USING THE OBJECTIVE RATING METHOD (ORM) AS A QUALITY ASSESSMENT TOOL FOR PHYSICAL TESTS, TEST METHODS, AND MATHEMATICAL MODELS

Linda Eriksson
Håkan Sundmark
Autoliv Sverige AB, Vårgårda
Sweden
Harald Zellmer
Autoliv B.V. & Co. KG, Elmshorn
Germany
Kurt Fograscher
Bernhard Drexl
Autoliv B.V. & Co. KG, Dachau
Germany
Marc Van Slagmaat
Autoliv France, Gournay-en-Bray
France

Paper Number 09-0163

ABSTRACT

There is a daily need to assess the quality of our work. On the crash track, the tests should be repeatable, the chosen test method should fulfil the test purpose and every result should have an explanation. The tests performed may also be used to validate mathematical models, the accuracy of which must then be assessed, or, to show whether a new design or method influences the performance or not. Regardless of which, there is a need of a quality assessment tool. By applying the Objective Rating Method on performed rear-end sled tests, Autoliv has previously shown that the BioRID II dummy allows for both repeatable and reproducible testing. Here, the ORM has been evaluated on frontal, side impact and component tests and the corresponding models.

For frontal impacts, test repeatability has been assessed, and correlation between physical tests and mathematical models are shown. For side impacts, the test repeatability, test method predictability and mathematical model predictability have been assessed. The repeatability of frontal sled tests is comparable with that presented for rear-end sled tests, while the side impact sled test repeatability is generally somewhat lower.

Although the ORM has to be used with care and knowledge, it is a useful tool, especially for assessments regarding test repeatability and reproducibility. The ORM allows for agreement, in advance, on a quality level for tests and mathematical models. Beneficial is that the ORM not only compares peak values but also curve shapes. Furthermore, the ORM compares two tests; many methods require several tests and that is normally not available in daily work.

INTRODUCTION

There are several methods to assess the repeatability or reproducibility of duplicated tests, or the predictability of simulation models or test methods. One common method in daily work is to visually compare results for two tests and decide if they are sufficiently similar or not. This method is often fast to carry out, but the decision may differ depending on personal experience and opinion. Therefore, this kind of method is not solitary suitable as a quality assessment tool for scientific or professional purposes. A useful correlation quality assessment tool must be suitable for many types of inputs; for test and simulations issues it should be possible to compare scalars such as criteria and peak values, and curve shapes. Also, the rating from poor to good correlation must be obvious. Further, the result must be independent of the user and it should be easy to evaluate and understand. Preferable, the tool should be easy and fast to use.

Autoliv has used several test quality assessment tools during the years. In a recent study ([1]), the repeatability and reproducibility of rear-end impact tests with the BioRID II were assessed by using the ORM ([2]). For that purpose, the ORM was found to be a good tool. Therefore, this paper will continue the evaluation of the ORM; this time applied on conducted frontal impact, side impact, and airbag module component tests and corresponding simulations from four Autoliv sites in Europe. Also, the ORM was used to evaluate the correlation between test methods. Thresholds for good and poor correlation for a few types of comparisons will be proposed along with a method to define which signals to included in a comparison.

METHOD

The ORM

The Objective Rating Method, ORM, assesses the correlation between duplicated tests (repeatability or reproducibility) or between a test and a simulation (predictability). The ORM enables comparison between scalars, such as criteria, minimum and maximum peak values, and their occurrence times, and between curve shapes.

The ORM scalar correlations are calculated according to Equation 1. This expression is called the Factor Method and calculates the correlation between the reference test and the comparison test. The results range from 0 to 100%, where 100% represents a perfect match. The curve shape correlation is calculated according to Equation 2. This expression is called the Weighted Integrated Factor Method and is a combination of the Factor Method and the Root Mean Square Addition Method. This means that the correlation in each time step contributes to the total correlation just as the function value would contribute to the total area underneath the curve. The δ is very small and used to avoid division by zero. *r* and *c* are used as abbreviations for *reference* and *comparison*, respectively.

$$ORM_{scalar} = \frac{\max(0, reference \cdot comparison)}{\max(reference^2, comparison^2)} \quad \textbf{(1)}$$

$$ORM_{shape} = 1 - \sqrt{\frac{\int \left(\max(|r(t)|,|c(t)|) \cdot \left(1 - \frac{\max(0, r(t) \cdot c(t))}{\max(\delta, r(t)^2, c(t)^2)} \right)^2 \right) dt}{\int \max(|r(t)|,|c(t)|) dt}} \quad \textbf{(2)}$$

ORM values are calculated not only for the scalars and the curve shapes, but also for groups of scalars and curve shapes. The contribution of each ORM value in its group is defined by a weight factor, W. Equation 3 is used to calculate the ORM value for each group. Further, the groups are arranged into one single ORM value that is the correlation for the complete system. The contribution of the group ORM values to the complete ORM value is defined by weight factors, W. Equation 4 is used to calculate the ORM value for the complete system.

$$ORM_{group} = 1 - \sqrt{\frac{\sum \left(W_{scalar\ or\ shape} \cdot (1 - ORM_{scalar\ or\ shape})^2 \right)}{\sum W_{scalar\ or\ shape}}} \quad \textbf{(3)}$$

$$ORM_{complete} = 1 - \sqrt{\frac{\sum \left(W_{group} \cdot (1 - ORM_{group})^2 \right)}{\sum W_{group}}} \quad \textbf{(4)}$$

A study was conducted in order to better understand the correlation of the ORM values to engineering judgements and to identify possible weaknesses of the ORM. A typical acceleration curve was used (Figure 1) as the reference curve, and an identical curve was used as the comparison curved but shifted or scaled in several steps in X (e.g. along the time axis) and Y (e.g. along the amplitude axis), or inverted. The curves were evaluated in the time range 0.01 to 0.12 s. Furthermore, another curve (Figure 2) was used to assess the correlation between a filtered (CFC60) and an unfiltered curve. For each comparison, the correlation of the maximum peak, its occurrence time, and the curve shape were calculated. Also, the group ORM value for these three components were calculated, all with equal weight. The group ORM values were then presented along with correlations assessed by engineering judgement in order to propose thresholds in ORM for correlation assessments. The engineers used a four-level scale: OK/acceptable/NOK/–. OK means okay, NOK means not okay, and the dash (–) shows that the curves are very dissimilar.

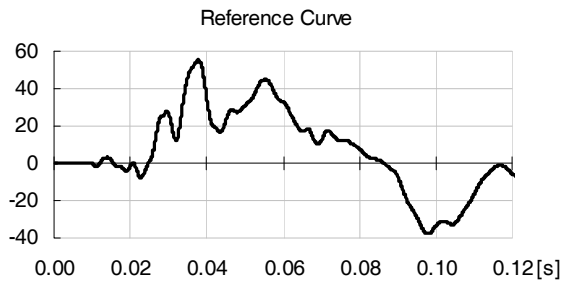


Figure 1. The acceleration curve used as the reference curve in the ORM evaluation.

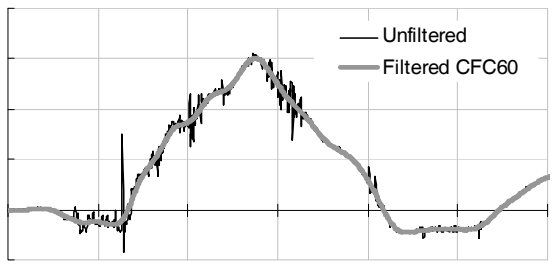


Figure 2. The curves compared to assess ORM values for a filtered versus an unfiltered curve.

Frontal impacts

<u>Series 1 – Evaluation of Included Signals</u> In order to evaluate how the selection of included signals influences on the complete ORM values one pair of USNCAP driver side sled tests were analyzed. Seven different sets of signals were composed.

- Set 1 contained the criteria HIC36 and Chest3ms, and these were equally weighted to compose the complete ORM value.
- Set 2 contained one group (weight 1/2) with head, chest, and pelvis accelerations, lumbar spine, femur, and belt loads, and chest deflection; and one group (weight 1/2) with only sled x-acceleration. Mainly peak values of the signals were compared, and a few of their occurrence times were added with half weight. The belt force

peaks were given half weight as well. No curve correlations were included in this set.

