A global measure of patient-reported outcomes after injury – life back on track

Gang Chen\(^a\), Maame Esi Woode\(^a\), Kah Ling Sia\(^a\), Nina Ellis\(^b\), Cassie Citroen\(^b\) and Anthony Harris\(^a\)

\(^a\)Centre for Health Economics, Monash University, Melbourne, Australia; \(^b\)Transport Accident Commission, Melbourne, Australia

**ABSTRACT**

**Purpose:** Assess the psychometric properties of the Life Back on Track (LBoT) measure, a novel self-reported single-item global measure of the trajectory of wellbeing after a transport accident.

**Materials and methods:** Evaluated the validity, reliability, sensitivity, and responsiveness using four survey waves (\(n = 1556\) in wave 1), and two repeated cross-sectional surveys (\(n = 5238\) and \(n = 1964\)), of individuals injured in a transport accident in Victoria.

**Results:** There were statistically significant differences in the distribution of the LBoT scores by the respondent depression or pain scores, return to work status, financial ability to get by, ability to cope, and ability to bounce back (all \(p < 0.001\)). The LBoT measure was a statistically significant (\(p < 0.001\)) and reasonable predictor of future work status, and was moderately correlated (\(r = 0.67\)) with the EQ-5D-3L (concurrent validity). Retest reliability (ICC \(\geq 0.76\)) and sensitivity (effect sizes \(\geq 1.52\)) were supported, and it was moderately responsive to change (standard response mean statistics 0.4–0.8).

**Conclusions:** LBoT is a valid measure to track the individual’s trajectory of subjective wellbeing in the context of recovery after a trauma, and is potentially useful as an indicator to track the performance of commissioned providers, and to monitor or evaluate the value of service outcomes.

**IMPLICATIONS FOR REHABILITATION**

- There is a demand to develop a simple metric to measure the impact of injury, the effectiveness of rehabilitation and the degree of recovery from trauma.
- Life Back on Track (LBoT) is a valid single-item measure to track an individual’s trajectory of subjective wellbeing after trauma.
- It has the potential to complement clinical measures where a routine collection of a simple measure is desirable.
- It is suitable as an indicator of service outcomes for organisations that commission services.

**Introduction**

Patient reported outcome measures (PROMs) have been used extensively in research and are increasingly being used in clinical registries and quality improvement activities in health services [1]. Their use in clinical practice helps to ensure that therapeutic and service management remains patient-centred. PROMs capture patients’ own opinions on the impact of their condition, and its treatment, on their life. Questionnaires are therefore usually designed to focus on one or more specific elements of a patient’s wellbeing. While some PROMs measure a single dimension of health, for example, physical activity, others specifically measure a combination of physical, mental and social aspects, collectively known as health-related quality of life (HRQoL), while still others measure a global concept of health or wellbeing. Most multi-dimensional HRQoL measures focus on health-related domains (physical and mental health components) describe a generic set of attributes of wellbeing and recovery. For example in head injury from a transport accident, the 20-item quality of life (QoL) questionnaire covers domains related to satisfaction, perception, health and social, and economic [2].

There is no agreement on how best to measure and track the HRQoL or the subjective wellbeing (SWB) of transport accident survivors [3]. There is, however, a clear demand to develop and use a valid simple (non-monetary) metric to integrate changes in years of life with changes in disability, functionality, pain, and emotional distress in order to measure the impact of injury, the effectiveness of rehabilitation and the degree of recovery from trauma [4]. A systematic review of the literature on PROMs in an outpatient rehabilitation setting confirmed earlier studies that found several barriers in using PROMs in rehabilitation settings including the length of required completion time and the inability for patients to independently complete [5]. Moreover for the most part there are questions over the general validity, reliability and responsiveness of current measures that are often not subject to a patient-centric development [1].

Following extensive consultation and discussion with clients, the Victorian Transport Accident Commission (TAC) and the Social Research Centre at the Australian National University devised a single-item global scale, Life Back on Track (LBoT), to measure individual self-appraisal of recovery following a transport accident. It asks a simple single question based on the everyday language of clients: “How would you rate the extent to which you have...”
been able to ‘get your life back on track’, on a scale from 1 to 10?‘ An analysis of the content of the LBoT measure, based on client survey responses, confirms that LBoT for clients measures progress towards an acceptable life that is influenced among other things by work status, living standards, social relationships, and HRQoL. (Supplementary Material 2).

