Benchmarking and Evaluating the Interpretation of Bibliographic Records

**Joffrey Decourselle · Fabien Duchateau · Trond
Aalberg · Naimdjon Takhirov · Nicolas Lumineau**

Abstract In a global context which promotes the use of explicit semantics for sharing information and developing new services, the MAchine Readable Cataloguing (MARC) format that is commonly used by libraries worldwide has demonstrated its limitations. The semantic model for representing cultural items presented in the Functional Requirements for Bibliographic Records (FRBR) is expected to be a successor of MARC, and the complex transformation of MARC catalogs to FRBR catalogs (FRBRization) led to the proposition of various tools and approaches. However, these projects and the results they achieve are difficult to compare on a fair basis due to a lack of common datasets and appropriate metrics. Our contributions fill this gap by proposing the first public benchmark for the FRBRization process.

Joffrey Decourselle
LIRIS, UMR5205
Universit Claude Bernard Lyon 1
Lyon, France
Tel.: +33
E-mail: jdecours@liris.cnrs.fr

Fabien Duchateau
LIRIS, UMR5205
Universit Claude Bernard Lyon 1
Lyon, France
Tel.: +33
E-mail: fduchate@liris.cnrs.fr

Trond Aalberg
NTNU
Trondheim, Norway
Tel.: +
E-mail: trondaal@idi.ntnu.no

Naimdjon Takhirov
Westerdals - Oslo School of Arts, Communication and Technology
Faculty of Technology, Oslo, Norway
Tel.: +
E-mail: taknai@westerdals.no

Nicolas Lumineau
LIRIS, UMR5205
Universit Claude Bernard Lyon 1
Lyon, France
Tel.: +33
E-mail: nluminea@liris.cnrs.fr

Note that this paper is an extension of a previous version published in TPDL 2016 [1].

1 Assessing strengths and weaknesses - results by test

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by test.
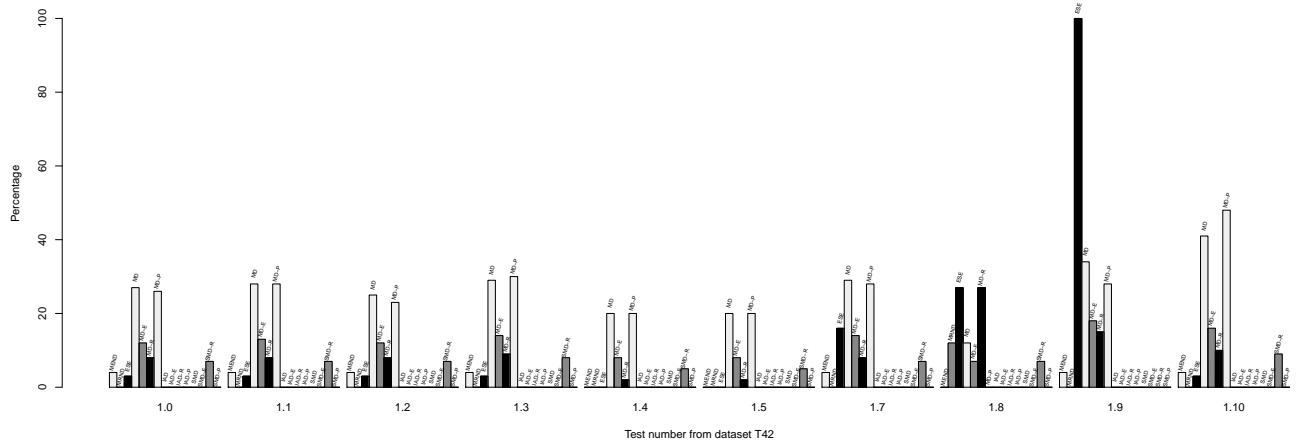