- Set 3 contained one group (weight 1/11) with head and neck signals, one group (weight 2/11) with chest signals, one group (weight 1/11) with lumbar spine signals, one group (weight 2/11) with pelvis signals, one group (weight 1/11) with femur signals, one group (weight 2/11) with belt signals, and one group (weight 2/11) with sled x-accelerations. For all signals, all available components were included, and in the chest, lumbar spine, and pelvis groups the signals acting in the dummy sagittal plane were given double weight.
- Set 4 contained one group (weight 4/14) with sled x-acceleration, belt signals and dummy placement signals, and five groups with most of the measured dummy signals: head (weight 3/14), upper neck (weight 1/14), chest (weight 3/14), lumbar spine (weight 1/14), and pelvis and femur (weight 2/14). For the head and chest the main peaks and the curve shapes were included. For the other included signals, only the main peaks were included.
- Set 5 contained one group (weight 2/12) with sled x-accelerations, one group (weight 2/12) with belt signals, and six groups with most of the measured dummy signals: head (weight 2/12), upper neck (weight 1/12), chest (weight 2/12), lumbar spine (weight 1/12), pelvis (weight 1/12) and femur (weight 1/12). All available components were included. Dummy signals acting in the dummy sagittal plane were given higher weights. Curve shapes were given higher weights than peaks and their occurrence times.
- Set 6 contained one group (weight 3/22) with sled x-accelerations, one group (weight 3/22) with belt signals, and six groups with most of the measured dummy signals: head (weight 3/22), upper neck (weight 1/22), chest (weight 3/22), lumbar spine (weight 3/22), pelvis (weight 3/22) and femur (weight 3/22). For almost all signals, the main peaks, their occurrence times, and the curve shapes were included, with the exceptions of the upper neck and sled for which only the curve shapes were included.
- Set 7 contained one group (weight 4/5) with dummy signals: head, chest, and pelvis x- and z-accelerations, upper neck loads, chest deflection, and femur z-forces, and one group (weight 1/5) with belt signals. For almost all signals one third of the components were significant peaks, one third their occurrence times, and one third curve shapes. The sled pulse curve shape correlations were evaluated separately and not included in the complete ORM value.

The two compared tests were analysed visually by engineers in order to rate the frontal sled test repeatability on a five-level-scale: dissimilar, nearly dissimilar, fairly similar, similar, and almost identical.

Series 2 – Correlations between Sled Tests Sled test repeatability was assessed for tests with a belted Hybrid III 50 %-ile seated on a padded R16 seat. Two tests with the EuroNCAP pulse and three tests with the USNCAP pulse were compared; in total were four comparisons performed. The signals included in the comparisons were selected with the same method as used in Set 7: head, chest and pelvis x-and z-accelerations, upper neck and lumbar spine F_x, F_z, and M_y, chest deflection, torso displacement, femur z-forces, belt forces, and belt displacement. For all signals the curve shapes correlations during 200 ms, the minimum or maximum peak value and occurrence time were included; for five signals both minimum and maximum peaks and occurrence times were included. All signals were contributing with the same weight to the complete ORM value. The sled pulse correlation was not included in the complete ORM value, but checked separately. The four pairs of compared tests were analysed visually by test engineers in order to rate their overall repeatability on a five-level-scale: dissimilar, nearly dissimilar, fairly similar, similar, and almost identical.

Series 3 – Correlations between Sled Tests Five USNCAP sled tests were included in Series 3. Sled tests #1 to #3 were conducted with the Hybrid III 50%-ile and sled tests #4 and #5 with the Hybrid III 5%-ile. The ORM correlations were assessed for #1 versus #2, #1 versus #3, #2 versus #3, and #4 versus #5. The signals included in the ORM analyses were selected according to the method used in Set 7. The five pairs of compared tests were analysed visually by test engineers in order to rate their repeatability on a five-level-scale: dissimilar, nearly dissimilar, fairly similar, similar, and almost identical.

Series 2 – Correlations between Simulations and Sled Tests The tests conducted in Series 2 were aimed as verification tests for Madymo models of belt systems. The models of the belt systems were built up of validated belt component models and tuned to mimic the chosen load limiter levels and test conditions. The sled test predictability of the final Madymo system models were assessed by the ORM. A similar signal set were used as those used to correlate the tests in Series 2, the main differences were that the lumbar spine loads and the torso displacement were omitted, and the time interval differed. Due to the aim of verifying the Madymo belt system models predictability in normal crash conditions, the correlations were assessed during the first 85 ms of the USNCAP simulations and during the first 110 ms of the EuroNCAP simulations. After that the dummy chest hit its legs and that should not be mimicked by the Madymo models. The usefulness of the Madymo models was assessed by simulation engineers.

Series 3 – Correlations between Simulations and Sled Tests Sled tests #2 and #3 conducted in Series 3 were used as verification tests to a Madymo system simulation. The sled test predictabilities of the final Madymo model were assessed by the ORM. A similar signal set as used to correlate the tests in Series 3 were used. The usefulness of the Madymo model was assessed by simulation engineers.

Side Impacts

The ORM was used to guide in an EuroNCAP side impact project, from component development to full scale crash tests. On component level, the ORM was used to assess that sufficiently predictability levels were reached. Later, the predictability of the complete systems mathematical models were assessed by the ORM. The project included sled tests and full scale crash tests for the MDB and the pole load cases. In order to assess the predictability of the sled tests to the full scale crash test the ORM was used. Also, ORM was used to confirm that changes of the safety components did influence the dummy performance. All steps and correlations will not be covered in this study. In all cases the ORM was applied the repeatability or predictability was also rated by engineering judgement.

Mathematical models of side impact airbags were built up in LS-Dyna and tuned to four component tests: static deployments for airbags with no vent holes and equipped with ordinary vent holes and dynamic tests with an impactor at 7 m/s and a total weight of 12 kg for the pelvis and thorax parts and with an impactor at 4 m/s and a total weight of 20 kg for the pelvis and thorax parts. Two static tests of each type, and three dynamic tests of each type were conducted and the repeatability level were assessed by the ORM. For all tests, signals in the pelvis part and the thorax part were collected. For the static tests, the two bag pressures were compared from 10 to 60 ms. For each pressure signal, the peak, its occurrence time, and the curve shape correlation were given equal weight and repeatabilities were assessed by the ORM. For the dynamic tests, the two bag pressures and the displacement, velocity, and acceleration of the impactors were compared from 10 to 60 ms or to strike through. For the accelerations and pressures, peak values, their occurrence times, and the curve shapes were included with equal weight and for the displacement and velocities only the curve shapes were included and these were given the same weight as all other components. One side impact airbag mathematical model was compared by the ORM to all conducted tests, and the same signals sets were used as in the repeatability assessments.

Mathematical models of the side impact safety components were fitted into mathematical models of full scale crash tests and the components were optimized to perform well in the EuroNCAP MDB and pole tests. Sled tests were then conducted, and these aimed to mimic the mathematical models of full scale crash tests. The repeatability of the sled tests were assessed by the ORM. Then, one full scale MDB crash test and one full scale pole crash test were conducted and the sled test predictabilities of the full scale crash tests were assessed by the ORM. The sled test set-ups were then improved to better mimic the full scale crash tests and these improvements were assessed by the ORM. The repeatabilities were assessed by the ORM for these sled tests as well. The mathematical models were updated and used to improve the EuroNCAP performance further, and sled tests were conducted to assess these improvements. The predictability and the repeatability of the sled tests were assessed by the ORM, and the ORM was also used to assess if the changes of the safety component resulted in significant changes of the dummy readings in the sled tests. Finally full scale crash tests were performed and the sled test predictabilities of these were assessed by the ORM. The ORM sets used in the side impact project where selected according to the method used in Set 7 in this study. In this paper, only the ORM values for the MDB tests and simulations will be presented.

RESULTS

The ORM

ORM values for an extensive set-up of curves were compared with the correlations assessed by engineering judgements. The filtered versus the unfiltered curve were compared, and the ORM value was 97% for the maximum peak correlation, 99% for the occurrence time correlation, and 89% for the curve shape correlation. The group ORM value was 93% and the curves were judged to have okay correlation (OK). For the normal versus the inverted curve, the ORM value for the peak occurrence time correlation was 100%, but the other two correlations were zero, resulting in a group ORM value of 18%. These curves were rated as very dissimilar (–). The ORM values for the comparisons in which one of the two curves was shifted or scaled are given in Table 1 along with the engineering judgements. As can be seen in all groups with increasing shifting or scaling, the scalar, the curve shape and the group ORM values are continuously decreasing which is fundamental for a consistent correlation assessment tool.

