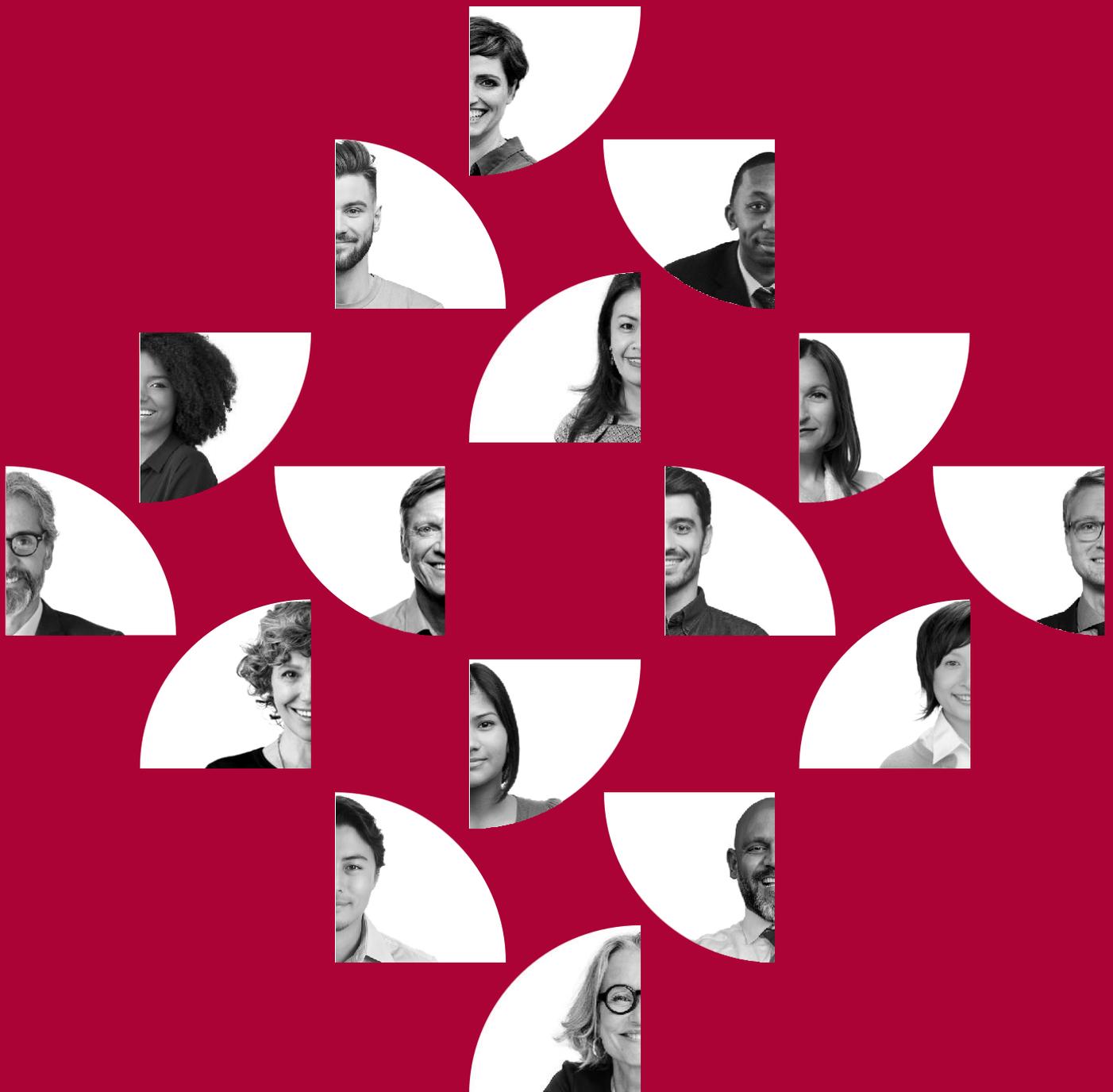# EDGE UNEXPLAINED GENDER PAY GAP (UGPG) ANALYSIS

**METHODOLOGY V1.1**
**1 JULY 2020**

**CONTENTS**

## 1. PURPOSE

The purpose of this document is to describe in detail the method prescribed by the EDGE Certified Foundation to determine whether there are unexplained pay gaps between women and men.