**Fig. 1** Quality results for FRBR-ML with dataset T42 tests 1.x
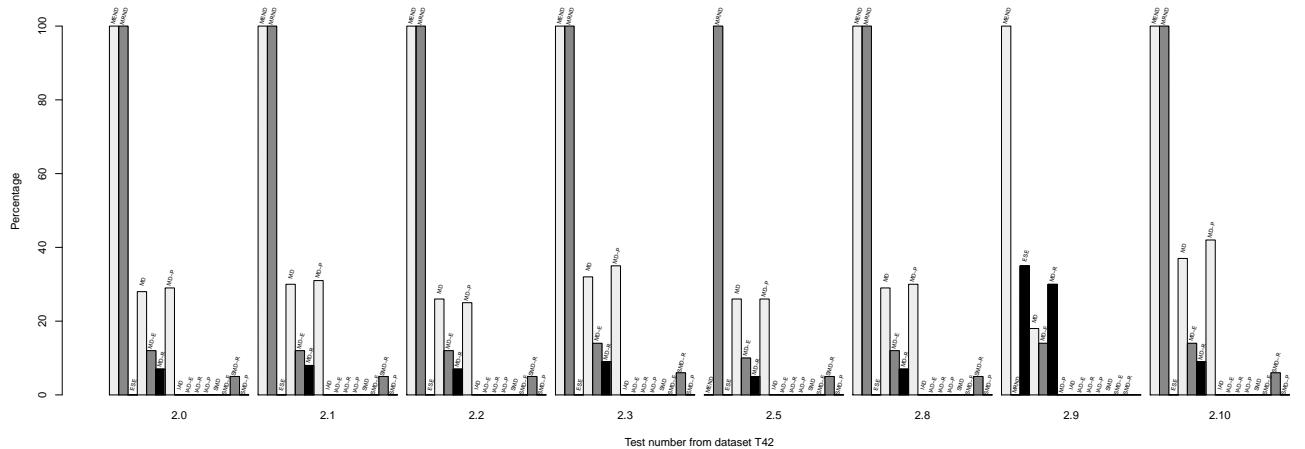

Fig. 2 Quality results for FRBR-ML with dataset T42 tests 2.x
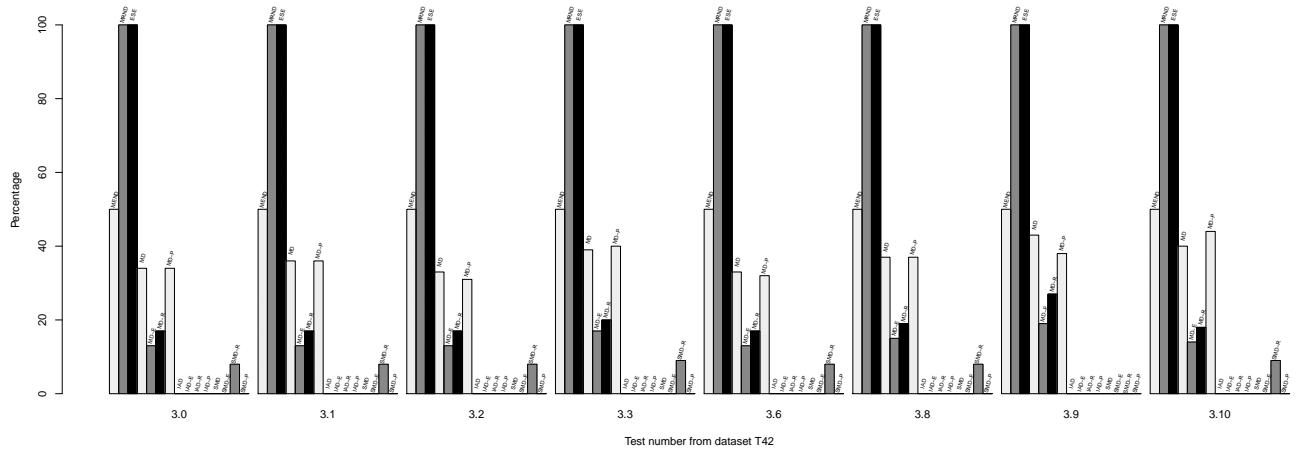
Fig. 3 Quality results for FRBR-ML with dataset T42 tests 3.x
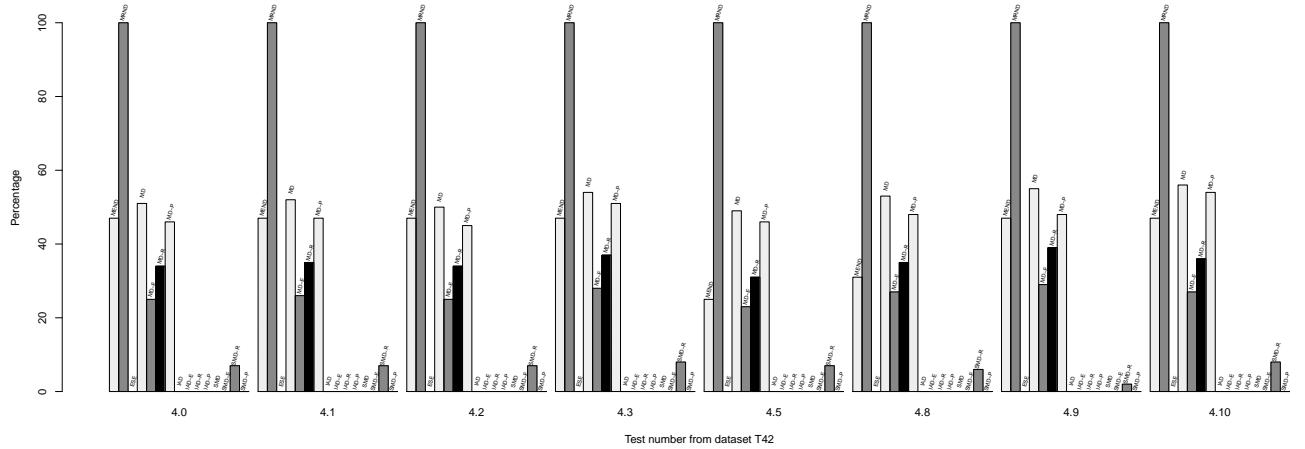
Fig. 4 Quality results for FRBR-ML with dataset T42 tests 4.x
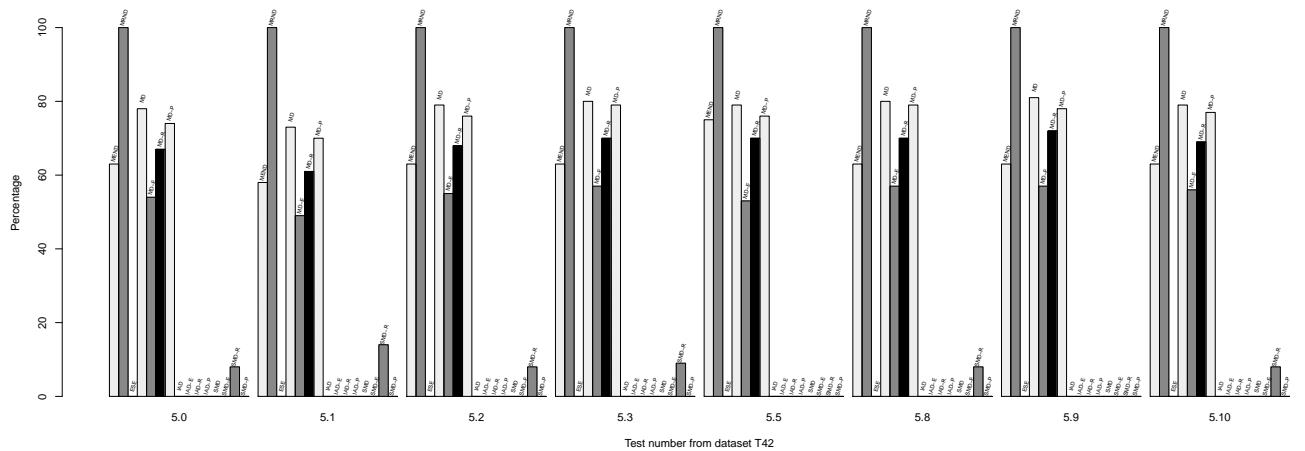
Fig. 5 Quality results for FRBR-ML with dataset T42 tests 5.x
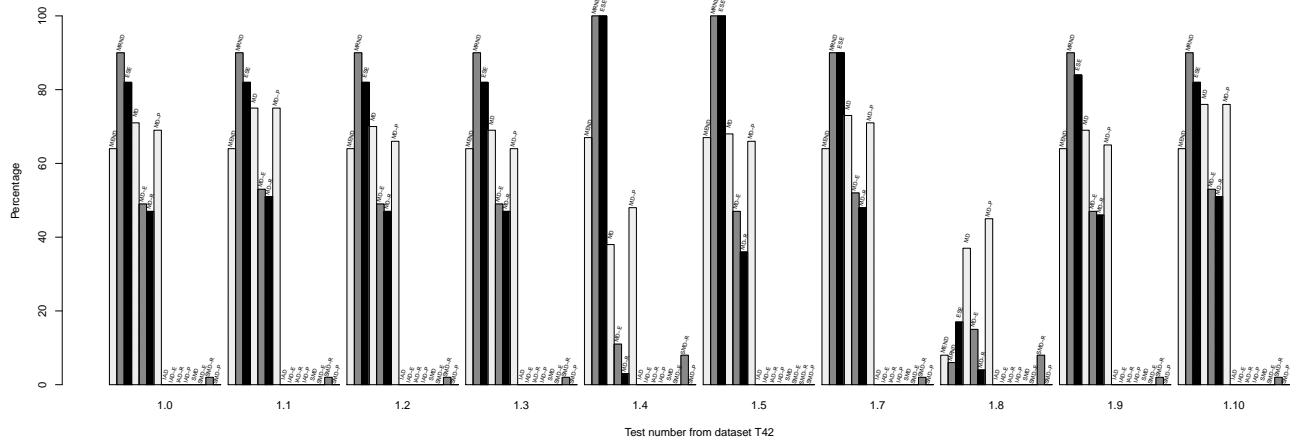
**Fig. 6** Quality results for Variation/VFRBR with dataset T42 tests 1.x
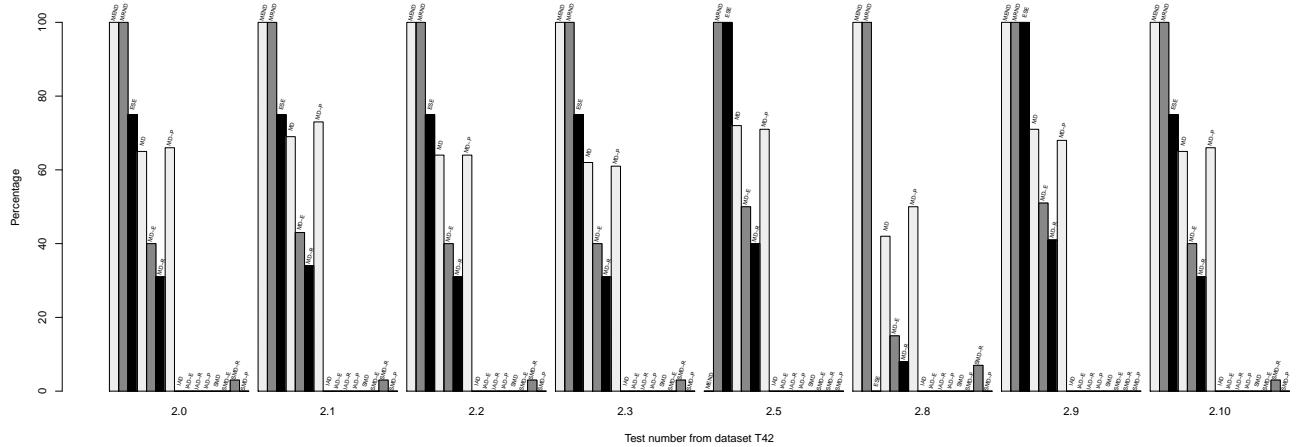
Fig. 7 Quality results for Variation/VFRBR with dataset T42 tests 2.x
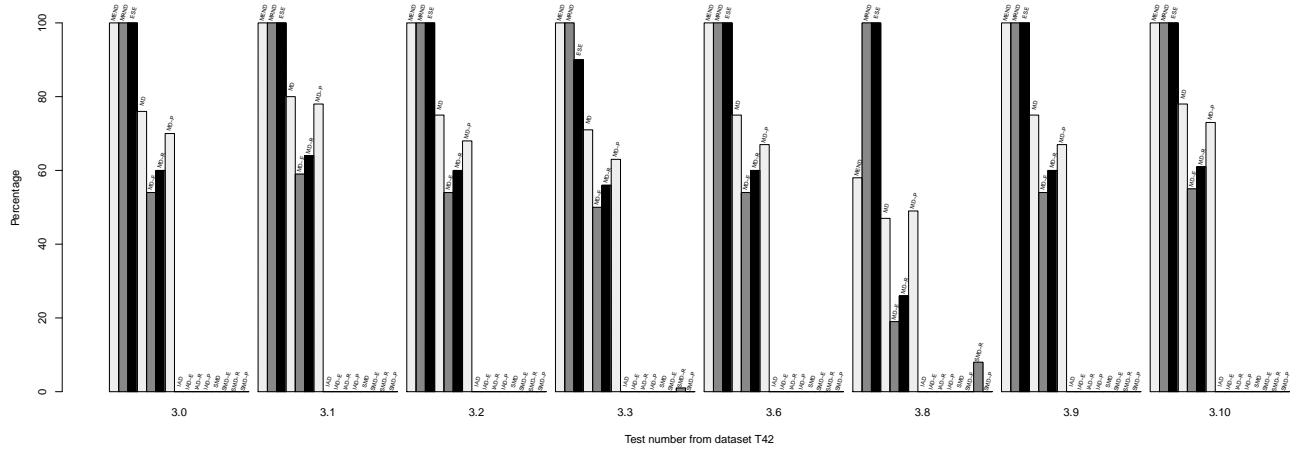
Fig. 8 Quality results for Variation/VFRBR with dataset T42 tests 3.x