Frontal impacts

Series 1 – Evaluation of Included Signals The complete ORM value was 96% for Set 1 and Set 2, 84% for Set 3, Set 4, and set 6, 73% for Set 5, and 88% for Set 7. It can be concluded that the choice of signals and components included in the ORM analysis largely influence on the complete ORM values. The sled tests were judged as almost identical by several test engineers.

Table 1.
ORM values in percent for the comparisons in which one of the curves was adjusted, along with the correlations assessed by engineering judgements. OK means okay, Acc. means acceptable, and NOK means not okay. – means that there is no correlation due to very dissimilar curves.

Adjustment of the comparison curve	Peak ORM value	Peak time ORM value	Curve shape ORM value	Group ORM value	Engineering judgement
Shift Y + 1	98	100	91	95	OK
Shift Y + 5	92	100	72	83	OK
Shift Y + 10	85	100	57	74	Acc.
Shift Y + 20	73	100	39	62	NOK
Shift Y + 40	58	100	23	49	–
Shift Y + 80	41	100	14	37	–
Scale Y*1.1	91	100	91	93	OK
Scale Y*1.2	83	100	83	87	OK
Scale Y*1.3	77	100	77	81	OK
Scale Y*1.4	71	100	71	77	Acc.
Scale Y*1.5	67	100	67	73	Acc.
Scale Y*1.8	56	100	56	64	NOK
Shift X+ 0.001	100	97	79	80	OK
Shift X+ 0.005	100	88	50	70	Acc.
Shift X+ 0.01	100	79	35	61	NOK
Shift X+ 0.02	100	65	21	50	–
Shift X+ 0.04	100	49	6	38	–
Shift X+ 0.08	100	32	1	31	–
Scale X*1.1	100	91	44	67	NOK
Scale X*1.2	100	83	31	59	NOK
Scale X*1.3	100	77	29	57	NOK
Scale X*1.4	100	71	26	54	–
Scale X*1.5	100	67	23	51	–
Scale X*1.8	100	56	16	45	–
Scale X*1.1 & Y*1.1	91	91	43	66	NOK
Scale X*1.5 & Y*1.5	67	67	21	48	–
Scale Y*1.1 & X*1.5	91	67	22	45	–
Shift X + 0.005 & Y + 5	92	88	47	68	NOK
Shift X + 0.03 & Y + 30	65	56	20	44	–
Shift X + 0.005 & Y + 30	65	88	27	53	–

Series 2 – Correlations between Sled Tests First, the two EuroNCAP tests were compared. The sled pulse curve shape correlation was very high, 93%. The complete ORM value was 87%. The ORM value was 86% for the group of dummy signals, and 91% for the group of belt signals. For all included signals, the average peak value correlation was 93%, the average peak value occurrence time correlation was 95%, and the average curve shape correlation was 85%. Then, three USNCAP tests were compared in pairs. The sled pulse curve shape correlations were all very high, these ranged from 94 to 96%. The complete ORM values were 81, 85, and 86%. The ORM values were 79, 84, and 84% for the groups of dummy signals, and 90, 90, and 96% for the groups of belt signals. For all included signals in each comparison, the average peak value correlations were 91, 94, and 94%, the average peak value occurrence time correlation was 97, 98, and 98%, and the average curve shape correlation was 77, 81, and 82%. The EuroNCAP sled tests were judged as almost identical, and the USNCAP sled tests as similar by several test engineers.

Series 3 – Correlations between Sled Tests The crash pulse in sled test #1 differed significantly from those in sled tests #2 and #3: the amplitude differed about 4g and the pulse ended about 10 ms later. In ORM values the crash pulse curve shape correlation was 70 and 71% for the pairs including sled test #1, and 95% for sled test #2 versus #3. The crash pulse curve correlation was 93% for #4 versus #5. The complete ORM values were 76% for sled test #1 versus #2, 77% for #1 versus #3, 86% for #2 versus #3, and 81% for sled test 4# versus #5. The comparisons including sled test #1 were rated as nearly dissimilar, test #2 versus #3 as similar, and #4 versus #5 as fairly similar.

Series 2 – Correlations between Simulations and Sled Tests The Madymo model of the EuroNCAP system predicted the three EuroNCAP tests in Series 2 with 75, 76 and 78% in terms of complete ORM values. For the groups of dummy signals the predictabilities were 73, 73, and 74%, and the predictabilities were 86, 86, and 87% for the group of belt signals. The Madymo model of the USNCAP system predicted both USNCAP tests in Series 2 with 75% in terms of complete ORM values. For the groups of dummy signals the predictabilities were 72 and 73%, and the predictabilities were 87 and 88% for the group of belt signals. The Madymo models were judged as useful by simulation engineers.

Series 3 – Correlations between Simulations and Sled Tests The complete ORM values for the predictabilities were 68 and 69% to sled test #2 and #3, respectively. The average peak value predictions were 81% for both comparisons, the average peak value occurrence time predictions were 81% for both comparisons, and the curve shape correlations were 60 and 61%, respectively. The Madymo models were judged as useful by simulation engineers.

Side Impacts

The ORM was used to assess the repeatability of static and dynamic airbag component tests and the mathematical model predictabilities of these tests. Two static tests with no vent hole were conducted and reached 93% in complete ORM value. These tests were rated as similar by engineers. For the static tests with ordinary vent holes, the complete ORM values was 80% and the test repeatability rated as poor by engineers since the pressure curves in the pelvis part were unlike from 15 to 30 ms due to a squeezed tube; the pelvis pressure curve shape correlation was 68%. Three dynamic tests were conducted at 7 m/s. The complete ORM values were 86 and 89% for the tests judged as fairly similar and 91% for the test judged as similar. Three dynamic tests were conducted at 4 m/s. Two pairs of tests were judged as fairly similar due to different pelvis accelerations. The complete ORM values for these were 85 and 86%. One pair of tests were judged as almost

identical and reached a complete ORM value of 95%. The mathematical model predictabilities were 87 and 90% for the static tests with no vent holes, and 86% for the static test with ordinary vent whole. No predictability ORM value was calculated for the test with the squeezed tube. The predictabilities of the dynamic tests at 7 m/s were 82 to 85%, and the predictabilities of the tests at 4 m/s were 71 to 76%. Overall, the mathematical model was judged to predict the tests sufficiently well.

Sled tests were conducted with settings estimated by mathematical modelling. The sled test repeatability was 85% and judged as similar for almost all included signals. The sled crash pulse shape repeatability was 87% and the door crash pulse shape repeatability was 89%. The mathematical models predicted both sled tests with only 46% in terms of complete ORM values. However, the most important signals resulted in higher ORM values: the predictabilities of the abdomen were both 60% and the predictabilities of the rib deflections were 74 and 78%. The sled tests predicted the full scale crash test with 55 and 56% in terms of complete ORM values. For the mathematical models, the predictability of the full scale crash test was 50% for the abdomen and 57% for the rib deflections. The predictability of the full scale crash test was judged as insufficient for both the sled tests and the mathematical models. Hence, a new loop of tests and simulations were performed.

The repeatability of the next pair of sled tests was 71% and these tests were judged as nearly dissimilar. The sled and the door crash pulse shape correlations were 82% and 79%, respectively. It was possible to assess which of the two that was more accurate than the other, and that test will be used in the coming comparisons. The sled tests predicted the full scale crash tests with 62%; the predictability of the abdomen was 60% and the predictability of the rib deflections was 80%. The corresponding number for the mathematical model predictability was 51% for the abdomen and 56% for the rib deflections. Two pairs of sled tests with improved safety components were then conducted. The repeatabilities within these pairs were 77 and 80%. The correlations between the pairs were 73 and 64%, indicating that the changes of the safety components resulted in changes of the dummy readings. The final safety components were then used in a full scale crash test. The sled test predictability of this full scale crash test was only 51%: the rib deflections predictabilities were 75% and 76%, and the abdomen predictabilities were 27 and 28%. The very low values of the abdomen predictability were due to that the front and middle forces measured in the sled tests were around or just below zero and a bit above zero in the full scale crash test which resulted in ORM values close to zero. The correlation between the abdominal rear forces, which were the important signals, was 96%.

