



■ HIP

Modelling and estimation of health-related quality of life after hip fracture

A RE-ANALYSIS OF DATA FROM A PROSPECTIVE COHORT STUDY

**N. Parsons,
X. L. Griffin,
J. Achten,
T. J. Chesser,
S. E. Lamb,
M. L. Costa**

Warwick Medical
School, University of
Warwick, Coventry,
United Kingdom

■ N. Parsons, BSc, MSc, PhD, Statistician, Statistics and Epidemiology, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK.
■ X. L. Griffin, FRCS (Tr & Orth) PhD Medicine, Associate Professor of Orthopaedic Trauma Surgery,
■ J. Achten, MSc PhD, Research Manager,
■ S. E. Lamb, DPhil, Kadoorie Professor of Trauma Rehabilitation; Director, Centre for Statistics in Medicine,
■ M. L. Costa, FRCS (Tr&Orth) PhD, Professor of Orthopaedic Trauma Surgery, University of Oxford, Oxford Trauma, Kadoorie Centre, John Radcliffe Hospital, Oxford, Oxfordshire OX3 9DU, UK.
■ T. J. Chesser, MBBS FRCS (Tr and Orth), Consultant Trauma and Orthopaedic Surgeon, North Bristol NHS Trust, Southmead Hospital, Southmead Road, Westbury-on-Trym BS10 5NB, UK.

Correspondence should be sent to N. Parsons;
email: nick.parsons@warwick.ac.uk

doi: 10.1302/2046-3758.71.BJR-2017-0199

Bone Joint Res 2018;7:1–5.

Objectives

This study investigates the reporting of health-related quality of life (HRQoL) in patients following hip fracture. We compare the relative merits and make recommendations for the use for two methods of measuring HRQoL; (i) including patients who died during follow-up and (ii) including survivors only.

Methods

The World Hip Trauma Evaluation has previously reported changes in HRQoL using EuroQol-5D for patients with hip fractures. We performed additional analysis to investigate the effect of including or excluding those patients who died during the first four months of the follow-up period.

Results

The dataset included 503 patients, 25 of whom died between 30 days and four months of injury. There was a statistically significant difference in 30-day HRQoL between those alive (mean 0.331 and standard deviation (SD) 0.360) and those dead (mean 0.156 and SD 0.421) by four months (independent-samples *t*-test; *p* 0.022). The estimated difference of 0.175 in HRQoL (95% confidence interval 0.025 to 0.325) was also highly clinically significant.

Conclusion

When reporting HRQoL for patients after a hip fracture, excluding patients who die during follow-up leads to an overestimate of the effects of the intervention or treatment pathway. We would recommend that death-adjusted estimates should be used routinely when reporting HRQoL in this population.

Cite this article: *Bone Joint Res* 2018;7:1–5.

Keywords: Hip fracture, Health-related quality of life, EQ-5D, Longitudinal models

Article focus

■ Analysis of health-related quality of life (HRQoL) after hip fracture including patients who died during follow-up versus survivors only.

Key messages

- When reporting HRQoL using EuroQol-5D for patients after a hip fracture, excluding patients who die during follow-up leads to overly-optimistic estimates of patient outcomes and the effects of the treatment pathway.
- We would recommend that ‘death-adjusted’ estimates should be used routinely when reporting HRQoL in this population.

Strengths and limitations

- This is a large study that reports highly significant differences between HRQoL outcomes in this population.
- The main limitation is that all of the data were reported from a single trauma centre.

Introduction

Health-related quality of life (HRQoL) is now the most widely used primary outcome measure for studies reporting outcomes for patients after hip fracture.^{1–5} The EuroQol-5 Dimensions (EQ-5D) has become the preferred measure to determine HRQoL in the United Kingdom and in many other countries.^{1,2,6–9} EQ-5D at four months after the

fracture is part of the UK Core Outcome Set for hip fracture studies,¹⁰ which has been adopted by the National Institute for Health and Care in its most recent Hip Fracture Guidelines.¹¹

