Ministry of Justice

Methodology review report: Statistical reporting from a diversity perspective

George Papadopoulos, Georgina Eaton and Aidan Mews
Statistical Project Delivery Team, Justice Statistics Analytical Services

Contents

Introduction

This review considers Ministry of Justice (MoJ) Data and Analytical Services Division (DASD) / Judicial Appointments Commission (JAC) statistical reporting from a diversity perspective. It accompanies the move to combine the Judicial Office (JO) and JAC statistical publications and to include relevant data from the legal professions from September 2020. These recommendations are intended for usage within this new combined statistics report and the JAC-commissioned 'deep dive' (which looks to explain interactions between key characteristics and the application process). In addition, the review also makes recommendations for MoJ DASD statistical reports more generally, with consideration focused on the following ones that currently contain the Relative Rate Index (RRI)[1]:

- Race and the Criminal Justice System Statistics: https://www.gov.uk/government/collections/race-and-the-criminal-justice-system
- Women and the Criminal Justice System Statistics: https://www.gov.uk/government/collections/women-and-the-criminal-justice-system
- Black, Asian and Minority Ethnic disproportionality in the Criminal Justice System in England and Wales: https://www.gov.uk/government/publications/black-asian-and-minority-ethnic-disproportionality-in-the-criminal-justice-system-in-england-and-wales
- HMPPS workforce statistics (the "Recruitment diversity experimental statistics" report and tables are listed below the main bulletin report and tables): https://www.gov.uk/government/collections/national-offender-management-service-workforce-statistics
- Youth justice statistics (Youth Justice Board): https://www.gov.uk/government/collections/youth-justice-statistics
- The journey of the child (Youth Justice Board): https://www.gov.uk/government/news/the-journey-of-the-child

The focus of the review is on:

- the current use of the RRI and alternative analytical approaches (including more nuanced ones e.g. to tweak the present approach to address any anomalies caused by it) and their relationship to other measures used;
- how best to communicate the story in such statistical publications

It has been led by the Statistical Project Delivery Team within Justice Statistics Analytical Services, reporting to the MoJ Chief Statistician. A wide range of stakeholders have been consulted including JAC personnel, users of the JAC statistical reports, methodology experts, policy makers and authors of other MoJ DASD statistical outputs that include the use of the RRI. The full list of consultees is provided in Appendix A.

---

[1] The RRI provides a comparison of the success rate for one (typically minority or under-represented) group with that of another (typically majority or over-represented) group.

To aid clarity, the content of this report has been arranged in five sections. The first four sections (A-D) list the issues and corresponding recommendations relating to the JAC statistical outputs. Section E provides recommendations relating to MoJ DASD statistical outputs more generally. The sections are as follows:

A: Coverage of the JAC statistical outputs
B: Key statistical indicators within JAC statistical outputs
C: Accompanying metrics/rules within JAC statistical outputs
D: Presentational features within JAC statistical outputs
E: MoJ DASD statistical reports recommendations

In addition, for ease of use all recommendations relating to the JAC statistical outputs are listed in Appendix B.

A: Coverage of the JAC statistical outputs

Issues

1. As a key duty of the JAC is to encourage a diverse range of applicants, it is vital from a policy perspective to consider applications, shortlisting and recommendations within the context of the eligible pool. The eligible pool is the pool of people who meet minimum eligibility requirements for the roles being recruited for. While the eligible pool is often considered, the most recent JAC statistics publication in June 2019[2] places a disproportionately large focus on the application stage onwards (applications, shortlisting, recommendations for appointment). For example, the RRIs are only shown from application to recommendation.

2. Where statistics include coverage of the eligible pool, they may be less informative than they could be because the eligible pool may include people who meet minimum eligibility requirements but have little chance of applying successfully.

3. The most recent JAC statistics publication in June 2019 covers each diversity characteristic separately but not the intersectionality between the characteristics (e.g. ethnicity and gender).

4. The new combined JAC statistics publication could provide insight into emerging trends over time (e.g. where there are no statistically significant disparities in any single year but the trend is moving closer to or away from parity of outcomes).

Recommendations

1. We recommend that applications, shortlisting and recommendations should be presented within the context of the eligible pool (of potential applicants who meet the minimum eligibility criteria). This is to ensure we cover the full scope of the JAC's work and for the statistics to be as relevant as possible.

---

[2] See https://www.judicialappointments.gov.uk/jac-official-statistics

2. Where eligible pool figures are not available:

    a. clear and appropriate warnings should be included within the report that care should be taken when drawing conclusions based on partial coverage of the statistics.
    b. further work should be undertaken to enable the availability of eligible pool data.

3. Further analysis should be considered to ascertain whether the eligible pool includes people who meet minimum eligibility requirements but have little or no chance of applying successfully, and if so the impact(s) of this. For example, it is possible that in practice some of those meeting minimum eligibility requirements do not have enough experience in the field to be recommended for appointment. Further analysis could determine this and potentially produce statistics after adjusting for years in the field. In addition, further analysis of more detailed or robust data from the professional legal bodies could provide greater insight into why certain eligible individuals may not be applying as well as any potential 'blockers' further on in the recruitment process. Such further analyses would seem to fit initially within future 'deep dive' work but with a view to inclusion within the new combined JAC statistical publication at a later point.

4. Including analysis of intersectionality in the JAC statistics (e.g. looking at the combination of ethnicity and gender) will greatly benefit the overall storytelling of the publication as it is pivotal to enhancing understanding of where there may be disparities. Additional intersectionality analyses should be considered along with providing indications of emerging trends where there are no statistically significant disparities in any single year, and where such disparities would become statistically significant if a few years data were rolled together. As with the analyses in 3. above, such analyses initially fit within 'deep dive' work but should be with a view to inclusion within the new combined JAC statistical publication at a later point.

