severe). These are summed and doubled to give a total ranging from 0 to 100. Raw scores can be converted to a disability severity classification devised by Vernon et al, with scores of 8 or less corresponding to no disability, 10-28 mild disability, 30-48 moderate disability, 50-68 severe

disability and greater than 68 complete disability<sup>19</sup>. These categorisations are a commonly used approach in analyses using of the NDI<sup>20</sup>. Multiple longitudinal NDI measurements can be converted into a categorisation of patient recovery trajectories defined by Sterling et al, with injuries classed as mild, moderate or chronic-severe<sup>21</sup>.

Participants completed resource use questionnaires at 4, 8 and 12 months, asking for details of all hospital inpatient, outpatient, and community healthcare and social services used, diagnostic tests undertaken and medications prescribed<sup>14</sup>. National unit costs (2009 prices, UK£) were applied to these data<sup>22</sup>, which were added to patient recorded expenditures on privately purchased healthcare to give total per patient societal costs<sup>14</sup>.

MINT contained two measures of HRQoL, the SF-12 and EQ-5D. From the SF-12, a six dimension health-state classification called the SF-6D can be constructed, containing questions on physical functioning, role limitations, social functioning, pain, mental health and vitality<sup>23</sup>. The SF-6D describes 7,500 health states with utility values, calculated using the standard gamble technique from 611 members of the UK general population, ranging from 0.345 to 1.0<sup>24</sup>.

The first five EQ-5D items form a health state classification system asking respondents to describe their mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Responses to each of these are divided into three ordinal levels: (1) no problems; (2) some/moderate problems; (3) severe/extreme problems, generating a total of 243 possible health states. Utility scores can be produced using a UK specific tariff<sup>25</sup>, calculated from a time trade-off study of 3,336 people from the UK general population, taking values between 0.594 and 1.0. Pre-injury utility values were not available for the MINT dataset and were therefore estimated from published general population health estimates, adjusted for age<sup>26</sup>.

## *Statistical Analysis*

Descriptive statistics were estimated for economic costs and health state utilities. Regression modelling was then used to estimate relationships between WAD severity and costs/health state utilities, adjusting for age, sex and pre-injury utility. In healthcare cost data, a significant proportion of individuals will often incur no costs and, for those that do, the distribution of costs is often highly skewed, with a small number of individuals incurring costs considerably higher than the mean<sup>27</sup>. A two-part model was thus used, the first a logistic regression model for the probability of incurring any costs. The second part was a generalised linear model (GLM) for costs incurred conditional on the cost being greater than zero. A log link function in the GLM means parameters have a multiplicative effect on cost, helping to model the skewed distribution of the data<sup>28</sup>. Models were produced for costs over the first four months of the follow-up period (with disability severity at 4 months as a predictor), and for costs between months 8 and 12 (using disability severity at 12 months). Since it was not possible to disentangle intervention costs from standard care, treatment allocation at initial randomisation was added as a covariate.

Ordinary least squares (OLS) models for EQ-5D and SF-6D utility values (with injury severity, age, sex and pre-injury utility as covariates) were also produced. Utility scores at 4 months (acute injury) and 12 months (chronic injury) were regressed on disability severity. However, such models allow for the prediction of values outside the range of the EQ-5D/SF-6D. A high proportion of respondents also gave responses implying perfect health (a utility score of 1.0), something difficult to estimate using OLS models<sup>29</sup>. One solution is censored regression, which remove predictions outside the range and can lead to a cluster of predictions at the limit of the range<sup>30</sup>. However, these models assume that utility data are censored at 1.0 (i.e. higher values are possible, we simply cannot measure them) whereas in reality 1.0 is a genuine upper bound. Thus, we produced similar two-part models for utilities

as for cost data, with a logistic model for the probability of having a utility less than 1.0, followed by a GLM for total disutility, conditional on the utility score being less than 1.0. We did not include treatment allocation as a covariate, since any improvements in health from trial interventions should not affect the relationship between the different measures.

To make use of the longitudinal data available, total costs (the sum of the three follow-up periods) and quality-adjusted life-years (QALYs) (estimated using the area under the EQ-5D utility curve, assuming linear interpolation between measurements) accumulated over the follow-up period were regressed on the Sterling trajectories defined above, including length of follow-up as a covariate to adjust for differential follow-up lengths between participants. The model for total costs was a two-part model, including treatment allocation as a covariate, whilst that for QALYs was an OLS model, as there are no longer fixed upper and lower bounds on this value.