As a direct global measure of subjective personal wellbeing, the LBoT measure has similarities to measures of life satisfaction or happiness such as the single domain Satisfaction with Life Scale (SWLS) [6], and single-item global life satisfaction measure used in large surveys (e.g., Household, Income and Labour Dynamics in Australia survey; German Socio-Economic Panel) and its conceptual basis has much in common with the consumer-oriented definition of psychological recovery in the mental health [7] that has been described as the establishment of a fulfilling, meaningful life and a positive sense of identity founded on hopefulness and self-determination. LBoT is consistent with this approach insofar as it monitors recovery status and not just symptomology.

Like other SWB measures, LBoT allows a within-person QoL construct dynamism, as what matters to the individual and how much it matters (the standard by which they judge QoL) can change over time [8]. To the extent that the LBoT measure has been designed explicitly in the context of a trajectory of recovery with a pre-injury reference point, the standard by which an individual judges wellbeing may be more stable than in other measures of SWB. There are a number of multiple domain SWB measures, but only those with a small number of items are coherent [9]. There is some evidence that single-item global life satisfaction measures perform very similarly compared to the multiple-item SWLS [10]; in health-related quality of life, there is evidence that a global rating showed satisfactory construct validity in terms of measurement of health perception and physical functioning [11]. It remains unclear however if a global SWB measure is sufficiently reliable or responsive to be useful in clinical, program, or service performance evaluation.

This study evaluated the psychometric performance of the LBoT measure (criterion and construct validity, reliability, sensitivity, and responsiveness) in a sample of people who have made an injury claim to a universal public transport injury insurer in the state of Victoria (TAC) following a transport accident. The main aim of the study is to assess whether the LBoT measure can provide a simple but reliable measure of the patient value of the outcomes from services and supports provided to those who have experienced a traffic injury. Having such a client-based value of outcomes is important if insurers and service providers are to monitor and evaluate the patient-relevant outcomes of those services and supports and when costs rise, whether they provide value for money. The potential use of such a metric range from the evaluation of the outcomes of the use of technologies such as telehealth, the comparative performance of contracted provider groups over time, through to a broad assessment of the impact on outcomes of inequities in geographical access to specialist services and supports. The advantage of a global measure such as this is that it is easy to measure routinely and the potential advantage of the LBoT measure, in particular, is that it is framed in a way that is meaningful not only for those injured in a traffic accident but also for a wider group of people recovering from trauma.

Methods

Data

The assessment of psychometric properties of the LBoT measure is based on 3 telephone surveys of Victorians with a managed TAC claim for injury in a transport accident: four waves of a Longitudinal Study survey 2012–2016 with 1556 respondents in wave 1; and two repeated cross-sectional surveys conducted on behalf of the TAC from October 2011 to 2017, the Client Outcome Survey (COS) with 5238 respondents and the Client Experience Survey (CES) with 1964 interviews from 604 unique respondents. While the Longitudinal and COS surveys focused on recovery outcomes, the CES survey focused on perceptions of service and included respondents who had more enduring disabilities. The three surveys are described in detail in Supplementary Material 1.

The life back on track (LBoT) measure

Respondents were asked to rate whether they considered their lives to be back on track as:

- ‘In other research, TAC clients often talk about trying to ‘GET THEIR LIFE BACK ON TRACK’ following a transport accident. This can mean different things to different people. Thinking about your own circumstances right now (today), how would you rate the extent to which you have been able to ‘get your life back on track’, on a scale from 1 to 10, where 1 means ‘not at all’, and 10 means ‘completely back on track’?’ In the CES for respondents who had more enduring disabilities, the recall time was 2 weeks preceding the interview.

Analyses

Construct validity

Construct validity is the degree to which the LBoT measure captures what it intends to measure – in this case, general SWB in recovery from a transport accident. A known-groups validation was conducted, based on the principle that certain specified groups of TAC clients are expected to score LBoT differently from others, and the LBoT measure should be sensitive to these differences.

The known-groups were identified based on construct analysis of qualitative survey data (Supplementary Material 2) and represented by the indicators: (a) self-reported injury severity levels (a 5-point scale: Very Severe, Severe, Moderate, Slight and Very Slight); (b) depression subscale of the Depression Anxiety Stress Scales (DASS)-21 [12,13] (which contains 7 items, each with a four-point severity scale: none of the time, some of the time, a lot of the time, all of the time); (c) pain; (d) financial ability to get by (a 4-point scale: with great difficulty, with some difficulty, very easily); (e) ability to cope with their injuries given its nature (a 5-point Likert scale: very poor, poor, moderate, good, and very good) [14,15], and (g) ability to bounce back from the accident (on a 10-point scale ranging from strongly disagree to strongly agree).