Hip fracture affects an older and often frail population. Griffin et al² report conservative estimates of mortality in the United Kingdom population of approximately 12% at four months and 20% at one year for patients aged > 80 years. In many areas of healthcare, patients who die before completing a study are often excluded from the primary analysis. This is, in general, not a particularly important issue if the number of patients who die is low. However, since the number of patients who die in the months following a hip fracture is relatively high, this approach inevitably leads to a loss of data and therefore a loss of precision when estimating outcomes. It also potentially leads to biased analyses, as by excluding those patients who die early we are likely to produce unduly optimistic estimates of HRQoL. One potential advantage of EQ-5D is that it provides, through the associated health utilities, a natural value for study participants who died prior to an outcome assessment; EQ-5D is anchored at 1 for full health and 0 for death. By including the important sub-group of patients who die in the months following hip fracture in the primary analysis of outcomes, we could therefore increase the precision of hip fracture studies. Hereafter, we refer to the inclusion of patients who died during follow-up as a 'death-adjusted' EQ-5D estimate, as opposed to a 'complete-case' estimate, which is based on only those patients alive at the index assessment occasion (four months). However, if the patients who die early had high HRQoL before their death, then assuming that their EQ-5D score was zero risks underestimating the potential benefits of an intervention, even if the patient subsequently dies before reaching their four-month assessment.

In this study, we investigate the use of a death-adjusted EQ-5D score in the analysis of outcomes following hip fracture. We undertake additional analysis of the data available from the World Hip Trauma Evaluation (WHiTE),^{1,2,4} which provided routine EQ-5D at four weeks, in addition to the four-month timepoint. These early outcome data allow us to model temporal changes in EQ-5D during the recovery phase, and to compare the relative merit of a death-adjusted *versus* a complete-case EQ-5D estimation and make recommendations about which to use.

Materials and Methods

Data. We conducted a prospective longitudinal cohort study to assess HRQoL at four weeks, four months and one year after hip fracture,⁴ referred to as the WHiTE study. In addition, a small convenience sample of the WHiTE study participants provided HRQoL assessments immediately post-injury. All patients, or proxy respondents where appropriate, provided informed consent or agreement,

Table 1. Age, gender split and baseline EuroQol 5 Dimensions (EQ-5D) for participants from the full World Hip Trauma Evaluation population (n = 741) and for those available for this study (n = 503)

	Full population (n = 741)	Study population (n = 503)
Mean age, yrs (SD)	83.1 (8.7)	82.8 (8.3)
Gender, female:male (% F)	503:186 (73.0)	362:117 (75.6)
Mean baseline EQ-5D (SD)	0.559 (0.348)	0.574 (0.337)

SD, standard deviation

respectively. The study is registered with Current Controlled Trials (ISRCTN63982700) and full protocols have previously been published.^{4,12} The data in this study comes from participants who presented with a hip fracture at a single major trauma centre in England between January 2012 and March 2014; those who were aged less than 60 years or who were managed nonoperatively were excluded from the study. A full description of the totality of data collected is available elsewhere.² Here we focus exclusively on the primary outcome measure, which was the EQ-5D score (EQ-5D-3L),^{6,7} a generic health utility instrument used to measure HRQoL. EQ-5D is a validated, cross-disciplinary standardized instrument that is widely used to assess HRQoL after hip fracture. It has two parts: a visual analogue scale (VAS), which measures self-rated health and a health status instrument consisting of a three-level response (no problems, some problems and extreme problems) for five health domains related to daily activities. These health domains are mobility, self-care, usual activities, pain and discomfort and anxiety and depression. More recent data collected as part of the WHiTE study uses the 5L version of EQ-5D that provides greater sensitivity than the 3L. Responses from the EQ-5D health classifications were converted into an overall score using a published utility algorithm for the population of the United Kingdom.¹³

Statistical analysis. Study data were summarized using means and standard deviations (SD), and visualized by box plots and strip plots to show variation in outcomes. Independent-samples *t*-tests were used to draw inferences on mean differences between selected sub-populations. The complexity of the setting we describe here is that early death of study participants postoperatively caused dropout (loss to follow-up), since no EQ-5D measurements were available after the terminal event for the four-month timepoint. If this dropout is non-random, then this is likely to cause bias in any analysis that ignores the dropout process. To obtain valid inferences, we use methods that allow fitting of joint models of longitudinal and time-to-event (survival) data.¹⁴ The longitudinal model for the temporal changes in EQ-5D postoperatively was a mixed-effects model that had a random additive participant effect, fixed effects for the baseline (pre-injury) EQ-5D and a quadratic (second-order polynomial) model for the log-transformed postoperative time. The survival data