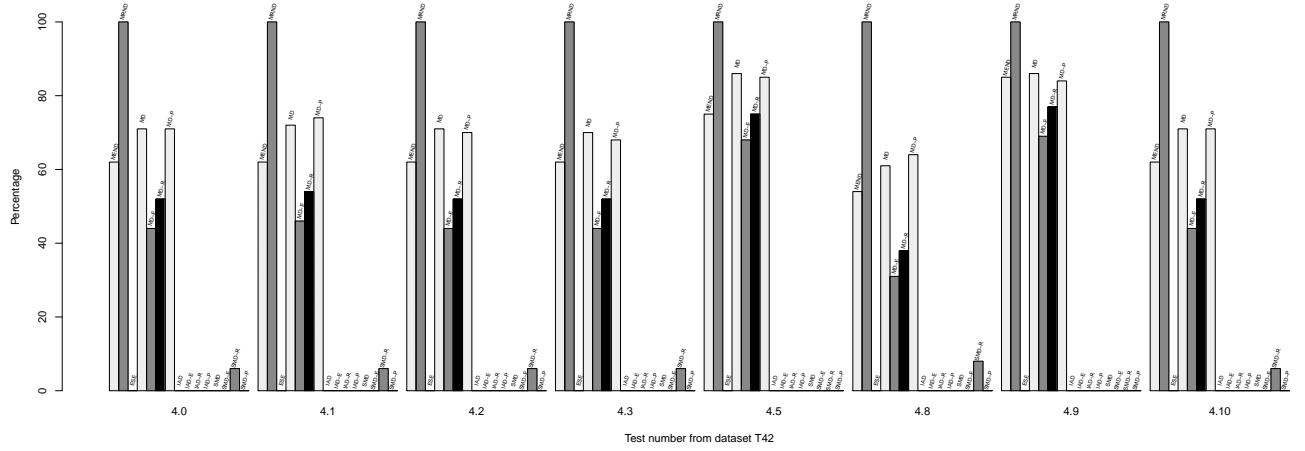
**Fig. 9** Quality results for Variation/VFRBR with dataset T42 tests 4.x
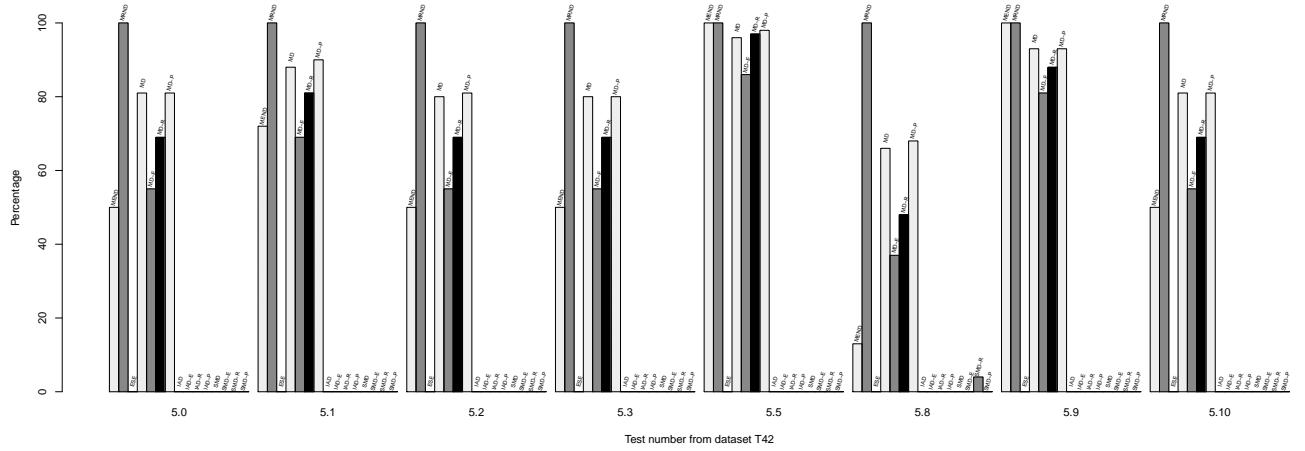
**Fig. 10** Quality results for Variation/VFRBR with dataset T42 tests 5.x
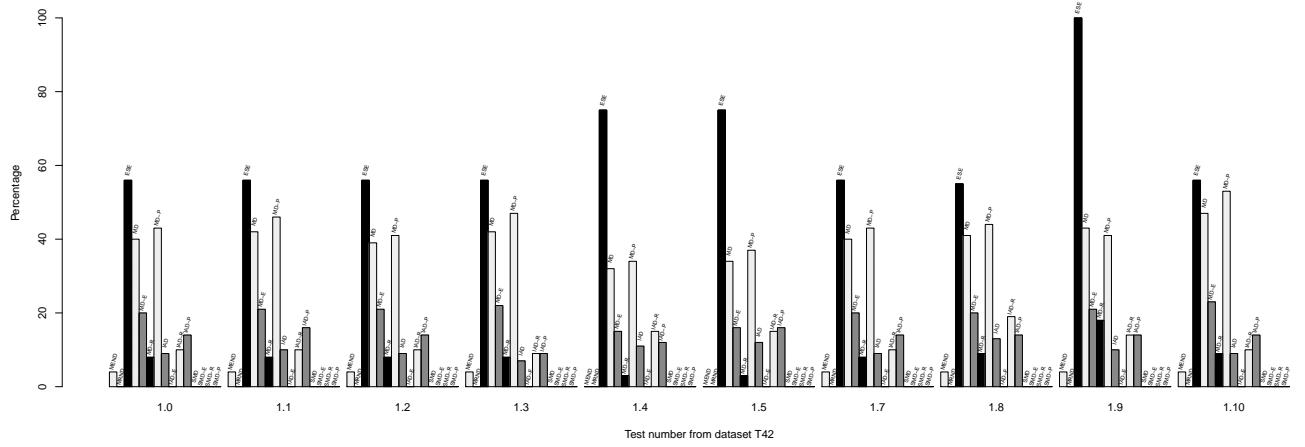
Fig. 11 Quality results for Extensible Catalog with dataset T42 tests 1.x
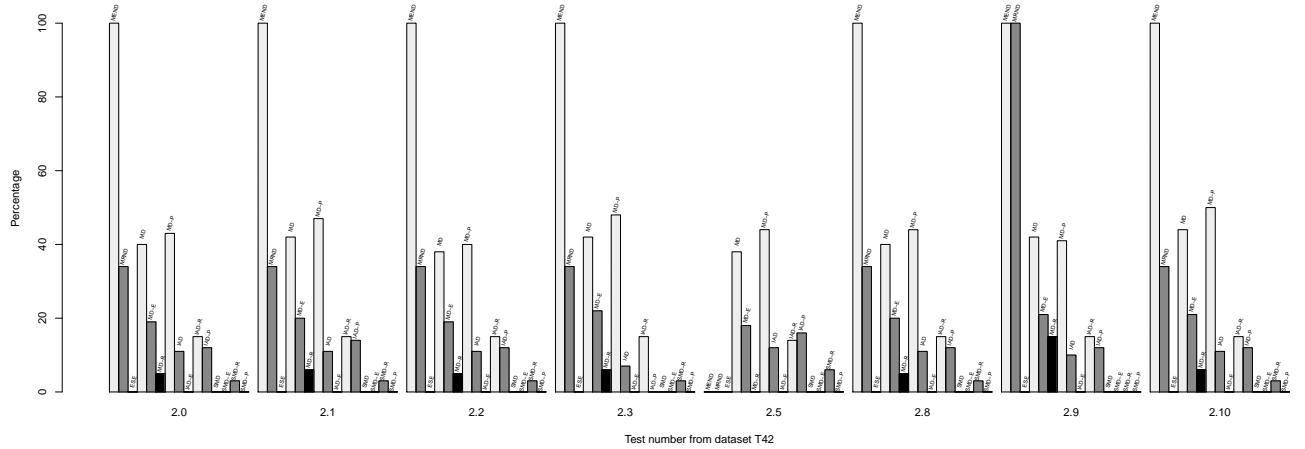
Fig. 12 Quality results for Extensible Catalog with dataset T42 tests 2.x
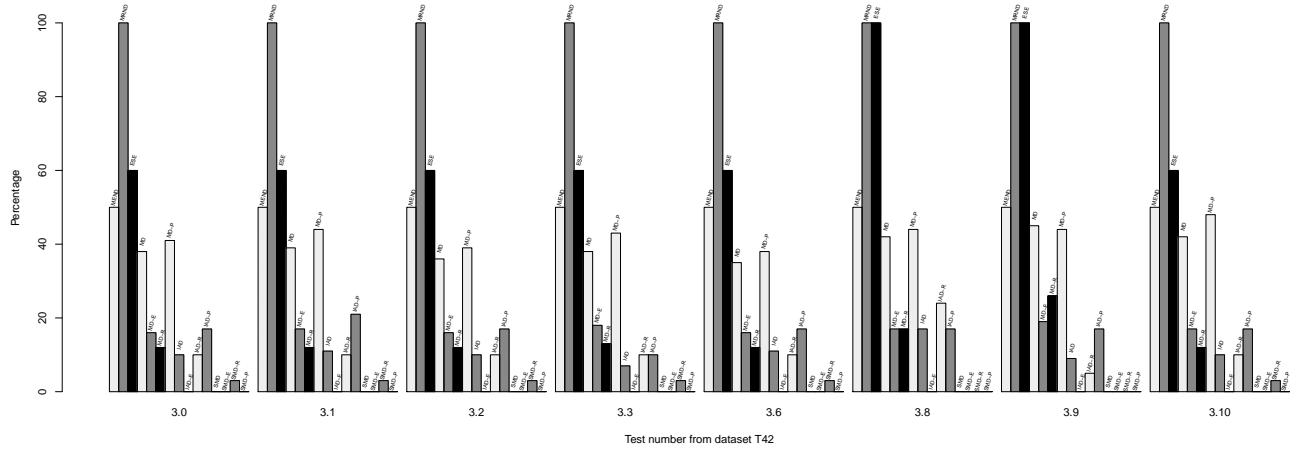
Fig. 13 Quality results for Extensible Catalog with dataset T42 tests 3.x
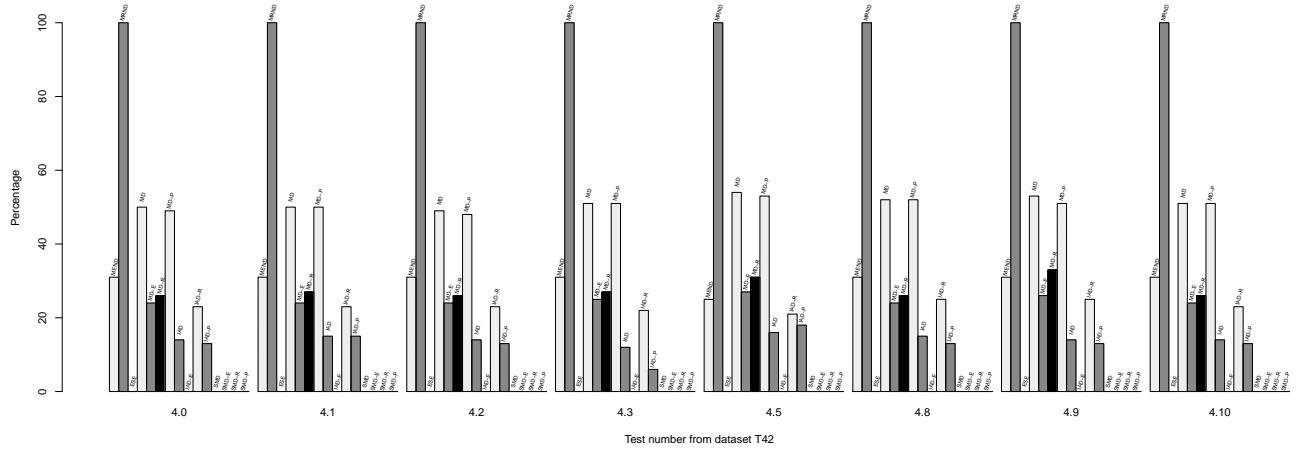
Fig. 14 Quality results for Extensible Catalog with dataset T42 tests 4.x
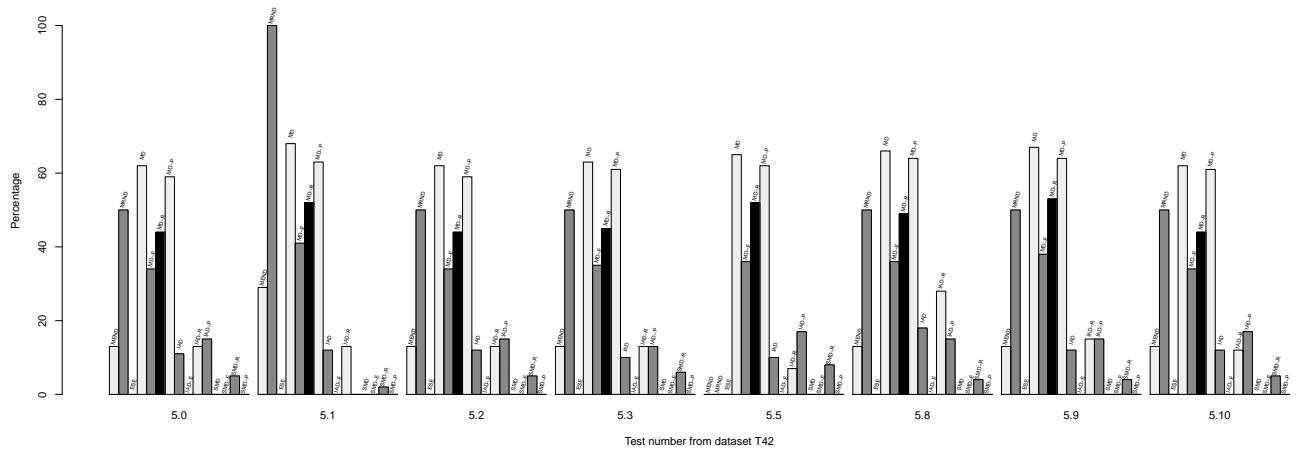
Fig. 15 Quality results for Extensible Catalog with dataset T42 tests 5.x