Statistical analyses were performed in R version 3.0.1<sup>31</sup>.

### *Sensitivity analysis*

Baseline analyses of costs used a societal perspective, including all cost items collected in the study. Sensitivity analyses were conducted, running the same models but restricting to NHS and personal social services (PSS) costs, the perspective recommended for economic evaluations in England by the National Institute for Health and Care Excellence<sup>22</sup>.

## Results

Descriptive statistics are presented in table 1. At baseline, using the Vernon classifications, 29.4%, 42.0%, 20.8%, 6.5% and 1.3% of people with complete NDI data had no, mild, moderate, severe and complete neck-related disability, respectively. Baseline questionnaire response rates varied from 78.6% (SF-6D utility) to 89.1% (EQ-5D utility), whilst response rates at the end of the follow-up period ranged from 50.4%-69.2%. There were significant differences (at the 95% level) between responders and non-responders in age, sex, initial pain intensity and treatment group, with responders tending to be older (37.91 vs 36.65), more likely to be female (61% vs 53.4%), have a lower pain intensity at baseline (1.92 vs 1.78) and be randomised to the control group (44.9% vs 40.3%)<sup>32</sup>. However, there were no significant differences (again at the 5% level) in WAD grade at presentation, baseline EQ-5D utility, baseline SF-6D utility or costs between people who later went on to become responders or non-responders.

Table 2 presents descriptive statistics for costs and health state utilities, stratified by NDI classification. NDI scores improve over time, with an increase from 29.4% to 48.5% of participants with no disability, with simultaneous decreases in the proportion of participants in other categories. There is also a consistent improvement in utility scores over time. This is driven by two distinct factors, the movement of people to less severe disability classifications, and an improvement in the average utility score reported by people with the same disability status. Both cost perspectives generated similar results, with more severe disability associated with higher costs (societal costs between baseline and 4 months averaged £99.55 for those with no disability compared to £668.53 for those with complete disability) and costs decreasing as time since injury increases (societal costs averaged £234.15 over the first 4 months, decreasing to £127.51 between 8 and 12 months). This

decline was more pronounced for NHS and PSS costs, which comprised 73.9% of total costs between baseline and 4 months, but only 51.8% between 8 and 12 months.

Table 3 reports the two-part models for costs, where AIC-based model selection indicated the gamma family as the correct specification for the mean-variance relationship in the GLM.

Increasing disability levels were monotonically associated with an increased probability of incurring costs, and an increase in the total cost given that one was incurred. Average societal costs, given one was incurred, were similar at 4 and 12 months (£166.67 vs. £174.40 for a 40 year old woman in the intervention group with a baseline utility of 0.925 and no disability), but there was a considerably higher probability of incurring a cost during the first 4 months (0.564 vs. 0.195 for the same person as above).

Table 4 presents the OLS regression models for utility scores. Disability levels were again highly significant predictors, with higher disability associated with lower utility scores (e.g. complete disability was associated with an EQ-5D decrement of 0.899 at 4 months and 0.888 at 12 months compared to no disability). SF-6D scores were lower than EQ-5D scores for people with low levels of disability (0.856 vs. 0.948 at 4 months for the characteristics listed above), but considerably higher in people with higher levels of disability (0.543 vs. 0.039 for the same characteristics, but with complete disability).

Table 5 summarises two-part models for the EQ-5D and SF-6D utility scores, respectively. Again, a gamma variance function was selected for the GLM, except for the 4 month SF-6D model, where a Gaussian distribution was selected using the AICs. Again, greater disability is associated with an increase in the probability of the utility score being less than 1.0, and the magnitude of that deviation, with the impact greater on probability than magnitude. There is also evidence of a difference in utility scores between men and women, though this is only present with the SF-6D.

Finally, table 6 reports models of total costs (with a gamma variance function in the GLM) and QALYs accrued over the follow-up period, using the Sterling trajectories as explanatory variables. For a woman with the characteristics listed above followed up for 1 year, both probability of cost and level of cost increase with more severe recovery trajectories (0.795 and £353.76 for mild trajectory versus 0.986 and £633.10 for moderate and 0.989 and £1378.33 for chronic-severe, using societal costs), whilst number of QALYs decrease (0.790 for mild versus 0.635 for moderate and 0.483 for chronic-severe).