Table II. Age, gender split and baseline EuroQol 5 Dimensions (EQ-5D) for participants who were alive ($n = 478$) and for those dead ($n = 25$) by the four-month timepoint

	Alive at 4 mths ($n = 478$)	Dead at 4 mths ($n = 25$)	p-value
Mean age, yrs (SD)	82.7 (8.2)	85.8 (8.5)	0.062*
Gender, female:male (% F)	346:110 (75.9)	16:7 (69.6)	0.464†
Mean baseline EQ-5D (SD)	0.581 (0.339)	0.434 (0.254)	0.034*

*independent-samples *t*-test

†Fisher's exact test

was summarized using a proportional-hazards (Cox) model,¹⁵ adjusting for the baseline EQ-5D score. Joint models were fitted in the R package, JM (R Foundation, Vienna, Austria),^{16,17} using a piecewise-constant baseline risk function and the Gauss-Hermite method for integral approximation.

Results

The totality of data available ($n = 741$) has been described previously and in full by Griffin et al;² there were 118 deaths reported during the course of study, with age, gender, American Society of Anesthesiologists (ASA) grade,¹⁸ and preoperative Abbreviated Mental Health Score (AMTS)¹⁹ all being statistically significant predictors of survival.² Postoperative trends in EQ-5D varied by AMTS score (≤ 8 and > 8) and age-group (≤ 80 years and > 80 years); in summary, recovery was generally worse for those in the older group and those with lower AMTS.² EQ-5D scores did not recover to baseline (pre-injury) levels but followed a characteristic trajectory up to 12 months, with little or no improvements after four months.² Therefore, we focus our modelling work on this four-month period.

Population characteristics. The full (baseline, four weeks, and four months) or partial (one or more of the values present) time courses of EQ-5D data were available from 503 of the WHITE study participants. The age distribution, gender split and baseline EQ-5D for this group was comparable with the full study population (Table I).

The study population consisted of 478 participants who survived to four months, and 25 who died after providing four-week EQ-5D data but before reaching the four-month follow-up timepoint. Table II shows the characteristics of these two groups.

A *t*-test indicated that the baseline EQ-5D was statistically significantly lower for those participants who were dead at four months than for those who were alive at this timepoint ($p = 0.034$); estimated difference 0.146 (95% CI 0.01 to 0.282).

Figure 1 shows strip plots and box plots of four-week EQ-5D data; medians and interquartile ranges (IQRs) for the two groups are 0.290 (IQR 0.055 to 0.640) and -0.040 (IQR -0.170 to 0.625). Evidence from previous analyses is that Gaussian approximations for EQ-5D are reasonable for this population;^{1,2} focusing on means and SDs, a *t*-test shows that there was a statistically significant difference

in four-week EQ-5D between those alive (mean 0.331 and SD 0.360) and those dead (mean 0.156 and SD 0.421) by four months (*t*-test; $p = 0.022$). The estimated difference 0.175 (95% CI 0.025 to 0.325) was also highly clinically significant (the minimum clinically important difference for EQ-5D is 0.074),²⁰ and indicated (together with the significant difference in baseline EQ-5D) that low EQ-5D was strongly associated with postoperative death.

Complete-case analysis. The most widely used and recommended endpoint for EQ-5D in this population is at four months.² When reporting results of this outcome for a randomized controlled trial (RCT), one approach to analysis is to simply report summary statistics (e.g. means and SDs) based on the population of patients who are alive at the four-month timepoint. If we are willing to accept that withdrawals and losses due to participant deaths are not related to the interventions, then comparing, for instance, group means at four months should provide an appropriate analysis, all else being equal.

Using EQ-5D data from only those participants alive at the four-month timepoint ($n = 478$) provides an estimate of the mean EQ-5D at four months of 0.454 (95% CI 0.414 to 0.495).

Model-based prediction. It is clear from Figure 1 and Table I that the characteristics of those participants who die early (before four months) are different from those who survive to provide EQ-5D assessments. Therefore, we proceed to fit joint models that enable us to explicitly allow for the effects of the underlying longitudinal EQ-5D outcome on the risk of death. In these models, we implicitly make the assumption that the complete EQ-5D longitudinal response (to the study endpoint at four months) is meaningful for all participants, including EQ-5D observations that would have been collected after death for those participants who died early.