B: Key statistical indicators within JAC statistical outputs

Issues

It is worth considering which generic key statistical indicators (ignoring the specific way these have been calculated in the JAC statistical reports up to now) are the most appropriate ones to use. The most recent JAC statistics publication in June 2019 included the following key indicators:

- Representation percentages - the percentages of those at a particular stage (e.g. application) who declare specific characteristics (e.g. the percentage of applicants who are BAME).
- Success rates - the proportions with a specified declared characteristic that then go on to reach a further stage in the selection process (e.g. the proportion of BAME applicants that are recommended for appointment).

- Relative Rate Index (RRI) - this provides a comparison of the success rates. It is calculated by dividing the success rate for one (typically minority or under-represented) group by the success rate for another (typically majority or over-represented) group. An RRI of one would indicate the success rates are the same.

The further deep dive analysis that aims to explain interactions between key characteristics and the application process also makes use of the RRI. An alternative would have been to use the odds ratio which is similar to the RRI but compares the success rates in terms of odds rather than proportions.

Recommendations

1. To enable informative statistical reporting of JAC appointments from a diversity perspective, it is necessary to provide meaningful comparisons of success rates for the different diversity strands (e.g. for females compared to males, BAME compared to white etc.). There are two main methods of doing this:

   - The absolute difference between the success rates e.g. the proportion of the female eligible pool who apply *minus* the proportion of the male eligible pool who apply.
   - The relative difference between the success rates (or RRI) e.g. the proportion of the female eligible pool who apply *divided* by the proportion of the male eligible pool who apply.

Of the two, the relative difference between the success rates (or RRI) is by far the more informative. To demonstrate this:

   - if 10% of the female eligible pool and 5% of the male eligible pool make applications the key summary point is that females applied at twice the rate of males. This is shown by the relative difference which is 2 rather than the absolute difference which is 5% points.
   - if 50% of the female eligible pool and 45% of the male eligible pool make applications the key summary point is that females are slightly (11%) more likely to apply than males. This is shown by the relative difference which is 1.11 rather than the absolute difference which is 5% points.

The relative difference between success rates (or RRI) is used elsewhere as an indicator of disparities and in the context of recruitment[3]. Given this background, and in the context of the observations and recommendations made in this report,

---

[3] For example NHS workforce Race Equality Standard 2018 Data Analysis Report for NHS Trusts (https://www.england.nhs.uk/publication/workforce-race-equality-standard-data-reporting-december-2018/), U.S. Equal Employment Opportunity Commission guidance document (see sections 11 and 12 of https://www.eeoc.gov/policy/docs/qanda_clarify_procedures.html), 'A test for racial discrimination in recruitment practice in British cities' produced by the National Centre for Social Research on behalf of the Department for Work and Pensions (http://www.bollettinoadapt.it/old/files/document/3626ATESTFORRACIALDI.pdf), 'Are employers in Britain discriminating against ethnic minorities?' produced by the Growth, Equal Opportunities, Migration and Markets project (http://csi.nuff.ox.ac.uk/wp-content/uploads/2019/01/Are-employers-in-Britain-discriminating-against-ethnic-minorities_final.pdf)

this measure can be used to enable the statistical reporting of JAC appointments to be more informative.

The relative difference between the success rates (or RRI) is therefore recommended as the key summary indicator to flag up potential disparities between those with different diversity characteristics.

2. As mentioned in section A "Coverage of the JAC statistics outputs" it is necessary for applications, shortlisting and recommendations to be presented within the context of the eligible pool. It is therefore recommended that the key RRI findings in the new combined JAC statistical publication must at minimum include coverage of the RRI from the eligible pool (denominator) to recommendation (numerator) and where disparities are found (that are of practical and statistical significance - see section C) provide further information about where in the interim stages of the process disparities exist. Consultation with stakeholders also revealed a desire for the key RRI findings to by default include the RRI from application to recommendation. This could be included perhaps along with the complementary RRI from eligible pool to application.

3. The statistical tables in the new combined JAC publication should include all RRIs mentioned in 2. above; from eligible pool to recommendation, application to recommendation, and for each interim stage in the selection process (eligible pool to application, application to shortlisting, and shortlisting to recommendation).

4. To aid interpretation of the relative difference between the success rates (or RRI) as the key summary indicator, some underlying statistical context could be provided by showing success rates on which the RRI is based; and if user consultation deems necessary the representation percentages at each stage.

5. While statistical properties of the odds ratio enable it to be an easy to use metric for some deep dive analyses[4], it is much more difficult to understand than the RRI and therefore much easier to misinterpret. Any usage should therefore be accompanied by an illustration about what a particular odds ratio value means in practice.

C: Accompanying metrics/rules within JAC statistical outputs

Issues

Consultation discussions focused around the following mechanisms that could be used to supplement the key statistical indicators:

- The 4/5ths rule – the benefit of this is in flagging up whether an RRI value has practical as opposed to statistical significance. Only if an RRI value is outside the tolerance zone between 0.8 and 1.25 is there considered to be a disparity of outcomes from a practical (or material) perspective.

---

[4] For example, those involving the use of logistic regression.

