UC ADVANCES MODEL - FAIRNESS ASSESSMENT

Executive summary

It is the Department's assessment that there are minimal concerns of discrimination, unfair treatment or detrimental impact on legitimate claimants arising from the Advances model. Therefore, it remains reasonable and proportionate to continue operating the Advances model as a fraud prevention control. However, the model will be retrained since improved alignment to Advances fraud risks could reduce the measured inconsistencies between statistical disparities. Retraining the model is an activity the Department undertakes in the normal course of maintaining and improving the model to ensure it remains optimised to identify fraud risk.

This fairness assessment is informed by statistical fairness analysis covering the period 1st April 2024 and 31st March 2025, which is set out in the Statistical Annex. This is experimental analysis so we will continue to iterate the process and methodology in future years.

Purpose of the Fairness Assessment

1. The purpose of this Fairness Assessment is to consider the results of the statistical fairness analysis alongside other factors and review the extent to which any measured statistical disparity may represent risk of discrimination, unfair treatment or detrimental impact on claimants. Other factors taken into consideration in this Fairness Assessment include:

    - The size and nature of Advances fraud.
    - The consideration of using machine learning as a control to mitigate the identified fraud risk.
    - The safeguards in place for the design, development and operation of the model.
    - The impact of the model on claimants, including the timeliness of Advances payments for requests referred by the model.

2. Considering the results of the statistical analysis alongside these wider factors allows us to make an assessment in relation to the continued operation of the model as a reasonable and proportionate fraud prevention control.

Size and nature of Advances fraud

3. The Department provided 1.4 million UC advances to new UC claimants in 2024/2025, with a total value of £0.8bn. As set out in the Official Statistics, we estimate that for 2024/ 2025 the monetary value of fraud and error on UC advances lies between £0m and £60m[1]. Advances fraud was the subject of a National Audit Office report in March 2020[2] setting out the challenge of UC

[1] Background Information and Methodology: Fraud and error in the benefit system, Financial Year Ending (FYE) 2025 - GOV.UK
[2] Universal-Credit-advances-fraud.pdf

Advances fraud. The machine learning model has been part of the response to this type of fraud and contributed to more than halving the estimated scale of the problem. The persistent nature of the fraud threat requires appropriate controls to tackle Advances fraud risk, such as the Advances fraud prevention model.

Consideration of the use of machine learning as a fraud control

4. The Department has a clear obligation to protect public money and tackle fraud and error. It is also a Government manifesto commitment to reduce waste and safeguard public money.

5. It is not an option for the Department to do nothing where fraudsters seek to steal money from DWP. At the other extreme the Department could mandate all citizens provide evidence to DWP, for scrutiny and verification, of all relevant circumstances before an Advance is paid. Given the number of people who access support from the Department each day, this would be inefficient for DWP and add additional steps in the process for all legitimate claimants.

6. The Department is therefore developing machine learning capability to tackle fraud and error. The UC Advances model is designed to risk assess requests for Advances. It is the only machine learning model currently deployed at scale into live service.

7. This approach enables the Department to focus its efforts on reviewing and verifying UC Advances assessed to have the highest risk of fraud. This risk-based approach:
   - Minimises impact on the customer experience of legitimate claimants
   - Optimises the use of taxpayers' money in delivering the fraud control
   - Optimises the effectiveness of the control at tackling fraud

8. The performance information for 2024/25 demonstrates the model is around 3 times more effective at identifying fraud risk than a randomised control group sample.

Design, development and operation of the model

9. Departmental governance requirements have been adhered to throughout the design, test, implementation and operation of the Advances model to ensure compliance with legal obligations, including having due regard to the public sector equality duty. The Department's Personal Information Charter (privacy policy) informs claimants their data may be used for the prevention and detection of fraud and protecting public funds. This includes detailing the type of data which may be used and a specific reference to artificial intelligence as a method by which data may be processed.