One could argue that in this setting, as the terminating event is death, it makes no sense at all to consider the value of EQ-5D after death. However, for the purposes of exposition we proceed to fit models to the observed data and make predictions on future trends in EQ-5D scores for those study participants who died early. Figure 2 shows observed data and model fits for the full population, for participants alive at four months, and for participants who were dead at four months. The projected EQ-5D for the population of early deaths (under the assumption that they did not die but progressed to the

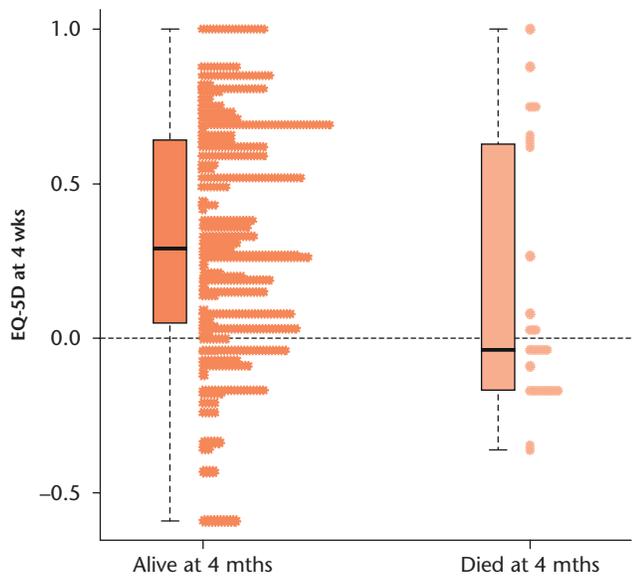


Fig. 1

Strip plots and box plots of four-week EuroQol 5 Dimensions (EQ-5D) data. For those World Hip Trauma Evaluation study participants alive ($n = 478$) and those who died ($n = 25$) at four months postoperatively. Stacked bars show numbers of participants for each EQ-5D score and box plots show interquartile range (IQR; box), median (solid line) and whiskers at 1.5 times the IQR.

four month timepoint) (Fig. 2c) indicates that, for this group, EQ-5D was likely to remain lower than that of those patients who we know survived to four months. The predicted EQ-5D score at four months for the dead group ($n = 25$) was 0.349 (95% CI 0.260 to 0.438) and for the alive group was 0.445 (95% CI 0.425 to 0.464); there was a statistically significant difference between groups based on an independent-samples *t*-test of predictions at four months (p -value = 0.034).

Using EQ-5D data from the full population ($n = 503$), then building a model to predict and project how those participants who died early may have progressed if they did not die, provides an estimate of the mean EQ-5D at four months of 0.440 (95% CI 0.421 to 0.459).

Death-adjusted analysis. Rather than attempt to model and project changes in EQ-5D scores from months one to four for those patients who did not survive to the study endpoint (the dashed line in Fig. 2c), a simpler approach is to assume that EQ-5D becomes zero at death, and then carry this observation forward to subsequent assessment occasions. We call this 'death-adjusted' EQ-5D.²

Undertaking this analysis for the WHITE study population provides an estimate of the mean EQ-5D at four months of 0.424 (95% CI 0.384 to 0.464).

Discussion

Three methods of summarising EQ-5D at four months postoperatively have been presented for patients after hip fracture. The first method (complete-case analysis) summarizes outcomes at four months using data only for those patients who were alive at this timepoint. Estimates

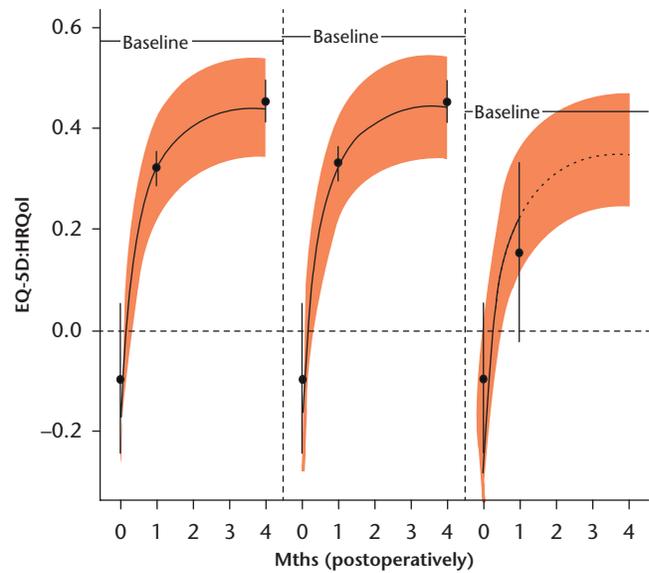


Fig. 2a

Fig. 2b

Fig. 2c

Longitudinal models for postoperative EuroQol 5 Dimensions (EQ-5D) from the full World Hip Trauma Evaluation study population ($n = 503$) (a), for participants alive at four months ($n = 478$) (b) and for participants who had died at four months ($n = 25$) (c). Observed means are plotted, with 95% confidence intervals (bars) and also fitted curves with 50% confidence regions.