- 95% confidence interval and statistical significance – the greater the statistical variability around an RRI value the less certain we can be about it. Only if the confidence interval excludes the value of one (or parity) does the RRI indicate a statistically significant disparity of outcomes.
- Indication of potential statistical bias due to non-reporting of diversity characteristics – the JAC statistical publication in 2019 included an indication of the scope for statistical bias as provided by the declaration rate[5] and the related number of unknowns[6], although this is only done at aggregated level where exercises are grouped together. It could also include some quantification of the impact of the potential statistical bias on success rates / the RRI.
- The declaration rate threshold[5] – at present key statistical indicators are only reported in the JAC statistical publication for diversity characteristics which are declared by more than 60% of applicants.
- Disclosure control – in the JAC statistics exercises with fewer than 10 recommendations are currently aggregated into groups of exercises so applicants cannot be personally identified. There was some user concern about the relevance of these groupings. In addition, some suppression takes place due to statistical volatility rather than confidentiality concerns. Percentages are suppressed where the denominator is less than 10 before excluding any unknown values, and RRIs are suppressed where either the numerator or denominator percentages are suppressed or zero.

Recommendations

1. The 4/5ths rule to indicate whether an RRI value is sufficiently far from one (or parity) to reflect a disparity of outcomes is considered very useful from a policy perspective. Although it may originally have been developed as a quick and approximate guide to determine whether a disparity is statistically significant, its usefulness in the JAC statistical publication is to provide an indication of what is significant from a practical point of view. The zone of tolerance lower and upper bounds of 0.8 (4/5) and 1.25 (5/4) are currently the de facto 'industry standard', with the 4/5ths rule being used elsewhere in a recruitment context (along with the RRI)[3] as well as in a criminal justice context. Moreover, there is an analogy with the 'range of equivalence' (also from 0.8 to 1.25) used by regulatory bodies in drug trials[7,8,9]. Given this review did not uncover any evidence that suggested the bounds should either be increased or decreased, it is recommended that the tolerance zone remains unchanged for the new combined JAC statistical

---

[5] Declaration rates measure the proportion that declare a particular characteristic (e.g. the proportion of applicants that declared their religion or belief), with those selecting any 'prefer not to say' or 'prefer not to share data outside the JAC' option counting as unknown for this purpose.

[6] The number of unknowns included those selecting any 'prefer not to say' or 'prefer not to share data outside the JAC' option.

[7] https://en.wikipedia.org/wiki/Bioequivalence

[8] https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/bioequivalence

[9] https://www.fda.gov/media/89135/download

publication until any conclusive evidence is uncovered that suggests otherwise. Consideration should be given to reviewing this at some point in the future.

2. It is important for the JAC statistical publication to flag up any disparities that can be considered statistically significant[10]. Therefore, for a disparity to be regarded as a high priority for further action, we recommend it is both statistically significant and outside the tolerance zone (presently 0.8 to 1.25) – see 1. above. A suggested guide to interpretation, which could be finessed with user testing, is as follows:

    a. Inside the tolerance zone (presently 0.8 to 1.25) – this signifies no practical disparity
    b. Outside the tolerance zone (presently 0.8 to 1.25) but not statistically significant – this implies that the disparity could well be due to chance
    c. Outside the tolerance zone (presently 0.8 to 1.25) and statistically significant – this implies that we can have confidence that a disparity exists and that our best estimate suggests that it is of a size that is important practically.

3. Some indication of the potential impact of statistical bias due to non-reporting of diversity characteristics should be provided in the JAC statistical publication. While further work could be done to quantify the impact of any statistical bias on key metrics, the estimation of upper and lower-case scenarios tends to provide somewhat speculative worst-case scenarios that tend to be unhelpful in practice. The total number, declaration rate and the related number of unknowns should continue to be published. Where possible these should be accompanied by a breakdown of the number of unknowns to understand the extent to which people are choosing any 'prefer not to say' and 'prefer not to share data outside the JAC' options and so are engaging with the monitoring process rather than simply not responding. Where such options exist but the related data are unavailable, further work should be undertaken to try and obtain these.

4. Where exercises are grouped together in the JAC statistical publication it would be helpful for the total numbers, declaration rates and numbers of unknowns (where possible with the additional breakdowns described in recommendation 3. above) to be provided separately at the application stage (instead of only at aggregated grouped level) for each exercise involving ten or more applications.

5. The potential for statistical bias is best dealt with at source. While declaration rates are consistently high overall[11], they vary by characteristic. Only one declaration rate in the 2019 JAC statistics publication tables was less than 70% and this was also less than 60%. The declaration rate threshold could therefore be raised from 60% to 70% or even 80% to encourage even better reporting

---

[10] Utilising standard statistical methodology to take account of the sample size and the volatility present in the data.
[11] Being 80% or more at the application stage for gender, ethnicity, professional background, disability, social mobility, sexual orientation and religion.

which will in turn lead to more accurate statistics. The level of this threshold should be reviewed every few years.

6. Exercises with fewer than ten recommendations should continue to be aggregated into groupings for confidentiality reasons so applicants cannot be personally identified.

7. Where exercises are aggregated and the eligible pools are known, an overall group eligible pool could be estimated by summing the eligible pools for each exercise. In decision making about which exercises to group together, consideration should be given to the relevance of groupings from both a policy and statistical perspective. User consultation could determine whether the groupings can be more relevant in future. Statistical anomalies could be avoided by checking the trend in the aggregated exercises is consistent with trends seen in the individual exercises.

8. Unless there are overriding confidentiality, presentational, or statistical concerns, there is no need for figures to be suppressed. Confidentiality concerns are addressed by 6. above. Together with the high declaration rates, the provision of confidence intervals (or indications of statistical significance) enables users to gain a good understanding of the uncertainty around key statistical indicators. While the calculation of standard confidence intervals[12] needs careful modification for small groups, an alternative approach is recommended that renders such modifications unnecessary[13]. Any presentational concerns arising from small groups leading to particularly high RRIs should be monitored and dealt with on a case by case basis until such a time that a useful rule can be put in place. If the current practice of suppressing figures where denominators are less than 10 is continued, it should be applied after first excluding any unknown values.

