Age Check
Certification Scheme

Technical Requirements for Age Estimation Technologies

ACCS 1: 2020

Superseded



Age Check Certification Services Ltd

UKAS PRODUCT CERTIFICATION 21037

© Age Check Certification Scheme Ltd (2021) All Rights Reserved



This is a publicly available specification created by the Age Check Certification Scheme Ltd. It is subject to the intellectual property rights of the Scheme and may not be copied, used in a retrieval system or utilised without the express consent of the Scheme, save that it may be mentioned by name as a reference document with appropriate attribution and a link to the document itself.
© Age Check Certification Scheme Ltd 2021. All rights reserved.

Superseded





Superseded

Contents

Introduction 3

1. Scope 5

2. Normative References 9

- Legal Provisions 10
- National and International Standards 10
- Other Documents 10

3. Terms and Definitions 13

4. Target of Evaluation 19

- Scope of the Age Estimation Technology Under Evaluation 20
- Hardware Requirements for Age Estimation Technology Under Evaluation 20
- Environmental Requirements for the Target of Evaluation 21
- Assessment of Results 22

5. Testing Protocols 25

- Sample Size 26
- Sample Population 26
- Acceptable Test Results 27
- Mean Absolute Error 27
- Upper Tolerance Level 28
- Absolute Tolerance Level 28
- Inherent Bias 28

Annex 1 – Sample Size Calculation (Normative) 32

Annex 2 – Table of Tolerance Levels (Informative) 33

Annex 3 – Example Test Results (Informative) 34

- Example 1 35
- Example 2 36
- Example 3 37
- Example 4 38
- Example 5 39
- Example 6 40

Superseded

Introduction

Age Estimation is the analysis of a citizen's age based on biometric features, such as facial image analysis, although age estimation can be accomplished using different biometric traits. This standard is intended to provide assurance for users of age estimation technologies when deployed as age check gateways, that they will work as intended.

Traditionally, age estimation is undertaken by natural persons, in a wide range of settings, all of the time. These natural persons make a judgment, based on a person's appearance, whether they are old enough to purchase age-restricted goods, content or services.

Some examples of products whose sale is controlled by age restriction are:

- alcohol
- cigarettes, including E cigarettes (and vaping products)
- lottery tickets and scratch cards
- dangerous weapons, e.g. crossbows, knives

With the advancement of technology, the ability to undertake age estimation electronically is fast emerging. The Age Check Certification Scheme (ACCS) has put together this standard to determine if age estimation technology is fit for deployment for a given challenge age category. For example, a Challenge 25 category means that anyone younger than 25 should be challenged for ID to ensure that they are over 18.

The standard envisages that age estimation technology is rapidly advancing and accuracy levels are always improving. In setting requirements around accuracy levels, these are assessed on the basis that of the minimum 'challenge age' that the technology is fit and safe to be deployed in is identified. So, for instance, a particular age estimation technology may 'pass' and be certified as fit for use at 'Challenge 25' or 'Challenge 28' or indeed any other age.

It is worth noting that the applicable tolerance levels are much wider the older the challenge age, so it is intended that users seeking to commission this type of technology as a part of their age verification processes, can have greater confidence in those certified with a lower challenge age category.

The methodology used to assess the accuracy of the technology have been developed in conjunction with Chartered Statisticians and considered by regulators, trade bodies and interested parties as an appropriate methodology.

Superseded



Superseded



1

Scope

Scope

Superseded



1. Scope

The Age Check Certification Scheme (ACCS) Standards are applicable to Scheme Clients submitting age estimation technologies for certification from ACCS and wish to have access to use of the ACCS certification mark as a mark of conformity for the technology.

This part of the ACCS Standards:

- establishes the definitions applicable to age estimation technology testing
- identifies the target of evaluation for age estimation technology testing
- sets the test protocols for biometric data subject capture
- identifies the key metrics of mean absolute error, upper tolerance level and absolute tolerance level
- sets the sample sizes to be undertaken for given levels of confidence and reliability required
- establishes the requirements for age estimation technologies to demonstrate conformity and be fit for deployment
- secures that inherent bias on the grounds of gender or skin tone in age estimation technology is below an acceptable tolerance

The suite of ACCS Standards are applied as applicable to the activities of clients certified by ACCS. This will be dependent on the scope of operations of the providers and the services that they provide.

Superseded



Superseded





Superseded



2

Normative References

Normative References

Superseded



2. Normative References

Legal Provisions

National and International Standards

ISO 4246:1994 – Cinematography – Vocabulary.

ISO/IEC 19794-1:2006 – Information technology – Biometric data interchange formats – Part 1: Framework.

ISO/IEC 19795-6:2012 – Information technology – Biometric performance testing and reporting – Part 6: Testing methodologies for operational evaluation.

Other Documents

Fitzpatrick, T. B. (1975). "Soleil et Peau" [Sun and Skin]. Journal de Médecine Esthétique (in French) – The 'Fitzpatrick Scale'.

Superseded



Superseded





Superseded



Terms and Definitions

Superseded



3. Terms and Definitions

In this document:

"shall" indicates a requirement

"should" indicates a recommendation

"may" indicates a permission

"can" indicates a possibility or a capability

GUIDANCE NOTES are shown in italic text and are intended to assist the reader with understanding provisions.

