Measuring changes in risk and need over time using OASys

Philip Howard and Robin Moore

Key points

- The Offender Assessment System (OASys) was designed to be administered repeatedly and National Standards for offender management suggest that reviews be completed for all assessed offenders. Analyses of initial assessments completed between July and December 2004, and all subsequent assessments over a two-year period, found an average of 4.4 assessments per offender, and 8% of offenders with no review assessment.

- Where re-assessment occurred, 31% of final assessments were unchanged on all 73 scored questions. Statistical modelling suggested that around one-third of these assessments were not properly considered reviews. Quality assurance procedures should check the thoroughness of review assessments with no score changes.

- On average, seven of the 73 scored questions changed, and risk of serious harm ratings were also often revised, showing that OASys assessments can change considerably over time. Scores on the more historic questions changed far less often, resulting in the relatively poor dynamism of the alcohol misuse and education, training and employability sections. The continued use of those questions whose scores seldom change will be considered as part of a wider review of the composition of OASys.

- It remains to be proven that changes in predictor scores sufficiently predict changes in reoffending rates. However, the dynamism of scores on all three OASys-based predictors of recidivism is encouraging. Furthermore, these initial findings indicate that offenders reconvicted at their third assessment were more likely than other offenders to have section- and total-score increases between the first and second assessments and less likely to have score reductions.

Context

The aim of the research was to establish whether the Offender Assessment System (OASys) measures changes in offenders’ risks and needs over time, and whether any measured changes predict reoffending. OASys is the principal risk assessment tool used by the National Offender Management Service (NOMS) of England and Wales. It is used to measure an offender’s likelihood of further offending; to identify any risk of serious harm issues; to develop an offending-related needs profile; to develop individualised sentence plans and risk management plans; and to measure progress and change over time. It was designed to be administered repeatedly, with NOMS National Standards (2007) dictating that assessments should be reviewed within 16-week periods in the community, with a termination assessment required at the end of the order or licence.
**Approach**

The study sought to address the following research questions:
1. How often are assessments being reviewed?
2. Do the reviewed assessments demonstrate change?
3. Which areas of the assessments change most?
4. Are the changes predictive of reoffending?

Initial probation OASys assessments completed from July to December 2004 were selected. ‘Initial’ assessments were completed at the Pre-Sentence Report (PSR) stage before a community sentence, at the start of a community sentence (if there was no PSR) or at the start of a custodial licence. Removing assessments which failed to meet data completion criteria or where OASys use was not obligatory left 17,824 assessments. The remaining assessments covered offenders serving prison sentences of at least 12 months and probation sentences involving supervision. This group had a higher likelihood of reoffending and risk of serious harm than those excluded.

All subsequent assessments completed within 24 months of these initial assessments were then identified, producing a ‘set’ of assessments for each offender. For the change analysis, it was determined that the ‘sequence’ of assessments should run from the initial assessment to either (i) the first termination assessment, (ii) the first assessment indicating reconviction, or (iii) the final assessment in the set if neither termination nor reconviction had occurred. The number of sequences where change could be studied was 16,222.

Net and absolute changes were both calculated, the latter enabling items where change seldom occurred to be distinguished from more dynamic items where increased scores for some offenders were balanced by decreases for others. Bearing in mind that some reviews could have been completed in a perfunctory manner where a more thorough review might have revealed some change, a zero-inflated Poisson regression model was fitted. When the number of scored item changes was zero, this model estimated the probability that the observation came from an ‘always zero’ group, i.e. a fully automated unthinking review, rather than being a properly considered review which happened to have zero score changes. The estimate was then used to weight zero-change assessments in the remaining results.

Finally, score changes between initial and second assessments were used to predict reconviction at third assessment. The available sample for this stage of the study was restricted to 1,862 sequences of assessments.

**Results**

**How often are assessments being reviewed?**
For the whole sample, there was an average of 4.4 assessments per offender with a mean interval between the first and second assessment of 118 days. The average number of assessments was higher for those offenders who had been reconvicted (38% of the sample) compared to those who had not been reconvicted: 5.9 compared to 3.5. Around half of the reconvicted offenders had four, five or six assessments. Overall, 8% of assessments were not reviewed (i.e. there was only one assessment in the sequence), while almost a fifth (19%) of sequences did not end in a termination or reconviction.

Where there was more than one assessment in the sequence, 69% demonstrated change in at least one scored question. The zero-inflated Poisson model results suggested that 31% of the sequences with no change were not completed well. The subsequent analysis therefore gave sequences with zero changes a weight of 0.69.