D: Presentational features within JAC statistical outputs

Issues

The main presentational issues relating to the 2019 JAC statistical publication as identified from the consultations were:

1. The main points summary at the start of the publication was not found to be user friendly due to the large amount of information and statistics.

2. The publication was found to be large and not easily navigable.

3. It could be easier for policy colleagues to find headline data for use in briefings.

---

[12] Those using a Normal Distribution approximation.
[13] Using 'bootstrapping'

4. Statistics and plots are not accompanied by sufficient explanation so that they can be understood without users having to refer to another part of the publication such as the methods of reporting section. An example of where this has been achieved is the report "BAME disproportionality in the Criminal Justice System"[14]. While providing explanations elsewhere in the document is helpful, many users will not in practice refer to them.

5. Presentation of the figures/plots could be improved.

6. Lay users found the guidance about the RRI difficult to understand, and some of the terminology is potentially misleading e.g. it is not clear whether the 4/5ths rule is about practical or statistical significance, and the declaration rate about unknowns or declarations. Where there are findings that appear contradictory (e.g. the results for aggregated groups seem to be inconsistent with those for the groups individually) there is a need for clear explanation.

7. Caveats (e.g. that the disparities may be explained by differences in other key factors) could be clearer and more strategically positioned.

Recommendations

The following presentational suggestions for the new combined JAC publication should be finessed by user consultation and consideration of the GSS Communicating quality, uncertainty and change guidance[15]. In particular:

1. For the summary section to be user friendly we advise it has a similar style to other Ministry of Justice publications[16] which use a bullet point structure with visual aids such as arrows and signs. The chosen approach should enable the publication to be easily readable with clear key messages. The summary of main points should ideally be one page and should contain one main point per section.

2. To enable the publication to be as navigable as possible we recommend including a table of contents. Efforts should be made to keep the report as concise as possible without any unnecessary graphs. Contextual information should be provided in an accompanying statistics definitions and measurements document and additional information provided in the appendices.

3. The key messages should be sufficiently well explained so that lay users don't have to refer to another part of the publication. This includes key points from plots and summary statistics relating to the RRI e.g. 'the following disparities were significant both practically (being outside the tolerance zone of no disparity) and

---

[14] See page 6 of:
https://www.gov.uk/government/publications/black-asian-and-minority-ethnic-disproportionality-in-the-criminal-justice-system-in-england-and-wales
[15] https://gss.civilservice.gov.uk/policy-store/communicating-quality-uncertainty-and-change/
[16] For example, the summary section of Criminal Justice System statistics can be viewed in:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/867102/criminal-justice-statistics-quarterly-september-2019.pdf

statistically: females in the eligible pool were 43% less likely to be recommended than males in the eligible pool with them having a success rate of 20% compared to 34% for males'.

4. All key messages should also include clear and well positioned caveats to avoid the potential for misleading conclusions to be drawn. For instance:

   a. In the summary that disparities may be explained by differences in other factors.
   b. where no eligible pool figures are available that care should be taken when drawing conclusions based on partial coverage of the statistics.
   c. where annual comparisons need to be approached with caution due to changes in exercises over the years.

5. The key headline data should be easy for publication users to find and use. This could be achieved by including tables showing the key RRI statistics for each diversity strand.

6. Presentation of the figures/plots should be reviewed. For instance, forest plots[17] with accessible colouring[18] could be used to show the RRIs accompanied by 95% confidence intervals and the zone of tolerance. An example can be seen below in Figure 1. Any (horizontal) bar charts could include the actual percentage figures along with accessibly coloured bars.

---

[17] https://en.wikipedia.org/wiki/Forest_plot
[18] The GSS "Effective tables and charts in official statistics" document provides guidance on the use of colours in charts: https://gss.civilservice.gov.uk/wp-content/uploads/2018/03/Effective-Tables-and-Charts-in-official-statistics-Edition-2.1-February-2018-4.pdf

Figure 1: Example forest plot with illustrative data showing the RRIs accompanied by 95% confidence intervals and the zone of tolerance. Where in light (rather than dark) blue, the disparity shown by the RRI is significant both practically (being outside the tolerance zone) and statistically (the 95% confidence interval not including one).
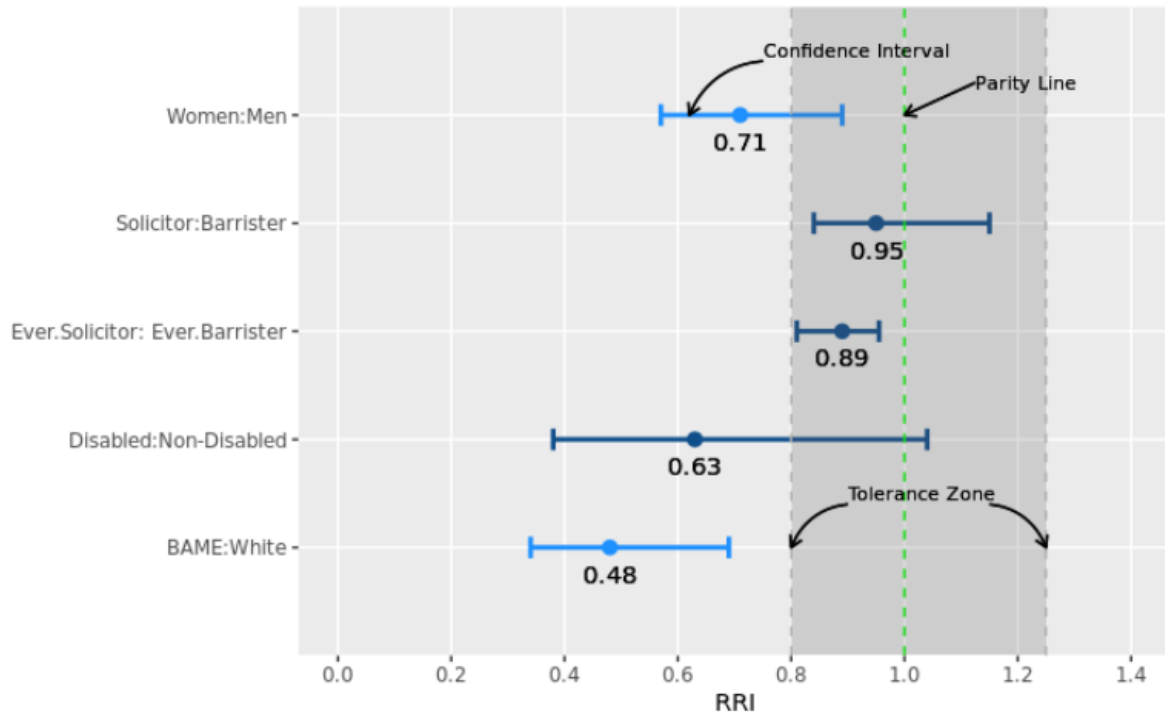
Interpretation:
- The RRI of 0.71 for 'Women:Men' shows that the success rate for women is 0.71 times the success rate for men (or to put another way 29% lower for women than men).
- As the accompanying 95% confidence interval does not include the value of one, we can be confident there is a disparity. As the RRI is outside the tolerance zone it is statistically and practically significant.