This method is generally applied as part of the preparation for EDGE Certification. It is also used independently to investigate gender pay equity and demonstrate an organization's compliance with the relevant country-level legislation on pay equity, such as in Switzerland.

Throughout this document, the method prescribed by the EDGE Certified Foundation is referred to as an EDGE-compliant analysis. The Unexplained Gender Pay Gap (UGPG) is referred to as UGPG. For further information concerning the method, please contact:

Email: aunguresan@edge-cert.org
Website: www.edge-cert.org
Mailing address:      Ms Aniela Unguresan
                    Co-founder
                    EDGE Certified Foundation
                    Vorstadt 2, 6300 Zug
                    Switzerland

### 1.1. Scope of Methodology

The method prescribed by the EDGE Certified Foundation has been developed for the purposes of achieving EDGE Certification. An organization is encouraged to also apply the methodology independently, to determine whether there are any unexplained pay gaps between women and men and to demonstrate compliance with the relevant country-level legislation on pay equity.

When applying the method prescribed by the EDGE Certified Foundation to demonstrate compliance with legislation on pay equity, organizations may adapt the following elements of the method to comply with regulatory/legislative criteria:

- the selection of a reference period;
- which employees are included in the analysis, and
- the inclusion/exclusion criteria for and categorization of base pay and bonuses/benefits.

However, adapting these elements of the methodology may result in the need for an organization to conduct two different analyses, one for the purposes of achieving EDGE Certification and one for the purposes of complying with relevant country-level legislation on pay equity.

## 2.  DATA REQUIREMENTS

### 2.1.  Selection of a reference period

To perform an EDGE-compliant analysis, the organization chooses the most recent 12-month reference period against which the assessment will be conducted. The reference period may be a calendar year, fiscal year or some other period that is relevant to the organization for internal reporting purposes. The first reference period becomes the baseline for assessing organization progress. Therefore, the months chosen for the reference period should not change in subsequent reference periods unless there are exceptional circumstances like a merger or acquisition that results in a change to an organization's reporting period, or in the case of a global approach, the alignment of multiple country-entity reference periods.

To be considered EDGE-compliant, no more than 12 months should have passed since the end of the reference period used for the EDGE-compliant analysis.

### 2.2.  Employees included in the analysis

To ensure there is sufficient statistical power to achieve credible results, an EDGE-compliant regression analysis requires a minimum of 100 observations, e.g., uses relevant data on at least 100 employees.

Anyone on the payroll of the organization (e.g., the country-level entity) undergoing the analysis that is employed on a work contract of longer than six months during the reference period should be included in the analysis, including employees of acquired organizations, where the acquired organization has been fully integrated on a legal basis into the acquirer, who are employed on the same policies and practices as the organization.

### 2.3.  Dependent variables included in the analysis

The regression analysis shall be run as a set of two regression analyses. The first shall consider 'Salary' (base salary) as the dependent variable and the second shall consider 'Pay' (base salary plus bonuses and other cash benefits) as the dependent variable. Inclusion and exclusion criteria for Salary and Pay are shown in Table 1 below.

As a general principle, if applicable, the organizations shall follow regulatory/legislative criteria for inclusion/exclusion and categorization of base pay and bonuses/benefits for relevant jurisdiction.

1.   Salary (base salary) and Pay (base salary plus bonuses and other cash benefits) shall be actual (realized) remuneration – e.g., remuneration that has actually been paid to employees during the Reference Period, not target salary or pay (bonuses/benefits) not yet paid.