When referring to the ACCS Standards, refer to the ACCS Standard, followed by the year of issue, followed by the provision – such as **ACCS 0:2021, 4.3.**

3.1 Absolute Tolerance Level	means the level at which the age estimation technology will not return an age estimation above during the evaluation.
3.2 Age Estimation Technology	means a system or process that captures inherent biometric features of a citizen and provides an estimation of the age of that citizen by reference to an artificial intelligence algorithm.
3.3 Algorithm	means a sequence of instructions that tell a biometric system how to solve a particular problem, in this context the estimation of the age of the biometric data subject [ISO/IEC 19794 1:2006, 3.3 modified for context].
3.4 Ambient Lighting	means the illumination within an area produced by scattered light or general lighting conditions [ISO 4246:1994, 15]. ***Note:*** *Ambient light can be natural or artificial light. Ambient light is generally uncontrolled and can be highly variable, posing a possible risk to image capture quality.*
3.5 Biometric Data Subject	means an individual whose individualised biometric data is captured by the age estimation technology.
3.6 Challenge Age	means a term used to describe the age at which a provider of age restricted goods, content or services may cease to require a potential customer to prove their age by means of producing evidence of their age. ***Note 1:*** *The term 'Challenge 25' is broadly used to describe a situation where a customer appearing to be under 25 should be asked to provide proof that they are over 18 to purchase age restricted goods, content or services.* ***Note 2:*** *This is not always referred to as 'Challenge' - some industries use 'Think 21', 'Measure 25', Assess 30' or various other combinations.*

Superseded

3.7 Confidence Interval	means the range of analytical error expected to contain the true value with a stated uncertainty as estimated from a statistical model of the measurement process.
3.8 Fitzpatrick Scale	means a scale of skin tones established by Thomas B. Fitzpatrick in Fitzpatrick, T. B. (1975). "Soleil et peau" [Sun and skin]. Journal de Médecine Esthétique (in French). ***Note:*** *The Fitzpatrick Scale is widely used in biometric data analysis to describe the skin tones of biometric data subjects.*
3.9 Inherent Bias	means the systematic error of an age estimation system to return an accurate result due to the existence of a protected characteristic of the biometric data subject, such as gender or skin tone. ***Note:*** *Local legislation is likely to define protected characteristics in equalities legislation, but for the purposes of this standard, the relevant protected characteristics are likely to be gender and race (in so far as skin tone is a proxy of race).*
3.10 Mean Absolute Error	means a measure of errors (specifically the absolute difference) between paired observations expressing the same phenomenon.
3.11 Mean Absolute Error Tolerance	means a factor of the absolute tolerance level as calculated in accordance with paragraph 5.11.
3.12 Nominal Age	means the age at which the age estimation technology is aiming to estimate with accuracy.
3.13 Performance Throttling	means any pre-programmed action by the age estimation technology that excludes or nullifies results generated by the algorithm that exceed stated tolerances.
3.14 Protected Characteristic	means an aspect of a person's identity that is given legal protection to prevent discrimination, such as gender or race.

Superseded

3.15 Presentation Attack Instrument (PAI)	means a biometric characteristic or object used in a presentation attack, such as a mask, mannequin or a genuine instrument that has been amended.
3.16 Target of Evaluation (ToE)	means the age estimation technology that is the subject of the evaluation.
3.17 Test Crew	means the selected biometric data subjects whose use of the target of evaluation is controlled or monitored as part of the evaluation [ISO/IEC 19795-6:2012(en), 4.17, modified for context]. ***Note:*** *In an operational evaluation, test subjects can be subjects of the operational system or they can be members of a test crew using the system specifically for evaluation purposes.*
3.18 Tethering	means the connection of a device to a power or data source, by physical connection or wireless connection (such as Bluetooth ® or Wi-Fi).
3.19 Upper Tolerance Level	means a factor of the absolute tolerance level as calculated in accordance with paragraph 5.13.

Superseded



Superseded





Superseded



4

Target of Evaluation

Target of Evaluation

Superseded



4. Target of Evaluation

Scope of the Age Estimation Technology Under Evaluation

4.1 The scope of the age estimation technology under test shall be defined and recorded on the certification agreement.

4.2 The scope shall include:

a) the name given to the age estimation technology by the client

b) the operating parameters of the age estimation technology including:

a. any process limitations

b. any prerequisites for operation

c. any tethering required (such as connectivity) for operation

d. any technological requirements

e. any performance throttling applied to the output of the technology (such as disregarding results through uncertainty assessments)

c) the version or deployment reference of the target of evaluation

4.3 The nominal age of the required test crew shall be identified.

Note 1: *The nominal age is the age at which the age estimation technology is aiming to estimate with accuracy. So, for instance, if the age gateway restriction is '18', then the nominal age will be 18.*

Superseded

Hardware Requirements for Age Estimation Technology Under Evaluation

4.4 A Scheme Client may:

a) supply their own hardware for the system under test

b) utilise hardware provided by the scheme

The hardware utilised during the test shall be recorded.

4.5 Where the Scheme Client is intending that the age estimation technology will be reliant upon the deployment of their own image capture equipment, then they may submit this equipment to be used to perform the test.

4.6 Where the Scheme Client is intending that 3rd party image capture equipment will be utilised (such as a camera on a smart phone or integrated to IT equipment) then the Scheme Certification Officer will select suitable testing equipment which shall have:

a) a minimum pixel count of 20 Megapixel;

b) a pixel size of at least 1.5 µm; and

c) a focal length of at least 15mm, but not exceeding 30mm.

