

EDWARD J. PINTO and TOBIAS J. PETER
AEI Housing Center

ALEX J. POLLOCK
R Street Institute

September 26, 2019

Department of Housing and Urban Development
Regulations Division
Office of the General Counsel
451 7th Street SW
Washington, DC 20410
Submission via www.regulations.gov

Dear Sir/Madam:

Re.: Docket No. FR-6111-P-02; RIN: 2529-AA98
HUD's Implementation of the Fair Housing Act's Disparate Impact Standard

Thank you for the opportunity to comment on this proposed rule on the Disparate Impact Standard. The authors of the comment have many years of experience in housing finance, as operating executives, analysts, and students of housing finance systems and their policy issues. We believe this rulemaking has the potential to significantly improve the existing standard.

Our fundamental recommendation is that the consideration of disparate impact issues must be able to include credit outcomes, i.e. default rates, not only credit underwriting inputs. Specifically:

1. Mortgage lenders, including smaller lenders, should have the option to use a credit outcomes-based statistical approach, as defined below, which qualifies as a valid defense under the Disparate Impact rule. This would improve the fairness, operation, and statistical basis of the rule.
2. HUD should develop a credit outcomes-based statistical screening approach that allows it to assess with a high degree of confidence, whether differences in mortgage lending results raise disparate impact questions for further review.

In both cases, the ability to use credit outcomes would enhance clarity and reduce uncertainty.

Problems with the Pure Input Approach

Applying its credit standards in a non-discriminatory way, regardless of demographic group, is exactly what every lender should be doing. Typically, the question of whether this is being carried out has been approached by looking only at inputs to a lending decision. This results in a focus on differing credit

approval/credit decline rates between protected and non-protected classes. The argument is then made that the existence of differing credit approval/credit decline rates between classes is evidence of discrimination even if a lender applies exactly the same set of credit underwriting standards to all credit applicants.¹

Some discussions of the disparate impact issue have analyzed different demographic groups by household income or other credit factors available at the time of loan application. While these factors can be indicators of future default rates, they are not the experienced, actual default rates. These actual rates may then be adjusted for differences in risk factors at loan origination (“risk-adjusted default rates”).² The fundamental insight is that there cannot be credit discrimination if actual risk-adjusted default rates are higher for the protected class alleged to be discriminated against than for the non-protected class. This should be a valid defense against disparate impact claims. The traditional pure input approach ignores the actual default rates, making it a flawed and inadequate measure. Moreover, it is difficult and expensive for a lender to defend against a claim of discrimination. A credit outcomes approach, using risk-adjusted default rates, has none of these flaws and therefore should be used.

The solution is to combine the credit outcomes approach with the input categories

The input-based approach relies heavily on HMDA approval-decline data. An outcomes-based approach would combine HMDA approval-decline data with the matching risk-adjusted default rates on closed mortgages.³ This would greatly enhance clarity and understanding and reduce uncertainty in the operation of the disparate impact rule. The default data should be organized by the same demographic categories as used in HMDA reporting.

If a protected class has a lower credit approval rate and therefore a higher credit decline rate than a non-protected class, a comparison of risk-adjusted default rates is in order.

In principle, there are three possible outcomes:

1. If a protected class has the *same* risk-adjusted default rate as the non-protected class, then the underwriting procedure was effective and the differing approval-decline ratios were appropriate and fair, since they resulted in the same default outcome. Controlling and predicting defaults is the whole point of credit underwriting. In this case, there is no evidence of disparate impact.
2. If the risk-adjusted default rate for a protected class is *higher* than for a non-protected class, that indicates that in spite of the fact that a protected class had lower credit approval and higher decline rates, it was in fact being given easier credit standards. Indeed, the process was

¹ [Alex J. Pollock and Edward J. Pinto](#), Reconsideration of HUD’s Implementation of the Fair Housing Act’s Disparate Impact Standard, August 17, 2018

² Risk factors that should be taken into account include: credit score, combined LTV, total DTI, loan type, loan purpose, tenure, loan term, documentation, and location. The AEI Mortgage Risk Index takes all but location into account. Location may be controlled for by taking account census tract income relative to area median income.