The pain was measured using the Numerical Rating Scale (NRS) in the TAC Longitudinal Study. This validated scale asks respondents to rate their level of pain on a scale of 0–10, where 0 is no pain at all and 10 worst possible pain [16,17]. The rating was then recoded as none (0), mild (1–2), moderate (3–5), strong (6–8), severe (9–10). In the COS, respondents were asked “the amount of bodily pain they had in the past 7 days” and had to choose between 6 options: 1 – none, 2 – very mild, 3 – mild, 4 – moderate, 5 – severe, and 6 – very severe.

The Kruskal-Wallis H test was used to test for statistically significant overall differences between groups, and Dunn’s test for
differences in item pairs between groups (Supplementary Material Tables 3.2–3.6). A p-value < 0.05 was taken as statistically significant in this and all other hypothesis tests in the paper.

**Criterion validity**
Criterion validity, in this context, refers to the extent to which the LBoT measure correlates with an external standard measure. Criterion validity is commonly assessed through the investigation of the concurrent validity and predictive validity of the measure. Concurrent validity involves comparing the LBoT measure to a standard measure of well-being in recovery at the same time point. There is no gold standard for wellbeing in recovery and based on the availability of data we focused on a closely related concept of HRQoL as measured by the EQ-5D-3L – the most widely used preference-based HRQoL instrument in the world [18]. The EQ-5D-3L contains five dimensions with each dimension measured using one item (which includes mobility, usual activities, self-care, pain/discomfort, and anxiety/depression) and three response levels (no problem, some problems, and extreme problems), as well as a single-item EQ-VAS [19]. The EQ-VAS lies on a scale of zero (worst imaginable health state) to 100 (best imaginable health state). The EQ-5D-3L was scored by using the original UK tariff [20] on a 0-1 scale where 1 is full health and 0 is dead. Concurrent validity was measured using the Spearman correlation coefficients between LBoT scores and overall HRQoL scores from the EQ-5D-3L utility scores and the EQ-VAS (visual analogue scale).

Predictive validity was assessed using LBoT scores in wave 1 of the Longitudinal Study in a logit model predicting a return to the same job with the same duties and employer, conditional on being in employment prior to the accident and controlling for injury severity, age, gender, education, employment type and country of birth. Validity was assessed as the ability of the logit regression to classify post-accident employment status correctly using the c statistic (area under the receiver operating characteristic curve, AUC). A value closer to 1 and further from 0.5 suggests greater discrimination and therefore stronger validity [21].

**Reliability**
Reliability refers to the extent to which LBoT scores are affected by random error. We focus on whether LBoT is consistent across time. The Longitudinal Study was used to identify clients who were in a stable condition across two survey time periods (waves 1 and 2). As the time interval between the two waves was relatively long (i.e., ≥ 3 months), we constructed samples with TAC clients who were in a relatively stable condition between waves as indicated by individual and combinations of scores on measures of pain, financial ability to get by, DASS score, a single-item global health rating from the Short-Form 12 Health Survey (SF-12) [22], main labour market activity, and vocational status.

Patients were defined as being in a stable condition if they gave the same pain rating, the same DASS group (Normal, Mild, Moderate, Severe, Extremely Severe), same SF-12 rating and same vocational status in waves 1 and 2. Reliability was measured by the intraclass correlation coefficient (ICC). An ICC of 0.75–0.90 is generally classified as good while an ICC larger than 0.9 is classified as high or excellent reliability [23].

**Sensitivity and responsiveness**
The sensitivity of a measure is the ability to detect differences in response from those who reported that they had recovered to those who said they had not. We chose pain levels and employment status to measure responsiveness, and examined those whose condition had improved in terms of pain rating or vocational status. To adjust for potential bias due to non-response, scores were weighted by the inverse probability of non-response from wave 1 to each of the subsequent waves. Two sets of inverse probability weights were used, one based on the average non-response to that time, and the other based on the probability of non-response for all three periods after wave 1 (Supplementary Materials Table 3.7). The probability of non-response was estimated with a logit model with age, gender, education, area of residence (rural vs metro), longitudinal survey cohort, involvement in an accident (road user), injury type, injury severity, recovery expectation, language spoken at home and country of birth as covariates. We further report the responsiveness results according to the injury severity to reveal potential heterogeneity among patients.