2 Assessing strengths and weaknesses - results by metric

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by metric.
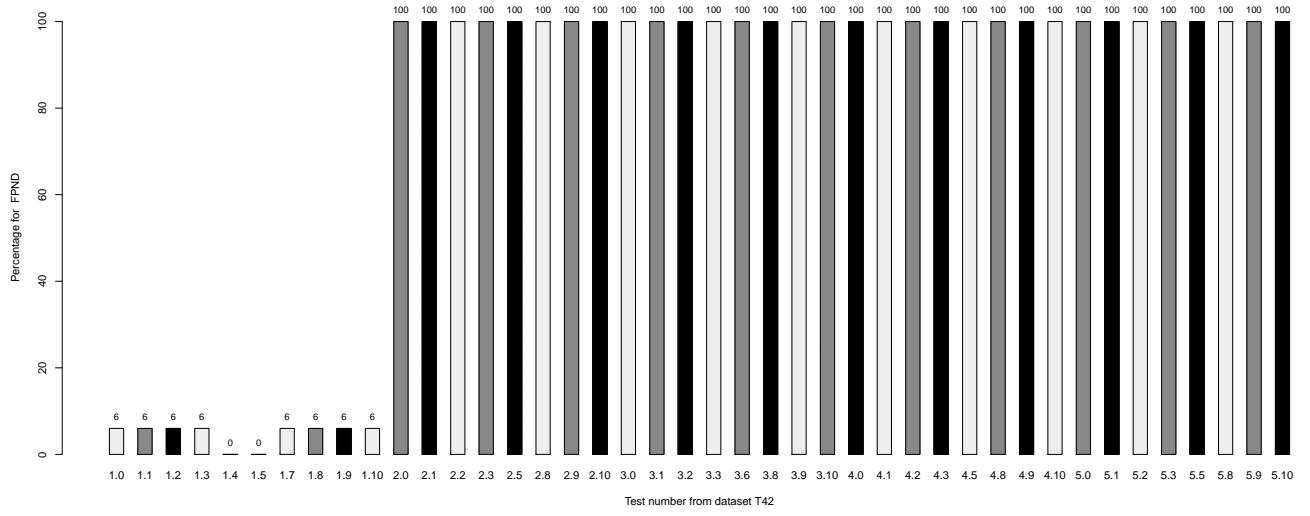
Fig. 16 Quality results for FRBR-ML with dataset T42 and metric FPND
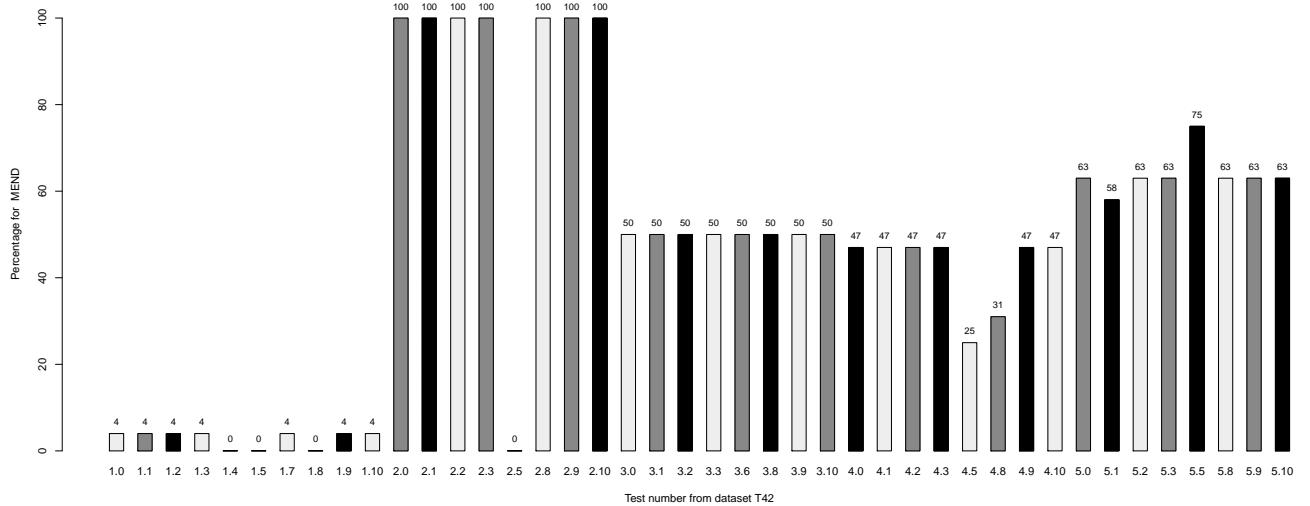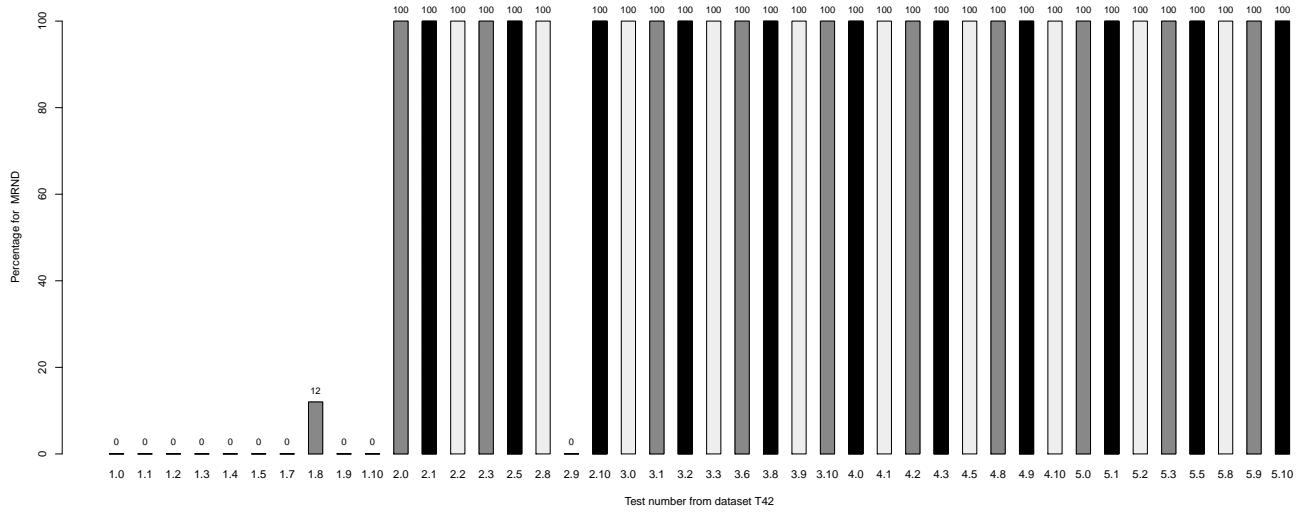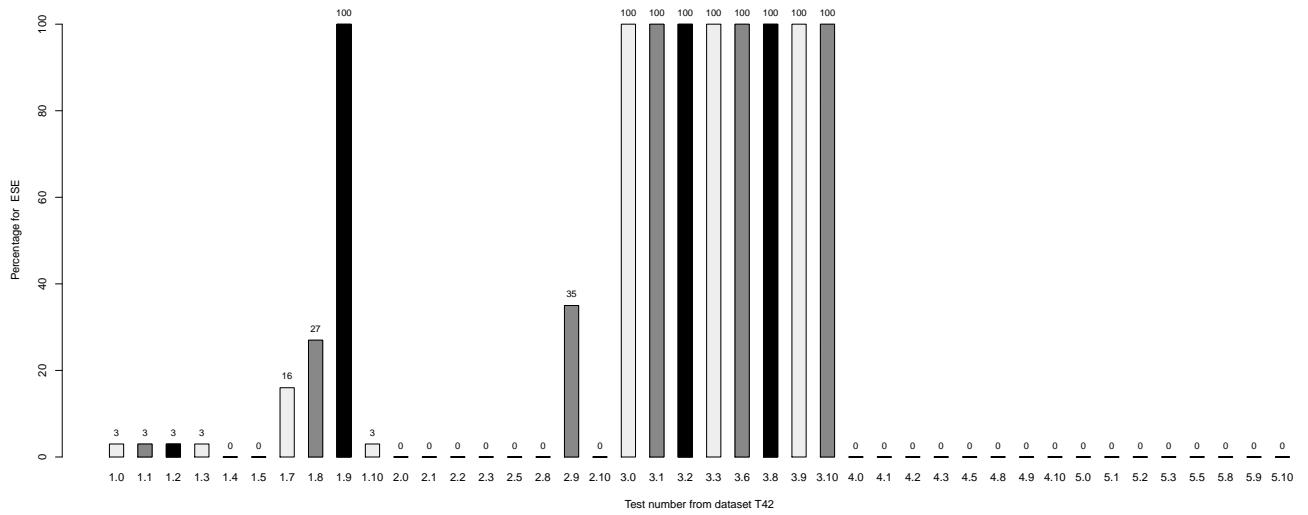
Fig. 17 Quality results for FRBR-ML with dataset T42 and metric MEND

