

METHOD FOR AGGREGATING GRADES

SUMMARY

This paper sets out the method used to aggregate component statutory performance indicator and baseline grades from Her Majesty's Inspectorate of Constabulary (HMIC) for the publication 'Police Performance Assessments 2004/05'. Assessments and further guidance is available from:

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

OVERVIEW OF ASSESSMENT

The performance of police forces in England and Wales is assessed in seven key areas: Reducing Crime, Investigating Crime, Promoting Safety, Providing Assistance, Citizen Focus, Resource Use and Local Policing.

In summary, two assessments are made in each of the seven areas, both of which are based on a combination of performance data and professional judgement. Each headline grade is an aggregate of other component grades. As such, a force with an excellent grade in a performance area will have many strengths but may have a few areas of relative weakness, likewise a force with a poor grade may have some areas of relative strength. Since assessments cover the period April 2004 – March 2005 they are not necessarily indicative of current performance.

When making assessments special rules are used which take account of:

- force compliance with national standards for recording crime;
- any instances where a force has not provided data to be assessed or where the data provided is incomplete / inaccurate in some way; and
- the national key priorities for policing set out in the National Policing Plan 2004-2007.

Further details on each area of assessment are provided at Appendix 1, the principles underpinning assessment are set out at Appendix 2 and a worked example is provided at Appendix 3. Finally, key points of clarification are listed at Appendix 4.

Assessment 1: Delivery

The first assessment concerns the performance delivered by a police force over the last year. Typically, this judgement is made by comparing the performance achieved by a force to that achieved by a group of similar forces (its peers). Of course, police forces are not identical but this 'like-for-like' comparison helps identify those forces which are performing better or worse relative to their peers. Forces delivering better performance are assessed as *excellent* or *good*, forces delivering performance similar to their peers are assessed as *fair*, and forces delivering performance worse than their peers are assessed as *poor*.

Assessment 2: Direction

The performance delivered by a force in any one year says much but it does not say if performance is getting better or worse over time. For this reason a second assessment is made on direction by comparing the performance achieved by a force in one year to that achieved by the same force in the previous year.

Forces performing much better than previously are assessed as *improved*, forces performing much the same are assessed as *stable* and forces performing much worse than previously are assessed as *deteriorated*.

Example

Take a force which is assessed as *good* and *improved* for investigating crime. This means the force is better than its peers (*good*) and is also better than it was one year earlier (*improved*). Each grade adds context to the other.

DETAILS OF ASSESSMENT

Step 1: ALLOCATE COMPONENTS TO PERFORMANCE AREAS

Individual statutory performance indicators (SPIs) and HMIC baselines form the basic components of assessment. All SPI and baseline assessments for 2004/05 are allocated to one of the seven performance areas but not all are assessed (see Appendix 1).

Step 2: GRADE INDIVIDUAL COMPONENTS

SPIs and baselines receive a grade for delivery (*excellent, good, fair, poor*) and, if possible, a grade for direction (*improved, stable, deteriorated*). A related paper sets out in detail the statistical methods for grading SPIs and HMIC has issued guidelines with respect to assessing baseline areas.

For 2004/05 four SPIs plus one baseline have not been assessed. The reasons are provided at Appendix 1: typically it is because data cannot be assessed robustly using a comparative method. Also, where an SPI or baseline is new (or substantially revised) in 2004/05 then it is not possible to compare performance to 2003/04 so no direction assessment is possible but a delivery assessment for 2004/05 is possible.

Step 3: ADDRESS MISSING, INCOMPLETE OR INACCURATE DATA

Rules are used which award or modify an SPI grade if underpinning data is missing, incomplete or inaccurate. Typically, this is because a force has either:

- not provided any data for the SPI; or
- provided data which is known to be incomplete or inaccurate.

In such cases it is not possible to calculate a robust grade for the SPI so a grade has to be awarded instead (otherwise there may be a perverse incentive not to supply data). The protocols used for awarding a grade are shown below.