2.   Salary (base salary) and Pay (base salary plus bonuses and other cash benefits) paid during the Reference Period for specified periods of time of less than or in excess of a year shall be pro-rated to annualized amounts based on contractual hours.

   For example, person A was employed on 1st March at 100% with an annual base salary of 60,000. Her annualized salary for the full calendar year is 60,000.

3. Salary (base salary) and Pay (base salary plus bonuses and other cash benefits) paid during the Reference Period for less than full-time work shall be pro-rated to Full-Time Equivalent amounts based on contractual hours.

For example, person B is employed part time, at 70%, with an annual base salary of 68,000. His annual base salary pro-rated to 100% is

$$\frac{68,000}{7} * 10 = 97,143.$$

Table 1. Salary and Pay inclusion and exclusion criteria

| INCLUDED | | EXCLUDED |
|---|---|---|
| **SALARY** | **PAY** | |
| ✓ In principle, all that is considered "standard" or "ordinary" for that kind of job or position | ✓ In principle, Base Salary plus all that is negotiable, variable and/or on top of the Base Salary; including recurring payments | ✗ In principle, one-off incentives or bonuses, e.g., signing bonus, "golden hello", or lump sum payment at beginning of a work contract to compensate for lost potential bonus at previous employer; redundancy or termination payments |
| ✓ Gross salary specified in employment contract, e.g., before tax, before other deductions (e.g., employee pension contributions, salary sacrifice/salary packages) | ✓ If unable to distinguish between "bonus" payments and "overtime" and/or "shift" pay/premiums, then include overtime and/or shift pay/premiums in bonus calculation | ✗ Overtime or shift pay/premiums<br>✗ Expense reimbursements<br>✗ Non-'cash' benefits (e.g., company cars; health (e.g., private) insurance) |
| Extras/Allowances that are considered standard for the type of job/position:<br>✓ On-call allowances/extras (e.g., for emergency workers)<br>✓ Allowances and/or awards that are non-negotiable and applied uniformly to employees for that kind of position | Bonuses based on performance (organization's and/or employee's):<br>✓ Incentives<br>✓ Commissions<br>✓ Awards with a monetary value (e.g., vouchers, securities & securities options/interests –shares, bonds, debentures and futures)<br>✓ Full amount of bonuses awarded in a particular year but payable over multiple years | ✗ Deferred bonus plans (e.g., not incurring a charge to income tax during the Reference Period)<br>✗ Allowances that are applied to ALL eligible employees AND are non-negotiable AND are non-job/non-position related (e.g., per child allowances; location-specific weightings/allowances) |
| ✓ Maternity/Parental/Paternity pay<br>✓ Sick Pay<br>✓ Holiday Pay<br>✓ Other special leave pay (e.g., long-service leave; carer's leave; etc.) | ✓ Extras/Allowances that are negotiable | ✗ Pay in lieu of annual or other leave |

2.4. Explanatory variables included in the analysis

There are two possible levels of complexity permitted when it comes to performing EDGE-compliant regression analyses:

1.　　Assessing potential unexplained gender pay gaps using a ***Standard Regression Analysis***. A ***Standard Regression Analysis*** is an EDGE-compliant regression analysis that measures salary and pay against the common set of variables described below under ***Standard Variables***.

2.　　Assessing potential unexplained gender pay gaps using a ***Customized Regression Analysis***. A ***Customized Regression Analysis*** is an EDGE-compliant regression analysis that measures salary and pay against the common set of ***Standard Variables*** and additional organization-specific variables – which <u>shall only be permitted</u> if an organization performs the Standard Regression Analysis first or in conjunction with the customized analysis.

   a) Performing a Customized Regression Analysis with more variables than the list of standard variables, that are tailored to an organization's specific characteristics, means that a pay equity model can take account of more factors that might affect pay at the organization under examination. Statistically, this is likely to reduce the number of issues that may produce an unexplained gender pay gap.

   b) In the case of a Customized Regression analysis, the EDGE Standard tolerance thresholds shall be reduced – the requirements for which are set out in section 3.3 below.