Environmental Requirements for the Target of Evaluation

4.7 The ambient lighting level during the testing shall be recorded (in lux, to within +/- 50 lux).

4.8 The ambient light shall either be:

a) a broadly consistent level of lighting (+/- 50 lux) with an absence of glare, flicker, stroboscopic effects, veiling reflections or interference or

b) deliberately controlled lighting, the parameters of which shall be recorded that tests the efficacy of the age estimation technology under:

a. **glare –** which occurs when one part of the visual field is much brighter than the average brightness to which the detection device is adapted

b. **colour effects –** different artificial light sources, or daylight under changing sky conditions, can negatively affect the detection device and cause variance in colour

c. **monochromatic light sources –** low-pressure sodium discharge lamps can stop colours being identifiable and affect the performance of the detection device

d. **stroboscopic effects –** occur when Presentation Attack Instruments appear to be stationary or moving in a different manner, confusing the detection device

e. **flicker –** is light modulation at lower frequencies (about 50 Hz or less) which can sensitise the detection device. It is especially detectable at the edges of the visual system's field of view

f. **veiling reflections –** high luminance reflections (which may be sharp-edged or vague in outline) which overlay the detail of the Presentation Attack Instrument and can affect detection device performance

g. **infrared and ultraviolet radiation –** some lamp designs produce significant emissions at infrared and ultraviolet wavelengths, affecting the detection device.

4.9 The Scheme Certification Officer shall record any other environmental factors that may have a material impact on performance of the target of evaluation.

Superseded

Assessment of Results

4.10 A test crew subject shall be presented to the age estimation technology for a maximum of 8 seconds.

4.11 The age estimation technology shall present the result of its algorithm in single whole years.

> ***Note 1:*** *The user interface for the technology should state the age, in whole years, that it estimates the test subject to be. This does not prevent a Scheme Client from having a more granular age estimation (such as years, months or days) or the assessment being conducted at a more granular level, but that granularity is not required for the performance of the test and statement of results.*

> ***Note 2:*** *A user interface that provides age classification (i.e. the subject is 'over 30') may be sufficient for deployment, but shall not be sufficient for evaluation of the system performance under this standard.*

4.12 Unless otherwise stated by the certification officer, all results of tests undertaken under this standard shall be stated to:

a) with a level of reliability and confidence as determined by the sample size (see paragraph 5.2)

b) with such operating parameter limitations as identified in paragraph 4.2 (b)

c) with such description of the ambient lighting of the test as identified in paragraph 4.8

d) with such description of any other material environmental factors as identified in paragraph 4.9

4.13 The outcome of the test shall state the absolute tolerance level that the age estimation technology is capable of operating to in whole years.

> ***Note 1:*** *The absolute tolerance level is described in section 5.15.*

> ***Note 2:*** *The absolute tolerance level may sometimes be described as the 'challenge age', i.e. the age at which a provider of age restricted goods, content or services may have set that it would seek to ask an individual apparently under that age to provide proof of age to demonstrate that they are over the nominal age as a part of their age verification policy. This is sometimes known in industry as 'Challenge 25' or 'Think 21' etc.*

Superseded



Superseded





Superseded



5

Test Protocols

Test Protocols

Superseded



5. Testing Protocols

Sample Size

5.1 The Scheme shall maintain access to a test crew, the identity and age of which has been verified by reference to official documentation in accordance with the Scheme Rules.

5.2 The sample size shall be determined by reference to Annex 1 – Determination of Sample Size.

5.3 The sample size determines the level of confidence and reliability that can be applied to the results of the test.

> ***Note 1:*** *A sample size of 2,995 presentations will mean (if the parameters of the test results are within the tolerances set out in this standard) that, estimated with 95% confidence, the likelihood of the age estimation technology falsely showing a person of the nominal age (i.e. 18) as being older than the absolute tolerance level (i.e. 25) is between zero and 1 in 1,000 presentations.*
>
> ***Note 2:*** *Whilst ultimately it is a matter for the courts to determine, a level of accuracy as set out in Note 1 would be likely to be considered as exercising all reasonable precautions and exercising all due diligence if the age estimation technology were to be deployed in circumstances where local legislation provided for such tolerances to be applied.*
>
> ***Note 3:*** *Statistically, it is not possible to state with absolute certainty that a false result would never be achieved (unless the whole population were tested). The higher the sample size, the lower the risk of false results (assuming the parameters of the test results are within the tolerances), but the sample size needs to be kept proportionate and reasonable to the purposes and use of the target of evaluation.*

Superseded

5.4 A minimum of 10% of the test crew used during the testing procedure shall be members of a test crew identified by the Scheme Certification Officer and utilise images or presentations not previously seen by the age estimation technology.

> ***Note 1:*** *It is possible for the Scheme Client to participate in a test supervised by a certification officer utilising images of a test crew held by the Scheme Client (up to 90% of the images used). Where this is the case, the images utilised by the certification officer shall be analysed to ensure that there is no substantial deviation from the overall test crew utilised. This provision is intended to accelerate the process of testing, minimise the use of personal data and reduce costs.*

Sample Population

5.5 The test crew population shall be representative of the population in terms of both gender and skin tone so that the results reflect what the age estimation technology is expected to observe in the population to which the technology will be applied.

5.6 The gender and skin tone of the test crew shall be recorded.

Note 1: *Gender records shall be in the form M, F or X.*

Note 2: *Skin tone records shall be in accordance with the Fitzpatrick Scale Type I – VI.*

5.7 The true age of the test crew shall be not more than +/- 12 months of the nominal age applied to the target of evaluation.