To measure both sensitivity and responsiveness, the effect size and the standard response mean (SRM) were used. A minimum effect size of 0.41 is recommended with an effect size of 1.15 considered moderate and 2.70 considered as strong [24]. A standard response mean of 0.5 is generally considered as indicating moderate responsiveness with a value of 0.8 and above indicating strong responsiveness.

The existence of ceiling and/or floor effects can threaten responsiveness and cause measurement inaccuracy. The potential ceiling and floor effects of the LBoT measure were examined based on the full sample as well as by gender and age groups. Ceiling or floor effects are taken as evident if ≥ 15% of respondents scored the best or the worst of a measure [25].

**Results**

**Sample characteristics**
The socio-demographic characteristics of the respondents in wave 1 of the Longitudinal Study are presented in Table 1. The age distribution of the respondents was roughly even, with more males than females. A little over 75% of the respondents had less than a bachelor’s degree. A little over half of the respondents (54%) were drivers of a vehicle during the accident, with slightly less than a fifth (18%) being motorcyclists during the accident. Respondent characteristics in the other two surveys are similar (Supplementary Materials Table 3.1).

**Psychometric properties of the LBoT measure**

**Descriptive statistics of the LBoT measure**
The mean, standard deviation (SD), and the distributions of the LBoT scores across the different time periods and across the three surveys are presented in Figure 1. The mean of the LBoT scores was higher in the COS and wave 1 of the Longitudinal Study compared to respondents in the CES. The mean LBoT score increased over time with most of the change occurring by the second wave (3 months later) when the median increased from 7 to 8. There is an increase in the proportion of respondents with a score of 10, but given the 41% loss to follow-up by wave 4, this trend should be interpreted with caution.

**Known-Groups validity**
We expected that respondents who reported more severe injuries, depression, or pain would report significantly lower LBoT scores, while those who reported having better ability to cope with their injuries and to bounce back from their injuries would report
higher LBoT scores. These hypotheses were supported by the Kruskal-Wallis $H$ test statistics ($p$-value < 0.001) (Table 2), and Dunn's test (Supplementary Materials Tables 3.2–3.6) with robust results using COS or CES data.

### Concurrent validity

Table 3 shows that there was a moderate correlation between the LBoT measure and the EQ-5D-3L at all time periods (TAC Longitudinal Study, wave 1 to wave 4) and for the COS data. The magnitude of correlations was stronger between LBoT and EQ-5D-3L utility scores ($r = 0.671–0.732$) than with EQ-VAS ($r = 0.567–0.647$), across all surveys.

### Predictive validity

LBoT scores at wave 1 were predictive of return to work at each subsequent time (Table 4). Individuals with a higher LBoT score were generally more likely to have returned to work in later periods. All coefficients are statistically significant (all $p < 0.001$), with AUC values of 0.74 in wave 2 to almost 0.8 in wave 4 suggesting the LBoT measure is a reasonable predictor of future work status post-injury.

### Test-Retest reliability

In the 3 months between waves 1 and 2 respondents who had stable response measures of pain, financial ability to get by, DASS score, global health rating, main labour market activity, and vocational status reported consistent LBoT scores, with an ICC of at least 0.76 across all combinations tested.

### Sensitivity

Those who considered themselves to have recovered gave a higher LBoT score than those who believed that they had not
### Table 2. Known-groups validity: Kruskal-Wallis test results for TAC longitudinal study (wave 1).