10. There is a suite of safeguards to minimise the risk of unfair treatment or detrimental impact on legitimate customers, irrespective of their protected characteristics, including:

- A blend of high risk and control group referrals are sent for human intervention to mitigate against human bias. In addition, the DWP employee delivering the intervention does not receive a risk rating in the referral nor are they made aware the referral has been generated by the model.
- A decision to decline an Advance request does not prevent the same claimant from making a further Advance request and does not automatically result in the associated new UC claim being refused. Decisions on eligibility and entitlement for a new UC claim are a separate consideration.
- Regular fairness assessment, including statistical analysis, is conducted to identify any concerns of unfair treatment or detrimental impact on customers.
- We also conduct analysis of any impact on payment timeliness on legitimate Advances requests.

11. The ultimate safeguard in place is that there is always a human intervention and decision, with no automated decision making by the model. Advance requests identified as high risk by the model are referred to a DWP employee, who reviews all available and relevant information, to decide whether to approve or decline the request.

Performance of the model as a fraud prevention control

12. Performance monitoring of the Advances model confirms it is an effective fraud prevention control and more efficient than an untargeted approach. The model is around 3 times more effective at identifying high risk advances than a control group sample. It has delivered and continues to deliver measurable savings. Therefore, the model enables the Department to reduce fraud and protect the public purse effectively.

13. Analysis confirms the payment of Advance requests predicted as high risk by the model and subsequently approved by a human decision maker are not unduly delayed. The median payment delay is 1 day longer compared to Advance requests that are approved automatically, which is in line with the delay experienced by Advance requests that are subject to other fraud controls that are distinct from the model.

Assessment of statistical fairness analysis

14. In traditional equality analysis, a ‘good’ outcome is typically defined as one in which all groups have an equal likelihood of experiencing a given outcome, assuming all other factors are equal. This would mean no measurable disparity between groups, which may indicate that the process or system is operating without bias. However, the UC Advances model is designed to assess claim characteristics associated with higher fraud risk. Due to fraudsters misrepresenting their circumstances and because fraud risk is not evenly distributed across all claim types, some statistical disparities between groups may arise. These disparities do not imply that any group is inherently more likely to commit fraud and all disparities are subject to fairness monitoring and review.

15. This year, the statistical analysis for the UC Advances model uses two metrics to support our understanding of what a 'good' outcome looks like:
    - Risking disparity - proportion of Advances risk-scored by the model that were predicted to be high risk
    - Outcome disparity – proportion of Advances predicted as high risk by the model that were confirmed to be fraudulent by a human decision maker.

16. DWP's assessment is that, where the metrics are consistent, this is an indication of a 'good' outcome. Specifically, it suggests that Advance requests made by people who share a protected characteristic were predicted by the model to be high risk at a rate which is consistent with the prevalence of fraudulent Advance requests appearing in that cohort of claims. To note, fraudsters misrepresent their true circumstances and therefore a fraudulent Advance request may share common claim characteristics with legitimate requests from a specific group. A higher Risking and Outcome disparity, relative to the comparator group, does not imply any group is more likely to commit fraud.

17. Where there are inconsistencies between the measured disparities, that is a signal to explore the disparity further. It might be the disparity is considered reasonable because we assess the model is successfully targeting the fraud risk found within a group, again noting fraudsters misrepresent their circumstances and fraudulent Advance requests may share common claim characteristics with legitimate requests within that group. Alternatively, a disparity may be an indication it is appropriate to consider further action, e.g. retrain the model.

18. To determine whether there are differences between groups, we are adopting 'relative likelihoods' as the chosen statistical methodology, following best practice recommended by the Cabinet Office on equality analysis[3]. Further information on the methodology can be found in the Statistical Annex.

19. Due to the limited availability of protected characteristic data for the high-risk cohort, age is the only protected characteristic that has been measured. For other protected characteristics, no direct proxies could be identified. To better understand the behaviour of the model, fairness analysis has also been undertaken based on whether the claim is from a UC couple contract, if the claimant reported an illness and if the claimant is a UK national or not. Further information on the data used can be found in the Statistical Annex.

Results

20. For the characteristics 'UC couple contract' and 'reported illness', there was consistency between referral and outcome disparity for claimants. The metrics are moving in the same direction; therefore, this is an indication that the model is working effectively to identify fraud risk – we have assessed this as a 'good' outcome.