Note 1: *So where the nominal age under test is 18, the members of the test crew shall have reached their 17th birthday and shall not have passed their 19th birthday in order to be eligible to be used in the test crew.*

5.8 The Scheme Client shall provide evidence (from their internal testing) that the accuracy of the age estimation technology is approximately constant across different nominal ages.

Note 1: *This is an assessment of whether or not the error is similar for (say) a 12 year old, as it is for an 18 year old.*

Note 2: *Scheme Clients are not expected to have structured test data for this check and are not expected to have results for infants or elderly citizens.*

Superseded

Acceptable Test Results

5.9 In order to pass the test, the target of evaluation shall achieve results that show:

a) the mean absolute error and the predicted mean age is within acceptable tolerances

b) the number of results above the upper tolerance level are within the acceptable limits

c) no results above the absolute tolerance level

d) the inherent bias for gender and race is within acceptable limits

5.10 Annex 2 contains an illustrative table of the tolerance levels described in this standard and Annex 3 contains some illustrative examples of test results showing acceptable and unacceptable levels of estimation error.

Mean Absolute Error

5.11 The mean absolute error shall not be more than 0.25% of the square of the absolute tolerance level (rounded up to the next whole integer) and the predicted mean age shall not be more than +/- 0.25% of the square of the absolute tolerance level (rounded up to the next whole integer) from the mean age of the sample.

Note 1: *This means that the mean absolute error being tested against a 'Challenge 25' absolute tolerance level, shall be between 0 and 2 years (as 0.25% of the square of the absolute tolerance level age of 25 is 625 x 0.25% = 1.56 rounded to 2 years).*

Note 2: *This test is intended to show that, on average, the age estimation technology will operate to return results that are at or relatively near to the nominal age. Greater leeway is given for the mean absolute error the higher the absolute tolerance level (challenge age) is set.*

Upper Tolerance Level

5.12 No more than 1:40 results shall be higher than the upper tolerance level.

5.13 The upper tolerance level shall be the absolute tolerance level minus 0.25% of the square of the absolute tolerance level then rounded down to the next whole integer.

Note 1: *This means that the upper tolerance level being tested against a 'Challenge 25' absolute tolerance level, would be 23 (as 0.25% of the square of the absolute tolerance level age of 25 is 1.56 and 25 – 1.56 = 23.44 rounded down to 23). Thus, no more than 1:40 results from the test could be over 23.*

Note 2: *This test is intended to show that the estimated results of the age estimation technology are not skewed towards meeting the absolute tolerance level whilst disregarding the accuracy of the other ages (making the deployment of the technology riskier).*

Absolute Tolerance Level

5.14 No results shall be higher than the absolute tolerance level.

5.15 The absolute tolerance level can be taken as the next whole integer above the highest estimated result identified during the test.

Inherent Bias

5.16 The predicted mean age for each protected characteristic shall not vary from the predicted mean of the whole sample by more than +/- 0.25 years.

Note 1: *This is assessing whether there is an overall bias across the whole sample rather than the risk of a bias at an individual level.*

5.17 The certification officer shall be satisfied that the distribution of results between the protected characteristics shall be broadly similar.

Superseded

5.18 The protected characteristics under test shall be:

a) Gender Male or Female

b) Skin Tone Type I and II

c) Skin Tone Type III and IV

d) Skin Tone Type V and VI

Note 1: *Test crew subjects with gender marked as X can be excluded from tests of inherent bias.*

Note 2: *The Fitzpatrick Skin Tone analysis can be grouped.*

Superseded



Superseded



Annex

Annex

Superseded



Annex 1 – Sample Size Calculation (Normative)

A1.1 The primary aim of the testing procedure is to ensure that the age estimation technology does not assess a real-life person aged near to the nominal age as being over the challenge age. This would be deemed as a failure of the technology.

A1.2 Since it is not possible to test that the technology has a 0% failure rate using a sampling approach, we determine an appropriate risk-based sample size using the Bayes Success-Run Theorem.

A1.3 By using the Bayes Success-Run Theorem to calculate a sample size, it is possible to make a confidence statement about the rate at which the age estimation technology will meet the critical test, conditional on there being no failures in the samples presented. This rate represents the reliability of the technology; it is the proportion of times that the technology will perform satisfactorily when used correctly (i.e., not assess a person aged between 16 and 19 years as being over the challenge age).

A1.4 For example, if both the confidence is set at 95% and reliability 99% and the appropriate sample is found to be failure free, it is estimated with 95% confidence that the technology will function satisfactorily 99% of the time. We can also interpret this in terms of a failure rate; a reliability of 99% means that the failure rate of the technology is estimated to be between 0 and 1% with 95% confidence.

A1.5 The formula for the sample size calculation is as follows:

N = ln(1-C)/ln(R)

where C represents the confidence level, R the reliability and ln the natural logarithm. In the table below, we present the sample sizes required for a range of reliability levels.

Superseded

Confidence (C)	Reliability (R)	Failure Rate	Sample Size (N)
95%	99%	0 – 1% (1 in 100)	299
	99.9%	0 – 0.1% (1 in 1,000	2,995
	99.99%	0 – 0.01% (1 in 10,000)	29,956
	99.999%	0 – 0.001% (1 in 100,000)	299,572
	99.9999%	0 – 0.0001% (1 in 1,000,000)	2,995,732

A1.6 The table above demonstrates that as the reliability increases towards 100%, the sample size becomes very large. In order to be able to demonstrate the exercise of reasonable precautions and due diligence, a reliability of 99.9% is considered a suitable level which balances an acceptable maximum failure rate with the practicalities of implementing the sample size.