7. Explanations of the RRI and other technical terms should be reviewed to ensure they are understood by lay readers and that the terminology is clear and accurate. In particular the 4/5ths rule (or tolerance zone) should be described as a measure of practical rather than statistical significance (see section C recommendation 1) while the declaration rate is about declarations rather than unknowns. Consideration should be given to moving these sections to an appendix with concise wording developed for the main part of the report (an example of the latter would see a bullet point or two following an RRI graph which summarises what it is telling you – see example in section D recommendation 6. above).

E: MoJ DASD statistical report recommendations

Coverage

1. To be as relevant as possible it is necessary for the statistics to provide full coverage of the process from start[19] to end[20] and for the start population to take as full account of any eligibility criteria as possible. Where important milestone data (in particular at the start and end of the process) are unavailable:

    *   clear and appropriate warnings should be included that care should be taken when drawing conclusions based on partial coverage of the statistics.
    *   where feasible, further work should be undertaken to enable the availability of the data.

    Where milestone data exist but take inadequate account of eligibility criteria:

    *   clear and appropriate caveats should be included.
    *   where feasible, further work should be undertaken to take full account of the eligibility criteria.

2. The impact(s) of instances where the data for a milestone before the end of the process includes those who have little or no chance of making it to a later milestone should be assessed as part of future 'deep dive' work. If any key characteristic is uncovered as part of this further work, consideration should be given to producing statistics after adjusting for it.

3. Where appropriate, consideration should in the future be given to including:

    *   intersectional analysis (e.g. looking at the combination of ethnicity and gender) in addition to focusing on each diversity characteristic separately.
    *   indications of emerging trends where there are no statistically significant disparities in any particular year and where such disparities would be statistically significant if a few years data were rolled together.

Key statistical indicators

1. The relative difference between the success rates (or RRI) is recommended as the key summary indicator to flag up potential disparities between those with different diversity characteristics in MoJ DASD statistical publications. The key RRI findings should, where possible, include coverage of the RRI from the 'start of process' milestone (denominator) to the 'end of process' milestone (numerator) and where disparities are found (that are of practical and statistical significance - see accompanying metrics/rules 1. and 2.) provide further information about where in any interim stages of the process disparities exist. Where data for main interim stages are unavailable, further work, where feasible, should be undertaken to enable the availability of these data. The accompanying statistical

---

[19] This can be referred to as the eligible population or the population at risk.
[20] This can be referred to as the population who experienced the target event.

tables should include the RRIs from start to end of process and for each main interim stage in the process.

2. To aid interpretation of the relative difference between the success rates (or RRI) as the key summary indicator, some underlying statistical context could be provided by showing success rates on which the RRI is based and where deemed appropriate the representation percentages at each stage. User testing should be obtained in future decision making about this.

3. While the odds ratio is an easy to use metric for some deep dive analyses, it is much more difficult to understand than the RRI and therefore much easier to misinterpret. Any usage of the odds ratio (e.g. in future publications containing 'deep dive' analyses) should be accompanied by an illustration about what a particular odds ratio value means in practice.

Accompanying metrics/rules

1. Where not already used, consideration should be given to the RRIs being accompanied by a tolerance zone which provides a useful guide to indicate what is significant from a practical viewpoint. The lower and upper bounds of the tolerance zone should be 0.8 (4/5) and 1.25 (5/4), the de facto 'industry standard', unless there is evidence to suggest otherwise. This should be reviewed every few years.

2. It is helpful for disparities that are statistically significant to be flagged up so users know when they can be confident that a disparity really exists. For a disparity to be regarded as a high priority for policy action, we recommend it is both statistically significant and outside the tolerance zone. A suggested guide to interpretation, which could be finessed with user testing, is as follows:

    a. Inside the tolerance zone – this signifies no practical disparity
    b. Outside the tolerance zone but not statistically significant – this implies that the disparity could well be due to chance
    c. Outside the tolerance zone and statistically significant – this implies that we can have confidence that a disparity exists and that it is of a size that is practically important.