Situation	Resolution
<p>1. A force does not provide any SPI data that is used in assessment.</p> <p><i>This excludes SPIs which, for one reason or another, are not subject to assessment in 2004/05 (see Appendix 1)</i></p>	<ul style="list-style-type: none"> • Delivery graded as poor and deteriorated but with consideration to capping grades to fair and stable depending on reason for lack of data. • Where data has not been returned for three or more SPIs then the grade for HMIC 'Performance Management' cannot be better than fair and stable (again with consideration to be given to grading poor and deteriorated).
<p>2. A force provides SPI data used in assessment which is known to be incomplete / incorrect.</p>	<ul style="list-style-type: none"> • <u>Generally</u>: A grade will be to be awarded on a case-by-case basis. Minor problems will be overlooked and the SPI assessed as normal. Major problems will be treated as a lack of data, as situation #1. • <u>Specifically</u>. One quarter missing - assess as normal (but to be capped at fair and stable for <u>2005/06</u>). Two quarters missing - assess as normal but capped at fair and stable. Three or four quarters missing - capped at poor and deteriorated.
<p>3. A force provides SPI data as required but it is not possible to calculate a robust assessment (eg too few data points, confidence intervals too wide)</p>	<ul style="list-style-type: none"> • The data may be graded fair and stable since it is not possible to calculate a robust grade. HMIC baseline can be used to provide context. • For user satisfaction data, where confidence intervals are greater than $\pm 4\%$, no action will be taken for 2004/05. Where the survey <u>method</u> is materially non-compliant a force may be downgraded one grade (assessed on a case-by-case basis).

In any event, published grades will be reviewed in the light of subsequent audit reports since assessments are made prior to statutory audit. A clear statement to this effect will be provided in publications and the website will indicate where SPI grades have been awarded rather than calculated. The rules related to data quality may be revised in subsequent assessments (ie from 2005/06).

Step 4: AGGREGATE COMPONENT GRADES

A numeric method is used to aggregate component grades in each area to produce an initial delivery grade and a direction grade. The initial grades are a 'stepping stone' to determining final grades.

A value is given to each component grade and the average point score across all the components in an area determines the overall grade, as set out in the two tables below. As noted earlier, in calculating the average point score, any new indicators where direction of travel cannot be assessed are excluded.

Grade	Component Value	Criteria for the domain grade
Excellent	3	If the average point score is more than 2 (but capped at <i>good</i> if one or more components are graded <i>poor</i>)
Good/Excellent*	2.5	n/a
Good	2	If the average point score is 1.5 or more but less than or equal to 2
Fair	1	If the average point score is 2/3 or more but less than 1.5
Poor	-1	If the average point score is less than 2/3

*The grade *good/excellent* is used for SPIs when the distinction between *good* and *excellent* is statistically difficult to determine (survey data and some proportionality data).

Grade	Component Value*	Criteria for the domain grade
Improved	1	If the average point score is more than $\frac{1}{4}$
Stable	0	If the average point score is between $\frac{1}{4}$ or less and minus $\frac{1}{4}$ ($-\frac{1}{4}$) or more
Deteriorated	-1	If the average point score is less than minus $\frac{1}{4}$ ($-\frac{1}{4}$)

*Since delivery grades (EGFP) and direction grades (ISD) are not combined it does not follow that 'improved' (value 1) is equivalent to 'fair' (where the value also 1).

To score SPI 10a (fear of crime), the individual point scores of its three components are averaged. This combines the assessments of its three sub-measures in a way which gives SPI 10a as a whole no greater weight than other components.

Step 5: REFLECT NATIONAL STANDARDS FOR CRIME RECORDING

Special rules are used which may modify the initial aggregate grades for reducing crime and investigating crime depending on force compliance with national crime recording standards. If compliance is a significant problem, indicated by an audit finding 'red' for data quality, then affected forces are precluded *excellent* and *improved* grades for these two domains (ie capped at *good* and *stable*). The situation will be reviewed for 2005/06 and, possibly, a stricter rule used.

Step 6: REFLECT NATIONAL PRIORITIES

The National Policing Plan 2004-2007 made a commitment to reflect national key priorities in assessment and the first year of this plan is the year being assessed (2004/05). Consequently, some components have been given 'priority' status to reflect the five key national priorities. These components are as follows:

- | | |
|--|-------------------------------|
| (i) SPI 1e: satisfaction with overall service | - in Citizen Focus; |
| (ii) SPI 10b: perceptions of anti-social behaviour | - in Promoting Safety; |
| (iii) HMIC baseline: reducing volume crime | - in Reducing Crime; |
| (iv) HMIC baseline: tackling level 2 criminality | - in Investigating Crime; and |
| (v) SPI 7a: overall sanction detection rate | - in Investigating Crime. |

The grade for a priority component can affect the aggregate grade for a performance area since a force cannot receive an aggregate grade more than one step above that of the lowest graded priority component in that performance area.