Annex 2 – Table of Tolerance Levels (Informative)

Challenge Age Category	MAE Tolerance Level (MAET)	Upper Tolerance Level (UTL)	Absolute Tolerance Level (ATL)
21	1	20	21
22	1	21	22
23	1	22	23
24	1	23	24
25	2	23	25
26	2	24	26
27	2	25	27
28	2	26	28
29	2	27	29
30	2	28	30
31	2	29	31
32	3	29	32
33	3	30	33
34	3	31	34
35	3	32	35
36	3	33	36
37	3	34	37
38	4	34	38
39	4	35	39
40	4	36	40
41	4	37	41
42	4	38	42
43	5	38	43
44	5	39	44
45	5	40	45
46	5	41	46
47	6	41	47
48	6	42	48
49	6	43	49
50	6	44	50

Superseded

The MAET = (ATL2 x 0.25%) then rounded up to the nearest integer

The UTL = ATL - (ATL2 x 0.25%) then rounded down to the nearest integer

The ATL = the highest returned age estimation rounded up to the nearest integer

Annex 3 – Example Test Results (Informative)

Tolerance Levels (Examples)

A2.1 Some illustrative examples are given to show the type of test results that do and do not pass the different criteria. Each example shows two distributions: the red curve represents the actual ages of the participants in the test (defined as the test crew), and the blue curve shows what the age estimation technology predicted the participants' ages to be. If the predictions were perfectly accurate, we would expect the two curves to overlap exactly.

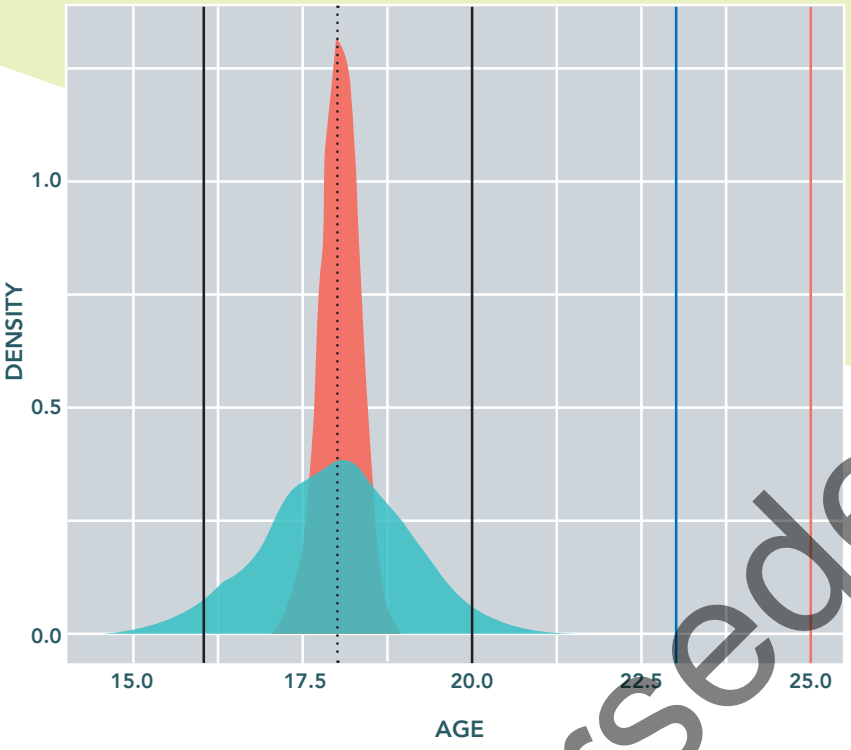
The vertical lines on the plots represent the three criteria that the predicted age distribution must pass. The black lines provide a range within which the average of the predicted ages (shown by the dashed black line) must sit (check 1). The blue line represents the upper tolerance (check 2) and the red represents the absolute tolerance (check 3).

The examples in this section illustrate that in order to pass all three criteria, the predicted age distribution must centre around the actual age distribution and its spread (how wide the distribution is) cannot be too large.

Superseded

Example 1

An age estimation system tested at a nominal age of 18 against a challenge age policy of 25 – comfortably conforming