Fig. 18 Quality results for FRBR-ML with dataset T42 and metric MRND

Fig. 19 Quality results for FRBR-ML with dataset T42 and metric ESE
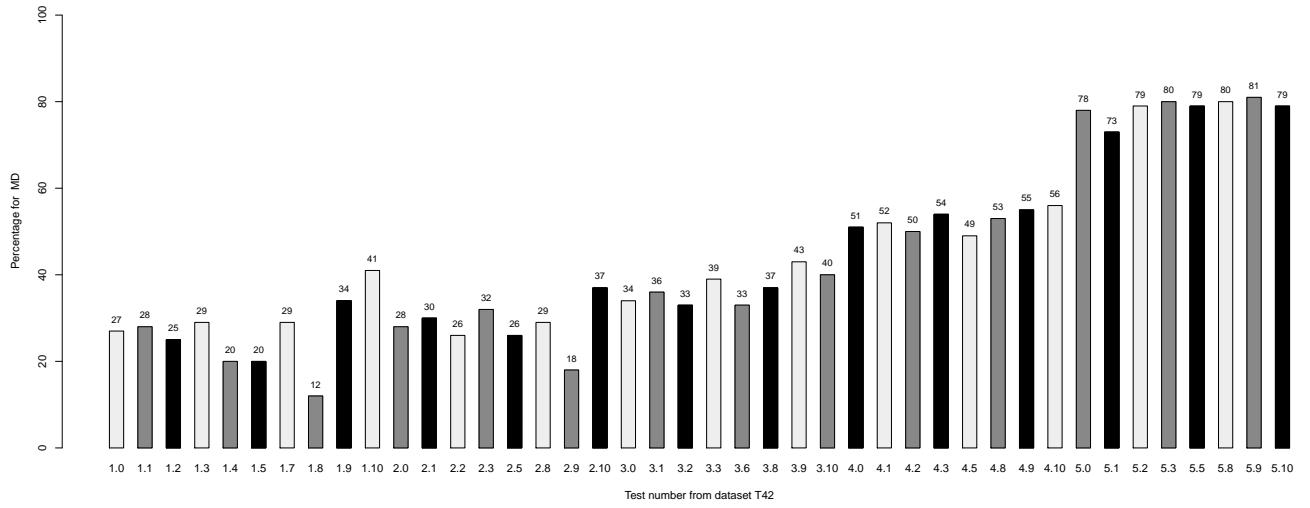
Fig. 20 Quality results for FRBR-ML with dataset T42 and metric MD
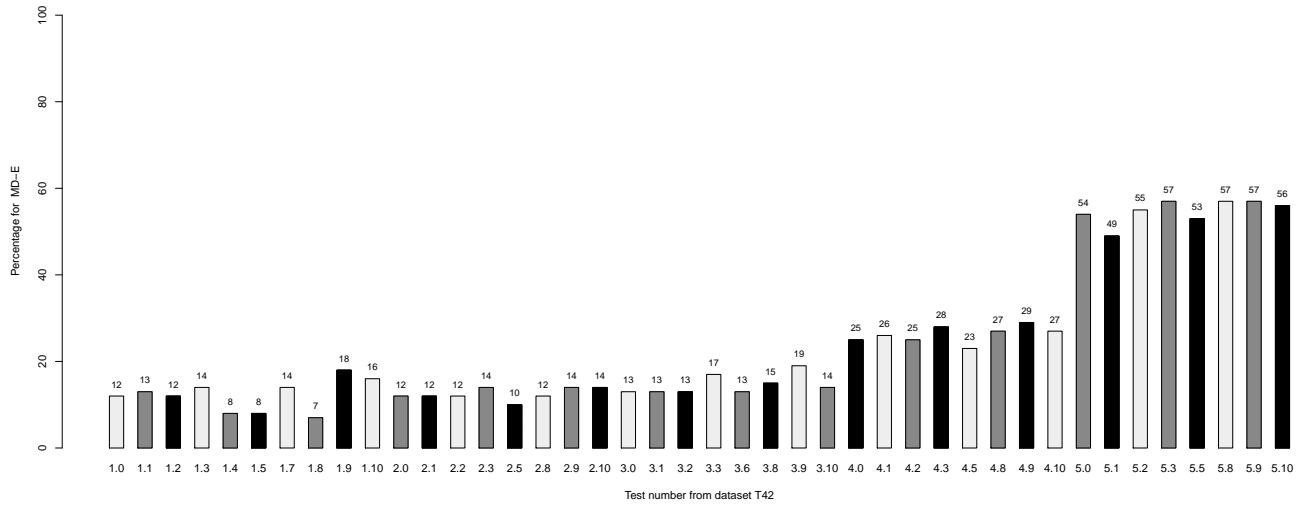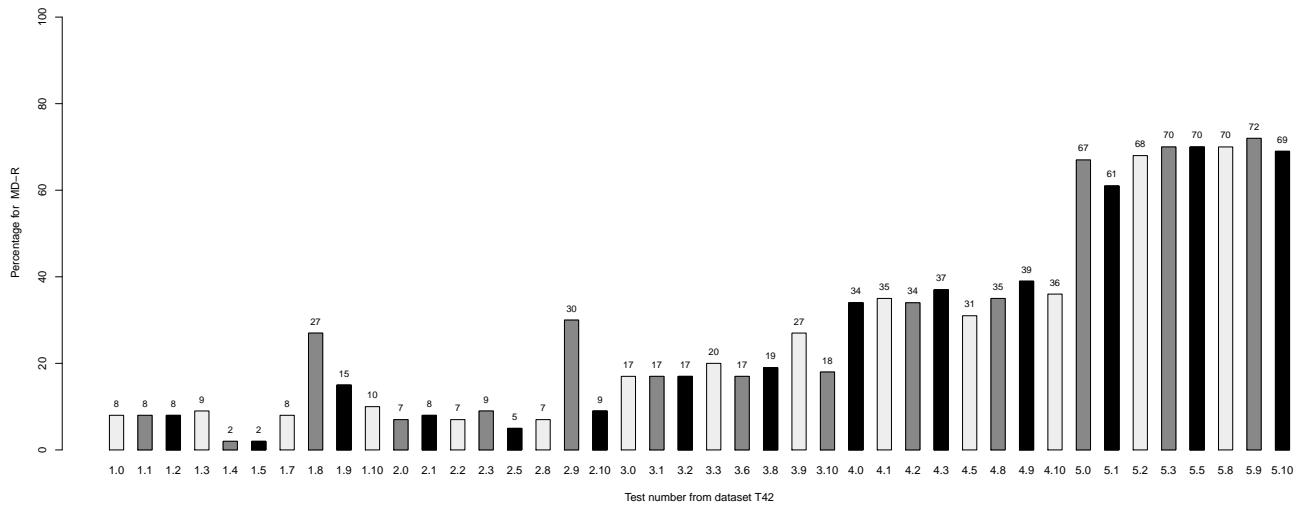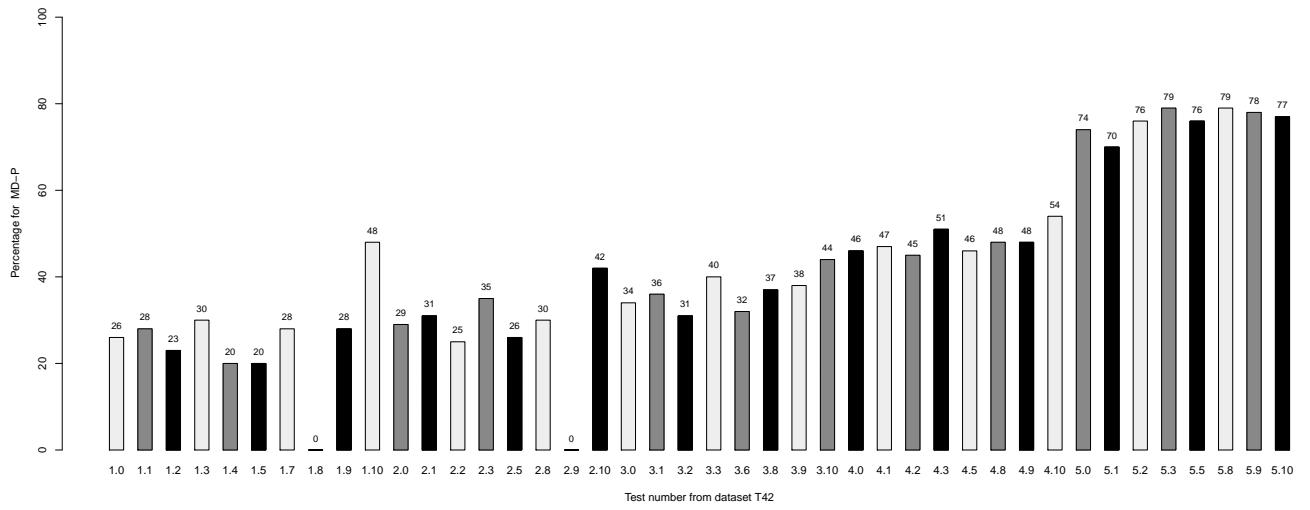
**Fig. 21** Quality results for FRBR-ML with dataset T42 and metric MD-E