3. The total number, declaration rate and related number of unknowns should be published to allow users to assess the potential impact of statistical bias due to non-reporting of diversity characteristics. Where applicable and the data are available, a breakdown of the number of unknowns should also be provided to enable users to understand the extent to which people are choosing "prefer not to say" rather than simply not responding. Where data relating to this extra breakdown are unavailable, further work should where possible be undertaken to obtain such data.

4. As the potential for statistical bias is best dealt with at source, consideration should be given to introducing a declaration rate threshold (below which the

results would not be reported) where one doesn't currently exist. Where a declaration rate threshold does exist, the level of it should be periodically reviewed and consideration given to whether increasing it could encourage better reporting and thereby lead to more accurate statistics. Suitable advanced notice should be given in advance of any declaration rate threshold introduction or increase to enable it to have its desired effect.

5. For results reported separately for different groups (e.g. different selection exercises):

    a. Where necessary to withhold findings for individual groups to maintain the confidentiality of individuals, consideration should be given to whether the findings for these groups can be suitably aggregated rather than merely withheld. In decision making about which groups to combine, consideration should be given to the relevance of combinations from both a policy and statistical perspective. Regarding the latter, it is desirable for the trend in the aggregated groups to resemble the individual trends for these groups.

    b. Where groups are combined and the beginning milestone is an eligible pool, a combined group eligible pool could be estimated by summing the eligible pools (even where overlapping) for each group. In circumstances where the eligible pools for the combined groups overlap this should be appropriately caveated.

    c. Where the findings of some groups are withheld or aggregated together for confidentiality reasons, summary information about the total number, declaration rate and number of unknowns should be provided individually for these groups as for other groups. This could be done at an early milestone in the process providing confidentiality can be maintained (e.g. for processes involving applications at the application stage where there are ten or more applications).

6. The provision of confidence intervals (or indications of statistical significance) and declaration rates enables users to gain a good understanding of the uncertainty around key statistical indicators. Unless there are overriding confidentiality, presentational or statistical concerns, there is no need for figures to be suppressed. While the calculation of standard confidence intervals needs careful modification for small groups, an alternative approach is recommended that renders such modifications unnecessary[21]. Any presentational concerns arising from small groups leading to particularly high RRIs could then be monitored and dealt with on a case by case basis until such a time that a useful rule can be put in place.

Presentational features

1. The key messaging (including main points from plots) should be concise but also sufficiently well explained so that lay users don't have to refer to another part of

---

[21] Using 'bootstrapping'

the publication. It should also include clear and well positioned caveats to avoid the potential for misleading conclusions to be drawn. For instance:

    a. that disparities may be explained by differences in other factors.
    b. where eligible pool (or start of process) figures are unavailable that care should be taken when drawing conclusions based on partial coverage of the statistics.
    c. where annual comparisons need to be approached with caution due to changes over time.

2. The key headline data should be easy for publication users to find and use. This could be achieved by including tables showing the key RRI statistics for each diversity strand.

3. Where appropriate consideration should be given to reviewing presentation of:
    a. related figures/plots e.g. to show RRIs accompanied by a tolerance zone and 95% confidence intervals along with accessible colouring[22] and a brief explanatory summary. An example forest plot[23] that could be considered is shown in Figure 1 (see section D).
    b. explanations of the RRI and other technical terms to ensure they can be understood by lay readers and that the terminology is clear and accurate (in particular the tolerance zone should be described as a measure of practical rather than statistical significance).

---

[22] The GSS "Effective tables and charts in official statistics" document provides guidance on the use of colours in charts: https://gss.civilservice.gov.uk/wp-content/uploads/2018/03/Effective-Tables-and-Charts-in-official-statistics-Edition-2.1-February-2018-4.pdf
[23] https://en.wikipedia.org/wiki/Forest_plot

Appendix A: Consultee list

The following people were consulted as part of the review:

Name	Organisation/Team
Dr Habib Naqvi MBE	NHS - Deputy Director - NHS Workforce Race Equality Standard
Owen Chinembiri	NHS - Senior Analytical Manager - NHS Workforce Race Equality Standard
Baljit Gill	Race Disparity Unit
Charles Lound	ONS Quality Centre
Wincen Lowe	MoJ - HR HMPPS Workforce statistics
Rachel Blake	MoJ - HR HMPPS Workforce statistics
Robert Reeve	MoJ - Criminal Justice Outcomes and Equalities
Miranda Crusco	MoJ - Criminal Justice Outcomes and Equalities
Sandy Rass	DWP - former Judicial statistics team leader in MoJ – (worked for JAC part time)
Richard Jarvis	JAC - Chief Executive
Alice Ripley	JAC - Deputy Chief Executive
Jessica Prandle	JAC - Head of Diversity and Engagement
Matt Walker	MoJ - Judicial statistics team leader (now former) – (worked for JAC part time)
Matthew Tranter	MoJ – Judicial statistics team leader (works for JAC part time)
Zalika Awuku	MoJ - Senior Policy Advisor, Judicial Operations: Appointments and Diversity Policy
Alistair Cook	MoJ - Head of Judicial Operations: Appointments and Diversity Policy
Andrea Coomber	Director of JUSTICE
Dr Yael Levy Ariel	Justice Team Policy Lead – Solicitor Judges Division
Tim Smith	JUSTICE council
David Blunt	MoJ - Head of Profession for statistics
Tracie Kilbey	MoJ - Head of Contracts and Offender Equalities Statistics
Abigail Higgins	MoJ - Development and Performance Lead
Nick Read	Youth Justice Board - Senior Information Adviser
Noah Uhrig	MoJ - Lead Analyst in the Lammy Review
Prof David Spiegelhalter	University of Cambridge
Prof Cheryl Thomas	University College London
Prof Dame Hazel Genn	University College London

Appendix B: List of all recommendations for new combined JAC publication[24]

Coverage

1. We recommend that applications, shortlisting and recommendations should be presented within the context of the eligible pool (of potential applicants who meet the minimum eligibility criteria). This is to ensure we cover the full scope of the JAC's work and for the statistics to be as relevant as possible.