³ [Yezer](#) (2010) discusses the history of both approaches, along with common statistical challenges and previous investigations of discrimination in FHA mortgages.

evidently biased in its favor, not against it—even if this was not intended. Again, there is no evidence of disparate impact.

3. If on the other hand, a protected class's default rate is *lower* than that of a non-protected class, that may indicate that the protected class is experiencing a higher credit standard, even if this is not intended. This may be evidence of disparate impact and merits further examination.

Thus in principle, if the risk-adjusted default rate of a protected class is equivalent or higher than that of the non-protected class, then the claim of disparate impact disappears. If it is lower, further examination is required.

A Valid Defense for Lenders

In a disparate impact complaint, it should be a valid defense if a lender is able to demonstrate, using an outcomes-based statistical screening approach, that no statistically significant pattern of lower default experience was experienced by the protected class. This outcomes-based statistical screening approach should be incorporated as a valid defense under HUD's Disparate Impact rule.

Implementing an outcomes-based statistical screening approach should qualify as a valid defense under the Disparate Impact rule because by adjusting for differences in risk factors present at the time of loan origination, it renders lending outcomes directly comparable between the protected and non-protected classes. Should a lender's portfolio of adequately risk-adjusted loans, show no statistically significant difference in the default outcomes between a protected class and a non-protected class, said lender cannot be found to be discriminating. The lender ought to be able to provide the results of such screening approach as a valid defense to any disparate impact liability.

Statistical Considerations

The statistical examination of credit outcomes must take account:

- The normal statistical variation ("noise") in any group of statistics, especially for smaller groups.
- An appropriate statistical confidence level for performing this screening examination. We have suggested a confidence level of 90%. This results in a range of risk-adjusted default rates based on lender volume, protected class origination percentage, and other relevant factors. This is because sample size matters when determining statistical certainty. The uncertainty range is wider for smaller sample sizes and much narrower for very large sample sizes.⁴

We summarize the statistical factors the actual data will involve.

An outcomes-based statistical screening approach must address three statistical factors:

1. **Low default rates early in a loan's life:** default rates are typically low until a loan is seasoned a number of years, and even then rates might be in the single digits or teens.

⁴ Tom Hopper, "Sample Size Matters: Uncertainty in Measurement", <https://tomhopper.me/2014/11/21/sample-size-matters-uncertainty-in-measurement/>

2. **Loans must be risk adjusted:** Loan performance varies based on differences in risk factors present at the time of loan origination.
3. **There are over 5,000 HMDA-reporting lenders:** Annual loan origination volumes range from over half a million to one loan. Unless a lender has a large origination portfolio or has a medium sized portfolio with significant aging, the analysis needs to separate the statistical variation (or noise) around the results from evidence supporting possible disparate impact.

Our Recommended Outcomes-Based Statistical Screening Approach (OBSSA) Addresses All of These Factors

OBSSA addresses each factor as follows:

1. It applies a rigorous statistical screening approach that uses ever-to-date delinquency rates. While our approach takes into account differing default rates that result from varying levels of seasoning, ever-to-date 60-day delinquency rates (rather than current rates) are commonly used in mortgage analysis because they provide a larger number of default incidents for any given lender, which allows for a better evaluation of loans with a lower level of seasoning.⁵
2. Differences in loan performance due to differing initial loan characteristics can be addressed by risk-adjusting loans using the methodology outlined in a FHFA working paper that assigns each loan a stressed default rate based on the actual default experience of similar loans originated in 2006 and 2007.⁶
3. The wide range of lender origination volumes and default levels are taken into account, since OBSSA separates the statistical variation (or noise) around the results from evidence of possible disparate impact. This allows for a similar level of confidence to be determined regardless of lender size or default experience.

A successful approach must be implementable by both HUD and individual lenders.

Below we describe how HUD and mortgage lenders might implement OBSSA.

Detailed Outline of the Outcomes-Based Statistical Screening Approach

Using a simple model that requires only a few easily-accessible inputs, HUD (including FHA) can statistically determine with a 90% degree of certainty whether a difference in lending outcomes for a given lender merits further desk review. The facts and circumstances around those lenders who do not

⁵ The Bureau of Consumer Financial Protection's *Notice of reopening of comment period and request for comment* from May 2012 states that "The Bureau believes that loan performance, as measured by delinquency rate such as [ever]-60 days or more delinquent, is an appropriate metric to evaluate whether consumers had the ability to repay those loans at the time made."