For example, a priority component graded *fair* would preclude the aggregate grade *excellent* (since *excellent* is more than one step above *fair*). The *fair* grade would not preclude the aggregate grade *good* (since *good* is not more than one step above *fair*).

EXAMPLE

A worked example for reducing crime is provided at Appendix 3.

**Police Standards Unit
Home Office**

APPENDIX 1

Domains, performance indicators and baselines

The performance of all 43 police forces in England and Wales is assessed in seven key performance areas (also known as domains):

1. REDUCING CRIME – This performance area focuses on the level of crime reported to the police directly and by the British Crime Survey (BCS).
2. INVESTIGATING CRIME – This performance area focuses on how crime is investigated reactively (eg robbery) and proactively (eg undercover work).
3. PROMOTING SAFETY – This performance area focuses on policing activity not linked directly to crime (eg reducing anti-social behaviour).
4. PROVIDING ASSISTANCE – This performance area focuses on general policing and responding to requests for support (eg front-line policing).
5. CITIZEN FOCUS – This performance area focuses on satisfaction and confidence in policing; fairness and equality plus community engagement.
6. RESOURCE USE – This performance area focuses on how a force manages itself as an organisation and the resources it has at its disposal.
7. LOCAL POLICING – For 2004/05 this assessment is focused on public confidence and neighbourhood policing. In due course this area will focus on local priorities for improvement.

COMPONENTS USED IN ASSESSMENT

Each domain is assessed using data from 'statutory performance indicators' (SPIs) plus judgements in 'baseline areas' made by HMIC. These components are listed in full below and those which are not assessed in 2004/05 are shown by strikethrough text and the reason given. Overall, there are 36 measures (32 of which are assessed in 2004/05) and there are 27 baselines (26 of which are assessed in 2004/05).

Please note that assessment publications use 'short titles' for SPIs and baselines which are more reader-friendly than the full technical descriptions provided below.

Points of clarification

Further details regarding the definitions and data used in 2004/05 assessments are provided at Appendix 4.

REDUCING CRIME**Crime level SPIs**

- 4 Using the British Crime Survey –
- (a) the risk of personal crime; and
 - (b) the risk of household crime.
- 5
- (a) Domestic burglaries per 1,000 households.
 - (b) Violent crime per 1,000 population.
 - (c) Robberies per 1,000 population.
 - (d) Vehicle crime per 1,000 population.
 - (e) Life threatening crime and gun crime per 1,000 population.

HMIC assessments

- A Reducing Hate Crime and Crimes against Vulnerable Victims
- B Volume Crime Reduction
- C Working in Partnership to Reduce Crime

INVESTIGATING CRIME**Offences brought to justice SPIs**

- 6
- (a) ~~Number of offences brought to justice. Not assessed since absolute value not rate~~
 - (b) Percentage of offences brought to justice.
 - (c) ~~Number of Class A drugs offences brought to justice per 10,000 population; of these the percentage each for cocaine and heroin supply. Not assessed since directionally ambiguous (ie cannot tell if a rise reflects good performance)~~

Detection SPIs

- 7
- (a) Percentage of notifiable offences resulting in a sanction detection.
 - (b) Percentage detected of domestic burglaries*
 - (c) Percentage detected of violent crime*
 - (d) Percentage detected of robberies*
 - (e) Percentage detected of vehicle crime*

*SPIs 7b - 7e are assessed using 'sanction detection' data not 'all detections' data.

Domestic violence SPIs

- 8
- (a) Percentage of domestic violence incidents with a power of arrest where an arrest was made related to the incident.
 - (b) ~~Of 8a, the percentage which involved partner-on-partner violence. Not assessed since absolute value not rate~~

HMIC assessments

- (A) Investigating Major and Serious Crime
- (B) Tackling Level 2 Criminality
- (C) Investigating Hate Crime and Crime against Vulnerable Victims
- (D) Volume Crime Investigation
- (E) Forensic Management
- (F) Criminal Justice Processes

PROMOTING SAFETY**Traffic SPIs**

- 9 (a) Number of road traffic collisions resulting in a person killed or seriously injured per 100 million vehicle kilometres travelled. **The 2004/05 assessment used the number of people killed or seriously injured (not the number of collisions) to align with ODPM and DfT*

Quality of life SPIs

- 10 Using the British Crime Survey –
 (a) fear of crime; *NB: the three sub-measures are aggregated to a single grade for the SPI*
 (b) perceptions of anti-social behaviour;

HMIC assessments

- (A) Reassurance
 (B) Reducing Anti-social Behaviour and Promoting Public Safety

PROVIDING ASSISTANCE**Frontline policing SPI**

- 11 (a) Percentage of police officer time spent on frontline duties.