2. Where eligible pool figures are not available:
   a. clear and appropriate warnings should be included within the report that care should be taken when drawing conclusions based on partial coverage of the statistics.
   b. further work should be undertaken to enable the availability of eligible pool data.

3. Consider further analysis initially within future 'deep dive' work but with a view to inclusion within the new combined JAC statistical publication at a later point:
   a. on whether the eligible pool includes people who meet minimum eligibility requirements but have little or no chance of applying successfully, and if so the impact(s) of this. For example, it is possible that in practice some of those meeting minimum eligibility requirements do not have enough experience in the field to be recommended for appointment. Further analysis could determine this and potentially produce statistics after adjusting for years in the field.
   b. of more detailed or robust data from the professional legal bodies which could provide greater insight into why certain eligible individuals may not be applying as well as any potential 'blockers' further on in the recruitment process.

4. Additional intersectionality analyses (e.g. looking at the combination of ethnicity and gender) should be considered along with providing indications of emerging trends where there are no statistically significant disparities in any single year, and where such disparities would become statistically significant if a few years data were rolled together. Such analyses initially fit within 'deep dive' work but should be with a view to inclusion within the new combined JAC statistical publication at a later point.

Key statistical indicators

1. The relative difference between the success rates (or RRI) is recommended as the key summary indicator to flag up potential disparities between those with different diversity characteristics.

---

[24] This appendix contains a summary of recommendations already listed earlier in the report.

2. It is necessary for applications, shortlisting and recommendations to be presented within the context of the eligible pool. It is therefore recommended that the key RRI findings in the new combined JAC statistical publication must at minimum include coverage of the RRI from the eligible pool (denominator) to recommendation (numerator) and where disparities are found (that are of practical and statistical significance - see below section) provide further information about where in the interim stages of the process disparities exist. Consultation with stakeholders also revealed a desire for the key RRI findings to by default include the RRI from application to recommendation. This could be included perhaps along with the complementary RRI from eligible pool to application.

3. The new combined JAC statistical tables should include all RRIs mentioned in 2. above; from eligible pool to recommendation, application to recommendation, and for each interim stage in the selection process (eligible pool to application, application to shortlisting, and shortlisting to recommendation).

4. To aid interpretation of the relative difference between the success rates (or RRI) as the key summary indicator, some underlying statistical context could be provided by the success rates on which the RRI is based; and if user consultation deems necessary the representation percentages at each stage.

5. While statistical properties of the odds ratio enable it to be an easy to use metric for some deep dive analyses[25], it is much more difficult to understand than the RRI and therefore much easier to misinterpret. Any usage should therefore be accompanied by an illustration about what a particular odds ratio value means in practice.

Accompanying metrics/rules

1. The 4/5ths rule to indicate whether an RRI value is sufficiently far from one (or parity) to reflect a disparity of outcomes is considered very useful from a policy perspective. Although it may originally have been developed as a quick and approximate guide to determine whether a disparity is statistically significant, its usefulness in the JAC statistical publication is to provide an indication of what is significant from a practical point of view. The zone of tolerance lower and upper bounds of 0.8 (4/5) and 1.25 (5/4) are currently the de facto 'industry standard', with the 4/5ths rule being used elsewhere in a recruitment context (along with the RRI)[3] as well as in a criminal justice context. Moreover, there is an analogy with the 'range of equivalence' (also from 0.8 to 1.25) used by regulatory bodies in drug trials[26,27,28]. Given this review did not uncover any evidence that suggested the bounds should either be increased or decreased, it is recommended that the

---

[25] For example, those involving the use of logistic regression.
[26] https://en.wikipedia.org/wiki/Bioequivalence
[27] https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/bioequivalence
[28] https://www.fda.gov/media/89135/download

tolerance zone remains unchanged for the new combined JAC statistical publication until any conclusive evidence is uncovered that suggests otherwise. Consideration should be given to reviewing this at some point in the future.

2. It is important for the JAC statistical publication to flag up any disparities that can be considered statistically significant[29]. Therefore, for a disparity to be regarded as a high priority for further action, we recommend it is both statistically significant and outside the tolerance zone (presently 0.8 to 1.25) – see 1. above. A suggested guide to interpretation, which could be finessed with user testing, is as follows:

   a. Inside the tolerance zone (presently 0.8 to 1.25) – this signifies no practical disparity
   b. Outside the tolerance zone (presently 0.8 to 1.25) but not statistically significant – this implies that the disparity could well be due to chance
   c. Outside the tolerance zone (presently 0.8 to 1.25) and statistically significant – this implies that we can have confidence that a disparity exists and that our best estimate suggests that it is of a size that is important practically.

3. Some indication of the potential impact of statistical bias due to non-reporting of diversity characteristics should be provided in the JAC statistical publication. While further work could be done to quantify the impact of any statistical bias on key metrics, the estimation of upper and lower-case scenarios tends to provide somewhat speculative worst-case scenarios that tend to be unhelpful in practice. The total number, declaration rate and the related number of unknowns should continue to be published. Where possible these should be accompanied by a breakdown of the number of unknowns to understand the extent to which people are choosing any 'prefer not to say' and 'prefer not to share data outside the JAC' options and so are engaging with the monitoring process rather than simply not responding. Where such options exist but the related data are unavailable, further work should be undertaken to try and obtain these.