⁶ Davis, Larson, and Oliner "Mortgage Risk Since 1990" (FHFA Working Paper 19-02).

pass the statistical screening should then be examined further in terms of discriminatory acts or practices, since there may be underlying explanatory facts or circumstances.

Required data:

Loan-level data for a given lender in a given year with the following information:

- Protected class status,
- Default experience--binary variable, can be defined (ever-to-date 60-day or more delinquency rates (ED60+), ever-to-date 90-day or more delinquency rates, etc., as needed), with at least one default for either class
- Borrower risk characteristics
- Note rate
- The property's census tract (or in particular, the census tract's income percentage relative to area median income.)

FHA has the required administrative data needed to match HMDA to its own loans to pull in HMDA categories for FHA loans. The loan performance data of the matched loans then need to be adjusted for differences in borrower note rates and the relevant ex ante risk factors. For that, the differences in risk factors present at the time of loan origination (such as credit score, debt-to-income ratio, combined loan-to-value ratio, etc.) should be grouped and interacted to capture the effects of risk layering.⁷ This methodology can be successfully tested by HUD using the FHA data.

The tract income percentage relative to area median income will allow for a comparison of default rates across geographies with different economic trajectories. The data should be grouped by tract income as well (i.e. low income $\leq 80\%$, moderate income $> 80\% \ \& \ \leq 120\%$, and high income $> 120\%$). Though of lesser importance during an economic expansion, this variable will control for idiosyncratic risk in certain localities more affected by job losses during an economic downturn. These data come from the Federal Financial Institutions Examination Council (FFIEC) and are already merged on to the HMDA dataset.

Regression analysis approach:

- 1) Run a logistic regression (logit) determining the relationship between a default outcome (binary: default or not default) and protected class status as defined by law. The regression equation should include a term for the mortgage risk score and a dummy variable for tract income group to control for risk characteristics and differences in locations.

⁷ See the data supplement in Davis, Larson, and Oliner "[Mortgage Risk Since 1990](#)" (FHFA Working Paper 19-02) for binned default tables by loan type, product type, income documentation, amortization status, credit score range, CLTV range, DTI range.

2) Translate the established relationships into a risk-and location-adjusted predicted default rate for each class using an average marginal effect.⁸

3) Screen whether the actual default rate, when risk-and location-adjusted for the protected class, is statistically lower than that rate for the non-protected class with a 90% degree of certainty (in other words, we can be 90% certain on a statistical basis whether the lender is discriminating or not).

For the logistic regression to run, there has to be at least 1 observed default for both the protected class and the non-protected class. The logit could also be run by pooling multiple years and/or lenders, which would then require including the lender-protected class status interaction term in the margins command. For more about the logit, please see Appendix A.

Real World Results for 2015 FHA Loans

We assembled a dataset of around 157,000 FHA purchase loans in 2015, which represents about 20% of FHA total purchase loans in that year. The data include the ED60+ and protected class status. We judge this dataset to be representative of the overall FHA book in that year. We find a default rate of 11.7% for the protected class and 7.4% for the non-protected class. These percentages change slightly after adjusting for differences in risk between both classes. It falls for the protected class, granting it more lenience for defaults because of the relatively higher risk index of that group, and it rises for the non-protected class, granting it less lenience for defaults because of the relatively lower risk index of that group (see Table 1).

Table 1: Estimated 2015 FHA purchase loan default rates and risk-adjusted rate: by class

	Default Rate	Risk-adjusted default rate
Protected class	11.7%	10.9%
Non-protected class	7.4%	7.7%

Note: Default rates are Ever D60+. Based on a 20% sample of 2015 FHA loans.

Table 1 shows there is no indication of discrimination for the entire FHA book, as the risk-adjusted default rate for the protected class is substantially greater than the risk-adjusted default rate of the non-protected class. In fact, the probability that there was discrimination in 2015 for FHA purchase loans is less than 1% if our data are representative.