HMIC assessments

- (A) Call Management
 (B) Providing Specialist Operational Support
 (C) Roads Policing

CITIZEN FOCUS**User satisfaction SPIs**

- 1 Satisfaction of victims of domestic burglary, violent crime, vehicle crime and road traffic collisions with respect to –
 (a) making contact with the police;
 (b) action taken by the police;
 (c) being kept informed of progress;
 (d) their treatment by staff;
 (e) the overall service provided.

Fairness, equality and diversity SPIs

- 3 (a) Satisfaction of victims of racist incidents with respect to the overall service provided.
 (b) Comparison of satisfaction for white users and users from minority ethnic groups with respect to the overall service provided.
 (c) Percentage of PACE searches which lead to arrest by ethnicity of the person searched.
 (d) Comparison of sanction detection rates for violence against the person offences by ethnicity of the victim.

HMIC assessments

- (A) Fairness and Equality
 (B) Customer Service and Accessibility
 (C) Professional Standards *Not assessed since subject to planned HMIC thematic*

RESOURCE USE

Human resource SPIs

- 12 (a) Proportion of police recruits from minority ethnic groups compared to the proportion of people from minority ethnic groups in the economically active population.
- (b) ~~Ratio of officers from minority ethnic groups resigning to white officer resignations. Not assessed since the number of resignations are too low to be assessed statistically~~
- (c) Percentage of female police officers compared to overall force strength.
- 13 (a) Average number of working hours lost per annum due to sickness per police officer.
- (b) Average number of working hours lost per annum due to sickness per police staff.

HMIC assessments

- (A) HR Management
- (B) Training and Development
- (C) Race and Diversity
- (D) Resource Management
- (E) Science and Technology Management
- (F) National Intelligence Model
- (G) Leadership
- (H) Strategic Management
- (I) Performance Management and Continuous Improvement

LOCAL POLICING

Confidence SPI

- 2 (a) Using the British Crime Survey, the percentage of people who think their local police do a good job.

HMIC assessments

- (A) Neighbourhood Policing and Community Engagement

APPENDIX 2

Principles of assessment

- a) THE METHOD SHOULD DELIVER A REASONABLE SPREAD OF GRADES WITHOUT INCENTIVISING FORCES TO DO EASY THINGS WELL
- This will also help identify forces performing well and those performing less well.
- b) THE FOLLOWING DEFINITIONS SHOULD BE USED FOR EGFP AND ISD GRADES
- *excellent* - if a force performs significantly better than its peers;
 - *good* - if a force performs better than its peers;
 - *fair* - if force performance is similar to its peers; and
 - *poor* - if a force performs much worse than its peers.

 - *improved* - if performance is much better than previously;
 - *stable* - if performance is much the same; and .
 - *deteriorated* - if performance is much worse than previously.
- c) THE METHOD SHOULD BE SIMPLE, ROBUST, CONSISTENT, OBJECTIVE AND FAIR.
- Simple: to facilitate transparency and increase the credibility of assessments (the principles of assessment should be understandable to a non-technical audience).
 - Robust: small changes in data should not cause major changes in grading.
 - Consistent: with policy statements.
 - Objective: clear rationale for use and application.
 - Fair: do not disadvantage one or more type of forces (other things being equal).
- d) THE SAME SET OF COMPONENT GRADES SHOULD DELIVER THE SAME AGGREGATE GRADE UNLESS THERE IS A CLEAR AND EXPLICIT REASON WHY THIS IS NOT THE CASE
- This will facilitate transparency and increase credibility of assessments. It for this reason that a two-stage process is used.
- e) THE INITIAL AGGREGATE GRADE RESULTING FROM A SET OF COMPONENTS CANNOT BE HIGHER THAN THE HIGHEST COMPONENT NOR LOWER THAN THE LOWEST COMPONENT UNLESS THERE IS A CLEAR AND EXPLICIT REASON WHY THIS IS NOT THE CASE
- This will help consistency in interpretation.