4. Where exercises are grouped together in the JAC statistical publication it would be helpful for the total numbers, declaration rates and numbers of unknowns (where possible with the additional breakdowns described in 3. above) to be provided separately at the application stage (instead of only at aggregated grouped level) for each exercise involving ten or more applications.

5. The potential for statistical bias is best dealt with at source. While declaration rates are consistently high overall[30], they vary by characteristic. Only one declaration rate in the 2019 JAC statistics publication tables was less than 70%

---

[29] Utilising standard statistical methodology to take account of the sample size and the volatility present in the data.
[30] Being 80% or more at the application stage for gender, ethnicity, professional background, disability, social mobility, sexual orientation and religion.

and this was also less than 60%. The declaration rate threshold could therefore be raised from 60% to 70% or even 80% to encourage even better reporting which will in turn lead to more accurate statistics. The level of this threshold should be reviewed every few years.

6. Exercises with fewer than ten recommendations should continue to be aggregated into groupings for confidentiality reasons so applicants cannot be personally identified.

7. Where exercises are aggregated and the eligible pools are known, an overall group eligible pool could be estimated by summing the eligible pools for each exercise. In decision making about which exercises to group together, consideration should be given to the relevance of groupings from both a policy and statistical perspective. User consultation could determine whether the groupings can be more relevant in future. Statistical anomalies could be avoided by checking the trend in the aggregated exercises is consistent with trends seen in the individual exercises.

8. Unless there are overriding confidentiality, presentational, or statistical concerns, there is no need for figures to be suppressed. Confidentiality concerns are addressed by 6. above. Together with the high declaration rates, the provision of confidence intervals (or indications of statistical significance) enables users to gain a good understanding of the uncertainty around key statistical indicators. While the calculation of standard confidence intervals[31] needs careful modification for small groups, an alternative approach is recommended that renders such modifications unnecessary[32]. Any presentational concerns arising from small groups leading to particularly high RRIs should be monitored and dealt with on a case by case basis until such a time that a useful rule can be put in place. If the current practice of suppressing figures where denominators are less than 10 is continued, it should be applied after first excluding any unknown values.

Presentational features

The following presentational suggestions for the new combined JAC publication should be finessed by user consultation and consideration of the GSS Communicating quality, uncertainty and change guidance[33]. In particular:

1. For the summary section to be user friendly we advise it has a similar style to other Ministry of Justice publications[34] which use a bullet point structure with visual aids such as arrows and signs. The chosen approach should enable the

---

[31] Those using a Normal Distribution approximation.
[32] Using 'bootstrapping'
[33] https://gss.civilservice.gov.uk/policy-store/communicating-quality-uncertainty-and-change/
[34] For example, the summary section of Criminal Justice System statistics can be viewed in: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/867102/criminal-justice-statistics-quarterly-september-2019.pdf

publication to be easily readable with clear key messages. The summary of main points should ideally be one page and should contain one main point per section.

2. To enable the publication to be as navigable as possible we recommend including a table of contents. Efforts should be made to keep the report as concise as possible without any unnecessary graphs. Contextual information should be provided in an accompanying statistics definitions and measurements document and additional information provided in the appendices.

3. The key messages should be sufficiently well explained so that lay users don't have to refer to another part of the publication. This includes key points from plots and summary statistics relating to the RRI e.g. 'the following disparities were significant both practically (being outside the tolerance zone of no disparity) and statistically: females in the eligible pool were 43% less likely to be recommended than males in the eligible pool with them having a success rate of 20% compared to 34% for males'.

4. All key messages should also include clear and well positioned caveats to avoid the potential for misleading conclusions to be drawn. For instance:
   a. In the summary that disparities may be explained by differences in other factors.
   b. where no eligible pool figures are available that care should be taken when drawing conclusions based on partial coverage of the statistics.
   c. where annual comparisons need to be approached with caution due to changes in exercises over the years.

5. The key headline data should be easy for publication users to find and use. This could be achieved by including tables showing the key RRI statistics for each diversity strand.

6. Presentation of the figures/plots should be reviewed. For instance, forest plots[35] with accessible colouring[36] could be used to show the RRIs accompanied by 95% confidence intervals and the zone of tolerance. An example can be seen in Figure 1 (section D). Any (horizontal) bar charts could include the actual percentage figures along with accessibly coloured bars.

7. Explanations of the RRI and other technical terms should be reviewed to ensure they are understood by lay readers and that the terminology is clear and accurate. In particular the 4/5ths rule (or tolerance zone) should be described as a measure of practical rather than statistical significance (see Accompanying metrics/rules recommendation 1 above) while the declaration rate is about declarations rather than unknowns. Consideration should be given to moving

---

[35] https://en.wikipedia.org/wiki/Forest_plot
[36] The GSS "Effective tables and charts in official statistics" document provides guidance on the use of colours in charts: https://gss.civilservice.gov.uk/wp-content/uploads/2018/03/Effective-Tables-and-Charts-in-official-statistics-Edition-2.1-February-2018-4.pdf

these sections to an appendix with short concise wording developed for the main part of the report (an example of the latter would see a bullet point or two following an RRI graph which summarises what it is telling you – see example in section D recommendation 6).

Contact

Press enquiries should be directed to the Ministry of Justice press office:
Tel: 020 3334 3536
Email: newsdesk@justice.gsi.gov.uk

Other enquiries and feedback on this report should be directed to the Justice Statistics Analytical Services division of the Ministry of Justice:
Aidan Mews, Head of Statistical Project Delivery Team
Ministry of Justice, 10 South Colonnade, London, E14 4PU
Email: aidan.mews@justice.gov.uk