DISCUSSION

The ORM

The Objective Rating Method, ORM, was published in 2005 by [2] as a tool for assessing the predictability of Madymo simulation models to mechanical tests. [2] stated that high correlation is 65% or above for mechanical test repeatability. This statement was based on component tests with one Hybrid III 50%-ile without arms and lower legs. Ten different tests were repeated ten times, and in each test thirty signals were recorded. All signals of the repeated tests were then compared to the first test in each test series. However, which signals compared or weight factors used, are not specified. According to the authors, special attention was given to positioning the dummy before each test to ensure good repeatability, and a well-defined environment was used in the tests. A couple of previous studies, [1] and [3], have shown that an ORM value of 65% most likely is a too low threshold to correctly rate two tests as highly repeatable, although the threshold is depending on which signals and components that are included. This study shows that correlations above 65% can be judged as poor, indicating that a higher threshold should be more proposed. Furthermore, there should preferably be different requirements on scalar and curve shape correlations.

In an earlier study ([1]) the ORM was applied to twenty-six pairs of rear-end sled tests in order to assess the BioRID II repeatability and reproducibility. The tests were conducted at two crash test sites. Four BioRID II dummies, five different seats, and three crash pulses were used. Both criteria and dummy readings were compared. The BioRID II repeatability, in terms of complete ORM values, ranged from 83 to 90%, and the reproducibility ranged from 74 to 78%. In this study, other types of tests were included, as well as mathematical models. For those of the comparisons conducted in this study that were conducted with the same set-up selection methodology as used in [1], i.e. Set 7, the complete ORM values for frontal sled test repeatability ranged from 81 to 88%, and the mathematical system model predictability ranged from 68 to 78%. Hence, the sled test repeatability is in the same range for rear-end and frontal impacts. In the BioRID II repeatability ORM values a few criteria are included, that may be one reason why the rear-end tests have slightly higher complete ORM values than the frontal impacts. The reproducibility ORM values for the rear-end sled tests are somewhat higher than the predictability ORM values for the frontal simulation models. That was expected, since two physical dummies in duplicated sled tests should mirror each other fairly well, while simplified mathematical models of the dummy and the sled test environment hardly can mirror its mechanical counterpart exactly. Nevertheless, the predictabilities of the mathematical

models are fairly good, and the models are judge as useful by engineers. For the side impact tests conducted in this study, the sled test repeatability ranged from 77 to 85% for the MDB sled tests that were properly conducted. For the pole sled tests conducted in the same side impact project, the repeatability ranged from 72 to 80%. Side impact tests are more complex than most of the rear-end and frontal impact tests, and therefore it might be more difficult to reach good repeatability in side impacts sled test. Additional studies to this one are needed to define thresholds for good correlation in terms of ORM values for several types of tests, simulations and purposes within the car safety area.

The ORM evaluation in this study, with shifted and scaled curves, showed a perfect match between ORM values and engineering judgement if the values for okay correlations are above 80%, acceptable correlations between 70 and 80%, not okay correlations between 55 and 70%, and for ORM values below 55% there are no correlations at all due to very dissimilar curves. Although the ORM values match the engineering judgement perfectly for these evaluations, it might be possible that these levels do not correlate for ORM values and engineering judgement for other types of curves and more realistic differences between curves than those formulated in this evaluation. The comparison in which one of the curves was shifted 0.001 s in X resulted in a curve shape ORM value of 79% (Table 1) although these curves are almost identical from an engineers point of view. This ORM value is a result of the curve shape correlation being calculated as a relative error in each time step (Equation 2), and consequently the ORM value will decrease for curves with steep slopes and for curves oscillating around zero in amplitude.

One application of a correlation rating method is to develop an automatic tuning tool that uses the calculated objective correlations as optimization targets for tuning mathematical models to mimic performed tests. Automatic tuning tools have been evaluated on module level for several types of airbags. The ranges of the tuning parameters were predefined to avoid unrealistic values. The models were tuned toward airbag pressure and impactor deceleration measured in several linear impactor or drop tests at different loading configurations. In some cases, also bag pressures measured in static tests have been used for model tuning. Not only the maximum peak correlation, its occurrence time correlation and the curve shape correlation from ORM have been evaluated, but also the Absolute Error Integral Criteria, the Error Integral Criteria, and the Curve Slope Criteria. Different weights on the six criteria have been evaluated in order to find a set of signals that result in models which with high accuracy are rated equally by the correlation tool and experienced engineers. Studies have been conducted to find appropriate weights to the criteria in order to achieve a proper performance of the models during the full test period. Up to date, no weight set-up has been found that works properly for all kind of airbag tests. However, for impactor tests it appears as criteria measuring the curve shape correlation should be more heavily weighted. Investigations of cutting off sections of signals that are known to be inaccurate measured or modelled, such as the deployment peak in an airbag pressure curve, and long sections of signals that are very low in amplitude compared to the peak values, have been made. Both these cuttings have been found beneficial, however these make the automatic tuning tool more complicated to use and require engineering experience. Applying these cuttings and extra weights on the curve correlation criteria, 80% have been identified as an appropriate threshold for good correlation for airbag model to impactor tests. Values between 70 and 80% are acceptable. For some airbag models values above 80% were not reached, likely due to deficiencies in physical representation of relevant parameters in the mathematical models or insufficient information of the test set-up or test repeatability. A prerequisite for automatic tuning is that the criteria used in the optimization must accurately rate the status of models the whole way from poor to good correlation, otherwise the correlation values cannot be used as optimization targets since the inaccuracy for poor and less good correlations may reduce the possibility for the optimization method to find the optimum. Further, in some cases, models have reached the same correlation values, although some were judged as sufficiently good and some were judged as insufficiently good by engineers. This indicates that both the ORM and the ORM with additional criteria need further development for the automatic tuning application. The correlation method needs to be further evaluated on a broad number of test cases in terms of correlation criteria included, weights used, time ranges used, signals and components included, and correlation criteria thresholds. Also, tuning parameters and optimization method for multiple loading conditions must be further evaluated.

Frontal Impacts

Series 1 – Evaluation of Included Signals Not only USNCAP sled test repeatability was evaluated in this part; one pair of USNCAP BiW driver tests and five pairs of EuroNCAP BiW driver and passenger tests were also analysed. For the BiW tests, the set-up of the sets were slightly different from those sets used for the USNCAP sled tests due to different requirements and measured signals. However, the methods of selecting the signals were similar. Hence, the complete ORM values from the different types of tests can be compared. Since the repeatabilities of

these tests also were rated by engineering judgement on the five-level-scale, a reliable set for ORM comparisons of duplicated frontal tests can be recommended. The complete ORM values from the different types of tests are given in Table 2 along with the engineering judgement of the test repeat-ability. These data are also plotted in Figure 3 to Figure 9, along with the linear trend lines and their R^2 values. The R^2 values show the quality of the line fits; the higher the R^2 value is, the better is the fit. Among the seven sets evaluated in this study, Set 7 had the highest R^2 value, 88%. The other sets had R^2 values between 27 and 78%. Hence, the set-up used in Set 7 is the set-up that most likely results in an ORM value that correlates to the engineering judgement. The method of selecting signals, components, and groups, and their weights according to Set 7 is further explained in the Recommendations section.

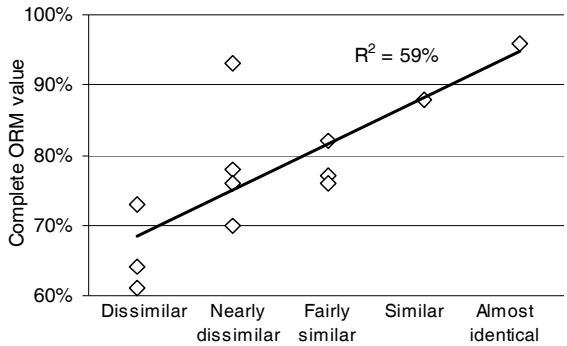


Figure 3. Complete ORM values calculated with signal Set 1 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

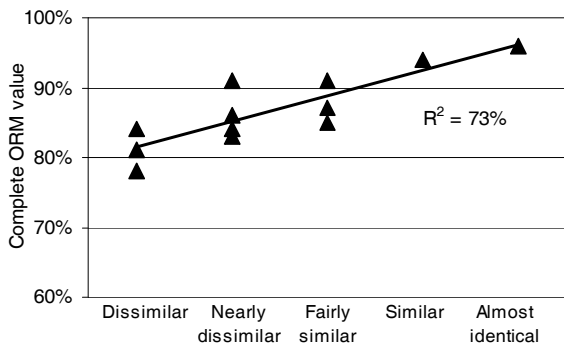


Figure 4. Complete ORM values calculated with signal Set 2 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