APPENDIX 3

Worked example for aggregating grades

This appendix presents an example of the rules used to aggregate individual component grades in a domain to produce a delivery grade and a direction grade for that domain. The example presents results for 'Northshire' using the aggregation method for reducing crime.

Step 1: ALLOCATE COMPONENTS TO DOMAINS

Reducing Crime focuses on performance with respect to crime directly reported to the police and crime reported by the British Crime Survey. The following SPIs and HMIC baselines are allocated to the domain.

REDUCING CRIME

Crime level SPIs

- 4 Using the British Crime Survey:
 - (a) the risk of personal crime; and
 - (b) the risk of household crime.

- 5
 - (a) Domestic burglaries per 1,000 households.
 - (b) Violent crime per 1,000 population.
 - (c) Robberies per 1,000 population.
 - (d) Vehicle crime per 1,000 population.
 - (e) Life threatening crime and gun crime per 1,000 population.

HMIC Baseline Assessments

- A Reducing Hate Crime and Crimes against Vulnerable Victims
- B Volume Crime Reduction
- C Working in Partnership to Reduce Crime

Step 2: GRADE INDIVIDUAL COMPONENTS

Each SPI and baseline area receives a grade for delivery (*excellent, good, fair, poor*) and, if possible, a grade for direction (*improved, stable, deteriorated*). A separate paper sets out in detail the statistical methods for grading SPIs and HMIC has issued guidelines with respect to assessing baseline areas.

For technical reasons SPI grades from survey results can only receive a delivery grade of either *good/excellent, fair* or *poor*. In this example this applies to SPIs 4a and 4b as they are taken from the British Crime Survey. The direction grade is unaffected.

The reducing crime results for Northshire are set out below. The names of components are provided as 'short titles' to improve readability compared to the 'long titles' used in legislation and technical guidance.

NORTHSHIRE		
Performance indicators	Delivery	Direction
4a. Comparative risk of personal crime	Fair	Improved
4b. Comparative risk of household crime	Good / Excellent	Improved
5a. Domestic burglary rate	Fair	Deteriorated
5b. Violent crime rate	Good	Stable
5c. Robbery rate	Poor	Deteriorated
5d. Vehicle crime rate	Excellent	Stable
5e. Life threatening and gun crime rate	Fair*	Stable*
Baseline assessments	Delivery	Direction
Reducing hate crime and vulnerable victimisation	Fair	Improved
Reducing volume crime	Good	Deteriorated
Working in partnership	Excellent	Stable

*Capped due to lack of data

Step 3: ADDRESS MISSING, INCOMPLETE OR INACCURATE DATA

Rules are used which award or modify an SPI grade if underpinning data is missing, incomplete or inaccurate as set out in the main paper.

In this example let us assume Northshire did not supply two quarters data for SPI 5e. In this case the grades for that component would have been capped at *fair* and *stable*. This is indicated by * and footnote in the table above.

Step 4: AGGREGATE COMPONENT GRADES

A numeric method is used to aggregate component grades in each domain to produce an initial delivery grade and a direction grade for that domain. The initial grades are a 'stepping stone' to determining final grades.

Each component grade is assigned a value as set out in the main paper. The values for each reducing crime component for Northshire force are as follows.

NORTHSHIRE				
Component	Delivery	Value	Direction	Value
4a. Comparative risk of personal crime	Fair	1	Improved	1
4b. Comparative risk of household crime	Good / Excellent	2.5	Improved	1
5a. Domestic burglary rate	Fair	1	Deteriorated	-1
5b. Violent crime rate	Good	2	Stable	0
5c. Robbery rate	Poor	-1	Deteriorated	-1
5d. Vehicle crime rate	Excellent	3	Stable	0
5e. Life threatening and gun crime rate	Fair	1	Stable	0
Reducing hate crime and vulnerable victimisation	Fair	1	Improved	1
Reducing volume crime	Good	2	Deteriorated	-1
Working in partnership	Excellent	3	Stable	0
	Total	15.5	Total	0
	Average	1.55	Average	0

In this example, since there are ten component grades, the average value for delivery is equal to **1.55** (ie 15.5 / 10) and the average value for direction is **0** (ie 0 / 10). Therefore:

- the initial delivery grade is *good* since (as set out in the main paper) the average value of 1.55 is more than 1.5 but less than or equal to 2; and
- the initial direction grade is *stable* since (as set out in the main paper) the average value of 0 is between $\frac{1}{4}$ or less and $-\frac{1}{4}$ or more.