| Categories                            | N   | LBoT score Mean | SD   | p-value |
|---------------------------------------|-----|-----------------|------|---------|
| Injury severity                       |     |                 |      |         |
| Very severe                           | 296 | 5.4             | 2.7  | <0.001  |
| Severe                                | 590 | 6.4             | 2.5  |         |
| Moderate                              | 465 | 7.5             | 2.2  |         |
| Slight                                | 105 | 8.4             | 2.1  |         |
| Very slight                           | 41  | 8.6             | 2.0  |         |
| General health (Wave 1 vs Pre)        |     |                 |      |         |
| Worse health                          | 885 | 6.0             | 2.5  | <0.001  |
| Same health                           | 521 | 7.8             | 2.2  |         |
| Depression (DASS)                     |     |                 |      |         |
| Normal                                | 1023| 7.4             | 2.3  | <0.001  |
| Mild                                  | 114 | 5.7             | 2.2  |         |
| Moderate                              | 127 | 4.8             | 2.3  |         |
| Severe                                | 42  | 4.4             | 1.9  |         |
| Extremely severe                      | 69  | 3.7             | 2.5  |         |
| Pain                                  |     |                 |      |         |
| None (0)                              | 392 | 8.4             | 2.1  | <0.001  |
| Mild (1–2)                            | 118 | 7.9             | 2.1  |         |
| Moderate (3–5)                        | 436 | 6.7             | 2.3  |         |
| Strong (6–8)                          | 500 | 5.5             | 2.3  |         |
| Severe (9–10)                         | 65  | 4.2             | 2.5  |         |
| Time to recovery                      |     |                 |      |         |
| Already recovered                     | 244 | 8.8             | 1.8  | <0.001  |
| Next few months                      | 569 | 7.2             | 2.2  |         |
| Within a year                         | 319 | 6.0             | 2.2  |         |
| Longer than a year                    | 92  | 5.4             | 2.4  |         |
| Will not recover                      | 32  | 3.7             | 2.8  |         |
| Return to work                        |     |                 |      |         |
| Paid work, same employer              | 617 | 7.8             | 2.0  | <0.001  |
| Paid work, different employer         | 49  | 7.4             | 2.4  |         |
| Retired/permanently unable to work    | 247 | 6.2             | 2.8  |         |
| Not working, pre & post               | 171 | 5.9             | 2.6  |         |
| Paid work, pre                        | 390 | 5.5             | 2.5  |         |
| Financial ability to get by           |     |                 |      |         |
| Same difficulty                       | 887 | 7.0             | 2.4  | <0.001  |
| Worse difficulty                      | 375 | 5.7             | 2.6  |         |
| Ability to cope                       |     |                 |      |         |
| Very poor                             | 60  | 5.0             | 2.8  | <0.001  |
| Poor                                  | 169 | 4.8             | 2.5  |         |
| Moderate                              | 483 | 6.3             | 2.3  |         |
| Good                                  | 527 | 7.3             | 2.3  |         |
| Very good                             | 254 | 8.2             | 2.2  |         |
| Ability to bounce back (recoded):     |     |                 |      |         |
| difficult to bounce back 1 =          | 484 | 8.4             | 2.0  | <0.001  |
| Strong disagreement                   |     |                 |      |         |
| 2                                    | 164 | 7.1             | 2.0  |         |
| 3                                    | 284 | 6.6             | 2.0  |         |
| 4                                    | 281 | 5.8             | 2.1  |         |
| 5                                    | 265 | 4.6             | 2.6  |         |

### Table 3. Concurrent validity: Spearman rank correlation.

|                           | N   | Mean  | SD    | Coefficient | p-value |
|---------------------------|-----|-------|-------|-------------|---------|
| LBoT wave 1               |     |       |       |             |         |
| EQ-VAS                    | 1506| 71.61 | 20.77 | 0.57        | <0.001  |
| EQ-5D-3L utility          | 1488| 0.62  | 0.32  | 0.67        | <0.001  |
| LBoT wave 2               |     |       |       |             |         |
| EQ-VAS                    | 1208| 75.89 | 19.81 | 0.64        | <0.001  |
| EQ-5D-3L utility          | 1189| 0.68  | 0.30  | 0.70        | <0.001  |
| LBoT wave 3               |     |       |       |             |         |
| EQ-VAS                    | 1063| 76.83 | 19.19 | 0.65        | <0.001  |
| EQ-5D-3L utility          | 1049| 0.70  | 0.30  | 0.73        | <0.001  |
| LBoT wave 4               |     |       |       |             |         |
| EQ-VAS                    | 833 | 0.72  | 0.30  | 0.65        | <0.001  |
| EQ-5D-3L utility          | 833 | 0.75  | 0.25  | 0.70        | <0.001  |
| COS                       |     |       |       |             |         |
| EQ-VAS                    | 5348| 0.67  | 0.32  | 0.64        | <0.001  |
| EQ-5D-3L utility          | 5438| 70.99 | 21.38 | 0.69        | <0.001  |

**Notes:** A logit model was used and an odds ratio (OR) was reported. Standard errors in parentheses: *p < 0.10, **p < 0.05, ***p < 0.01, AUC: Area under the Curve. In addition to what has been reported in the table, other control variables include respondents’ perceived injury severity, injury type, age, gender, education, employment at the time of the accident, type of household, country of birth, whether they live in the metro area or a rural area, their role in the accident (driver, passenger, motorcyclist, etc.) and the language they speak at home.
recovered. The absolute value of the effect size was 1.538–2.461 across the waves of the Longitudinal Study and 1.523 in the COS data. These results indicate that irrespective of the sample used, the LBoT measure was able to detect different recovery statuses with moderate effect.