**Do the reviewed assessments demonstrate change?**
On average, seven of the 73 scored items changed, and 1.0 of the ten criminogenic need indicators based on these items, with decreases slightly outweighing increases. The OASys weighted score changed by an average of ±7.3 points, with a net decrease of almost one point. This represented an average percentage change of ±16% from the initial mean for the dynamic elements (7.3/46).

In comparison, the OASys General Reoffending Predictor (OGP) was less dynamic with an average change from its initial dynamic mean of ±13% (1.6/12.8) but the OASys Violence Predictor (OVP) was more dynamic with an average change from its initial dynamic mean of ±18% (2.2/12.3). Mean OGP and OVP scores both fell slightly.
Which areas of the assessment changed most?
At individual item level, the prevalence of score changes ranged from 2% (Q4.8 learning difficulties) to 22% (Q3.4 suitability of accommodation). As shown by Figure 1, accommodation was the most dynamic section, with around 8% of offenders moving each way across the criminogenic need threshold. The (i) thinking and behaviour and (ii) lifestyle and associates sections both had over 11% absolute change. Alcohol misuse and education, training and employability (ETE) changed least often (6.5% and 7.1%). In the ETE section, the employability questions were quite dynamic but the education and training questions were very static, while the alcohol misuse need measure was rendered relatively static by two questions based on past behaviour. Thinking and behaviour showed the greatest net change: the proportion with this need fell by 3% from initial to final assessment.

Risk of serious harm to the community converged towards medium risk at final assessment. The risk to all four groups (known adults; children; general public; and staff) and the overall community risk saw more offenders move from low to medium than vice versa and more from high/very high to medium than vice versa. The overall risk level changed in approximately one in five (18%) of the assessments.

Did changes to the assessment predict reoffending?
Changes in the total OASys score between initial and second assessment predicted reconviction at third assessment. Non-reconvicted offenders’ scores fell on average by 2.4 points, while reconvicted offenders’ scores rose by an average of 0.8 points. The difference between reconvicted and non-reconvicted offenders was largest for those with moderate initial scores, possibly because these scores provided ample room for scores to increase or decrease with few ‘ceiling’ or ‘floor’ effects.

As shown by Figure 2, changes in each of the ten criminogenic need scores were also correlated with reconviction. The largest differences were in accommodation, ETE, financial management, and thinking and behaviour, where the mean changes in score for reconvicted offenders were 0.4 points greater than for non-reconvicted offenders. When controlling for charged but not convicted offences and the static Offender Group Reconviction Scale (OGRS) score (based upon offender demographics and offending history information), change in the thinking and behaviour score was the strongest predictor of reconviction.

Figure 1: Percentage of assessments with changes in criminogenic need status between initial and final assessment

![Bar chart showing percentage of assessments with changes in criminogenic need status between initial and final assessment](chart.png)
Implications

Many offenders were found to have several OASys assessments and some key OASys items were highly dynamic. However, not all offenders had patterns of assessment which met National Standards for assessment timing and frequency, and a substantial minority of repeat assessments reported no change in any of the 73 scored OASys questions. While deviation from National Standards will legitimately occur in a minority of cases, and it is possible that some offenders will demonstrate reaction to neither supervision nor external events, the prevalence of these outcomes suggests sub-optimal practice in many cases. A new NOMS Offender Management metric on the timely completion of termination assessments should have a positive effect. In order to maximise its impact upon the quality as well as the quantity of assessments, it should be accompanied by checks that some item scores have changed in the majority of cases. The OASys Quality Management Plan should thus scrutinise reviews, ensuring that they reflect changes whenever these have occurred.

Valid dynamic questions are important as they maintain the predictive validity of risk scores over time and allow the success or failure of offender management, interventions and the offender’s own efforts to change to be demonstrated. Individual item results confirm ‘common sense’ readings of the items, indicating that those which are based on historic or family factors (i.e. out of the offender’s control) rarely changed. There are strong moral and practical grounds for only including such items when absolutely necessary, and their continued use will need to be considered carefully alongside other research findings on the reliability and validity of OASys.

The dynamism of scores on all three OASys-based predictors of recidivism is reassuring. It remains to be proven that changes in predictor scores sufficiently predict changes in reoffending rates, but the initial findings set out in this report are encouraging with the mean scores rising for those who were reconvicted and falling for those who were not reconvicted.