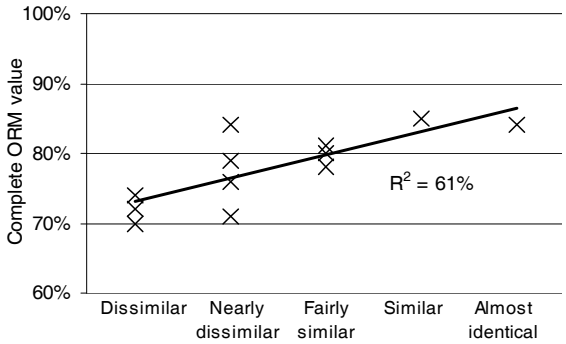


Figure 5. Complete ORM values calculated with signal Set 3 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

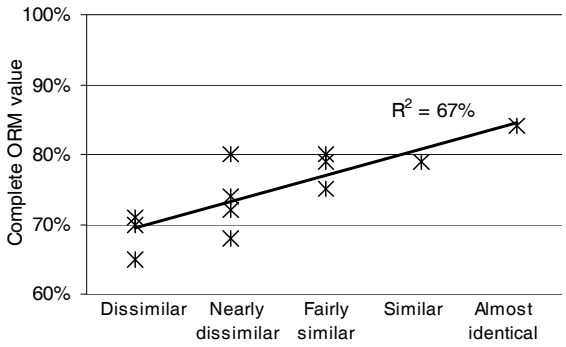


Figure 6. Complete ORM values calculated with signal Set 4 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

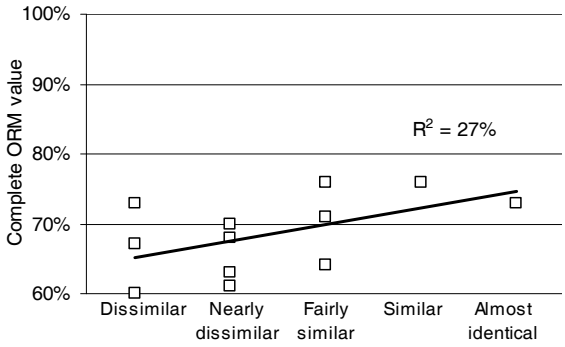


Figure 7. Complete ORM values calculated with signal Set 5 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

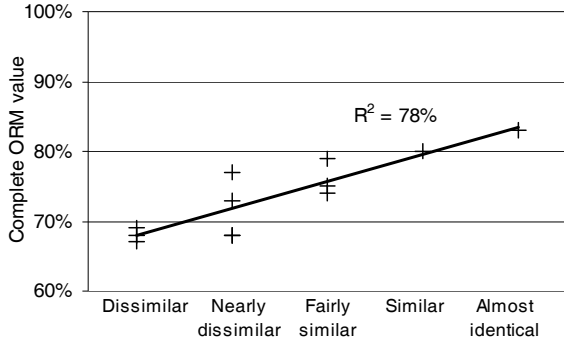


Figure 8. Complete ORM values calculated with signal Set 6 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is.

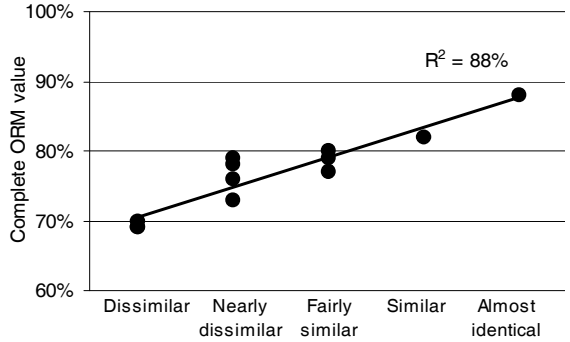


Figure 9. Complete ORM values calculated with signal Set 7 for the twelve compared pairs of tests in Series 1 versus the engineering judgement of the repeatability of each pair. The linear regression shows the trend, and the R^2 value shows how good the fit is

Table 2.
Complete ORM values in percent and engineering judgement for the comparisons in Series 1.

Test / Set-up	USNCAP Sled Tests	USNCAP BiW Tests	EuroNCAP BiW Tests D1	EuroNCAP BiW Tests D2	EuroNCAP BiW Tests D3	EuroNCAP BiW Tests D4	EuroNCAP BiW Tests D5	EuroNCAP BiW Tests P1	EuroNCAP BiW Tests P2	EuroNCAP BiW Tests P3	EuroNCAP BiW Tests P4	EuroNCAP BiW Tests P5
Set 1	96	93	78	73	64	76	77	61	82	70	76	88
Set 2	96	91	86	84	78	83	85	81	91	84	87	94
Set 3	84	71	76	70	74	79	80	72	78	84	81	85
Set 4	84	72	80	71	70	74	80	65	79	68	75	79
Set 5	73	63	68	73	67	70	71	60	64	61	76	76
Set 6	83	68	77	68	69	73	79	67	74	68	75	80
Set 7	88	78	79	69	70	76	79	69	80	73	77	82
Eng. judge.	Almost identical	Nearly dissimilar	Nearly dissimilar	Dissimilar	Dissimilar	Nearly dissimilar	Fairly similar	Dissimilar	Fairly similar	Nearly dissimilar	Fairly similar	Similar

<u>Series 2 – Correlations between Sled Tests</u> On the whole, the repeatability ORM values were high for the comparisons conducted for the tests in Series 2. Visual inspections of the tests by experienced test engineers assessed that the EuroNCAP sled tests were almost identical and the USNCAP sled tests were similar. Further, all signal correlations were rated visually in three grades: good/either/poor. Based on these grades, thresholds for ORM values for good and poor correlation can be recommended. Peak value correlations that were judged as good mostly had ORM values above 90%, and those judged as poor mostly had ORM values below 80%. Values between 80 and 90% were either good or poor, and it is recommended to assess these individually. For peak value occurrence time correlations ORM values above 90% are almost always good; values below 90% should be assessed individually. There are two not too rare reasons to low peak value occurrence time ORM values besides the case when the peaks occur with a long time apart. First, ORM values can be confusing for times early in the crash event since the ORM value is a ratio between the first and the last of the two compared times. Second, low peak time occurrence ORM values can be a result of double peaks with almost equal amplitudes, but the maximum peak in opposite order; this indicates a poor correlation by engineering judgement only if the peak occurrence time is critical. The curve shape correlations generally results in lower ORM values than scalar correlations. Curve shape correlation ORM values above 75% were almost always judged as good, while values below 60% were judged as poor. Curve correlations ORM values between 60 and 75% are recommended to be evaluated individually. Consequently, comparisons of sets with significant signals that contain about 30% curve shape correlations and 70% peak value and their occurrence time correlations have good correlations if the complete ORM values are above circa 80%. Independent of the complete ORM value, it is recommended to visually inspect all curves and corresponding component ORM values to uncover confusions. A drawback with the ORM is that the curve correlation (Equation 2) punishes signals with long intervals of amplitudes close to zero and signals that cross through the time axis numerous times unfairly hard. Consequently, the complete ORM value can be a bit too low compared to the grade given by engineering judgement. Further, the

complete ORM value is a weighted mean which may hide a few ORM values that are lower than wanted, an inspection will identify these. The crash pulse should preferable be analysed separately; ORM values above 90% for crash pulse shape corresponds to almost identical curves.

Series 3 – Correlations between Sled Tests The five tests in Series 3 were compared both visually and by aid of the ORM. The visual inspection of tests #1, #2, and #3, which were intended to be identical, showed that #1 versus #2 and #1 versus #3 were nearly dissimilar and the corresponding complete ORM values were 76 and 77%, respectively. The reason for the poor correlations of these tests is most likely that the pulse differed a lot for test #1; the corresponding crash pulse curve shape correlations were only 70 and 71%. Sled tests #2 and #3 were judged as similar; the corresponding complete ORM value was 86% and the pulse shape correlation was 95%. Sled test #4 versus #5 was fairly similar; the corresponding complete ORM value was 81% and the crash pulse shape correlation was 93%. For these four comparisons, the amplitudes of the component ORM values are given in Figure 10 to Figure 13. In each figure, the ORM components are sorted in amplitude, and the three patterns on the bars represent peak ORM values, peak occurrence time ORM values, and curve shape ORM values. Comparing Figure 10 and Figure 11 with Figure 12 and Figure 13, it can be seen that the ORM values in general are higher in the latter two figures than in the former two; as indicated by the complete ORM values and in line with the engineering judgements. Using the thresholds for good (≥90% for scalars and ≥75% for curve shapes) and poor (<80% for scalars and <60% for curve shapes), one can see that the two pairs of tests rated as nearly dissimilar only have 56 and 59% of the scalars rated as good, and only 22 and 28% of the curve shapes rated as good, while the two pairs of tests judged as nearly similar and similar have 75 and 78% of the scalars and 67 and 83% of the curve shapes rated as good. Hence, the tests compared in Figure 12 and Figure 13 have both about the same amount of scalar components rated as good but the latter has less curve shapes rated as good than the former. This partly explains why the complete ORM values and engineering judgement differed for these comparisons. Nevertheless, most component ORM values are good in Figure 13 making these tests sufficiently repeatable for most purposes. By plotting the bars in order, one can easily find out if just a few components cause a low complete ORM value, or if all components are low and therefore aggregate to a low complete ORM value. This kind of plot is also useful in order to find out which signals and components that may need to be further analysed.