**Responsiveness**

Overall, we observed a ceiling effect for LBoT scores for the three different surveys, across age groups and gender. The percentage of respondents who scored 10 was 18–26% across the Longitudinal and COS surveys. Only 4% in the first two samples and 7% of those classified as having a disability reported a score of 0.

**Figure 2** plots the estimated standard response mean values of the study sample whose pain or employment improved from wave 1 to the following three waves. The figure also shows a comparison with the EQ-5D-3L instrument, as well as the sub-sample analyses based on four injury severity levels. More detailed results can be found in Supplementary Materials Tables 3.8–3.10.

LBoT is more responsive over time with an SRM of 0.4 between waves 1 and 2, rising to 0.8 between waves 1 and 4 (16–21 months later). While LBoT was less responsive than EQ-5D-3L between waves 1 and 2, it was more responsive between wave 1 and the later waves. Using the conventional Cohen’s effect size threshold of ‘trivial (<0.2), ‘small’ (≥0.2 & <0.5), ‘moderate’ (≥0.5 & <0.8), or ‘large’ (≥0.8), the effect sizes would be regarded as moderate. Figure 2 also shows the heterogeneity of different severity groups across the recovery trajectory. From waves 1 to 2, the severe group was the most responsive, followed by very severe, moderate and (very) slight groups. Between waves 1 and 3, all four groups attained a similar level of responsiveness but between waves 1 and 4, the severe and very severe groups demonstrated higher responsiveness than the other two groups. Adjusting for potential attrition bias using inverse probability weightings did not affect these results irrespective of the time period used to create the weight.

**Discussion**

This study provides empirical evidence on the psychometric properties (validity, reliability, sensitivity and responsiveness) of a novel measure of wellbeing in recovery following a transport accident – Life Back on Track. As a post-injury outcome measure, the LBoT has the potential to be used for priority setting in injury prevention, treatment and rehabilitation. It is likely to be particularly useful where there is a focus on the long-term health and wider consequences of injury, and the ability of people to adapt and cope with the consequences of physical or emotional trauma.

A qualitative analysis of open-ended text responses in the survey (Supplementary Materials 2) suggests that the reported domains of the LBoT measure cover the concepts of recovery towards a normal life in terms of independence, control, happiness, work, social life with family and friends, pain, physical function, cognitive function, work and leisure activities, income, anxiety and depression. A smaller sample of clients that included a large number of individuals classified as having a significant enduring disability answered the question of “what the measure means to them,” reported a somewhat wider set of concepts including independence and living a normal life, where normal for some was life before the accident and for others appeared to be accepted as an altered life. For these clients, the distinction between being “on track” and “back on track” was more evident and brings into focus the performance of the measure as both an indicator of recovery and an indicator of the quality of life.

Overall, the empirical evidence indicates that the LBoT measure is a valid measure to track a given individual’s trajectory of SWB in recovery after a trauma. We found that the LBoT measure was able to distinguish the pre-defined known groups, such as injury severity and self-reported health status. The reliability of the LBoT measure was judged high by the ICC regardless of the way a stable cohort between the first two waves of the longitudinal data was constructed. The LBoT measure was found to be a moderately sensitive instrument when distinguishing self-evaluated recovery status. The psychometric properties of the LBoT measure are as strong as those measured for other single-item quality of life instruments. One single-item global life satisfaction scale reported high criterion and construct validity [10], while the linear analogue self-assessment (LASA) for overall quality of life in cancer patients undergoing treatment, showed convergent and discriminant validity at baseline, but the evidence was less clear under treatment [26].