[3] Using relative likelihoods to compare ethnic disparities - GOV.UK

21. The likelihood of non-UK nationals being referred by the model was higher than UK nationals, however the likelihood that the Advance was rejected by a human decision maker following a referral was equivalent to the likelihood for UK nationals.

22. For the age characteristic, there was consistency between Risking and Outcome Disparity for claimants aged 16-24 and 25-34, as the likelihood of being referred by the model and the likelihood of those referrals being correct are both higher compared to the 35- to 44-year-old age group.

23. However, for other age groups an increased likelihood of being referred by the model is inconsistent with the reduced likelihood of those referrals being correct. The evidence suggests the model is not working as effectively as we would expect.

24. To try and reduce measured inconsistencies between referral and outcomes metrics, the model will be re-trained and further fairness analyses conducted to measure the impact of this action on reducing age and nationality related disparities.

The full statistical analysis is set out in the Statistical Annex: Ad hoc statistical release of the fairness analysis for the Advances model.

Assessment

25. It is the Department's assessment it remains reasonable and proportionate to continue operating the UC Advances model because:
    - **Safeguards –** there are a suite of safeguards in place protecting all legitimate claimants irrespective of protected characteristics to prevent detrimental impact, crucially that a human always reviews the evidence and makes a decision
    - **Assessment of Statistical analysis findings** – for characteristics related to UC couple contracts and reported illness, where a higher referral rate is observed, so is the likelihood of those referrals being correct. This indicates the model is operating effectively. For some age groups and for non-UK nationals, an increased likelihood of being referred by the model is inconsistent with the reduced likelihood of those referrals being correct. This warrants further action, and we will retrain the model.
    - **Impact on payments -** there is minimal impact on payment timeliness for legitimate claimants. The median impact on payment of UC Advances referred by the model, and subsequently approved by a human decision maker, is 1 day (compared to an Advance not assessed as high risk by any fraud control measures). The payment timeliness of the UC claim is not affected by the model.
    - **Performance of the model -** the model minimises the impact on the whole Advances claimant population by focussing the additional fraud prevention intervention on the riskiest (rather than randomly selected) Advance requests, reducing unnecessary interventions for the broader claimant population. The

model is around 3 times more effective at identifying high risk advances than a random control group selection.

26. Given the factors set out above, the Department has determined it is reasonable and proportionate to continue operating the model as a fraud prevention control. There is minimal concern of discrimination, unfair treatment or detrimental impact on claimants. However, the model will be retrained to improve its effectiveness, which should reduce the statistical disparities and the inconsistencies between Risking and Outcome disparities which have been measured. Further statistical analysis will be completed, following the model being retrained, to evidence the impact on the measured disparities and inconsistencies. Retraining the model is an activity the Department undertakes in the normal course of maintaining and improving the model to ensure it remains optimised to identify fraud risk.

Statistical Annex: Ad hoc statistical release of the fairness analysis for the Advances model

Ad Hoc Statistics

1. This annex is a statistical publication, presenting the statistical analysis that has been conducted to understand the impacts on groups with protected characteristics. This is a complicated process and there are limited precedents available, hence we expect to further develop the methodology in the future. This ad hoc statistical release has been produced in accordance with the principles set out in the Code of Practice for Statistics, ensuring transparency, integrity, and quality.

Equality Act 2010

2. Under the Equality Act 2010, a protected characteristic refers to specific attributes or traits that are safeguarded against discrimination. There are nine protected characteristics: age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation. The definition of each protected characteristic is given in legislation. Characteristics or attributes that do not meet the definition of a protected characteristic are referred to as a 'non-protected characteristic' or simply a 'characteristic' in this report.

Equality data availability

3. Universal Credit claimants are asked to answer optional equality questions when making their claim. Disparity analysis conducted using a protected characteristic that has limited response data could be subject to significant non-response bias. We would only perform disparity analysis of a characteristic where we have at least a 70% completion rate, in line with the Department's approach to the use of UC equality data for published analysis. Of the claimants referred by the Advances model as high-risk, the proportion who responded to the UC equality questionnaire and who responded with an answer other than 'prefer not to say' did not meet the 70% threshold, and therefore we have not used this data source on this occasion.