In order to get a feeling for the ORM values, Figure 15 to Figure 19 contain the dummy readings from sled test #1 to #3 and the Madymo simulation in Series 3 and the corresponding ORM values.

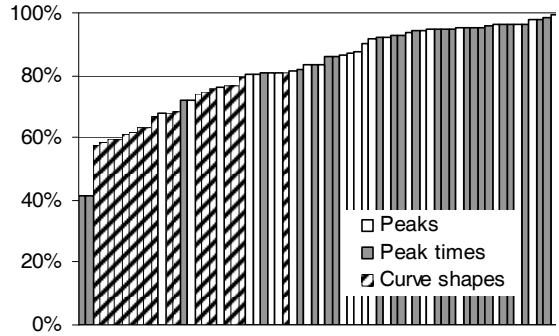


Figure 10. All component ORM values for comparison #1 versus #2 in Series 3. The ORM values are sorted in amplitude and the different patterns show the kind of component compared. The complete ORM value is 76%, and the crash pulse curve shape correlation is 71%.

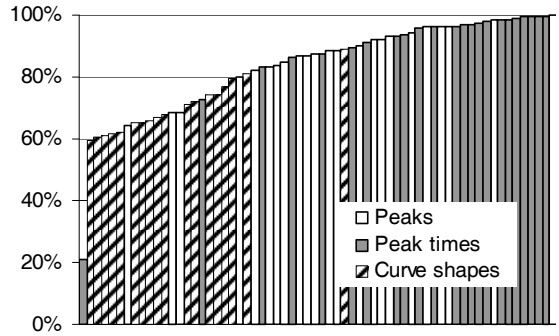


Figure 11. All component ORM values for comparison #1 versus #3 in Series 3. The ORM values are sorted in amplitude and the different patterns show the kind of component compared. The complete ORM value is 77%, and the crash pulse curve shape correlation is 70%.

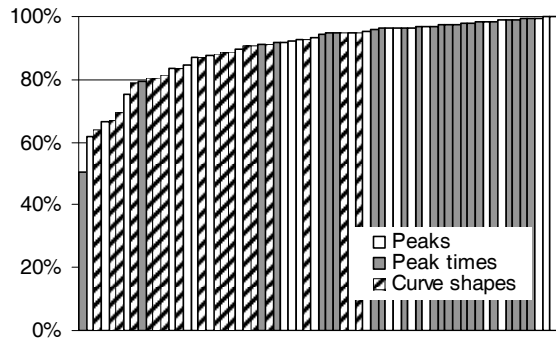


Figure 12. All component ORM values for comparison #2 versus #3 in Series 3. The ORM values are sorted in amplitude and the different patterns show the kind of component compared. The complete ORM value is 86%, and the crash pulse curve shape correlation is 95%.

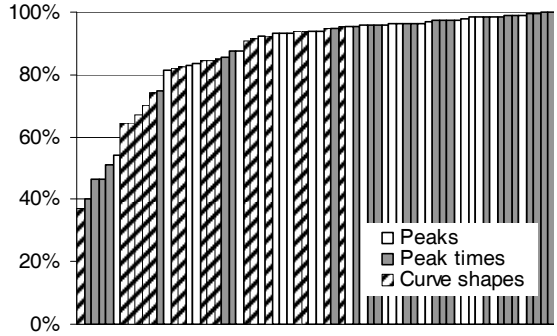


Figure 13. All component ORM values for comparison #4 versus #5 in Series 3. The ORM values are sorted in amplitude and the different patterns show the kind of component compared. The complete ORM value is 81%, and the crash pulse curve shape correlation is 93%.

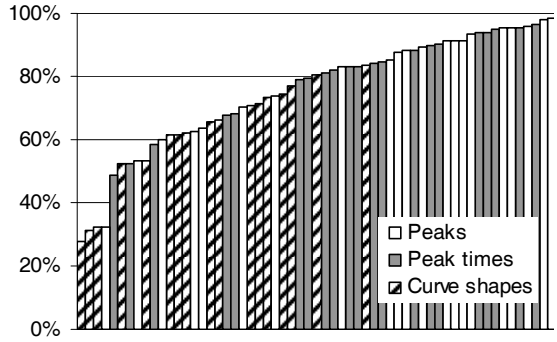
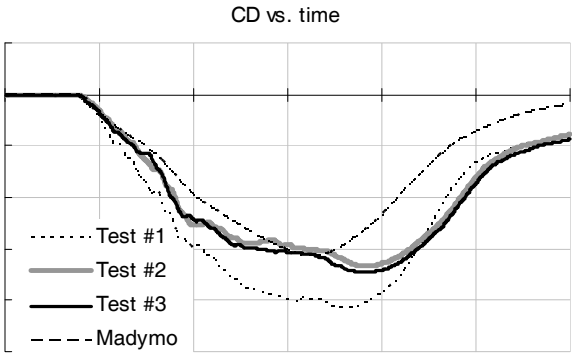
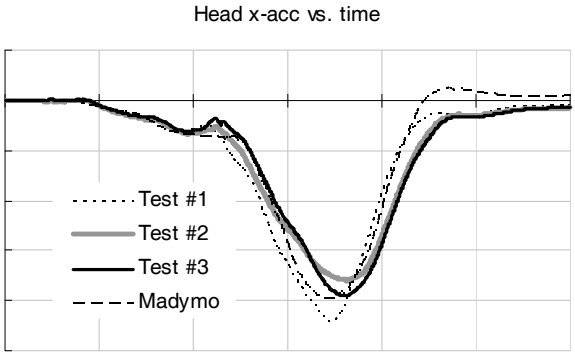


Figure 14. All component ORM values for comparison #3 in Series 3 vs. the Madymo simulation. The ORM values are sorted in amplitude and the different patterns show the kind of component compared. The complete ORM value is 69%, and the crash pulse curve shape correlation is 91%.

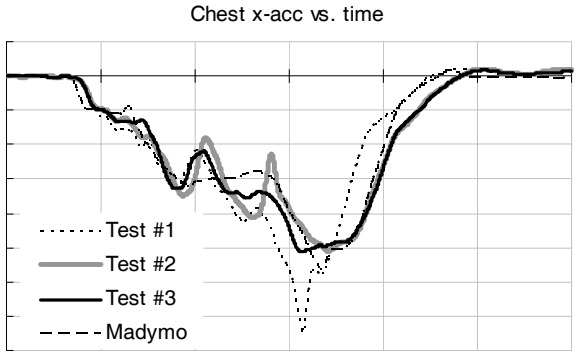


	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	83	93	81
#2 vs. #3	-	-	96	98	95
Madymo vs. #3	-	-	91	83	66

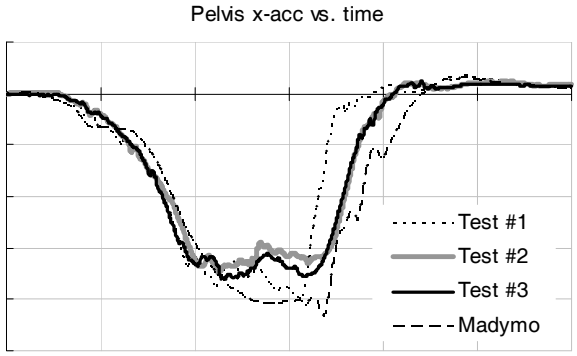
Figure 15. Chest deflection curves for sled test #1 to #3 and Madymo simulation and ORM values for the comparisons #1 versus #3 (different pulses), #2 versus #3, and Madymo simulation versus #3.