Step 5: REFLECT NATIONAL STANDARDS FOR CRIME RECORDING

Special rules are used which may modify the initial aggregate grades for reducing crime and investigating crime depending on compliance with national crime recording standards. Forces rated 'red' for data quality are precluded *excellent* and *improved* grades (ie capped at *good* and *stable*).

In this example we have assumed that Northshire received 'amber' for the data quality element of audit. As an 'amber' force, the capping rule does not apply for 2004/05 but stricter data quality rules may be used in subsequent assessments (eg from 2005/06).

Step 6: REFLECT NATIONAL PRIORITIES

In line with the National Policing Plan 2004-2007 a number of components have been given 'priority status' to reflect the five key priorities as set out in the plan (see main paper). Priority status means that a force cannot receive an aggregate domain grade more than one step above that of the lowest graded priority component in that domain.

One component in the reducing crime domain is a 'priority component' - *HMIC baseline: reducing volume crime* - so the grade for this component may affect the aggregate domain grade for Northshire.

Northshire's grades for this priority component are *good* and *deteriorated* and its initial aggregate grades for reducing crime are *good* and *stable* (see Step 4).

Neither of these aggregate grades are more than one step above that of the priority component. Therefore, the domain grades for Northshire are not affected by this rule and its final published grades for reducing crime will be *good* and *stable*.

If Northshire's baseline assessment for 'reducing volume crime' had been *poor* then the initial delivery grade *good* would have been downgraded (capped) to *fair* since *good* is more than one step above that of the lowest graded priority component. Its direction grade *stable* would have been unaffected.

APPENDIX 4

Points of clarification with respect to data and assessment for 2004/05

All SPIs

Short titles are used in publications to aid understanding rather than full statutory titles: no other change is intended by re-naming. SPIs will be subject to review as part of a lessons learned exercise. Data provided by forces was, where applicable, used on the understanding it complies with SPI technical guidelines and has been / will be treated by the force and its police authority as final SPI data subject to audit. Published data are not 'official national statistics' within the meaning of that term.

SPI 1a – 1e (user satisfaction)

The assessments use data weighted by user group. Locally published values which use unweighted data will therefore be different from those used in assessment.

SPI 2a (local confidence)

The SPI (along with HMIC baseline 'Neighbourhood Policing') is allocated to the Local Policing domain pending the introduction of a new method of assessing the local domain (wef 2006/07 performance).

SPI 3a (satisfaction of victims of racism)

No comment.

SPI 3b (satisfaction between different ethnic groups)

Data not weighed by user group since only pooled data available,

SPI 3c (stop/search)

The short title has been changed to better reflect the measure: '*parity of arrests from stop and search between ethnic groups*'.

The measure is based on the number of people arrested following 'stop/search' based on the self-defined ethnicity of the person stopped. Data from stops and subsequent arrests of people who choose not to state their ethnicity are excluded from assessment. Forces with a high 'not stated' rate may wish to examine the reason for this.

The statistical assessment is based on the hypothesis that the likelihood of arrest should be the same regardless of the ethnicity of the person stopped. In other words, the likelihood of arrest for a white person stopped should be the same as the likelihood of arrest for people from other ethnic groups.

The grades awarded should be interpreted as follows:

- **Poor:** for 2004/05, there is strong statistical evidence to suggest that the likelihood of arrest depends on the ethnicity of the person stopped. (*Chi-square¹ test significant at the 99% level*)
- **Fair:** for 2004/05 there is statistical evidence to suggest that the likelihood of arrest depends on the ethnicity of the person stopped. (*Chi-square test significant at the 95% level*)
- **Good/Excellent:** for 2004/05 there is no statistical evidence to suggest that the likelihood of arrest depends on the ethnicity of the person stopped. (*Chi-square test not significant at the 95% level*).

The assessment and data do not say why any difference in the likelihood of arrest exists, nor does the assessment depend on which arrest rate is the higher (white or VME). The assessments *fair* and *poor* can be used to say that there is evidence of a different likelihood of arrest when, in theory, there ought to be no difference.

SPI 3d (detections by ethnicity)

The short title has been changed to better reflect the measure: '*parity of detections for violent crime between ethnic groups*'.

The measure is based on the number of detected offences for violent crime based on the self-defined ethnicity of the victim. Data from victims who choose not to state their ethnicity are excluded from assessment. Forces with high a 'not stated' rate may wish to examine the reason for this.