ACCEPTED

## Discussion

This article reports economic costs, health state utilities and QALY profiles associated with differing disability levels during the recovery period following AWI. It uses a large dataset from a randomised controlled trial (MINT), which followed people through the recovery process<sup>14</sup>. The longitudinal dataset made it possible to consider costs and QALYs over the entire follow-up period, as well as at individual time points. Appropriate adjustments were made to address specific complexities that arise when using regression models on cost and HRQoL data.

The results show the considerable impact neck-related disability can have, both on costs (a 20 fold increase from the lowest to highest levels of disability) and health state utilities (an average drop of 0.9 in EQ-5D utility scores between those categories). Differences in cost captured the direct cost of managing the condition, indirect costs to the health system via the exacerbation of co-morbidities, and additional costs to the patients themselves. The sensitivity analyses, restricted to direct costs incurred by health and social care services, produced the same pattern of results.

There are differences between this and similar studies in other clinical areas<sup>33,34</sup>. First, most such studies use clinically defined events or clinical severity measures as explanatory variables, whereas we use another patient-reported outcome measure (PROM), the NDI. This was driven by the prevalence of the NDI in trials of WAD<sup>23</sup>, but presents both advantages and disadvantages. There is growing acceptance of the importance of PROMs in trials, and of their usefulness as a robust way of reflecting patient perspectives<sup>35,36</sup>. However, the reliance on patient completion leads to higher levels of missing data, which can both bias analyses and lead to difficulties in generalisation. One might expect participants in worse health to be less likely to return questionnaires, resulting in utilities appearing higher than is accurate, and

costs lower. However, there were no significant differences in our outcomes variables (EQ-5D utility, SF-6D utility, costs) between patients who went on to become responders and non-responders.

Secondly, whilst two-part models are now accepted as a technique for modelling of cost data, their use for utilities has been less common. Nevertheless, it has been recognised both that a significant proportion of participants will be in a state of “full health” at any time point, and that changes in mean utility scores for a population can be significantly driven by the move towards/away from the “full health” state<sup>37</sup>. This effect is particularly noticeable with the EQ-5D, due to the large gap between the (11111) state (responding 1 to every question, the best possible health state) and that with the next highest utility score<sup>38</sup>. Thus two-part models, which explicitly accommodate the switch from “full health” to other states, are equally useful in dealing with health utilities as costs. Other, more complex regression methods such as beta regression have also been proposed, but there is little evidence that they lead to improvements in model fit<sup>39</sup>.

Thirdly, all the usual caveats concerning data from clinical trials also apply in this context<sup>40</sup>. The study inclusion criteria mean the population may not be fully representative of the background population of WAD patients. However, since MINT was an extremely pragmatic trial, designed to reflect routine care as closely as possible, this should not be as significant a problem as with less representative trials. Finally, difference in healthcare practices and preference structures mean that the patterns of resource use (and thus cost), compensation payments and health utilities can vary considerably between different countries<sup>41</sup>. Thus, careful thought should be given to the necessary assumptions when applying these results to contexts outside of the UK healthcare system.

When considering correlations between initial disability and long term costs, it is important to note that the biological severity of the initial injury may not be the only cause. Other literature has suggested that patients may ascribe pre-existing or unrelated symptoms as being the effects of whiplash, thereby overestimating the impact of the event<sup>42</sup>. Further, there have long been concerns about the difficulties in defining injury mechanisms and disease categories in WAD, and the nature of the link between biological injury, high initial symptoms and long-term prognosis (e.g. patients with compensation claims and lawsuits have significantly worse long-term outcomes than those without, even when adjusting for initial severity<sup>42</sup>). However, since we have no reason to believe these effects would be greater in one arm of the trial than the other, they are unlikely to bias results derived from within trial comparisons.

Nevertheless, despite these limitations, this research should prove useful to both economists and health service researchers, by providing estimates of economic costs and preference-based HRQoL outcomes associated with different disability levels and recovery trajectories following AWI. These estimates can be used as inputs when modelling the cost-effectiveness of interventions for the management of WAD, and the budgetary impact of those interventions.

**Acknowledgements:**

This project benefitted from facilities funded through Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials Platform, with support from Advantage West Midlands (AWM) and the Wolfson Foundation.