4. We considered other sources of protected characteristics data that were available to us. The age of a claimant at the time they request an advance can be calculated from the date of birth they provide in their UC application. This source of age data directly corresponds to the definition within the Equality Act. We do not have enough data that directly corresponds to any other protected characteristic as defined in the Equality Act 2010 for claimants referred by the Advances model as high risk. Response rates for equality data in this release may be lower than those reported in other published statistics because the 70% completion rate must be reached for cases identified as high risk by the Advances model, as well as the majority of cases that are not.

5. The analysis was limited to one protected characteristic, as such we could not analyse the intersectionality of those characteristics. It is feasible that disparities measured for one characteristic may be a result of intersectionality with another.

6. In the interest of better understanding the characteristics of claims referred by the model we have conducted disparity analysis for some characteristics that are not protected. The disparity analysis of non-protected characteristics that are included in this report cannot be used as an approximation for any protected characteristics. The non-protected characteristics analysed are whether the claim is from a UC couple contract, if the claimant reported an illness or if the claimant is a UK national or not. Listed below are the reasons these characteristics do not meet the definition of a protected characteristic:

    - Whether the claimant is part of a UC couple contract or not is not an approximation for marriage and civil partnership. The relationship between the claimants in a couple contract may not have the legal status of marriage or civil partnership (as referenced in the Equality Act 2010).
    - The claimant's response to the question "Do you have any disabilities, illnesses, or ongoing conditions?" which is asked as part of the UC claim process, can include declarations of illness which may not meet the definition of disability under the Equality Act 2010.
    - Nationality is referenced in the Equality Act 2010 but it is not a protected characteristic.

Data used in analysis

7. All UC Advances risk scored by the model between 1 April 2024 and 31 March 2025 were considered in the analysis. Where a claimant or contract had more than one advance request during the period, one advance was randomly selected to ensure each observation was independent and to avoid bias from multiple requests by the same claimant.

8. We have considered two measures of disparity: Outcome and Risking. The dataset used to calculate Risking disparity contains UC Advances that were requested between 1 April 2024 and 31 March 2025 and risk scored by the model (limited to at most one advance per claimant and contract). The dataset used to calculate Outcome disparity contains UC Advances, that were requested between 1 April 2024 and 31 March 2025, risk scored by the model and were referred to a human decision maker because they were predicted to be high risk (limited to at most one advance per claimant and contract.

Relative likelihood and statistical significance

9. Relative likelihoods is a statistical technique carried out to identify any differences between groups and indicates the extent to which two groups differ in their likelihood of experiencing an event. It follows the methodology used as the standard approach for assessing disparities of outcomes by ethnicity, developed and recommended for use across Government by the Cabinet Office (Race Equality Unit) published in 2020. It is used in other Government departments for

considering potential disparities of outcomes by ethnicity and by DWP in its benefit sanctions statistics. To calculate a relative likelihood, we use the following formula: Relative likelihood = percentage (or proportion) of one group experiencing an outcome, divided by percentage (or proportion) of comparator group experiencing an outcome.

10. To calculate a measure of relative likelihood, a comparator group must be chosen. We have chosen the group with the largest number of UC advance requests as the comparator. It is noted in the report published by the Race Equality Unit that the estimates are more robust when the largest group is chosen as the comparator. While the comparator group provides a useful benchmark for assessing relative outcomes, it is important to note that it does not necessarily represent an ideal or optimal standard. Rather, it serves as a reference point to highlight differences or disparities. The comparator may itself be subject to limitations, biases, or systemic issues, and should not automatically be interpreted as the group to emulate.

11. The closer a relative likelihood is to 1, the greater equality there is between the two groups. A relative likelihood greater than 1 suggests the outcome is more likely in the group that was compared to the comparator group. A relative likelihood less than 1 suggests the outcome is less likely than in the comparator group.

12. We have used significance testing, in the form of 95% confidence intervals, to test whether a relative likelihood is statistically significantly different from parity with the comparator group. In simple terms, if a difference is statistically significant it represents a real difference rather than being solely due to chance.