Regression analysis uses independent variables to compute the relationship between the dependent variable (pay), the explanatory variable (gender) and a collection of different factors (characteristics related to the employee and the workplace). For statistical and computational purposes, independent variables are transformed into 'predictors' when writing the code for the analysis.

The total number of predictors will depend on whether a variable is considered:

   a) 'Numerical' – a real number, e.g., age or tenure. Each numerical variable counts as one predictor.

   b) 'Categorical' – a variable that uses two or more possible values, usually of a qualitative nature, e.g., level of responsibility or department:

   i. Where a categorical variable has only two possible values, it is coded as a single binary variable and is counted as one predictor.

   ii. Where a categorical variable has more than two possible values (categories), it is transformed into a collection of binary variables, and the number of additional predictors will be the number of categories minus one (K-1).

***Standard Variables***

A Standard Regression Analysis shall measure salary and pay (as defined above) against the following pre-defined variables:

1. *Gender* – the explanatory variable, coded as binary (0 for Male, 1 for Female), which counts as one predictor.

2. *Tenure* – a numerical variable usually inputted as years of service in the organization, which counts as one predictor.

3. *Age* – a numerical variable, which counts as one predictor.

4. *Age^2* – age squared – a numerical variable included for statistical purposes, which counts as one predictor.

5. *Responsibility of the role* – a categorical variable coded as either Individual Contributor (IC) or People Manager (PM), which counts as one predictor.

   i. An Individual Contributor is an employee who is a team member; frontline staff member; a solo operator like a consultant, lawyer, researcher, literary agent, software engineer; and so on: an employee who is not responsible for managing or supervising any other employees.

   ii. A People Manager is an employee in a supervisory role, responsible for managing one or more people, and could include being a line manager; performance manager; project or team leader; section, division, or department head; and so on.

6. *Type of performed function* – a categorical variable coded as either Support Function (SF) or Core Function (CF), which counts as one predictor.

   i. A Support Function is a supporting activity carried out by the organization in order to enable or to facilitate the core business functions (e. g. legal services, accounting, book-keeping and auditing, human resources management, payroll management, procurement functions). The outputs (results) of support business functions are not themselves intended directly for the market or for third parties.

   ii. A Core Function is defined as an activity of an organization yielding income: the production of final goods or services intended for the market or for third parties. Usually the core business functions make up the primary activity of the organization (e.g., client facing roles, production).

7. *Level of responsibility* – a categorical variable with up to five levels, which counts as four predictors. The organization may choose to code this variable with four levels, in which case it counts as three predictors.

   i. Top management – reporting formally and directly to the CEO.
   ii. Upper management – reporting formally and directly to the top management.
   iii. Middle management – reporting formally and directly to the upper management.
   iv. Junior management – reporting formally and directly to the middle management.
   v. Operational level – all other employees.

Therefore, dependent on whether four or five levels of responsibility are included, a Standard Regression Analysis shall have between 9 and 10 of the above pre-defined standard predictors, including the explanatory predictor of gender.

### *Customized Variables*

A Customized Regression Analysis shall measure salary and pay (as defined above) against the standard variables plus additional variables defined by the organization in line with the organization's compensation and benefits structure. For example, but not limited to:

1. *Level of responsibility* may be separated into more levels (and therefore more predictors) than the EDGE five levels of responsibility.

2. *Education / qualifications*

3. *Business unit*

4. *Department*

5. *Geographic area*

A Customized Regression Analysis shall not include any discriminatory variables, such as race, nationality or part-time, or any other variables that are not strictly related to the skills and/or competencies of the employee, or the nature of the job.