MINT was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme (project number 02/35/02) and has been published in full in Health Technology Assessment. The support of the Oxford CLARHC and Biomedical Research Unit is acknowledged (SEL, MW, EJW).

ACCEPTED

## References

- [1] Galasko CSB, Murray P, Stephenson W. Incidence of whiplash associated disorder. *BCMJ* 2002, 44:237-40.
- [2] Association of British Insurers: Tackling whiplash: prevention, care and compensation. 2008.
- [3] Richter M, Otte D, Pohlemann T, Krettek C, Blauth M. Whiplash-type neck distortion in restrained car drivers: frequency, causes and long-term results. *Eur Spine J* 2000, 9:109-17.
- [4] Versteegen G, Kingma J, Meijler W, ten Duis H. Neck sprain after motor vehicle accidents in drivers and passengers. *Eur Spine J* 2000, 9:547-52.
- [5] Spitzer W, Skovron M, Salmi L, Cassidy J, Duranceau J, Suissa S, Zeiss E. Scientific monograph of the Quebec Task Force on whiplash-associated disorders: redefining 'whiplash' and its management. *Spine* 1995, 20:1S-73S.
- [6] Barnsley L, Lord S, Bogduk N. Whiplash injury. *Pain* 1994, 58:283-307.
- [7] Miles K, Maimaris C, Finlay D, Barnes M. The incidence and prognostic significance of radiological abnormalities in soft tissue injuries to the cervical spine. *Skeletal Radiol* 1988, 17:493-6.
- [8] Quinlan K, Annest J, Myers B, Ryan G, Hill H. Neck strains and sprains among motor vehicle occupants-United States, 2000. *Accid Anal Prev* 2004, 36:21-7.
- [9] Galasko G, Murray P, Stephenson W: Incidence of whiplash-associated disorder. *BCMJ* 2002, 44:237-40.
- [10] Kuppia SM. Injury criteria and anthropomorphic test devices for whiplash injury assessment. National Highway Traffic Safety Administration, 2004.

[11] Carroll L, Holm L, Hogg-Johnson S, Côté P, Cassidy JD, Haldeman S, Nordin M, Hurwitz EL, Carragee EJ, van der Velde G, Peloso PM, Guzman J: Course and prognostic factors for neck pain in whiplash-associated disorders (WAD): results of the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *J Manipulative Physiol Ther* 2008, 32:S97-107.

[12] Hurwitz EL, Carragee EJ, van der Velde G, Carroll LJ, Nordin M, Guzman J, Peloso PM, Holm LW, Côté P, Hogg-Johnson S, Cassidy JD, Haldeman S: Treatment of neck pain: noninvasive interventions: results of the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *J Manipulative Physiol Ther* 2009, 32:S141-75.

[13] Torrance GW: Measurement of health state utilities for economic appraisal. *J Health Econ* 1986, 5:1-30.

[14] Lamb SE, Williams MA, Williamson EM, Gates S, Withers EJ, Mt-Isa S, Ashby D, Castelnovo E, Underwood M, Cooke MW, MINT Trial Group: Managing injuries of the neck trial (MINT): a randomised controlled trial of treatments for whiplash injuries. *Health Technol Assess* 2012, 16:1-141.

[15] Spitzer WO, Skovron ML, Salmi LR, Cassidy JD, Duranceau J, Suissa S, Zeiss E: Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining “whiplash” and its management. *Spine* 1995, 20:1S-73S.

[16] Waddell G, Burton K, McClune T: The whiplash book. TSO, London. 2005.

[17] Ware JE, Kosinski M, Keller SD: A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996, 34:220-33.

[18] The EuroQol Group: A facility for the measurement of health-related quality of life. *Health Policy* 1990, 16:199-208.