The size of disparity

13. Some difference in outcomes between groups are to be expected due to natural variation. As such it is essential to not only understand if such differences are statistically significant, and not the result of chance alone, but also whether the scale of impact of those differences is meaningful. With larger sample sizes, significance testing has the power to detect small effects. Considering the scale of the effect size allows us to assess whether such differences would have a notable materially meaningful impact. This then guides interpretation as to whether such differences require monitoring, investigation or action.

14. To test the effect size (i.e. the scale of the difference between the relative likelihood and parity with the comparator group), a four-fifths rule has been applied to the relative likelihood estimates. Any relative likelihood estimates that fall outside a range of 0.80 to 1.25 indicate the impact of the disparity is meaningful and is described in this report as “notable”. The four-fifths rule has been used by DWP in its benefit sanctions statistics and by other government departments such as Ministry of Justice.

Risking disparity

15. To measure disparities in the model risk scoring we have calculated the predicted positive rate for each characteristic. The predicted positive rate is calculated as the number of UC advances that the model predicted to be high risk, divided by the total number of advances risk-scored by the model.

16. High risk UC Advances are referred to a human decision maker. The relative likelihood of the predicted positive rate tells us how likely a group was to be referred by the model compared to the comparator group.

Outcome disparity

17. The UC advances identified by the model as high risk and subsequently denied payment by a human decision maker are considered correct referrals. To assess disparities in the proportion of referrals correctly predicted by the model, we calculated the true discovery rate for each characteristic. The true discovery rate is defined as the number of UC advances correctly predicted by the model to be high risk, divided by all the advances it predicted to be high risk.

18. The relative likelihood of true discovery tells us how likely a group was to be correctly referred by the model compared to the comparator group.

19. It is not unexpected for a fraud prevention model to have measurable disparities and it is currently unknown whether it is feasible to design an effective fraud prevention model where Risking disparity and Outcome disparity are identical at all times.

Claimant age

20. The age group that had the largest number of Advances risk scored by the model was the 35- to 44-year-old group, therefore we have chosen this as the comparator group.

Table 1: Relative likelihoods of referral by age group, compared to the 35-44 year old age group

Age Group	Relative Likelihood of Referral	Size of Disparity	Difference in Relative Likelihood
16-24	1.67	Notable	Statistically Significant
25-34	1.13	Not Notable	Statistically Significant
35-44	1	N/A	N/A
45-54	1.35	Notable	Statistically Significant
55-65	2.80	Notable	Statistically Significant
66+	49.24	Notable	Statistically Significant

Table 2: Relative likelihoods of correct referral by age group, compared to the 35-44 year old age group

Age Group	Relative Likelihood of Correct Referral	Size of Disparity	Difference in Relative Likelihood
16-24	1.21	Not Notable	Statistically Significant
25-34	1.22	Not Notable	Statistically Significant
35-44	1	N/A	N/A
45-54	0.86	Not Notable	Statistically Significant
55-65	0.58	Notable	Statistically Significant
66+	0.23	Notable	Statistically Significant

21. There is consistency between Risking and Outcome Disparity for claimants aged 16-24 and 25-34. For those claimants, the likelihood of being referred by the model and the likelihood of the Advance being rejected by a human decision maker following a referral are both higher compared to the comparator group. However, for other age groups an increased likelihood of being referred by the model is inconsistent with a reduced likelihood of the Advance being rejected by a human decision maker following a referral. From a fairness analysis perspective, those inconsistencies require further attention.

22. Claimants aged 66 and above were 49.24 times more likely to be referred by the model compared to claimants aged 35 to 44 years old. It should be noted that claimants aged 66 and above are an unusual group as they only include those eligible for UC through their partner or move to UC cases from tax credits, which represent only 0.1% of observations in the data analysed. As such, the results for this group should be treated with caution due to the small sample size.

23. In the interest of better understanding the characteristics of claims referred by the model we have conducted disparity analysis for some characteristics that are not protected characteristics under the Equality Act 2010. We cannot draw any conclusions about the disparity that may or may not exist for any protected characteristic from the following analysis.

Nationality

24. The group with the largest number of Advances risk scored by the model are UK nationals, compared to non-UK nationals. Therefore, we have chosen UK nationals as the comparator group.

Table 3: Relative likelihood of referral for non-UK nationals compared to UK nationals.

