

METHODS FOR GRADING SPIs

SUMMARY

This paper describes the methods of scoring and assessment for each Statutory Performance Indicator (SPI). Appendix 1 provides an overview of the grading methods indicator-by-indicator and a complementary paper sets out the method used for aggregating grades. Further information can be found at:

<http://police.homeoffice.gov.uk/performance-and-measurement/performance-assessment/>

INTRODUCTION

Data for the 32 SPIs are of different types, and thus cannot all be assessed on the same basis. Every indicator has the following information in common, regardless of the data type.

- a) data relating to the forces' performance;
- b) comparable data from other police forces are used or, for the direction of travel indicator, comparison uses appropriate historical data from the force itself;
- c) a score comparing the force and comparator data; and
- d) the score thresholds required to achieve a particular grade.

PROCEDURES FOR GRADING

Force Data

The data for each force is collected for the period 2004/05. This may be recorded crime data (e.g. 5b violent crime rate), results from User Satisfaction and British Crime Surveys (e.g. 10b perceptions of anti-social behaviour) or other sources (e.g. 11a front-line policing). Data for road traffic safety is only available for the 2004 calendar year.

Force Comparator Data

For the majority of SPIs, a force is compared to the aggregate performance of its most similar forces (MSFs). For this year's assessment, the MSF average includes the force itself.

In the Providing Assistance and Resource Use domains a national target, the same for all forces, has been set for SPIs 11a (front-line policing), 12c (female officer representation) and 13a&b (sickness absence). For these indicators the national average is used as the comparator in place of the MSF.

Some indicators require a force to compare data against itself. This could be comparing the satisfaction of White and Visible Minority Ethnic (VME) victims of crime with respect to the overall service provided, or comparing the arrest rates from Police and Criminal Evidence act (PACE) searches between White and VME.

The City of London does not have an MSF for comparison purposes. Page 4 outlines the specific methods used in this case.

Scoring a force

Recorded Data

Scores for most recorded data are calculated as the ratio of the force performance to its comparator. For recorded indicators where the MSF is the comparator, the scores for both delivery and direction are transformed using the logarithmic function to make the distribution more regular. This means that an equal percentage change in performance will result in the same change in score relative to the MSF average regardless of the initial level of performance.

SPI 5a measures the domestic burglary rate per 1,000 households. These assessments use a more precise figure for households in a force than those used in Home Office Statistical Bulletin 11/05, and thus may not match exactly.

The distribution of the force scores is then used to determine the grade boundaries. The distribution is assumed to be normal with a mean of 100% (i.e. average performance against the MSF comparator) and a standard deviation as calculated from the force scores.

The boundaries are selected such that the expected distribution of Excellent, Good, Fair and Poor grades would be 20%:25%:45%:10%. For example, the boundary between Poor and Fair is the boundary between the bottom 10% of the distribution and the upper 90% if the distribution was perfectly normal. These are expressed in terms of z-scores, or the number of standard deviations away from the mean.

The boundaries for each SPI will be different, being dependent on the standard deviation, or amount of spread of force scores.

For front-line policing, the excellent boundary is fixed at the 2007/08 target of 72.5%.

For SPI 12c, the Excellent/Good boundary is set at 28% representation. This is less than the economically active population, but represents a level which should be attainable by the best forces within the next few years.

For SPI 13a&b, the Good/Fair boundary is set at the national targets. The targets are 92 hours/yr for officers and 96 hours/yr for staff. This corresponds to scores of 125% and 128% respectively (that is, 25% and 28% greater than the average score across all forces).

The same method is applied to the direction of travel indicators, with the exception that the z-scores are based on the expected distribution of scores between Improved, Stable and Deteriorated being 33%:33%:33% and the centre of the stable band is set at zero relative change.

Forces' FLP measures may this year have been affected in part by changes to the way sickness data is reported to the Home Office. We have converted sickness hours data into a measure of sickness days using a factor of 8.

Survey Data

Results from surveys are only from a sample of the total population, and thus there is a degree of uncertainty surrounding the score, usually expressed in terms of the confidence limits. Due to this, a different approach has been taken to grade indicators based on survey data.

If the difference between the force and MSF is *not* statistically significant at the usual 95% level, then the force is graded as Fair. Performance significantly below the MSF average is graded as Poor. Significantly better performance is graded as Good/Excellent.

Since there are only three possible grades, Good and Excellent are combined into a single grade.

Direction of travel grades are made on the same basis. Significantly better performance compared to the previous year is Good/Excellent, significantly worse performance is Poor, and a Fair is awarded where there is no significant difference.