**Fig. 22** Quality results for FRBR-ML with dataset T42 and metric MD-R


Fig. 23 Quality results for FRBR-ML with dataset T42 and metric MD-P
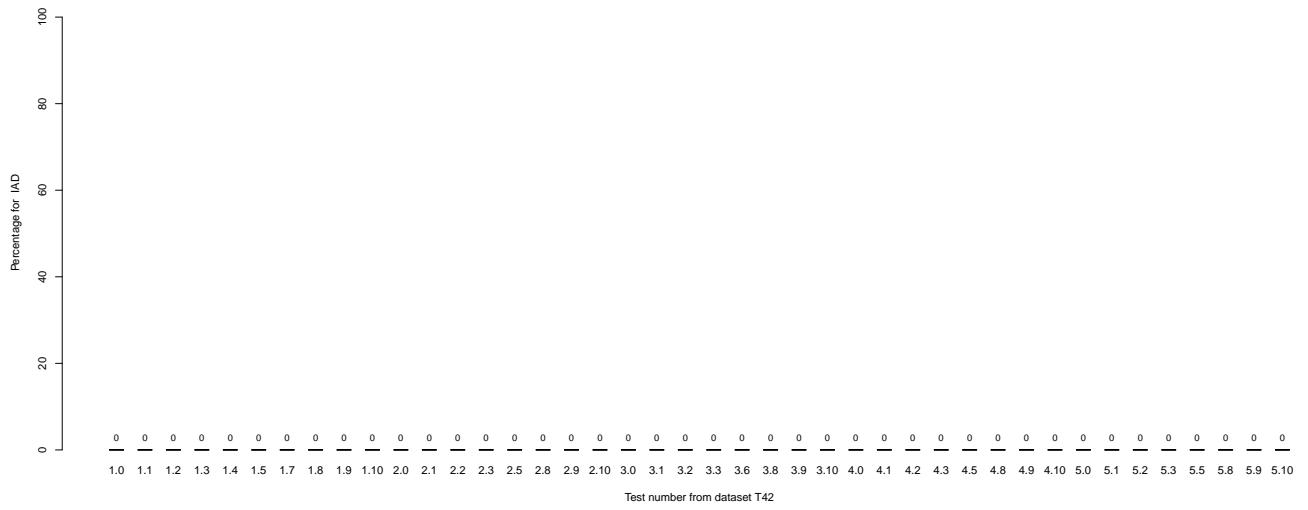
Fig. 24 Quality results for FRBR-ML with dataset T42 and metric IAD
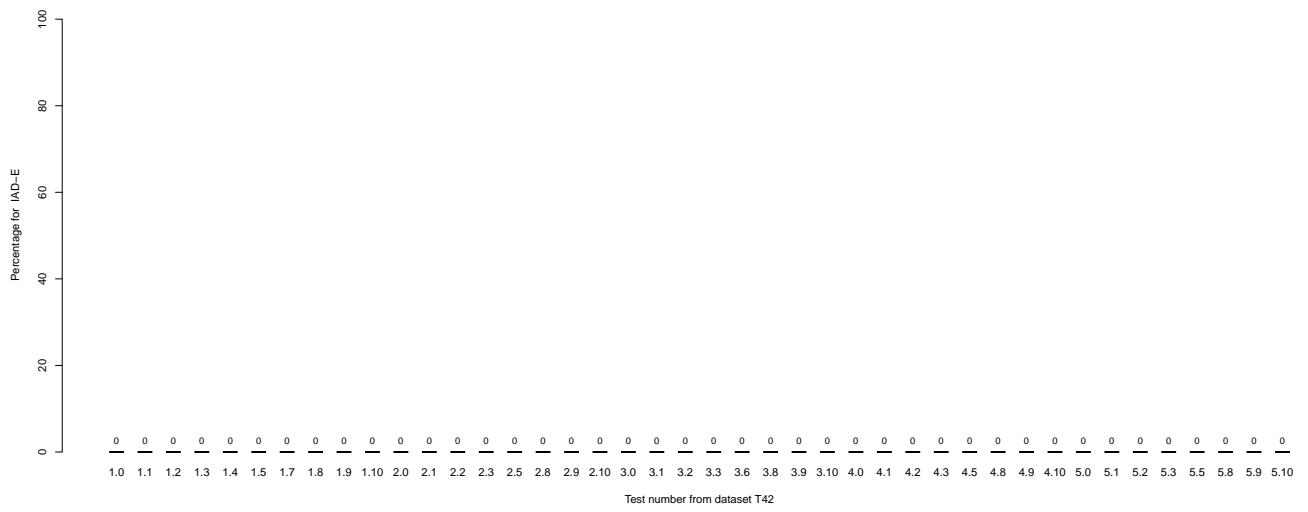
Fig. 25 Quality results for FRBR-ML with dataset T42 and metric IAD-E
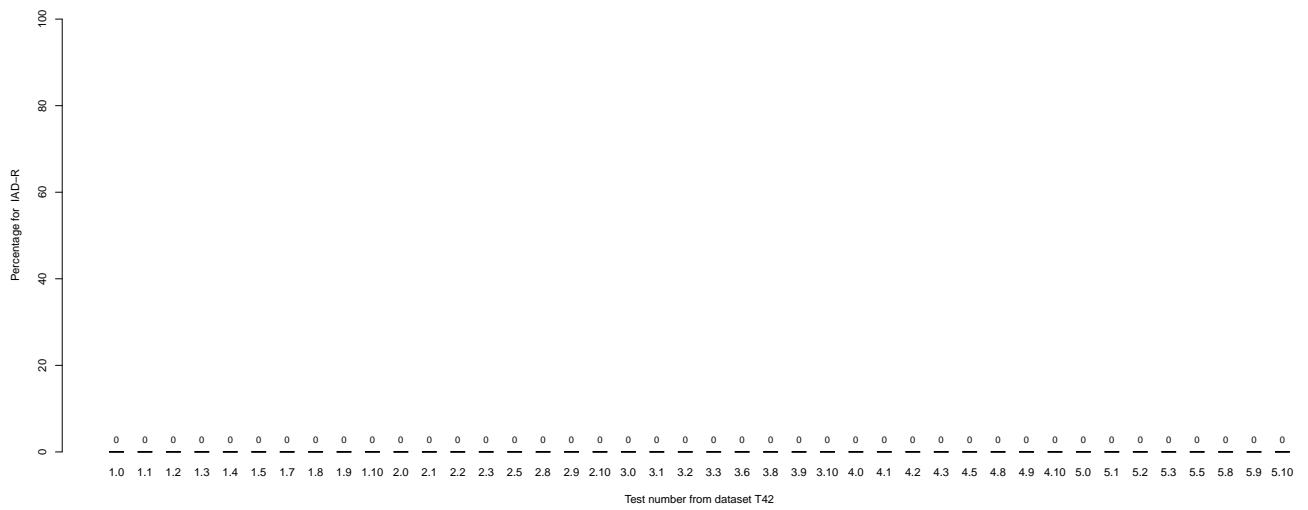
Fig. 26 Quality results for FRBR-ML with dataset T42 and metric IAD-R
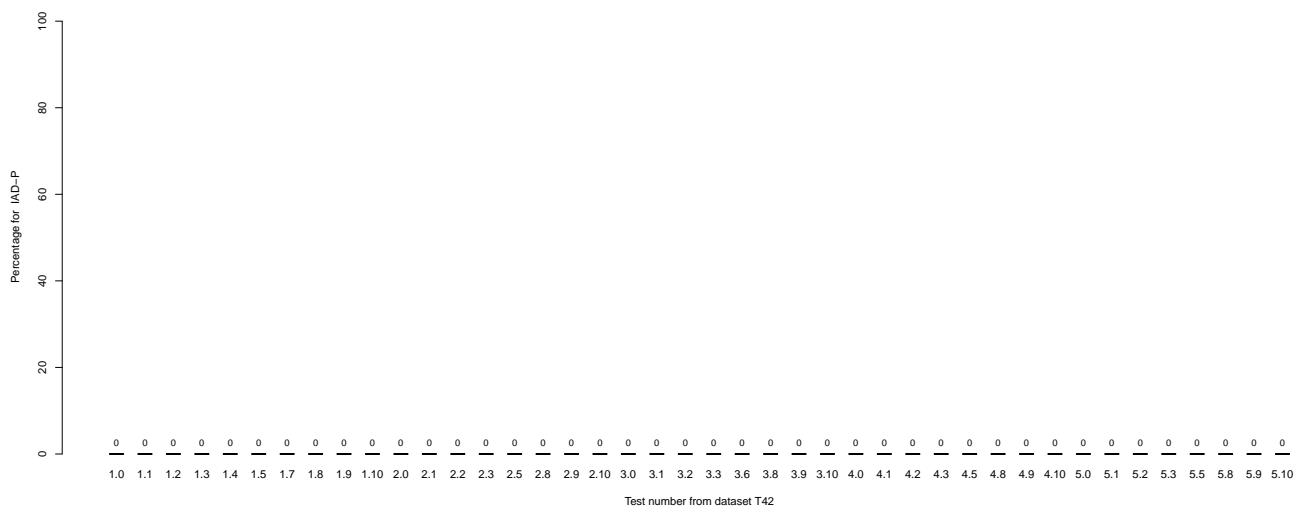