of mean EQ-5D using this method (0.454; 95% CI 0.414 to 0.495) were larger than the other two methods discussed. This is not an unexpected result, as we show that those patients who died before four months had significantly lower EQ-5D at the early four-week assessment. By modelling the observed temporal changes in EQ-5D, we attempt to predict what the EQ-5D would have been for these participants who died early, if they had survived to four months. This model-based prediction provided an estimate of mean EQ-5D at four months (0.440; 95% CI 0.421 to 0.459) that was lower than the complete-case method. Even if we assume that, for those who died, the cause of death was not directly related to the intervention, we conclude that the complete-case method provides positively biased estimates of EQ-5D (i.e. it tends to overestimate the measure). Therefore, we would not generally recommend the complete-case analysis unless the focus of a study is purely on outcomes for those participants who survived to the study endpoint. If this is the case, then some inflation to the sample size should be made to allow for postoperative losses due to death.

The death-adjusted analysis method provided estimates of EQ-5D (0.424; 95% CI 0.384 to 0.464) that were lower still than the model-based prediction method. Again, this is not unexpected, as we have replaced the missing four-month data for individuals who died prior to the four month assessment ($n = 25$) with values of zero, which were always lower than the model-based predictions. A paired *t*-test indicated that death-adjusted EQ-5D estimates were statistically significantly smaller than corresponding model-based

predictions (p -value = 0.037), although the difference (paired mean difference in EQ-5D is 0.027) was such as to be clinically unimportant.²⁰

Due to the lack of clinically important difference between these methods of estimation, because the death-adjusted analysis method is considerably simpler than the model-based method, and because the assumptions required for the model-based method are unlikely to be met (i.e. that participants could have survived to provide observations at the index four-month timepoint), we recommend that death-adjusted estimates should be routinely used for reporting HRQoL in this population.

One could argue that, for comparative analyses (i.e. comparing groups A and B in a RCT), in principle it is not necessarily important whether estimates of EQ-5D are biased, as we are only interested in differences between groups. This is a weak argument, as if one knows that estimates of EQ-5D are likely to be positively biased, then it is difficult to justify not attempting a correction. Reporting death-adjusted EQ-5D also provides the additional benefit of increasing the sample size, as we no longer need to discard participants who have died prior to the study endpoint. However, experience suggests that the death-adjusted estimator has a larger variance than the complete-case estimator, due to the inclusion of the zero scores, which are not typically located at the mean of the distribution. Therefore, some inflation of the SD, relative to previously reported values based on complete-cases,⁵ should be considered when calculating study sample sizes. It should be noted that the model-based approach provided tighter confidence intervals than either of the other methods discussed, as it uses full information for estimation and it imposes constraints on the form of the longitudinal model for EQ-5D.

In conclusion, when reporting HRQoL for patients after a hip fracture, excluding patients who die during follow-up leads to an overly optimistic estimation of the effects of the intervention or treatment pathway. We would recommend that death-adjusted estimates should be routinely used for reporting HRQoL in this population.