- [19] Vernon H, Mior S: The neck disability index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991, 14:409-15.
- [20] Vernon H: The neck disability index: state-of-the-art, 1991-2008. *J Manipulative Physiol Ther* 2008, 31:491-502.
- [21] Sterling M, Hendrikz J, Kenardy J: Compensation claim lodgement and health outcome developmental trajectories following whiplash injury: a prospective study. *Pain* 2010, 150:222-8.
- [22] National Institute for Health and Care Excellence: Guide to the methods of technology appraisal. NICE, London. 2008.
- [23] Ware JE, Kosinski M, Keller SD: A 12-item short-form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996, 34:220-233.
- [24] Brazier JE, Roberts J: The estimation of a preference-based measure of health from the SF-12. *Medical Care* 2004, 42:851-9.
- [25] Dolan P, Gudex C, Kind P, Williams A: A social tariff for EuroQol: results from a UK general population survey. Centre for Health Economics, University of York. 1995.
- [26] National Centre for Social Research and University of College London: Department of Epidemiology and Public Health HSfE. Health survey for England 2006. Colchester, Essex: UK Data Archive; 1991–2007.
- [27] Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health economics. *J Health Economics* 1998, 17:247-81.
- [28] Manning WG, Mullahy J: Estimating log models: to transform or not to transform? *J Health Economics* 2001, 20:461-94.

[29] Ruud P: An introduction to classical econometric theory. Oxford University Press, UK.

2000.

[30] Long JS: Regression models for categorical and limited dependent variables. Sage,

London. 1997.

[31] R Development Core Team: R: a language and environment for statistical computing.

Version 3.0.1. R Foundation for Statistical Computing, Vienna. 2013.

[32] Pink J, Petrou S, Williamson E, Williams M, Lamb SE. Properties of patient-reported outcome measures in individuals following acute whiplash injury. *Health Qual Life Outcomes* 2014, 12:38.

[33] Clarke P, Gray A, Legood R, Briggs A, Holman R: The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65). *Diabetic Medicine* 2003, 20:442-50.

[34] Clarke P, Gray A, Holman R: Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Med Decis Making* 2002, 22:340-9.

[35] McKenna SP: Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Medicine* 2011, 9:86.

[36] Fitzpatrick R, Davey C, Buxton MJ, Jones DR: Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998, 2:1-74.

[37] Clarke P, Gray A, Holman R: Estimating utility values for health states of type 2 diabetes patients using the EQ-5D (UKPDS 62). *Med Decis Making* 2002, 22:340-9.

[38] Brazier J, Roberts J, Tsuchiya A, Busschbach J: A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004, 13:873-84.

[39] Basu A, Manca A: Regression estimators for generic health-related quality of life and quality-adjusted life years. *Med Decis Making* 2012, 32:56-69.

[40] Feinstein AR, Horwitz RI: Problems in the “evidence” of “evidence-based medicine”. *Am J Med* 1997, 103:529-35.

[41] Chappuis G, Soltermann B, CEA, AREDOC, CEREDOC: Number and cost of claims linked to minor cervical trauma in Europe: results from the comparative study by CEA, AREDON and CEREDOC. *Eur Spine J* 2008, 17:1350-7.

[42] Malleon A. Whiplash and other useful illnesses. McGill University Press, 2002.

ACCEPTED

Table 1 - Baseline characteristics of the MINT study population

<b>Variable</b>	<b>N (%)</b>
Sex (male)	1661/3851 (43.1%)
Sex (female)	2133/3851 (55.4%)
Treatment group step 1 (usual care)	1598/3851 (41.5%)
Treatment group step 1 (active management)	2253/3851 (58.5%)
Treatment group step 2* (advice)	287/574 (50%)
Treatment group step 2* (physiotherapy)	287/574 (50%)
WAD grade at presentation (grade 1)	2088/3851 (54.2%)
WAD grade at presentation (grade 2)	1659/3851 (43.1%)
WAD grade at presentation (grade 3)	104/3851 (2.7%)
	<b>Mean (SD)</b>
Initial pain intensity	5.13 (1.89)
EQ-5D utility score - baseline	0.587 (0.298)
SF-6D utility score – baseline	0.647 (0.136)
NDI score – 4 months	21.03 (17.45)

Percentages may not add up to 100% due to missing data. \*The denominators are lower here as not all participants were randomised at step 2. SD: standard deviation. WAD: Whiplash associated disorder.