⁸ The exact procedure used for calculating the predicted default rates for the protected and unprotected class can be illustrated by focusing on loans of just the protected class. First, one needs to calculate the predicted default rate for every loan in the dataset using its actual mortgage risk and tract income percentage group, setting class to protected, and then taking the average of the predicted default rates. Then, one needs to repeat this process for the unprotected class. This process ensures that the predicted default rates show the marginal effects of changing class.

Simulation of the Credit Outcomes Approach

The findings in the previous section, however, do not rule out that individual lenders may have discriminated. Our dataset is lacking an identifier for individual lenders. Therefore, to test for the possibility of an individual lender discriminating, we turn to a simulation that aims to determine the minimum level of non-protected class risk-adjusted default rate that would allow HUD to establish with a 90% certainty whether a lender of varying size is discriminating.

The data are based on a loan-level dataset that is modeled closely around the 2015 FHA purchase loan book and this book's aggregate default experience (measured as ED60+). For more on the dataset and simulation procedure see the Appendix B.

Significance Testing

We begin with a hypothetical pattern of clear discrimination against the protected class by assuming a protected class default rate of 11.7% (the same as the aggregate rate for the 2015 FHA purchase loan ED60+ rate) and a non-protected default rate of a purely theoretical 100%. In this instance, this statistical screening result in lending outcomes for this hypothetical lender would obviously merit further review, given that the difference in risk-adjusted default rates between the classes is statistically significant with respect to the non-protected class. We then gradually reduce the non-protected class's risk-adjusted default rate at each step performing the statistical screening outlined above. Until the test shows that the difference in risk-adjusted default rates between protected and non-protected classes is no longer statistically significant. Then we stop and record the minimal number of non-protected class defaults required to prove statistical significance at the 90% level.

Results

Table 2 shows, the non-protected class risk-adjusted default rate required to determine discrimination with a 90% degree of certainty level for lender of varying sizes. For smaller lenders, the minimum threshold is larger than for larger lenders, which stems from wider confidence bands that arise from a smaller sample of loans, which introduce noise.⁹ For very large lenders, the minimum threshold is lower than for the smaller lenders, because the large number of loans provide less noise and therefore larger certainty in the findings.¹⁰ (For more on the simulation see appendix B.)

For example, for a lender with 10,000 loans, the non-protected class risk-adjusted default rate would have to be 11.4% or more to indicate discrimination at a 90% certainty for this lender. Given that we estimate a non-protected class default rate of 7.4% (albeit for the aggregate), it is very unlikely that many lenders – if any-- are currently discriminating against the protected class.

⁹ Supra, Tom Hopper

¹⁰ Ibid. Tom Hopper

Table 2: Results of the Simulation

Lender size (# of loans)	Protected class default rate	Minimum non-protected class risk-adjusted default rate required to determine discrimination with a 90% degree of certainty*
100	11.7%	20.0%
500	11.7%	14.7%
1,000	11.7%	13.5%
5,000	11.7%	11.8%
10,000	11.7%	11.4%
50,000	11.7%	10.9%
100,000	11.7%	10.7%
500,000	11.7%	10.6%

* We estimate an actual risk-adjusted default for the non-protected class of 7.7% for the aggregate. Note: results in column 3 are median values for 100 simulations for lenders with non-protected class shares of 20%, 30%, and 40% respectively.

Conclusion of the Analysis

For 2015 FHA purchase loans, it is highly likely that FHA lenders in the aggregate did not discriminate against the protected class. For both larger lenders and smaller ones, the analysis indicates that it is highly likely that they did not discriminate. While there may be some individual smaller lenders who may be discriminating, statistically this is a very small probability.

As discussed above, HUD with its own data, could identify instances where a difference in lending outcomes for a given lender merits further review.

In summary

We believe that the addition of a credit outcomes-based option would significantly improve the proposed rule.

We would be pleased to address any questions and the AEI Center on Housing is ready to assist with further statistical analysis.