Fig. 27 Quality results for FRBR-ML with dataset T42 and metric IAD-P
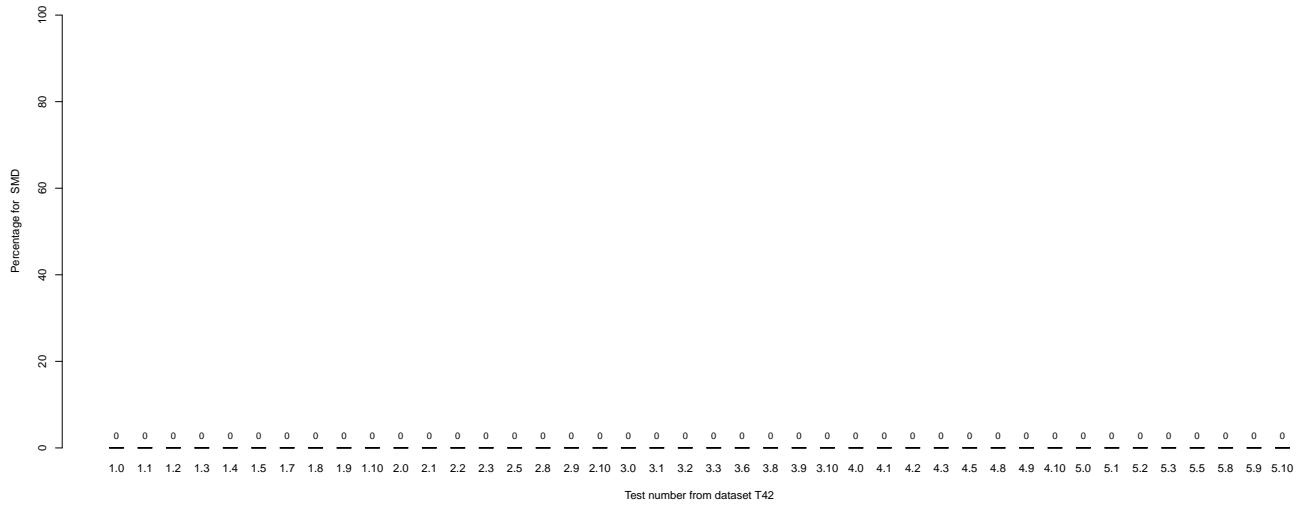
Fig. 28 Quality results for FRBR-ML with dataset T42 and metric SMD
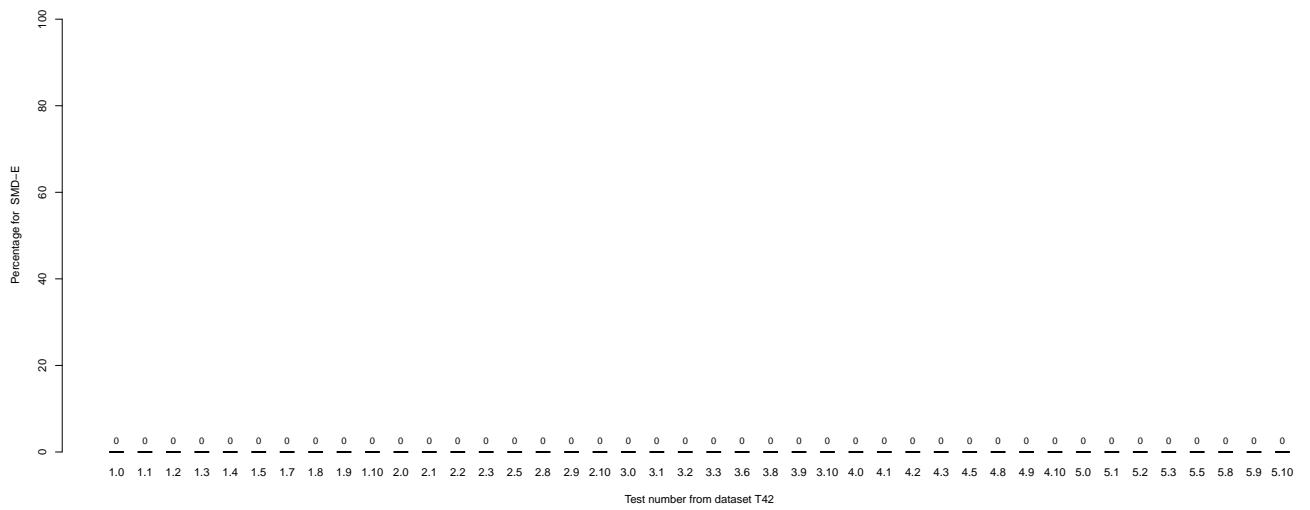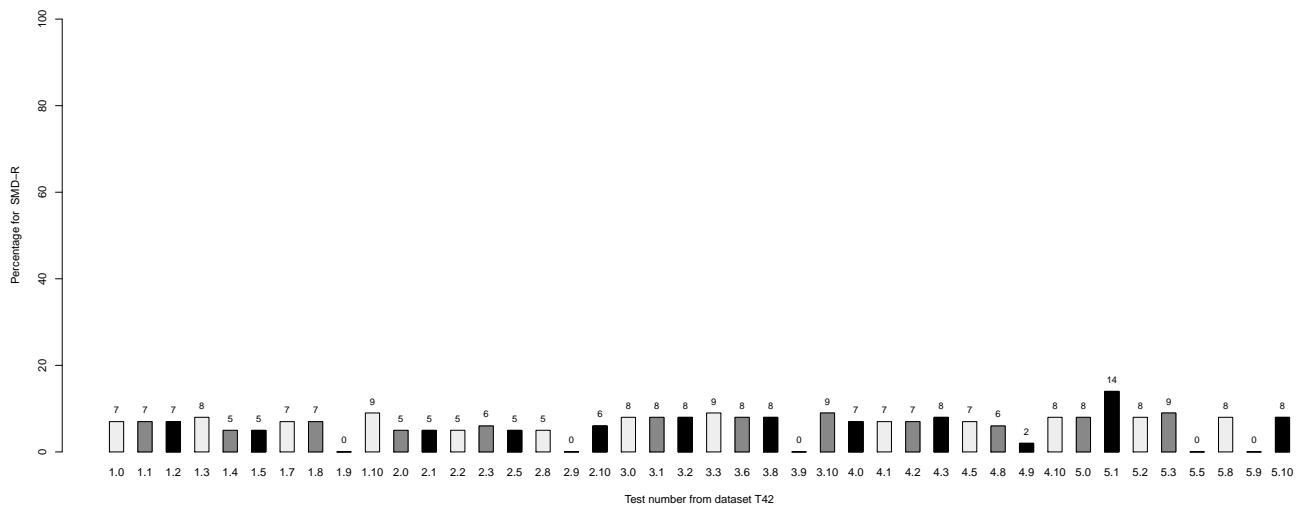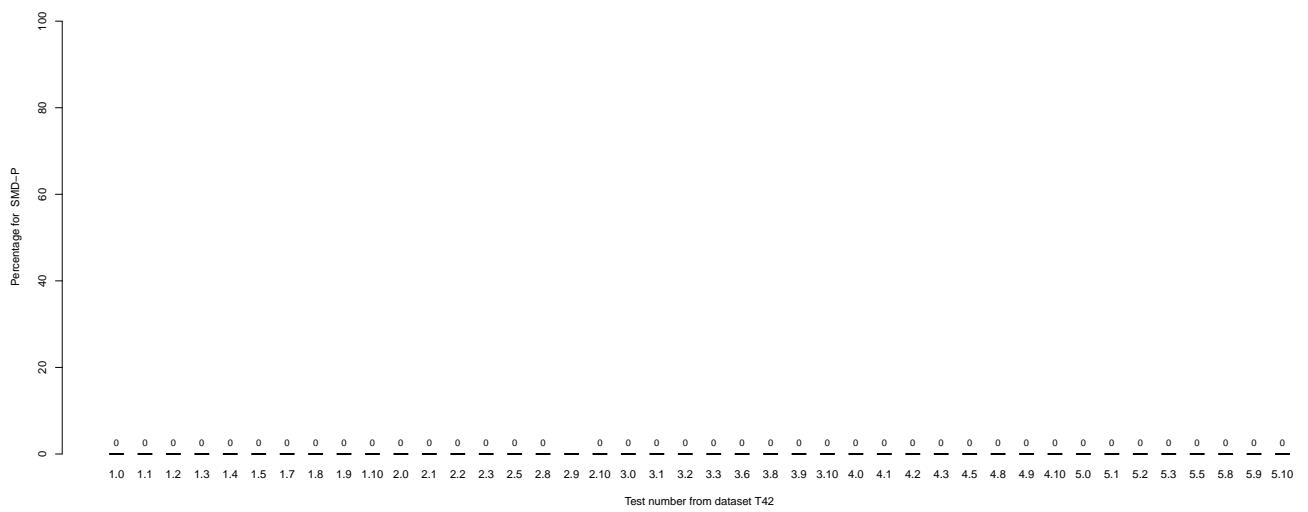
**Fig. 29** Quality results for FRBR-ML with dataset T42 and metric SMD-E