On the other hand, based on conventional rules of thumb, the LBoT measure was not found to be very responsive particularly in the earlier stages of recovery. However, this result should be interpreted with caution as measurement error introduced by the long gap between reports in later waves (1–2years; see Electronic Supplementary Material 1) may have biased downward the measurement of responsiveness to change compared to other studies with much shorter times between reports. In the context of the ability of the data to provide reliable evidence of change, we note that in the short run the LBoT was less responsive than the widely used quality of life measure EQ-5D-3L but in the long run it was more responsive than EQ-5D-3L in this data. This difference may potentially be due to the larger influence of pain on quality of life (i.e., pain is one of five dimensions of EQ-5D-3L) in the short run whereas in the relatively long run the LBoT was more capable to capture wider improvement on wellbeing. In general, there is currently limited evidence on the comparative responsiveness of single-item life satisfaction measures given that few longitudinal studies have included both single and multi-item measures of life satisfaction. In cancer patients, the responsiveness
of single-item LASA at 3 months after chemotherapy was comparable with indicators of physical wellbeing and coping [26], while a visual analogue scale based single item measure had moderate responsiveness at five weeks post-operatively (SRM = −0.47; effect size = −0.56) and responsiveness to detect clinically relevant changes showed a moderate correlation (r = 0.55) compared to multi-item questionnaires [11]. Moreover, the ceiling effect of the LBoT measure, i.e., the lack of discriminative ability (or precision) towards the highest end of LBoT distribution, may mean that we did not have enough power to detect a response based on the current sample size. It should be noted that the ceiling effect we observed is similar to that found for the EQ-5D-3L (i.e., scored as full health according to the health state classification system) in the same dataset (19% scored 10 in LBoT vs. 16% scored 1 based on EQ-5D-3L).

LBoT is a single-item global measure of the current experience of wellbeing, but in contrast to commonly used measures of current life satisfaction or happiness it has both an explicit life event reference point (the accident), and an explicit focus on a personal ideal life trajectory. The ability to get one’s life back on track following a trauma suggests a measure that encourages consideration of active aspects of satisfaction such as personal participation and personal control. Two people with radically different circumstances pre and post an accident may both reasonably state that their life was equally back on track; conversely, two individuals who are returned to the same level of functioning may report very different degrees of being “back on track.” It may be that the LBoT measure has elements of a measure based on a basic justice approach to measuring desirable outcomes – evaluating health and social care programs not only in terms of achieving an improvement in reported quality of life but also in regard to their effects on the capacity to achieve a preferred quality of life [27].

There are two main limitations of this study. First, the term Life Back on Track was initially developed from the language used by clients in previous qualitative research studies, and the underlying concept was explored in the current study in a content analysis of responses to simple questions on the meaning of the concept “life back on track.” While the concepts reported have face validity, it is not clear that this approach has resulted in a full understanding of the conceptual basis of the measure. Second, the analysis of the responsiveness of the LBoT measure was limited to proxy measures of the recovery status of the respondents; and for those with an enduring disability, although cross-section data showed similar results to other clients, an absence of repeated measurement meant that we were unable to test responsiveness adequately.

**Conclusion**

While further evidence on its responsiveness to change needs to be established, the LBoT measure is a valid measure to track an individual’s trajectory of SWB at least in the context of recovery after a transport accident. It is also a valid measure of average group outcomes and as such could be used as a performance outcome measure in the evaluation of rehabilitation service quality. The LBoT measure covers wider concepts than HRQoL domains that are traditionally captured in measures of value in health. The context differentiates the measure from standard single item SWB measures (life satisfaction and happiness), and the measure has a conceptual basis in ideas of recovery, physical and social functioning, wellbeing and employment. It is possible that respondents will adapt their attitudes and life aspirations to their changed circumstances and that changes in LBoT scores over time may in part reflect this. This kind of adaptation is not unique to this context, but the LBoT measure, by focussing on a trajectory anchored on a past ideal view of a good life, may be less susceptible to this source of shifts in response compared to other SWB measures.

The experience of the Transport Accident Commission is that the LBoT measure is useful as part of a suite of indicators of organisational performance. Following evaluation of the measure (including the current study), the TAC intends to use the LBoT measure, both as a performance indicator for the organisation over time and as a measure of client group outcomes as part of its value-based health care strategy. As that strategy unfolds, there will be opportunities to test further the validity of the LBoT measure.

**Ethical approval**

Ethics approval was granted by the Monash University Human Research Ethics Committee, Monash University, Australia (Reference no. 18278). The study adhered to the principles outlined in the Declaration of Helsinki.

**Disclosure statement**

Nina Ellis and Cassie Citroen are employed by the Transport Accident Commission. All other authors have no competing interests.