References

1. **Parsons N, Griffin XL, Achten J, Costa ML.** Outcome assessment after hip fracture: is EQ-5D the answer? *Bone Joint Res* 2014;3:69-75.
2. **Griffin XL, Parsons N, Achten J, Fernandez M, Costa ML.** Recovery of health-related quality of life in a United Kingdom hip fracture population. The Warwick Hip Trauma Evaluation--a prospective cohort study. *Bone Joint J* 2015;97-B:372-382.
3. **Haywood KL, Brett J, Tutton E, Staniszewska S.** Patient-reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability. *Qual Life Res* 2017;26:799-812.
4. **Costa ML, Griffin XL, Achten J, et al.** World Hip Trauma Evaluation (WHiTE): framework for embedded comprehensive cohort studies. *BMJ Open* 2016;6:e011679.
5. **Sims AL, Parsons N, Achten J, et al.** The World Hip Trauma Evaluation Study 3: Hemiarthroplasty Evaluation by Multicentre Investigation - WHITE 3: HEMI - An Abridged Protocol. *Bone Joint Res* 2016;5:18-25.
6. **Brooks R.** EuroQol: the current state of play. *Health Policy* 1996;37:53-72.
7. **EuroQol Group.** EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
8. **Marques A, Lourenço Ó, da Silva JAP, Portuguese Working Group for the Study of the Burden of Hip Fractures in Portugal.** The burden of osteoporotic hip fractures in Portugal: costs, health related quality of life and mortality. *Osteoporos Int* 2015;26:2623-2630.
9. **Honkavaara N, Al-Ani AN, Campenfeldt P, Ekström W, Hedström M.** Good responsiveness with EuroQol 5-Dimension questionnaire and Short Form (36) Health Survey in 20-69 years old patients with a femoral neck fracture: A 2-year prospective follow-up study in 182 patients. *Injury* 2016;47:1692-1697.
10. **Haywood KL, Griffin XL, Achten J, Costa ML.** Developing a core outcome set for hip fracture trials. *Bone Joint J* 2014;96-B:1016-1023.
11. **No authors listed.** National Institute for Health and Care Excellence. Hip Fracture: management. 2011. <https://www.nice.org.uk/guidance/cg124> (date last accessed 08 September 2017).
12. **Griffin XL, Achten J, Parsons N, et al.** The Warwick Hip Trauma Evaluation - an abridged protocol for the WHiTE Study: A multiple embedded randomised controlled trial cohort study. *Bone Joint Res* 2012;1:310-314.
13. **Dolan P.** Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095-1108.
14. **Rizopoulos DJM.** An R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw* 2010;35:1-33.
15. **Cox DR.** Regression models and life-tables. *J R Stat Soc B* 1972;34:187-220.
16. **R Core Team.** R Foundation. R: A Language and Environment for Statistical Computing. 2015. <https://www.R-project.org/> (date last accessed 08 September 2017).
17. **Rizopoulos D.** JM.: Joint modeling of longitudinal and survival data. *Journal of Statistical Software*. 2010. <https://cran.r-project.org/web/package=JM> (date last accessed 08 September 2017).
18. **Saklad M.** Grading of patients for surgical procedures. *Anesthesiology* 1941;2:281-284.
19. **Hodkinson HM.** Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age Ageing* 1972;1:233-238.
20. **Walters SJ, Brazier JE.** Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-1532.

Funding Statement

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Development Grants programme (Reference Number RP-DG-1210-10022). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This trial was co-sponsored by the University of Warwick, University Hospitals Coventry and Warwickshire NHS trust and subsequently University of Oxford.

The WHiTE Study Scientific Committee:

X. L. Griffin, N. Parsons, J. Achten, S. White, T. J. Chesser, C. Boulton, D. Metcalfe, A. Judge, M. Baxter, R. Pinedo-Villanueva, R. Lerner, S. Wallis, D. Haywood, S. E. Lamb, M. L. Costa.

WHiTE Study Principal Investigators:

M. Ancharya, P. Baker, C. Clark, J. Davison, P. Dixon, M. Farrar, P. Fearon, A. Gandhe, R. Handley, A. Mahmood, H. Majeed, A. McAndrew, I. McNamara, B. Olliviere, V. Patel, M. Reed, B. Rogers, G. Smith, J. Young.

WHiTE Study Trainee Principal Investigators:

J. Bassett, R. Begum, D. Burchette, R. Burton, A. Clarke, A. Das, S. Henning, R. Hutchinson, M-C. Killen, A. Kothari, K. Kulkarni, D. Lopez, B. Lord, A. Martin, V. Patel, R. Richardson, R. Sahemey.

Author Contribution

- N. Parsons: Data analysis, Writing the paper.
- X. L. Griffin: Conception of the study, Writing the paper.
- J. Achten: Critical review of the paper.
- T. J. Chesser: Critical review of the paper.
- S. E. Lamb: Critical review of the paper.
- M. L. Costa: Conception of the study, Writing the paper.

Conflicts of Interest Statement

- None declared

© 2018 Parsons et al. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, and not for commercial gain, provided the original author and source are credited.