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	88	97	72
#2 vs. #3	-	-	92	99	89
Madymo vs. #3	-	-	98	96	66



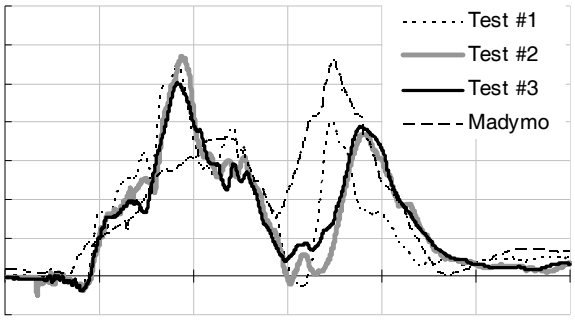
	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	68	100	71
#2 vs. #3	-	-	99	92	88
Madymo vs. #3	-	-	91	94	77



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	90	73	68
#2 vs. #3	-	-	96	99	91
Madymo vs. #3	-	-	83	68	71

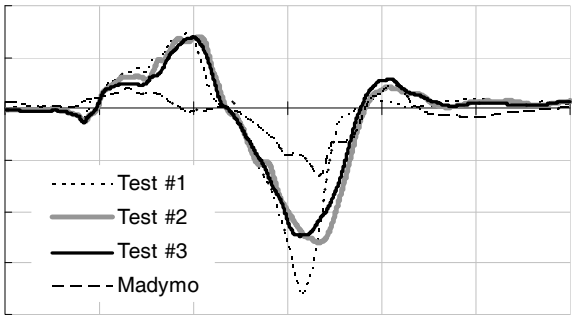
Figure 16. Head, chest, and pelvis x-acceleration curves for sled test #1 to #3 and Madymo simulation and ORM values for the comparisons #1 versus #3 (different pulses), #2 versus #3, and Madymo simulation versus #3.

Head z-acc vs. time



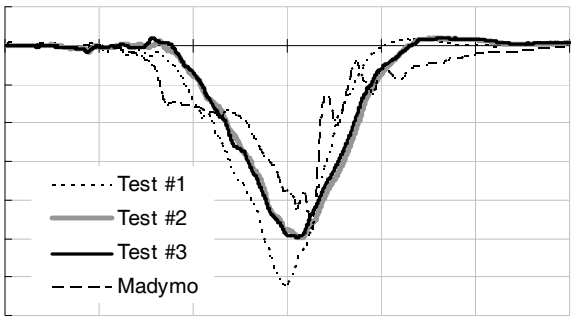
	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	92	99	-	-	61
#2 vs. #3	88	98	-	-	79
Madymo vs. #3	89	52	-	-	53

Chest z-acc vs. time



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	94	97	68	99	61
#2 vs. #3	100	96	95	94	80
Madymo vs. #3	33	49	53	94	31

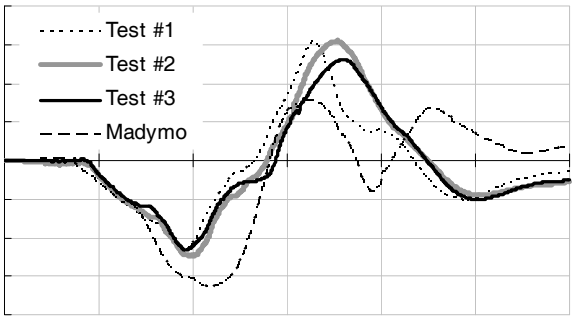
Pelvis z-acc vs. time



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	80	96	65
#2 vs. #3	-	-	100	98	89
Madymo vs. #3	-	-	96	95	52

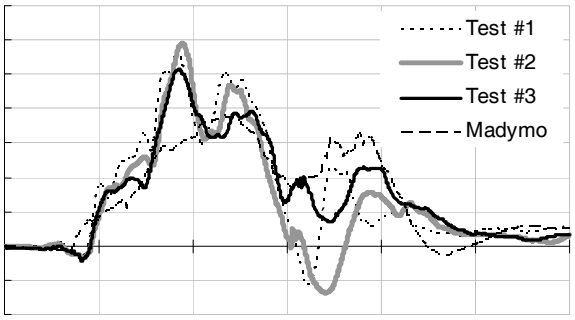
Figure 17. Head, chest, and pelvis z-acceleration curves for sled test #1 to #3 and Madymo simulation and ORM values for the comparisons #1 versus #3 (different pulses), #2 versus #3, and Madymo simulation versus #3.

Upper Neck Fx vs. time



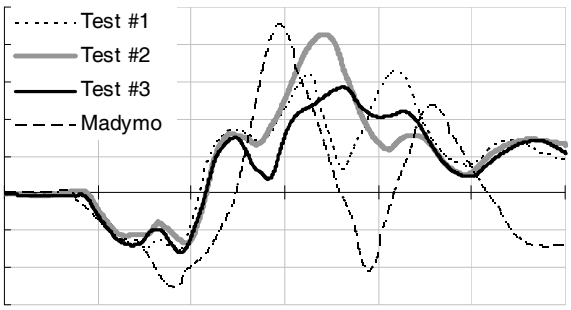
	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	85	91	98	98	60
#2 vs. #3	84	98	93	96	81
Madymo vs. #3	60	90	70	88	32

Upper Neck Fz vs. time



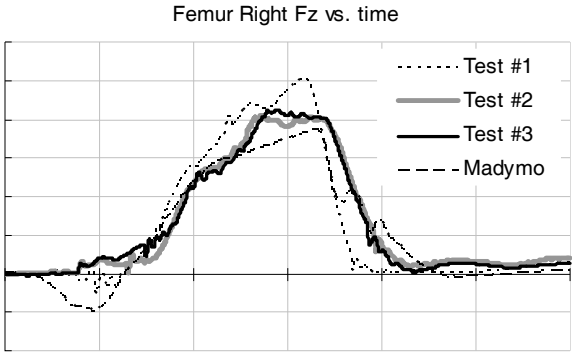
	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	92	100	-	-	61
#2 vs. #3	87	98	-	-	64
Madymo vs. #3	74	79	-	-	62

Upper Neck My vs. time

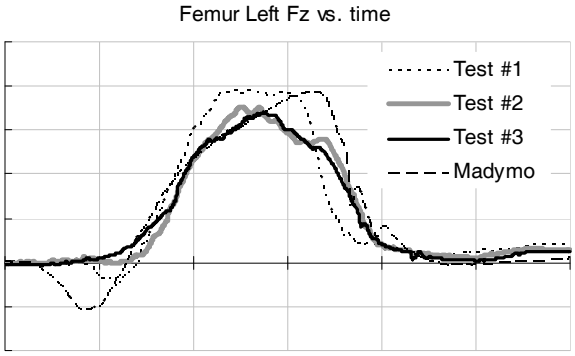


	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	87	86	97	99	65
#2 vs. #3	66	95	83	97	70
Madymo vs. #3	63	81	63	96	28

Figure 18. Upper neck F_x, F_z, and M_y curves for sled test #1 to #3 and Madymo simulation and ORM values for the comparisons #1 versus #3 (different pulses), #2 versus #3, and Madymo simulation versus #3.



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	84	88	62
#2 vs. #3	-	-	97	98	84
Madymo vs. #3	-	-	88	84	61



	Max Peak	Max Peak Time	Min Peak	Min Peak Time	Curve Shape
#1 vs. #2	-	-	87	96	67
#2 vs. #3	-	-	96	91	81
Madymo vs. #3	-	-	88	84	62

Figure 19. Femur right left F_z curves for sled test #1 to #3 and Madymo simulation and ORM values for the comparisons #1 versus #3 (different pulses), #2 versus #3, and Madymo simulation versus #3.

Series 2 and 3 – Correlations between Simulations and Sled Tests Although the ORM signal sets differed somewhat between the predictability and the repeatability assessments, these ORM values were compared. In general, and as expected, the Madymo models have lower predictability ORM values than the corresponding sled test repeatability ORM values. Nevertheless, the Madymo models were found to be useful. The Madymo models average peak value prediction ranged from 81 to 88%, the average peak value time occurrence prediction ranged from 81 to 95%, and the average curve shape prediction ranged from 60 to 69%. Corresponding values for the sled test repeatabilities are 91 to 94%, 95 to 98%, and 77 to 85%. The Madymo model of the EuroNCAP system in Series 2 predicts the three EuroNCAP tests with 75 to 78% and the corresponding sled test repeatability ORM values ranged from 81 to 86%. The Madymo model of the USNCAP system predicts both of the two USNCAP tests in Series 2 with 75% and the corresponding sled test repeatability was 87%. The Madymo model of the USNCAP sled tests with the Hybrid III 50%-ile in Series 3 predicts test #2 and #3 with 68 and 69% and the corresponding sled test repeatability was 86%. From a simulation engineering point of view, the Madymo models verified in this study were assessed as useful for product development and evaluation. Hence, Madymo models that reached about 80% or higher of the sled test repeatability ORM values were found to be useful. Nevertheless, higher ORM values are needed to fully predict tests with simulations. Depending on the purpose of the models, additional weight on the curve shapes can result in better understanding if the models predict trends adequately or not in cases when the peak values are poorly predicted. In order to attain useful ORM values of predictability, it is important to include mainly those signals which will be used in coming simulations; otherwise the ORM values will be confused by less predictive and irrelevant signals. Further, the recommended ORM values for models may differ depending on the type of test and system used, but also on modelling technique (MBS or FEM), and the purpose of the simulation.