GC designed the quantitative analysis and contributed to writing the manuscript. MW performed the quantitative analysis and contributed to writing the manuscript. KLS reviewed the literature and contributed to the writing of the manuscript. NE and CS were responsible for the data collection and initial instrument development with clients. AH analysed the qualitative data and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

**Funding**

The study is funded by the TAC Research Project (T019). We are grateful to the survey participants from the TAC Longitudinal Study, the Client Outcomes Survey (COS), and the Client Experience Survey (CES).

**ORCID**

Anthony Harris [http://orcid.org/0000-0003-1641-3320]

**Data availability statement**

Survey data used in the study contain confidential information and not released publicly

**Informed consent**

Informed consent was obtained from all individual participants included in the study.

**References**

[1] Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.
[2] Chaikoolvatana A, Sripech S. Quality of life of road accident head-injured patients after craniotomy. Southeast Asian J Trop Med Public Health. 2007;38(4):761–768.

[3] Rissanen R, Berg H-Y, Hasselberg M. Quality of life following road traffic injury: a systematic literature review. Accid Anal Prev. 2003;25(1):3–19.

[4] Segui-Gomez M, MacKenzie EJ. Measuring the public health impact of injuries. Epidemiol Rev. 2003;25(1):3–19.

[5] Briggs MS, Rethman KK, Crookes J, et al. Implementing patient-reported outcome measures in outpatient rehabilitation settings: a systematic review of facilitators and barriers using the consolidated framework for implementation research. Arch Phys Med Rehabil. 2020;101(10):1796–1812.

[6] Pavot W, Diener E. The satisfaction with life scale and the emerging construct of life satisfaction. J Positive Psychol. 2008;3(2):137–152.

[7] Anthony WA. Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. Psychosoc Rehabil J. 1993;16(4):11–23.

[8] Allison PJ, Locker D, Feine JS. Quality of life: a dynamic construct. Soc Sci Med. 1997;45(2):221–230.

[9] Cummins RA, Weinberg MK. Subjective wellbeing – multi-item measurement: a review. In: Global handbook of wellbeing and quality of life. Dordrecht: Springer; 2015.

[10] Cheung F, Lucas RE. Assessing the validity of single-item life satisfaction measures: results from three large samples. Qual Life Res. 2014;23(10):2809–2818.

[11] de Boer AGEM, van Lanschot JJB, Stalmeier PFM, et al. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? Qual Life Res. 2004;13(2):311–320.

[12] Lovibond S, Lovibond P. Manual for the depression anxiety stress scales. Sydney: Psychology Foundation Monograph; 1995.

[13] Henry JD, Crawford JR. The Short-Form version of the depression anxiety stress scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol. 2005;44(Pt 2):227–239.

[14] Lazarus R. From psychological stress to the emotions: a history of changing outlooks. Annu Rev Psychol. 1993;44:1–21.

[15] Hepp U, Moergeli H, Buchi S, et al. The long-term prediction of return to work following serious accidental injuries: a follow up study. BMC Psychiatry. 2011;11:53.

[16] Sendlbeck M, Araujo EG, Schett G, et al. Psychometric properties of three single-item pain scales in patients with rheumatoid arthritis seen during routine clinical care: a comparative perspective on construct validity, reproducibility and internal responsiveness. RMD Open. 2015;1(1):e000140.

[17] Thong ISK, Jensen MP, Miro J, et al. The validity of pain intensity measures: what do the NRS, VAS, VRS, and FPS-R measure? Scand J Pain. 2018;18(1):99–107.

[18] Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: J. CA, editor. Encyclopedia of health economics. San Diego (CA): Elsevier Science; 2014. p. 34–357.

[19] EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199–208.

[20] Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–1108.

[21] Hosmer DW, Lemeshow S. Applied logistic regression. New York (NY): John Wiley & Sons; 2000.

[22] 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(3):220–233.

[23] Koo TK, Li MY. A guideline of selecting and reporting intra-class correlation coefficients for reliability research. J Chiropractic Med. 2016;15(2):155–163.

[24] Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pract. 2009;40(5):532–538.

[25] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293–307.

[26] Bernhard J, Sullivan M, Hurny C, et al. Clinical relevance of single item quality of life indicators in cancer clinical trials. Br J Cancer. 2001;84(9):1156–1165.

[27] Sen A. Development as freedom. Oxford: OUP; 1999.