SPIs 1a-1e from User Satisfaction surveys have smaller confidence limits than for the British Crime Survey, and thus forces can be graded on a full four-point scale including Excellent. Where the force score is 5 percentage points less than the MSF average, a Poor grade is awarded. Provided that the difference is statistically significant, cases that are between 2 and 5 percentage points greater than the MSF are Good, and a grade of Excellent is given where the force score is 5 or more percentage points higher than the average. Otherwise the force is graded as Fair. User satisfaction results are not available for 2003/04, thus no direction of travel assessment will be made this year.

Although SPI 3a (satisfaction of victims of racism) is part of the User Satisfaction survey, the small number of respondents made it more appropriate to assess this indicator by significant difference as for BCS survey indicators.

SPI 10a (fear of crime) is assessed and graded as three separate indicators (fear of crime relating to burglary, car crime and violent crime). However, when aggregating to domain level each of these only has $\frac{1}{3}$ weight of a normal indicator.

Proportionality Data

Several indicators are based on proportionality measures. These indicators are based on the ratio of two sets of force data, and there are no external comparators such as an MSF average. A good score for these measures is either close to equality (i.e. for comparisons of satisfaction between white and VME groups) or a high ratio (i.e. ratio of ethnic minority recruits).

SPIs 3c and 3d (Fairness of stop and search towards minority ethnic groups, and investigating crimes against minorities proportionately and fairly) have been assessed using the standard chi-square statistical test. This test compares the arrest rates of white and VME groups - if there is no significant difference between the arrest rates at the 95% level of significance, then the force is graded Good/Excellent. Significant difference at the 95% level results the force being graded Fair, and a Poor is awarded where there is a statistical difference at the 99% level of significance.

While it is recognised that the standard this sets is almost certainly too low, it does highlight those who need to focus the most, and it does dramatically reduce the probability of highlighting a force in error. It is fully expected that this (and other proportionality standards) will be reviewed and revised over time.

For the measures in the Resource Use domain, the scores are calculated as normal as the ratio of the force performance to their comparator.

Several grade boundaries for these indicators have been calculated differently.

For 12a (ethnic minority recruits) the comparator is the proportion of ethnic minorities in the economically active population. The boundaries for this indicator are set as follows:

- Poor/fair boundary: National Average
- Good/Fair boundary: 100% (i.e. equal to the economically active population)
- Excellent/Good boundary: 200% (twice the economically active population).

All other grade boundaries for these indicators, including all change over time assessments are set using the normal z-score method described above.

Missing and Invalid Data

Where a force has not provided data, or where incorrect data has been supplied, that force is not graded for that indicator, and awarded a penalty grade accordingly. Data which is missing or substantially erroneous will be excluded from the distribution of force scores and all relevant MSF averages.

Forces who carried out a non-compliant user satisfaction survey have been scored as normal but penalised one grade. These scores have been included in the MSF averages.

City of London

As stated above, the City of London has no other force in its MSF group. Comparison has been made against the national average for selected indicators. These include user satisfaction, Offences Brought to Justice (OBTJ) and sanction detection measures, and also domestic violence and road traffic safety.

The City of London provides policing services to a far larger population than those who are resident in the area. With the exception of burglary, assessment of the recorded crime indicators (which are per 1,000 residents) is not meaningful. SPI 5a (Domestic Burglary rate) is measured by 1,000 households, so a comparison can be made for this indicator against the national average.

The BCS does not distinguish responses from the City of London and so grading of these measures is impossible.

Overall this means the City of London is only assessed on 23 indicators, and so is proportionately more reliant on HMIC assessments.

**Police Standards Unit
Home Office
October 2005**

Reducing Crime		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
4a	Comparative risk of personal crime	YES	MSF	Difference of force from MSF average	Fair unless significantly different from MSF average. Below average - Poor, better than average - G/E.	YES	Difference from previous year	Significant change	
4b	Comparative risk of household crime	YES	MSF	Difference of force from MSF average		YES	Difference from previous year	Significant change	
5a	Domestic burglary rate	NO	MSF	Difference of log(force) from log(MSF average). Log transformation makes the distribution of scores more regular.	Set using z score of distribution of score for all forces, centred on performance equal to MSF average.	YES	Difference of log(current) from log(previous) performance. Log transformation means an equal % change gives an equal change in score, regardless of initial level of performance.	Apply z score for distribution of changes across all forces, centred around no change.	
5b	Violent crime rate	NO	MSF			YES			
5c	Robbery rate	NO	MSF			YES			
5d	Vehicle crime rate	NO	MSF			YES			
5e	Life threatening and gun crime rate	NO	MSF			YES			

Investigating Crime		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
6a	Offences brought to justice	NO				YES			Assessed as part of SPI 6b
6b	Percentage of offences brought to justice	NO	MSF	Difference of log(force) from log(MSF average)	Set using z score of distribution of score for all forces, centred on performance equal to MSF average.	YES	Difference of log(current) from log(previous) performance	Apply z score for distribution of changes across all forces, centred around no change.	
6c	Class A drug supply offences brought to justice	NO				YES			No assessment. Data directionally ambiguous
7a	Percentage of sanction detections	NO	MSF	Difference of log(force) from log(MSF average)	Set using z score of distribution of score for all forces, centred on performance equal to MSF average.	YES	Difference of log(current) from log(previous) performance	Apply z score for distribution of changes across all forces, centred around no change.	
7b	Burglary sanction detection rate								
7c	Violent crime sanction detection rate								