Side Impacts

The ORM was used to guide in a side impact project. Both sled tests and full scale crash test were performed, and mathematical models were built up to predict the tests and improve the EuroNCAP results. The complete ORM values for the mathematical model predictabilities of the sled tests and the full scale crash tests, as well as the sled test predictabilities of the full scale crash tests, were rather low. Therefore, the dummy signals were divided into subgroups that made it possible to focus on those signals predominantly important for test purpose. Thus, it could be assessed that these signals were sufficiently well predicted. The ORM was also used to show that improvements of the safety components influenced the dummy performance by comparing ORM values from tests with and without the improved system to the repeatability values of truly duplicated tests. By analysing the signals individually, it was possible to find out which of the signals that were highly, moderately and insignificantly influenced by the change in the safety components. Overall, the ORM values assisted in the communication with the customer during this project.

The mathematical models of the airbags used were normally tuned to mimic two configurations of static tests and two configurations of dynamic test, and each configuration were repeated at least once. The mathematical models in the side impact project were tuned manually, and the ORM was applied afterwards to assess their predictabilities. For the airbag presented in the side impact part, the test repeatabilities were in all cases but one very good. Good

test repeatability and representative tests are crucial for tuning a mathematical model into an accurate model of a real airbag. In case of a low test repeatability ORM value, it is important to find out if the low value was caused by a test failure, a result of any ORM shortcomings, or just a spread in the tested components. If the low value was caused by a test failure, the mathematical model can be tuned to mimic just the accurate test. If the low value was caused by a shortcoming in the ORM and the signals correlations are judged better than the ORM value indicates, one may cut off sections of the signals or change the weights to better rate the repeatability. Otherwise, the mathematical model must be tuned to mimic both tests and will most likely return predictability ORM values lower than the repeatability for the corresponding tests.

CONCLUSIONS

ORM is a useful tool, especially for test repeatability and reproducibility assessments, although it should be used with care and knowledge. For all ORM correlations performed, it is recommended to check the included signals and components in order to identify scalar or curves correlations that might have been too low rated due to shortcomings of the ORM. The ORM allows for agreement, in advance, on a quality level for tests and mathematical models. Beneficial is that the ORM not only compares peak values but also curve shapes. Furthermore, the ORM compares two tests; many methods require several tests and that is normally not available in daily work.

This study shows that high ORM values correspond to good correlations, but for a few cases good correlations result in somewhat too low ORM values. For curve correlations rated as poor by ORM, the ORM does not sufficiently discriminate between curves that are poor but useful and poor but useless. This might be a drawback for simulation model predictability assessments, since the correlation level can be rather low although the models fulfil their purposes.

RECOMMENDATIONS

The usefulness of an ORM correlation is highly depending on the set of signals and components used. In this study, the Set 7 correlated to engineering judgement most accurately. The method of selecting signals and components according to Set 7 for a system correlation assessment is in detail described in A. to D. Normally, all signals and components should be compared up to the time when the crash is over. An example of a signal set selected with this method can be found in Table 3.

A. *Crash pulse group* Compare the crash pulse shape separately; it should not be included in the complete system comparison. Hence, set the group weight to zero.
B. *Criteria group* Criteria may be included in the complete ORM value, or analysed separately depending of the aim of the comparison. When the criteria are excluded in the complete ORM value, the comparison will focus more on overall repeatability, reproducibility or predictability. If the criteria group is excluded in the complete ORM value, set the group weight to zero. Otherwise, let the weight of this group be equal to the number of criteria.
C. *Dummy group* Include all kinematics and load cells signals acting in the motion plane, and avoid resultant signals. Signals in other directions than the motion plane are normally much lower and will most likely confuse the findings with less important data if they are included. For all included signals, include both minimum and maximum peaks if these are significant and not enormously different in amplitudes; otherwise include only the highest peak. For all included peaks, also include their occurrence times. Include the curve shape correlation for all included signals. For each included signal, set the total weight of the peak or peaks to 1, set also the total weight of the peak occurrence time or times to 1, and set the curve shape correlation weight to 1. Let the weight of the dummy group be equal to the number of included signals. In order to better focus on different parts of the dummy, the dummy group can be divided into subgroups.
D. *Safety group* Include all signals from the safety equipment that show how these parts perform. For all included signals, include both minimum and maximum peaks if these are significant and not enormously different in amplitudes; otherwise include only the highest peak. For all included peaks, also include their occurrence times. Include the curve shape correlation for all included signals. For each included signal, set the total weight of the peak or peaks to 1, set also the total weight of the peak occurrence time or times to 1, and set the curve shape correlation weight to 1. Let the weight of the safety group be equal to the number of included signals.

On signal level, good correlations are predominately associated with ORM values of 90% or above for peak value and peak value occurrence time correlations, and with ORM values of 75% or above for curve shape correlations. For extremely good curve correlation, as wanted for crash pulses, 90% is a preferable threshold. Peak value correlation ORM values below 80% and curve shape correlation ORM values below 60% are generally associated with poor correlation or no correlation at all.

Using an appropriate set of signals, in which about one third of the components are curve shapes, the proposal of ORM value representing good repeatability for frontal sled tests is 80% or above. Sled

test reproducibility, and more complex frontal tests, such as BiW and full scale crash tests, should be rated as good with a bit lower complete ORM values. The requirement on the predictability ORM values of a mathematical system model should be lower than the corresponding repeatability ORM values.

REFERENCES

[1] Eriksson, L., Zellmer, H. 2007 "Assessing the BioRID II Repeatability and Reproducibility by Applying the Objective Rating Method (ORM) on Rear-End Sled Tests" In proceedings of the 20th International Technical Conference of the Enhanced Safety of Vehicles (June 18-21, Lyon, France) Paper No. 07-0201-O.

[2] Hovenga, P. E., Spit, H. H., Uijldert, M., Dalenoort, A. M. 2005 "Improved Prediction of Hybrid-III Injury Values using Advanced Multibody Techniques and Objective Rating" In proceedings of the SAE 2005 World Congress & Exhibition (April, Detroit, MI, USA) Paper No. 05AE-222.

[3] Eriksson, L. 2008 "Repeatability and reproducibility ORM values for the BioRID II criteria and the dynamic assessment points used in the Euro NCAP rear-end testing protocol" In proceedings of the 2nd International Conference of Whiplash – Neck Pain in Car Crashes (Erding, Germany, Nov. 18-19). TÜV SÜD Akademie, Germany.

Table 3
Signals and weights used for a USNACP frontal sled test according to the selection method proposed in Recommendations section.

Group	W_{group}	Signal	Comp.	$W_{scalar/shape}$
Crash Pulse	0	Sled x-acc	Curve shape	1
Criteria	0	HIC36	Scalar	1
		Chest 3ms	Scalar	1
Dummy	12	Head x-acc	Min peak	1
			Min peak time	1
			Curve shape	1
		Head z-acc	Max peak	1
			Max peak time	1
			Curve shape	1
		Chest x-acc	Min peak	1
			Min peak time	1
			Curve shape	1
		Chest z-acc	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
		Pelvis x-acc	Min peak	1
			Min peak time	1
			Curve shape	1
		Pelvis z-acc	Min peak	1
			Min peak time	1
			Curve shape	1
		Chest defl.	Max peak	1
			Max peak time	1
			Curve shape	1
		Upper Neck F_x	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
		Upper Neck F_z	Max peak	1
			Max peak time	1
			Curve shape	1
		Upper Neck M_y	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
		Femur Left	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
		Femur Right	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
Safety	3	Webbing disp.	Max peak	0.5
			Max peak time	0.5
			Min peak	0.5
			Min peak time	0.5
			Curve shape	1
		Belt Force B3	Max peak	1
			Max peak time	1
			Curve shape	1
		Belt Force B6	Max peak	1
			Max peak time	1
			Curve shape	1