PASS

Age
- Actual Age
- Predicted Age

Tolerances
- Absolute Tolerance
- Mean Age Tolerance
- Upper Tolerance

Mean Age
- Mean Age Tolerance
- Predicted Age Tolerance

Superseded

Checks of Conformity

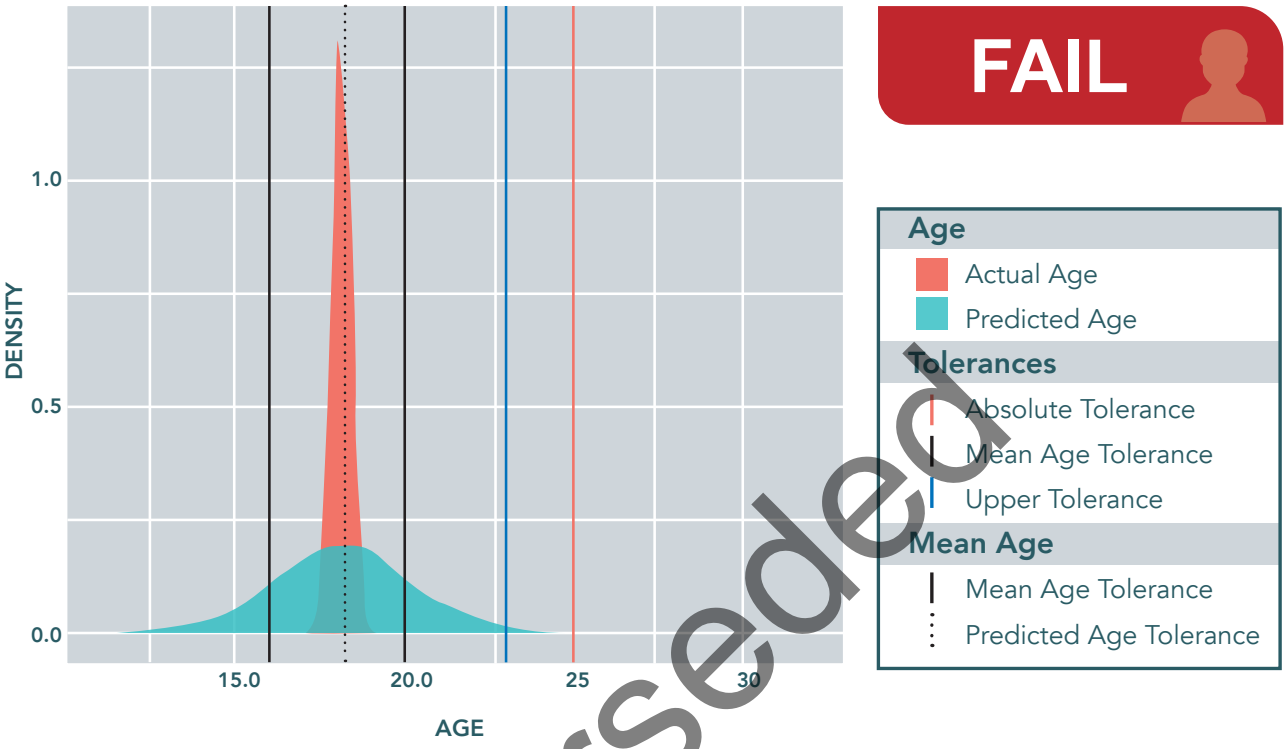
1	Both the mean absolute error (a measure of the average difference between the actual and predicted ages) and the mean age of the predictions (dashed black line) are within acceptable tolerances.	✓
2	There are fewer than 1 in 40 (or less than 2.5% of) test crew ages that have been estimated to be greater than the upper tolerance level (blue line). In fact, this product is comfortably conforming to this check as there are no results above this tolerance (the blue curve lies below the blue line).	✓
3	There are no results (on the blue curve) above the absolute tolerance level, therefore none of the test crew were predicted to be older than 25.	✓

Subject to also meeting the inherent bias tests set out below, this age estimation system would be fit for purpose for use in a 'Challenge 25' deployment where the nominal age (i.e. the age restriction in force) was 18.

It is possible that an age estimation technology displaying the results in this graph would be sufficiently accurate to be certified at an even lower challenge age category than 25.

Example 2

An age estimation system tested at a nominal age of 18 against a challenge age policy of 25 – non-conforming – too many results above the absolute tolerance level



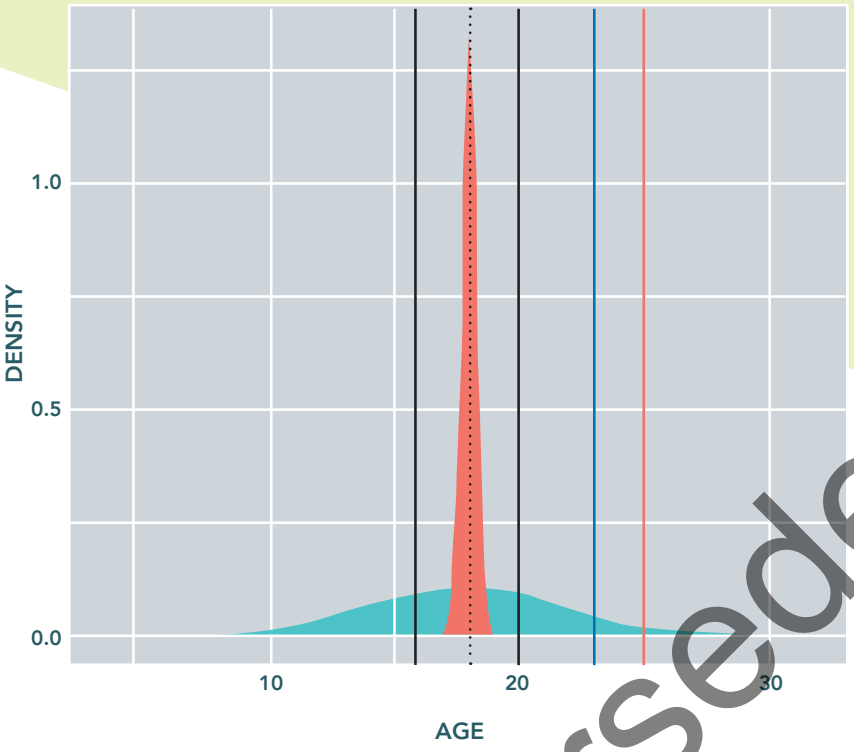
Superseded

Checks of Conformity	
1 Although the range of the predicted ages (the blue curve) is broader than example 1, the mean absolute error and the predicted mean age (black dashed line) are within acceptable tolerances, so these results conform with the first check.	✓
2 There are some, but no more than 1 in 40 (or less than 2.5% of) test crew ages that have been estimated to be above the upper tolerance level (blue line), so the results also conform with the second check.	✓
3 There are results that are above the absolute tolerance level, meaning some of the test crew were estimated to be older than 25.	✗

This age estimation system would not be fit for purpose for use in a 'Challenge 25' deployment where the nominal age (i.e. the age restriction in force) was 18, as check 3 did not conform. In order to conform, we would need to see all predicted ages within the absolute tolerance boundary.