## 3. EDGE-COMPLIANT ANALYSIS METHODOLOGY

An EDGE-compliant analysis shall be performed using Ordinary Least Squares (OLS) linear regression. Linear regression is a statistical method that can quantitatively demonstrate whether an organization has any pay gap which cannot be explained by any other factors than gender, e.g., unexplained gender pay gaps. The following section sets out the methodology behind this analysis.

### 3.1. Linear regression analysis

Linear regression is a statistical method that mathematically explains a chosen dependent variable (e.g., an employee's wage) in terms of a selection of independent predictors (e.g., gender, age, level of responsibility, etc.). Use of this method allows an organization to assess the extent to which an employee's salary or pay is influenced by gender after accounting for all other factors.

More specifically, linear regression expresses the dependent variable as a linear function of the independent predictors under certain assumptions about their statistical distribution. The general form of the regression equation is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

where:

- $y_i$ is the dependent variable for a given employee $i$, e.g., salary or pay.

- $\beta_0$ is a constant representing the average of the response $y_i$ when all independent predictors are set to 0.
- $\beta_1 \dots \beta_p$ are the coefficients of the regression line associated with the $1 \dots p$ independent predictors.
- $x_{1i} \dots x_{pi}$ are the values of the $p$ independent predictors for a given employee $i$, e.g., the gender, age, level of responsibility, etc. for a given employee.
- $\varepsilon_i$ is the error term of a given employee $i$, the difference between the predicted dependent variable and its true value (for example, between an employee's actual salary and their salary according to the linear regression equation).

The OLS approach calculates the values of the coefficients $\beta_1 \dots \beta_p$ required to minimize the error terms $\varepsilon_i$.

By making the statistical assumption that the distribution of the error terms $\varepsilon_i$ is normal with zero mean and constant variance, p-values and confidence intervals can also be used to indicate the level of uncertainty associated with the coefficient estimates.

### *Standard Regression Analysis*

The form of the regression equation adopted in the EDGE-compliant Standard Regression Analysis for salary shall be

$$\ln(Salary_i) = \beta_0 + \beta_1 Gender_i + \beta_2 Tenure_i + \beta_3 Age_i + \beta_4 Age_i^2 + \beta_5 Responsibiliy\ of\ Role_i + \beta_6 Type\ of\ Performed\ Function_i + \beta_7 Level\ of\ Responsibility_i + \varepsilon_i$$

where:

- $\ln(Salary_i)$ is the log-transformed salary of a given employee $i$.

- $Gender_i,\ Tenure_i,\ Age_i$ etc. are the values of the standard predictors described in section 2.4 above for a given employee $i$.

Similarly, the form of the regression equation adopted in the EDGE-compliant Standard Regression Analysis for pay is

$$\ln(Pay_i) = \beta_0 + \beta_1 Gender_i + \beta_2 Tenure_i + \beta_3 Age_i + \beta_4 Age_i^2 + \beta_5 Responsibiliy\ of\ Role_i + \beta_6 Type\ of\ Performed\ Function_i + \beta_7 Level\ of\ Responsibility_i + \varepsilon_i$$

where $\ln(Pay_i)$ is the log-transformed pay of a given employee $i$.

### *Customized Regression Analysis*

The form of the regression equation adopted in the EDGE-compliant Customized Regression Analysis shall be as in the standard regression analysis above, but with the addition of predictors defined by the organization in line with the organization's compensation and benefits structure.

### 3.2. Reported results of the regression analysis

An EDGE-compliant regression analysis shall report several metrics and statistics to inform the assessment of wage equality between men and women. These metrics and statistics are detailed in the following section.

### *Regression Coefficients*

Estimated values for the coefficients of the Standard and (if performed) Customized Regression analyses shall be reported. These values are used to calculate the ***Effect of gender or UGPG*** as described below.

### *Adjusted R-Squared*

The Adjusted R-Squared value is an output of most statistical software packages and is a measure of model fit. More specifically, the Adjusted R-Squared value measures the amount of variation in the data that is explained by the linear regression model. Values close to one indicate that variation in employees' salary and pay can be almost entirely explained by the predictors included in the Standard or Customized Regression analysis. On the other hand, values close to zero indicate that salary and pay cannot be adequately explained by the regression analysis and are likely to be driven by factors outside of the analysis.