7d	Robbery sanction detection rate	NO	MSF	Difference of log(force) from log(MSF average)	Set using z score of distribution of score for all forces, centred on performance equal to MSF average.	YES	Difference of log(current) from log(previous) performance	Apply z score for distribution of changes across all forces, centred around no change.	
7e	Vehicle crime sanction detection rate								
8a	Domestic violence arrest rate								
8b	Domestic violence partner on partner percentage	NO							No assessment. Not clear what constitutes good performance and deleted for 2005/06.
Promoting Safety		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
9a	Road traffic safety	NO	MSF	Difference of log(force) from log(MSF average)	Set using z score of distribution of score for all forces, centred on performance equal to MSF average.	YES	Difference of log(current) from log(previous) performance	Apply z score for distribution of changes across all forces, centred around no change.	Using 2004 calendar year data

10a	Residents fear of crime - burglary/car crime/violence	YES	MSF	Difference of force from MSF average	Fair unless significantly different from MSF average. Below average - Poor, better than average - G/E.	YES	Difference from previous year	Significant Change	Each crime type (burglary, car crime, violence) to be assessed separately but assessments combined in the aggregation to give 10a as a whole no greater weight than other components
10b	Perceptions of anti-social behaviour								
Providing Assistance		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
11a	Front line policing	NO	National	Difference of force from National average	Performance better than 2007/08 target (72.5%) is excellent. Other boundaries based on z score of distribution for all forces.	YES	Difference from previous year	Apply z score for distribution of changes across all forces, centred around no change.	Awaiting final data

Citizen Focus		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
1a	Satisfaction with making contact	YES	MSF	Difference of force from MSF average	Poor if user satisfaction for force is less than MSF by 5 percentage points or more, Good if 2 percentage points higher, and Excellent if 5 percentage points higher. If the difference is not statistically significant then an assessment of Fair is made	NO			For each SPI, survey results are combined to give equal weight to each user group. MSF average is calculated by pooling results across all forces for each user group then combining by applying equal weights.
1b	Satisfaction with action taken								
1c	Satisfaction with progress updates								
1d	Satisfaction with treatment by staff								
1e	Satisfaction with overall service								
3a	Satisfaction of victims of racism	YES	MSF	Difference of force from MSF average	Significant difference from average	NO			

3b	Comparative satisfaction of minority ethnic groups	YES		Ratio of satisfaction (VME / White)	Test of proportions (for equality between white and VME) A difference at 99% level of significance is Poor. A difference at 95% level of significance is Fair, otherwise Good / Excellent is awarded	NO			Pooled sample of all user groups (data only available as pooled sample for 2004/05)
3c	Fairness of stop and search towards minority ethnic groups	NO		Ratio of percentages (BME / White)		YES	Reduction in difference from equality (in either direction)	Centred on no change, with move towards equality as an improvement	
3d	Investigating crimes against minorities proportionately and fairly	NO		Ratio of percentages (BME / White)		NO			
Resource Use		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
12a	Minority ethnic recruitment	NO		For each force: proportion in force/ proportion in EAP	Performance better than National Average: Fair. Performance better than Economically Active Population: Good Performance better than twice the EAP: Excellent	YES, but not in identical form	Change in ratio from previous year (using current EAP)	Apply z score for distribution of changes across all forces, centred around no change.	

12b	Minority ethnic resignations	NO				YES			Assessment not proposed for this SPI. Sample sizes are too small for meaningful assessment
12c	Female representation	NO	National	Difference of force from National average	Excellent if percentage of female officers greater than 28%. Other boundaries based on z score of distribution centred on performance equal to national average	YES	Difference from previous year	Apply z score for distribution of changes across all forces, centred around no change.	On RDS advice, data published as National Statistics in the 04-05 Police Strength Bulletin substituted in total for incomplete and inconsistent data on the ADR442
13a	Officer sick absence				At least good if 2004/05 target achieved (92 hrs/yr per officer, 96 hrs/yr for other staff). Other boundaries based on z				

13b	Staff sick absence				score of distribution centred on performance equal to national average				
Local Policing		Survey based	EGFP Comparator	Delivery score calculation	Grade boundaries	Previous year data available	Direction score calculation	Grade boundaries	Comments
2a	Residents perception of police performance	YES	MSF	Difference of force from MSF average	Fair unless significantly different from MSF average. Below average is Poor, better than average is G/E.	YES	Difference from previous year	Significant Change	