Table 2 – Descriptive statistics (Vernon categories, utility scores and costs over time, stratified by disability status)

<b>Percentage of participants in each Vernon category by time point</b>				
<b>Vernon category</b>	<b>Baseline</b>	<b>4 months</b>	<b>8 months</b>	<b>12 months</b>
No disability	N/A	29.4	42.5	48.5
Mild disability	N/A	42.0	38.1	34.4
Moderate disability	N/A	20.8	14.1	12.9
Severe disability	N/A	6.5	4.2	3.3
Complete disability	N/A	1.3	1.0	0.9
<b>Mean (SD) EQ-5D utility scores at each time point, stratified by Vernon category at 4 months</b>				
<b>Vernon category</b>	<b>Baseline</b>	<b>4 months</b>	<b>8 months</b>	<b>12 months</b>
No disability	0.766 (0.210)	0.934 (0.128)	0.936 (0.131)	0.944 (0.127)
Mild disability	0.613 (0.232)	0.760 (0.141)	0.818 (0.171)	0.844 (0.177)
Moderate disability	0.463 (0.303)	0.612 (0.189)	0.671 (0.227)	0.699 (0.235)
Severe disability	0.226 (0.335)	0.327 (0.305)	0.465 (0.315)	0.534 (0.317)
Complete disability	0.080 (0.308)	0.033 (0.291)	0.109 (0.306)	0.213 (0.299)
All	0.587 (0.298)	0.739 (0.246)	0.792 (0.241)	0.817 (0.233)
<b>Mean (SD) SF-6D utility scores at each time point, stratified by Vernon category at 4 months</b>				
<b>Vernon category</b>	<b>Baseline</b>	<b>4 months</b>	<b>8 months</b>	<b>12 months</b>
No disability	0.735 (0.142)	0.858 (0.100)	0.868 (0.100)	0.873 (0.096)
Mild disability	0.634 (0.122)	0.766 (0.115)	0.800 (0.124)	0.809 (0.122)
Moderate disability	0.569 (0.086)	0.637 (0.100)	0.702 (0.126)	0.714 (0.131)
Severe disability	0.547 (0.076)	0.546 (0.076)	0.599 (0.118)	0.646 (0.120)
Complete disability	0.531 (0.069)	0.508 (0.054)	0.525 (0.080)	0.548 (0.079)
All	0.647 (0.136)	0.752 (0.142)	0.878 (0.140)	0.800 (0.135)
<b>Mean (SD) societal costs at each time point, stratified by Vernon category at 4 months*</b>				
<b>Vernon category</b>	<b>Baseline</b>	<b>4 months</b>	<b>8 months</b>	<b>12 months</b>
No disability	N/A	£99.55 (198.55)	£31.15 (121.27)	£23.02 (118.78)
Mild disability	N/A	£223.01 (305.64)	£80.95 (175.05)	£66.36 (360.83)
Moderate disability	N/A	£325.39 (340.50)	£236.64 (602.12)	£155.64 (276.18)
Severe disability	N/A	£535.89 (785.54)	£472.53 (1011.83)	£810.31 (5435.70)
Complete disability	N/A	£668.53 (352.73)	£412.94 (329.17)	£601.58 (853.23)
All	N/A	£234.15 (365.47)	£131.93 (420.74)	£127.51 (1292.73)
<b>Mean (SD) health and social care costs at each time point, stratified by Vernon category at 4 months*</b>				
<b>Vernon category</b>	<b>Baseline</b>	<b>4 months</b>	<b>8 months</b>	<b>12 months</b>
No disability	N/A	£72.49 (141.42)	£17.42 (93.05)	£12.33 (82.03)
Mild disability	N/A	£159.64 (228.34)	£47.27 (128.81)	£31.93 (113.73)
Moderate disability	N/A	£254.22 (269.74)	£140.64 (246.24)	£100.29 (197.41)
Severe disability	N/A	£376.31 (406.29)	£331.02 (613.89)	£294.06 (774.77)
Complete disability	N/A	£534.34 (395.78)	£340.49 (284.25)	£363.37 (506.64)
All	N/A	£172.96 (254.23)	£81.74 (230.90)	£66.08 (262.15)

\*Costs at a time point refer to those accumulated over the 4 months prior.