The statistical assessment is based on the hypothesis that the likelihood of detection should be the same regardless of the ethnicity of the victim. In other words, the likelihood of detection where the victim is white should be the same as the likelihood of detection for victims from other ethnic groups.

The grades awarded should be interpreted as follows:

- **Poor:** for 2004/05, there is strong statistical evidence to suggest that the likelihood of detection for violent crime depends on the ethnicity of the victim. (*Chi-square test significant at the 99% level*)
- **Fair:** for 2004/05 there is statistical evidence to suggest that the likelihood of detection for violent crime depends on the ethnicity of the victim. (*Chi-square test significant at the 95% level*)

¹ 'Chi-square' is the name of the analytical method used to decide if the difference in rates is 'statistically significant'.

- Good/Excellent: for 2004/05 there is no statistical evidence to suggest that the likelihood of detection for violent crime depends on the ethnicity of the victim. (*Chi-square test not significant at the 95% level*).

The assessment and data do not say why any difference in the likelihood of detection exists, nor does the assessment depend on which detection rate is the higher (white or VME). The assessments *fair* and *poor* can be used to say that there is evidence of a different likelihood of detection when, in theory, there ought to be no difference.

SPI 4a – 4b (BCS crime)

The assessments use a 95% confidence limit and compare force data to its MSF average (not the national average as used by published national crime statistics).

SPI 5a – 5e (recorded crime)

Assessments are based on mid-2003 estimates for population and residential data. The data for SPI 5e no longer double-counts gun crime and life-threatening crime. Following release of draft assessments to forces RDS issued updated data which was used in final assessments.

SPI 6a - 6b (OBTJ)

SPI 6a is not assessed but data will be provided on the website. SPI 6b is assessed.

SPI 6c (Class A drug supply)

SPI 6c is not assessed but data will be provided on the website.

SPI 7b – 7e (detections)

The data used in assessment is not the same as the SPIs: the assessments use sanction detection data not 'all detection' data. 'All detection' data will be provided on the website.

SPI 8a - 8b (domestic violence)

SPI 8a is assessed. SPI 8b is not assessed but data will be provided on the website. In some cases forces provided rates which could not be used (since base data needed).

SPI 9a (road traffic collisions)

The data used in assessment is not the same as the SPI. This is defined in 2004/05 as the number of collisions leading to someone killed / seriously injured (KSI). In 2005/06 the definition changed to the number of KSIs. Given that KSI data is available for 2004/05 it was decided to make the assessment using the 2005/06 definition to bring it in line with a shared local authority indicator. In other words, the 2004/05 assessment uses the number of KSIs not the number of collisions leading to a KSI. Calendar year data (both number and denominator) is used in line with DfT published statistics.

SPI 10a – 10b (perception measures)

SPI 10a is based on three elements. Each of these is graded with the 'sub-grades' being aggregated to establish a grade of the SPI overall. Each sub-grade therefore contributes one-third of the assessment for SPI 10a. Finally, SPI 10b used data based on the 'seven-strand' definition of anti-social behaviour not the 'five-strand' definition.

SPI 11a (front-line policing)

The layout of the spreadsheet issued with draft assessments was inadvertently confusing: the grading boundaries applied to the measure, not the ratio compared to the national average. The data used to make final assessments was as advised by the Home Office FLP team. This data reflected a policy decision that a retrospective re-allocation of officers to HMIC codes was not eligible for inclusion. The national target for 2007/08 was used to help establish grade boundaries for 2004/05 assessments.

SPI 12a (minority ethnic recruits)

Transferees are included in the definition for 2005/06, not for 2004/05.

SPI 12b (minority ethnic resignations)

SPI 12b is not assessed but data will be provided on the website.

SPI 13a (police officer sick absence)

Where forces provided revised hours (in line with assessment protocols) then this data was used. Rates could not be used since the denominator value was not stated.

Revised denominator: the number of police officers based on published headcount data for 2003/04 and 2004/05 (Home Office Statistical Bulletins). An average was taken based on the two end year positions for police ranks.

SPI 13b (police staff sick absence)

Where forces provided revised hours (in line with assessment protocols) then this data was used. Rates could not be used since the denominator value was not stated.

Revised denominator: the number of police staff based on headcount data for 2003/04 and 2004/05 (Home Office Statistical Bulletins). An average was taken based on the two end year positions for the sum of police staff, PCSOs, traffic wardens & designated officers.