Example 3

An age estimation system tested at a nominal age of 18 against a challenge age policy of 25 - non conforming – too many results above the upper and absolute tolerance levels



FAIL

Age
- Actual Age
- Predicted Age

Tolerances
- Absolute Tolerance
- Mean Age Tolerance
- Upper Tolerance

Mean Age
- Mean Age Tolerance
- Predicted Age Tolerance

Superseded

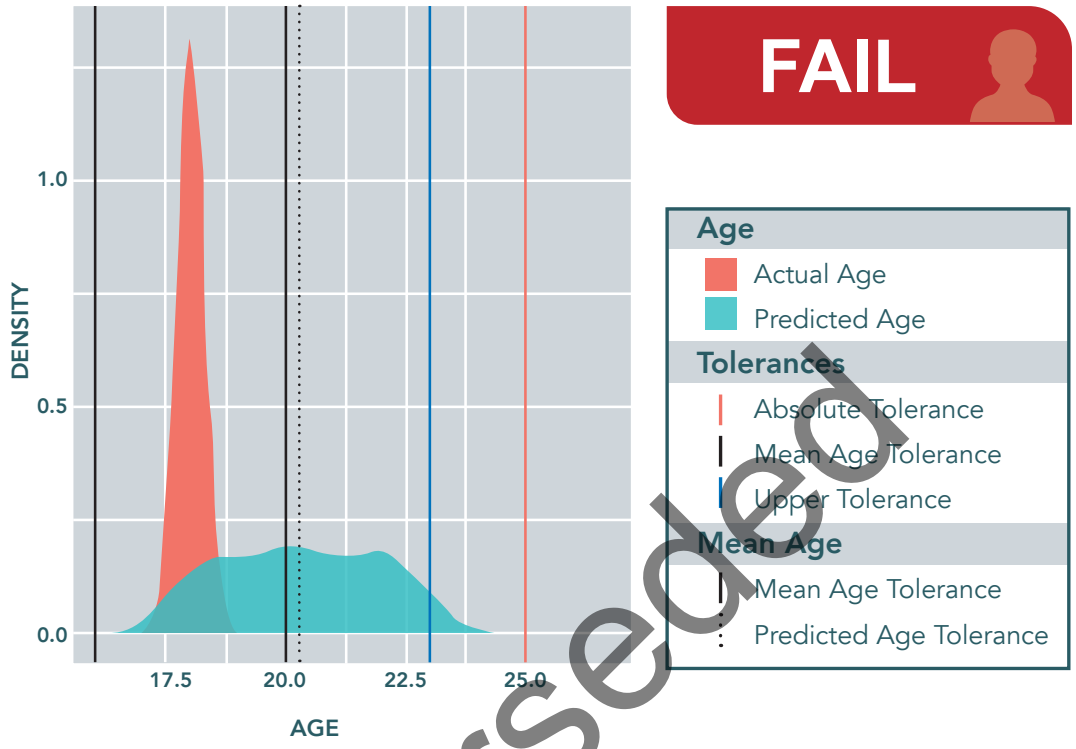
Checks of Conformity

1	Although the range of age results is much broader than the previous two examples, the predicted mean age (black dashed line) and mean absolute errors remain within the acceptable tolerances so the results conform to the first check.	✓
2	There are more than 1 in 40 results (on the blue curve) above the upper tolerance level (blue line), meaning that over 2.5% of the test crew's ages have been estimated to be over the upper tolerance level of 23. These results do not conform to the second check.	✗
3	There are results (on the blue curve) above the absolute tolerance level (red line) meaning that some of the test crew's ages have been estimated to be over 25. Therefore, there is a significant risk that this product would indicate that a citizen of 18 was older than the challenge age of 25.	✗

This result demonstrates what this standard is testing. This product has a mean absolute error that is well within the acceptable tolerance. However, within that data, it has results that exceed both the upper tolerance level and the absolute tolerance level indicating that, if it were deployed for decisions at that challenge age, it would let people through that are at or near the nominal age. This presents an unacceptable risk of non-conformance with age restrictions legislation.

Example 4

An age estimation system tested at a nominal age of 18 against a challenge age policy of 25 - non conforming – over-estimating the ages of the test crew



Superseded

Checks of Conformity	
1 The range of estimated age results (the blue curve) is much broader and older than the actual ages of the test crew (the red curve). Consequently, the mean absolute error and predicted mean age (dashed black line) are too high for the acceptable tolerances (solid black lines) and this is a non-conforming product.	✗
2 There are more than 1 in 40 results (greater than 2.5%) of the test crew ages that have been estimated to be greater than the upper tolerance level (blue line), indicating that the spread or range of the results is also too high to be a conforming product.	✗
3 There are no results (on the blue curve) above the absolute tolerance level (red line), meaning that the third check is passed. However, the failure of the other checks would mean that this product would fail overall.	✓

This result demonstrates why there are three aspects to the test. Whilst ostensibly, this product is demonstrating that all of the test subjects at the nominal age are being assessed as under the challenge age (in this case, 25), on average the age estimation technology is significantly over estimating the age, making the deployment too risky. In addition, there are too many results above the upper tolerance level, meaning that the likelihood of a result above the challenge age was increased.