Table 3 – Two-part models for societal and NHS cost data, modelling probability of cost for an individual being greater than 0, and total cost given that cost is greater than 0

<b>All costs</b>				
	<b>Model 1 (4 months)</b>		<b>Model 2 (12 months)</b>	
	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>
Intercept†	0.243* (0.112)	5.167*** (0.070)	-1.614*** (0.111)	4.925*** (0.196)
Sex (female)	-0.099 (0.110)	-0.047 (0.053)	0.119 (0.100)	0.050 (0.141)
Age	-0.014 (0.011)	-0.004 (0.005)	0.001 (0.010)	0.020 (0.013)
Baseline utility	0.045 (3.178)	-0.280 (1.502)	3.089 (2.843)	6.654 (3.847)
Treatment group	0.170 (0.108)	0.019 (0.052)	-0.001 (0.099)	-0.060 (0.138)
Mild disability	1.753*** (0.117)	0.382*** (0.068)	2.219*** (0.104)	0.080 (0.187)
Moderate disability	2.742*** (0.207)	0.701*** (0.078)	3.860*** (0.206)	0.642** (0.311)
Severe disability	3.586*** (0.512)	1.185*** (0.109)	4.391*** (0.469)	1.510*** (0.311)
Complete disability	16.40 (387.97)	1.388*** (0.207)	4.750*** (1.025)	3.280*** (0.518)
Sample size	2760	2243	2601	1218
<b>National health service and personal social services costs only</b>				
	<b>Model 1 (4 months)</b>		<b>Model 2 (12 months)</b>	
	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>
Intercept†	-0.317** (0.101)	5.148*** (0.066)	-2.669*** (0.142)	5.395*** (0.173)
Sex (female)	-0.068 (0.090)	-0.058 (0.048)	-0.023 (0.109)	0.025 (0.111)
Age	-0.015 (0.009)	-0.002 (0.005)	0.009 (0.010)	-0.006 (0.011)
Baseline utility	-1.021 (2.578)	0.583 (1.359)	3.915 (2.987)	-0.139 (2.982)
Treatment group	0.085 (0.089)	0.019 (0.048)	0.157 (0.107)	0.054 (0.109)
Mild disability	1.183*** (0.097)	0.296*** (0.065)	1.807*** (0.132)	-0.396* (0.169)
Moderate disability	2.142*** (0.141)	0.588*** (0.072)	3.220*** (0.162)	0.065 (0.177)
Severe disability	3.013*** (0.309)	0.906*** (0.096)	4.245*** (0.315)	0.771*** (0.226)
Complete disability	15.97 (235.39)	1.195*** (0.173)	5.763*** (1.028)	1.303*** (0.325)
Sample size	2760	1853	2601	681

\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001. †The reference case is a man, 36 years of age, receiving standard care, with a pre-injury utility of 0.9 and no neck-related disability.

Table 4 – Ordinary least squares models for EQ-5D and SF-6D health state utilities at 4 and 12 months post baseline

	<b>EQ-5D (4 months)</b>	<b>SF-6D (4 months)</b>	<b>EQ-5D (12 months)</b>	<b>SF-6D (12 months)</b>
Intercept†	0.933*** (0.007)	0.863*** (0.005)	0.951*** (0.006)	0.887*** (0.004)
Sex (female)	0.001 (0.006)	-0.008 (0.004)	-0.004 (0.006)	-0.023*** (0.005)
Age	0.0003 (0.001)	0.0002 (0.0004)	-0.002*** (0.001)	-0.0005 (0.0005)
Baseline utility	0.178 (0.180)	0.035 (0.128)	-0.353 (0.183)	-0.037 (0.133)
Mild disability	-0.175*** (0.007)	-0.093*** (0.005)	-0.165*** (0.007)	-0.094*** (0.005)
Moderate disability	-0.322*** (0.009)	-0.222*** (0.006)	-0.332*** (0.010)	-0.223*** (0.007)
Severe disability	-0.614*** (0.014)	-0.313*** (0.010)	-0.598*** (0.018)	-0.321*** (0.014)
Complete disability	-0.899*** (0.028)	-0.350*** (0.022)	-0.888*** (0.033)	-0.318*** (0.028)
Sample size	2899	2366	2584	1925

\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001. †The reference case is a man, 36 years of age, with a pre-injury utility of 0.9 and no neck-related disability.