Respectfully submitted,

Edward J. Pinto
Resident Fellow and Director
AEI Housing Center
Edward.pinto@aei.org
240-423-2848

Tobias J. Peter
Senior Research Analyst
AEI Housing Center
Tobias.peter@aei.org
202-419-5201

Alex J. Pollock
Distinguished Senior Fellow
R Street Institute
apollock@rstreet.org
202-900-8260

Appendices

Appendix A: Logit regression

Logit regression equation for an individual lender in a given year:

$$Default = b_0 + b_1 Protected + b_2 MRI + b_3 Location + b_4 Rate + e$$

where *Default* represent a binary default occurrence, whichever way defined, *Protected* is a dummy variable for protected class borrowers, *MRI* represent the mortgage risk index score, which combines a borrower's credit score, CLTV, and DTI into a single metric, *Location* is a set of dummy variables for the census tract income group, and *Rate* is the loan's note rate. This equation should be estimated as a logit regression with robust standard errors. The ensuing margins command should be estimated for the *Protected Class* dummy variable.

Logit regression equation for multiple lenders over various years:

$$Default = b_0 + b_1(Protected * Lender * Yr) + b_2 MRI + b_3 Location + b_4 Lender + b_5 Yr + b_4 Rate + e$$

Where *Lender* represents a set of dummy variables for each lender and *Yr* represents a set of dummy variables for the year the loan was originated. The ensuing margins command should be estimated for the *Protected * Lender * Year* dummy variables.

Appendix B: Simulation

Data:

The data for the simulation are modeled as closely as possible around what we believe 2015 FHA portfolio looks like. For that we used a sample of matched FHA 2015 purchase loans with actual default experiences (see appendix table 1). For simplicity, we assume that every lender originates loans in the same tract income group. The protected class default rate is set to 11.7% (comparable to the 2015 FHA purchase loan ED60+ rate) and initially the non-protected class default rate to 100%, which we will gradually lower.

Methodology:

With the given data, we perform the statistical screening outlined in the model above. If the lender is found to be discriminating against a protected class, we change one loan for the non-protected class from default to non-default. After each such reduction, we repeat the statistical test and if the lender is still found to be discriminating at a statistically significant level, we continue. Once the test shows that the difference in risk-adjusted default rates between protected class and non-protected class borrowers is no longer statistically significant, we stop and record the minimal number of non-protected class defaults required to prove statistical significance at the 90% level.

Robustness check:

For robustness of our results, we vary the share split of protected and non-protected classes by lender size, which we find to differ in the real world (see appendix chart 1). While the average protected class share in 2015 for all lenders was around 30%, smaller lenders tend to have smaller protected class shares, while larger lenders tend to have higher shares (see appendix table 2).

Results:

For smaller lenders, to prove discrimination, the risk-adjusted non-protected class default rate would have to be around 1.5 times as high as the protected class default rate. For larger lenders, the risk-adjusted non-protected class default rates would have to be a bit below the protected class default rate (see Table 2). The explanation for this difference lies in the wider confidence bands that arise from a smaller sample of loans. These results can vary a bit depending on borrowers’ risk profile but they provide a useful illustration of how even an appearance of discrimination by a small lender may in the end prove to be not statistically significant.

Appendix Table 1: Breakout of hypothetical purchase loan-level test file by tract income group, class status, and default experience

Default	Risk Score (MRI)		Risk Score (MRI) - Standard Deviation		Count		Default Rate	
	Protected class	Non-Protected class	Protected class	Non-Protected class	Protected class	Non-Protected class	Protected class	Non-Protected class
Y	28.9%	27.2%	9.5%	9.5%	28,083	41,248	11.7%	7.4%
N	24.3%	21.9%	9.9%	9.7%	211,917	518,752		

Note: Protected class defined as Black or Hispanic, while non-protected class defined as non-Hispanic white. Modeled around FHA’s 2015 loan-level book and default experience. Default is defined as Ever D60+.

Appendix Table 2: FHA 2015 Purchase Loan Lender Breakouts

Lender size	# of lenders	Median loan count	Median protected class share of originated loans	Median difference in non-protected class vs. protected class approval rate (in ppts.)
All	1325	133	23%	5.7
> 10k	9	11739	32%	6.8
> 5k & ≤ 10k	15	5612	40%	6.0
> 1k & ≤ 5k	144	1739	30%	6.2
> 500 & ≤ 1k	129	672	28%	6.0
> 100 & ≤ 500	450	202	23%	5.7
> 50 & ≤ 100	225	71	20%	5.1
≤ 50	353	27	17%	4.4

Note: For lenders with at least 10 originated loans. Approval rates are based on at least 10 originated protected class and at least 20 total loans.

Source: HMDA 2015