Inherent Bias (Examples)

A2.2 The assessment of inherent bias is based on the following:

a) Classification or outcome error parity: the model is fair if protected groups receive an equal proportion of positive outcomes, or an equal proportion of errors

b) Calibration: the model is well-calibrated if the predicted ages reflect the actual ages in real life. Equal calibration definitions of fairness say that a model should be equally calibrated between protected attribute groups

A2.3 If none of the samples have been estimated to be equal to or above the absolute tolerance level, then it is fair to say that the samples exhibit an equal proportion of errors (or outcome error parity) across the protected characteristics. However, it is important to note that our confidence in the error rate will vary across the protected characteristics as the sample size for each individually varies.

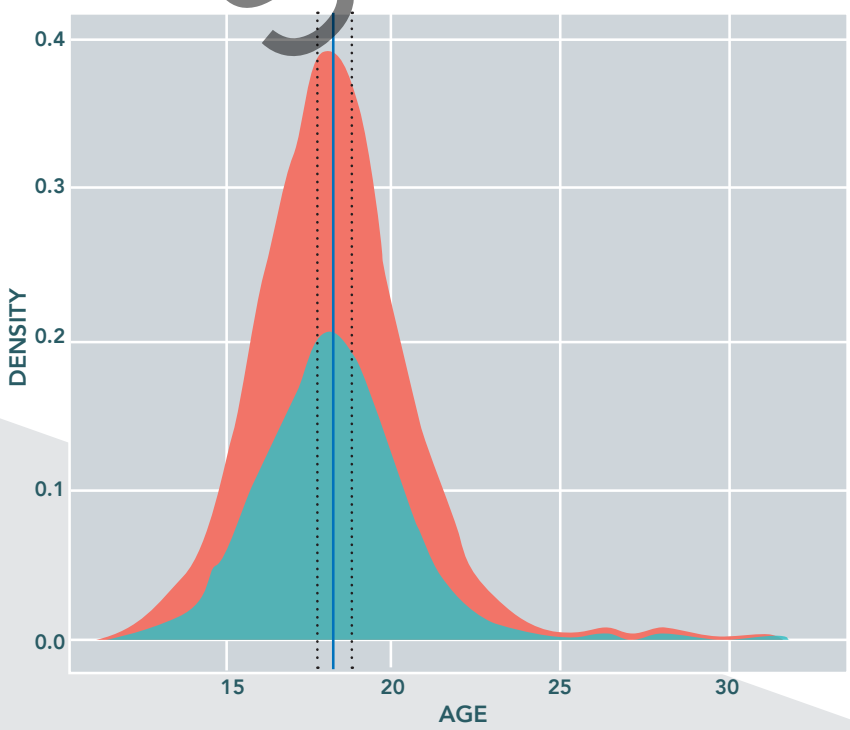
A2.4 To investigate how well the model is calibrated, the average predicted age for a protected characteristic is compared with the average predicted age of the whole sample.

Illustrative examples are provided below where the red curve represents the distribution of the predicted ages for the whole sample and the blue curve represents the distribution of the predicted ages for all the females in the sample. The dashed black lines represent the interval in which the average age of the females should lie in order to conform (within +/- 0.25 years of the mean predicted age of the whole sample).

Superseded

Example 5

An age estimation system tested for gender bias - conforming – the mean results of both genders are broadly aligned and within acceptable tolerances



PASS

Group
- Full Sample
- Female

Age Tolerances
- Full Sample Mean Age
- Mean Age Tolerance

Mean Age
- Female Mean Age
- Full Sample Mean Age

Checks of Conformity

1 The results for female predicted ages (the blue curve) are broadly similar to the results for the predicted ages across the whole sample (the blue curve) and the average age of the female samples does not differ by +/-0.25 years from the average of the whole sample.

✓

Example 6

An age estimation system tested for gender bias – non conforming – the average age of the female samples is lower compared to the whole sample, indicating an inherent bias



Superseded

Checks of Conformity

1 The average predicted age of the female samples (the blue line) is more than 0.25 years less than the average predicted age of the whole sample (it lies below the dashed lines), indicating an inherent bias.

✗



Superseded



© Age Check Certification Scheme

Visit us on social media...

@AgeCheckCert





Internationally recognised standards

Our scheme provides evidence that identity and age check practices meet international standards.



Highly qualified certification officers

Our locally trained certification officers are highly qualified professionals in each jurisdiction.



Evidence you can use to demonstrate compliance

Our certificates of conformity are internationally recognised by law enforcement.



Guaranteed independence & impartiality

Our independent board, impartiality committee and ISO processes guarantee impartial certification.

Superseded

Age Check Certification Scheme

www.accscheme.com

An independent not-for-profit company registered in England and Wales (11493870) at Unit 321B Broadstone Mill, Broadstone Road, Stockport, SK5 7DL

ISBN 978-1-7398513-1-6

0 638266 208931