From an EDGE compliance perspective, when an organization performs a Customized Regression after or in conjunction with a Standard Regression, the adjusted R-squared value for the Customized Regression *shall not* be lower than that produced by the Standard Regression.

If the inclusion of customized predictors lowers the adjusted R-squared value, then a different subset of customized predictors shall be selected and the regression re-run.

> a) If no subset can be found that leads to a higher adjusted R-squared value, the organization shall exclude the results and use only the results of the Standard Regression.

### *Observations*

The number of observations included in the Standard or Customized Regression Analysis shall be reported. This is the number of employees whose data were input for analysis and should correspond to the total numbers of employees in the organization as defined in section 2.2.

### *Number of predictors*

The number of predictors in the Standard or Customized Regression Analysis shall be reported. This value refers to the number of predictors used in the analysis, which corresponds to the number of variables that were included in the regression model. Categorical variables are treated as multiple predictors for computational purposes, the total number of which depends on the allowed values.

### *Effect of gender OR UGPG*

To estimate the UGPG, the estimated coefficient $\beta_1$ for the gender predictors should be transformed using the following equation:

$$(\exp(\beta_1) - 1) * 100$$

This formula expresses the effect of gender on salary or pay as a percentage. Its value can be interpreted as the average percent increase (or decrease) in salary or pay for men vs. women, where:

> a) A positive (+) value denotes a pay gap in favour of women.

b) A negative (-) value denotes a pay gap in favour of men.

For example, if the Regression Analysis is run and the coefficient for gender is found to be -0.15,

$$(\exp(-0.15) - 1) * 100 = -13.93.$$

This indicates that when all other personal and organization-related factors are accounted for, women are paid on average 13.93% less than men.

### *Threshold*

The Threshold is the EDGE Standard against which the effect of gender is compared, based on the type of regression (e.g., Standard or Customized) that was used. The EDGE Standard is described in section 3.3 below.

### *Effect of gender greater than threshold*

A statistical test at 5% significance level can be performed to determine whether the effect of gender is significantly outside of the EDGE Standard Threshold. As described below in section 3.3 below, the EDGE Standard does not require any statistical test for the significance of the UGPG.

### 3.3. The EDGE Standard for UGPG

To pass the EDGE Standard, the UGPG shall be calculated as a percentage, as described above, and compared to the Threshold.

1.  A positive value denotes a gender pay gap in favour of women.

2.  A negative value denotes a gender pay gap in favour of men.

For a Standard Regression Analysis, the Threshold shall be +5% or -5%. This means that for the results to pass the EDGE Standard, the percentage results for both regression analyses ('Salary' and 'Pay') must fall within this range.

The EDGE Standard does not require any statistical test for the significance of the unexplained gender pay gap. An organization may, however, find statistical tests of significance helpful in understanding to what extent the UGPG represents a meaningful gap.

For a Customized Regression Analysis, there are more organization-specific factors that are taken into account. Assuming these add statistically meaningful information to the analysis, logically the actual unexplained gender pay gap should be approaching zero. Therefore, the Threshold shall be reduced by 0.25% for every additional predictor added, up to a maximum of 20 additional predictors. This means that for the results to pass the EDGE Standard, the range of percentages for both runs ('Salary' and 'Pay') shall be narrower than +/-5%.

For example: if four (4) additional predictors are added to a Customized Regression Analysis, then the Threshold is reduced by 4 x 0.25% = 1% from +/-5%. This means that to pass the EDGE Standard the results must be within the range of +4% to -4%.

Similarly, if three (3) additional predictors are added, the Threshold is reduced by 3 x 0.25% = 0.75%, meaning that to pass the EDGE Standard the results must be within the range of +4.25% to -4.25%.