ACCEPTED

Table 5 – Two-part models for EQ-5D and SF-6D health state utilities, modelling probability that the utility of an individual is less than 1, and total disutility given the utility is less than 1

EQ-5D models				
	Model 1 (4 months)		Model 2 (12 months)	
	Probability of disutility	Disutility given utility < 1	Probability of disutility	Disutility given utility < 1
Intercept†	-0.924*** (0.099)	-1.494*** (0.032)	-1.392*** (0.093)	-1.460*** (0.037)
Sex (female)	0.040 (0.116)	0.001 (0.021)	0.202 (0.108)	-0.013 (0.030)
Age	0.016 (0.012)	-0.002 (0.002)	0.034** (0.011)	0.004 (0.003)
Baseline utility	2.619 (3.483)	-0.966 (0.575)	2.204 (3.074)	0.982 (0.775)
Mild disability	2.890*** (0.116)	0.197*** (0.033)	2.625*** (0.110)	0.167*** (0.038)
Moderate disability	5.642*** (0.456)	0.556*** (0.035)	4.629*** (0.329)	0.550*** (0.044)
Severe disability	18.425 (291.349)	1.108*** (0.045)	18.668 (425.340)	1.038*** (0.067)
Complete disability	18.405 (641.121)	1.455*** (0.081)	18.703 (795.660)	1.403*** (0.114)
Sample size	2899	2144	2584	1423
SF-6D models				
	Model 1 (4 months)		Model 2 (12 months)	
	Probability of disutility	Disutility given utility < 1	Probability of disutility	Disutility given utility < 1
Intercept†	1.811*** (0.159)	-1.846*** (0.027)	1.422*** (0.127)	-1.965*** (0.023)
Sex (female)	0.647** (0.229)	0.015 (0.016)	0.847*** (0.186)	0.089*** (0.025)
Age	0.040 (0.025)	-0.0004 (0.002)	0.027 (0.019)	-00.002 (0.002)
Baseline utility	5.266 (7.579)	0.028 (0.452)	-1.657 (6.084)	-0.496 (0.678)
Mild disability	2.466*** (0.340)	0.398*** (0.029)	2.354*** (0.350)	0.427*** (0.027)
Moderate disability	18.320 (810.372)	0.827*** (0.029)	17.502 (672.306)	0.873*** (0.037)
Severe disability	18.341 (1537.141)	1.051*** (0.032)	17.548 (1449.526)	1.115*** (0.072)
Complete disability	18.300 (3576.154)	1.132*** (0.050)	17.588 (1906.713)	1.113*** (0.141)
Sample size	2366	2277	1925	1779

\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001. †The reference case is a man, 36 years of age, with a pre-injury utility of 0.9 and no neck-related disability.

Table 6 – Two-part model for costs and quality-adjusted life-years (calculated from the EQ-5D) over full follow-up period

<b>Societal costs</b>		
	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>
Intercept†	1.087*** (0.130)	5.825*** (0.107)
Sex (female)	0.001 (0.125)	0.001 (0.096)
Age	-0.001 (0.012)	0.001 (0.009)
Baseline utility	-1.235 (3.648)	1.345 (2.688)
Treatment group	0.302* (0.123)	0.005 (0.095)
Trajectory (moderate)	2.933*** (0.508)	0.582*** (0.133)
Trajectory (chronic-severe)	3.157*** (0.508)	1.360*** (0.124)
Length of follow-up (days)	0.004 (0.002)	-0.002 (0.001)
Sample size	2265	1928
<b>National Health Service and personal social services costs</b>		
	<b>Probability of cost</b>	<b>Total cost given cost &gt;0</b>
Intercept†	0.255* (0.106)	5.629*** (0.075)
Sex (female)	0.112 (0.100)	-0.013 (0.067)
Age	-0.010 (0.010)	-0.003 (0.006)
Baseline utility	-2.075 (2.877)	0.301 (1.828)
Treatment group	0.234* (0.099)	0.040 (0.066)
Trajectory (moderate)	1.783*** (0.206)	0.568*** (0.089)
Trajectory (chronic-severe)	2.500*** (0.249)	1.090*** (0.082)
Length of follow-up (days)	0.002 (0.002)	-0.002 (0.001)
Sample size	2265	1614
<b>Quality-adjusted life-years</b>		
	<b>Parameters</b>	
Intercept†	0.802*** (0.006)	
Sex (female)	-0.013* (0.006)	
Age	-0.001* (0.001)	
Baseline utility	0.055 (0.170)	
Trajectory (moderate)	-0.155*** (0.009)	
Trajectory (chronic-severe)	-0.307*** (0.008)	
Length of follow-up	0.002*** (0.0001)	
Sample size	2256	

\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001. †The reference case is a man, 36 years of age, receiving standard care, with a pre-injury utility of 0.9, following a mild recovery trajectory and being followed up for one year.