Nationality	Relative Likelihood of Referral	Size of Disparity	Difference in Relative Likelihood
Non-UK national	2.27	Notable	Statistically Significant
UK national	1.00	N/A	N/A

Table 4: Relative likelihood of correct referral for non-UK nationals compared to UK nationals.

Nationality	Relative Likelihood of Correct Referral	Size of Disparity	Difference in Relative Likelihood
Non-UK national	0.97	Not Notable	Not Statistically Significant
UK national	1.00	N/A	N/A

25. The likelihood of non-UK nationals being referred by the model was higher than UK nationals, however the likelihood that the Advance was rejected by a human decision maker following a referral was equivalent to the likelihood for UK nationals.

UC couple contract

26. The group with the largest number of Advances risk scored by the model are requested from single claimant UC contracts, compared to couple contracts. Therefore, we have chosen single claimant contracts as the comparator group.

Table 5: Relative likelihood of referral for UC couple contracts compared to single claimant contracts

Couple Contract Status	Relative Likelihood of Referral	Size of Disparity	Difference in Relative Likelihood
Couple Contract	0.83	Not Notable	Statistically Significant
Single Claimant Contract	1.00	N/A	N/A

Table 6: Relative likelihood of correct referral for UC couple contracts compared to single claimant contracts

Couple Contract Status	Relative Likelihood of Correct Referral	Size of Disparity	Difference in Relative Likelihood
Couple Contract	0.24	Notable	Statistically Significant
Single Claimant Contract	1.00	N/A	N/A

27. There is consistency between Risking and Outcome Disparity for couple contracts compared to single claimant contracts. The likelihood of couple contracts being referred by the model and the likelihood of those referrals

resulting in the Advance being refused by a human decision maker are both lower compared to single claimant contracts.

Reported Illness

28. The group with the largest number of Advances risk scored by the model are requested by claimants who have not reported an illness; therefore, we have chosen that claimant group as the comparator group.

Table 7: Relative likelihood of referral for claimants who have reported an illness compared to claimants who have not reported an illness

Reported an Illness	Relative Likelihood of Referral	Size of Disparity	Difference in Relative Likelihood
Claimant has reported an illness	0.37	Notable	Statistically Significant
Claimant has not reported an illness	1.00	N/A	N/A

Table 8: Relative likelihood of correct referral for claimants who have reported an illness compared to claimants who have not reported an illness

Reported an Illness	Relative Likelihood of Correct Referral	Size of Disparity	Difference in Relative Likelihood
Claimant has reported an illness	0.55	Notable	Statistically Significant
Claimant has not reported an illness	1.00	N/A	N/A

29. There is consistency between Risking and Outcome Disparity for claimants that have reported an illness compared to those that have not. The likelihood of claimants with a reported illness being referred by the model and the likelihood of those referrals resulting in the Advance being refused by a human decision maker are both lower compared to claimants that have not reported an illness.

Other relevant statistics considered in the fairness assessment process:

30. In total, 1.4 million advances were awarded, amounting to £0.8 billion. These figures comprise New Claim advances and Benefit Transfer advances, with approximately 80% of the volume and 75% of the value attributed to New Claim advances.

31. A random sample of Advances predicted by the model to be low risk are sent to human decision makers to safeguard against confirmation bias and monitor false negatives. This is combined, in equal proportions, with a random sample of high risk Advances to produce a single stratified random sample. The proportion of Advances predicted by the model as high risk and confirmed to be fraudulent by a human decision maker is compared to the proportion of Advances in the random sample that are confirmed to be fraudulent. This comparison finds the model to be 2.9 times more effective at identifying Advances that are fraudulent.

32. Advances predicted as high risk by the model are referred to a human decision maker, who reviews all relevant and available information to decide whether to approve or decline the Advance. Payment is made if the Advance is approved. Advances requested between 1 April 2024 and 31 March 2025 that were risk scored by the model and not referred by any fraud control had a median payment delay of 0 days, so were paid on the same day as the Advance request. Whereas Advances identified as high risk by the model during the same period and approved by a human decision maker had a median payment delay of 1 day, so payment was made the day after the Advance request. This median payment delay is the same for Advances that were not referred by the model but were subject to other fraud controls.