Fig. 30 Quality results for FRBR-ML with dataset T42 and metric SMD-R

Fig. 31 Quality results for FRBR-ML with dataset T42 and metric SMD-P
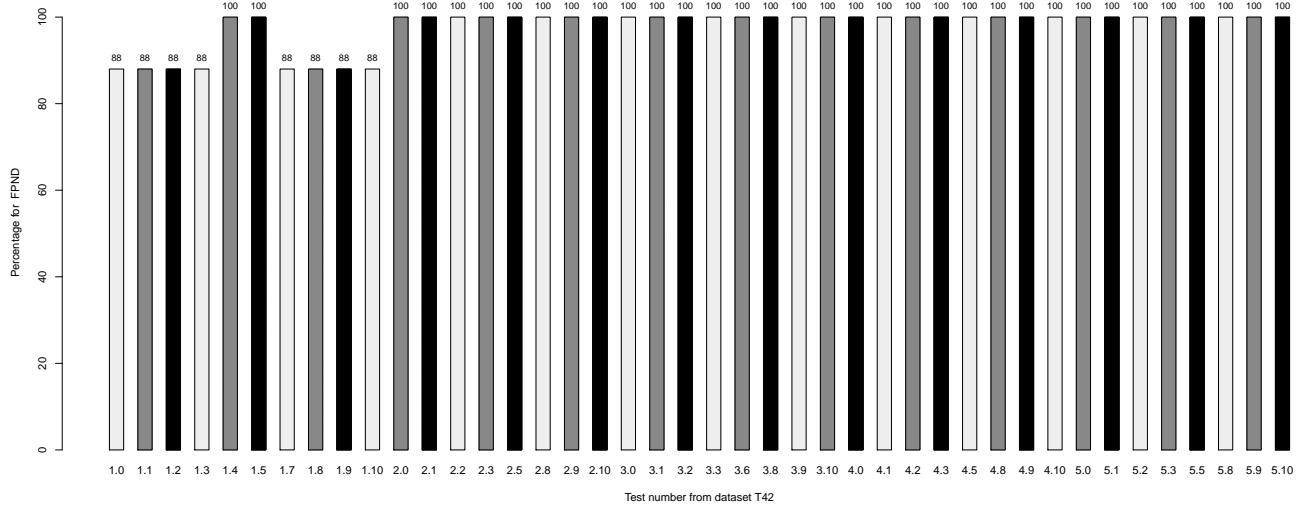
Fig. 32 Quality results for Variation/VFRBR with dataset T42 and metric FPND
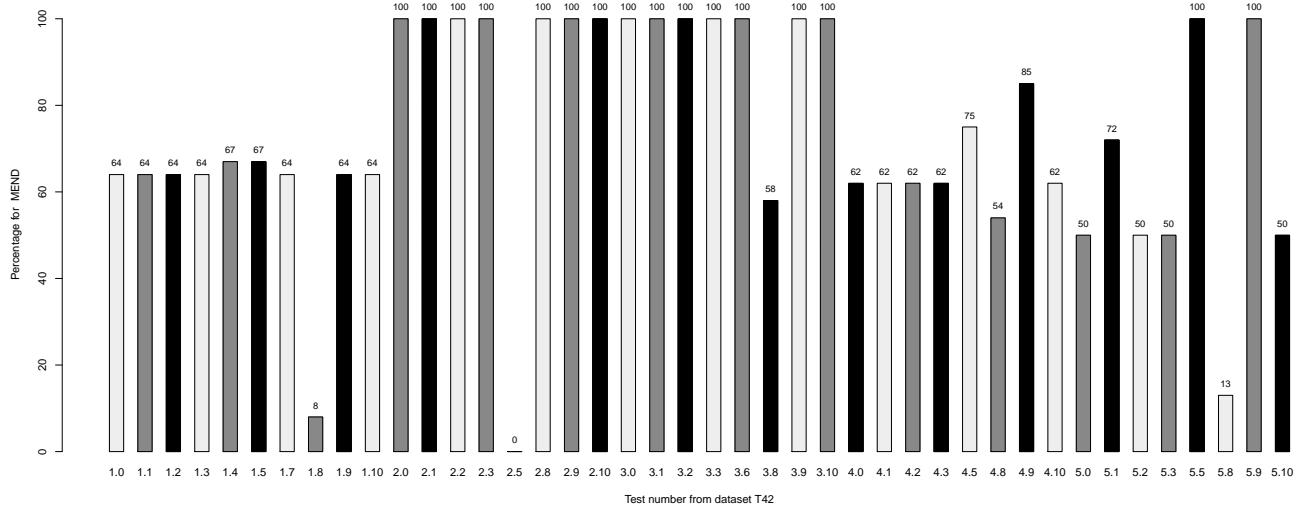
Fig. 33 Quality results for Variation/VFRBR with dataset T42 and metric MEND
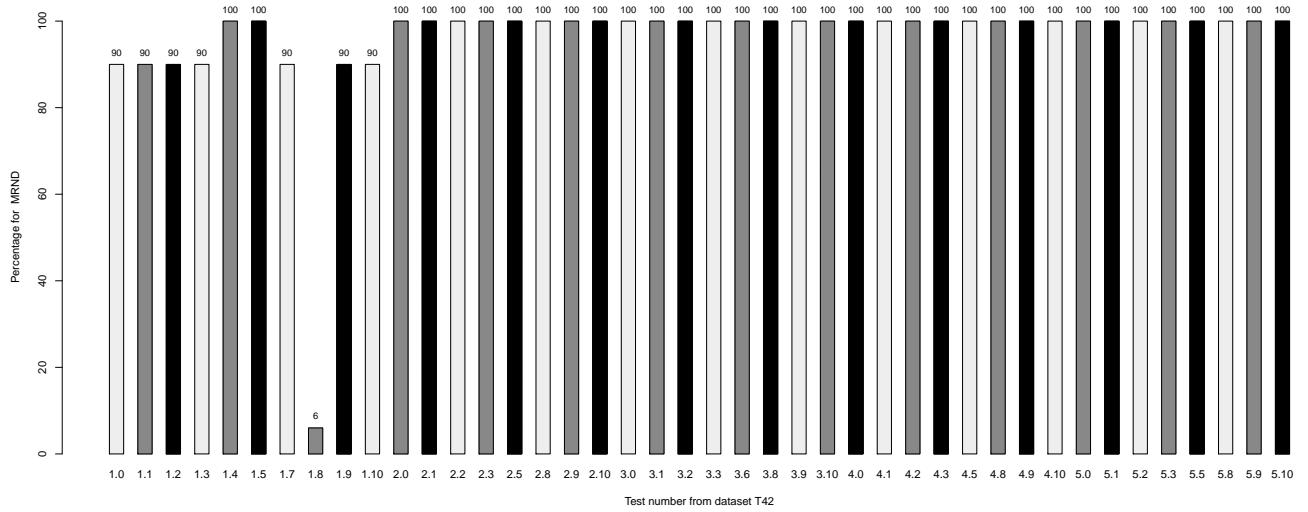
Fig. 34 Quality results for Variation/VFRBR with dataset T42 and metric MRND
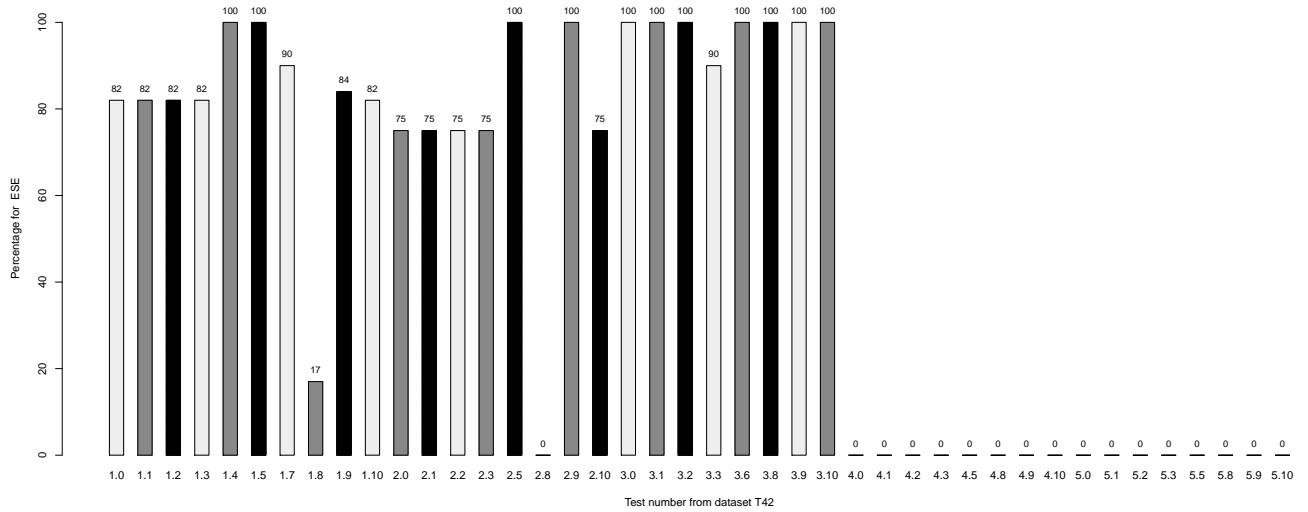
Fig. 35 Quality results for Variation/VFRBR with dataset T42 and metric ESE
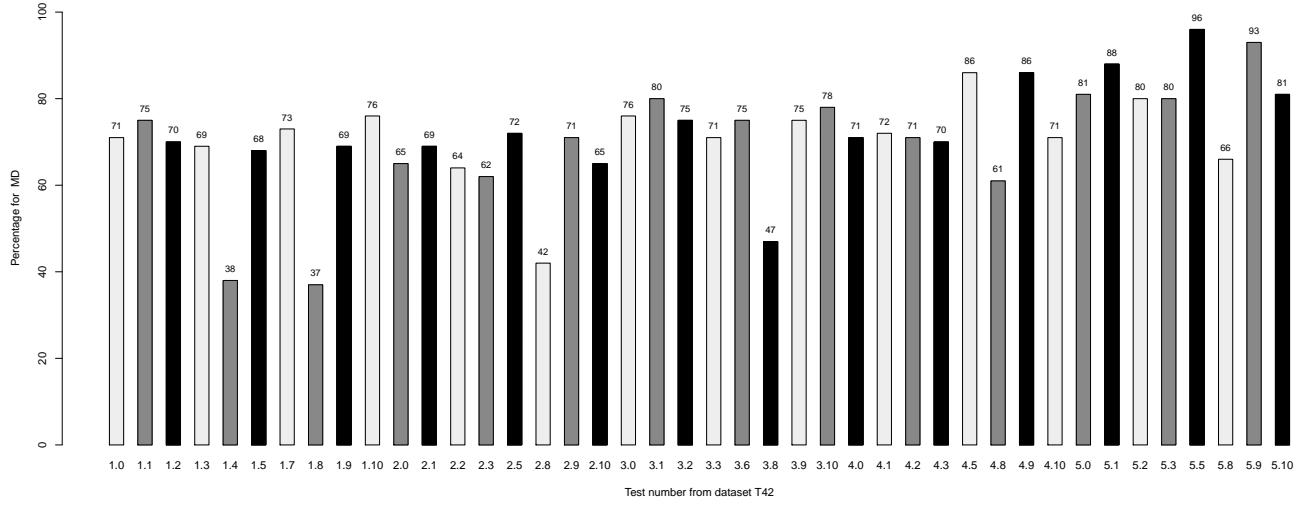
Fig. 36 Quality results for Variation/VFRBR with dataset T42 and metric MD
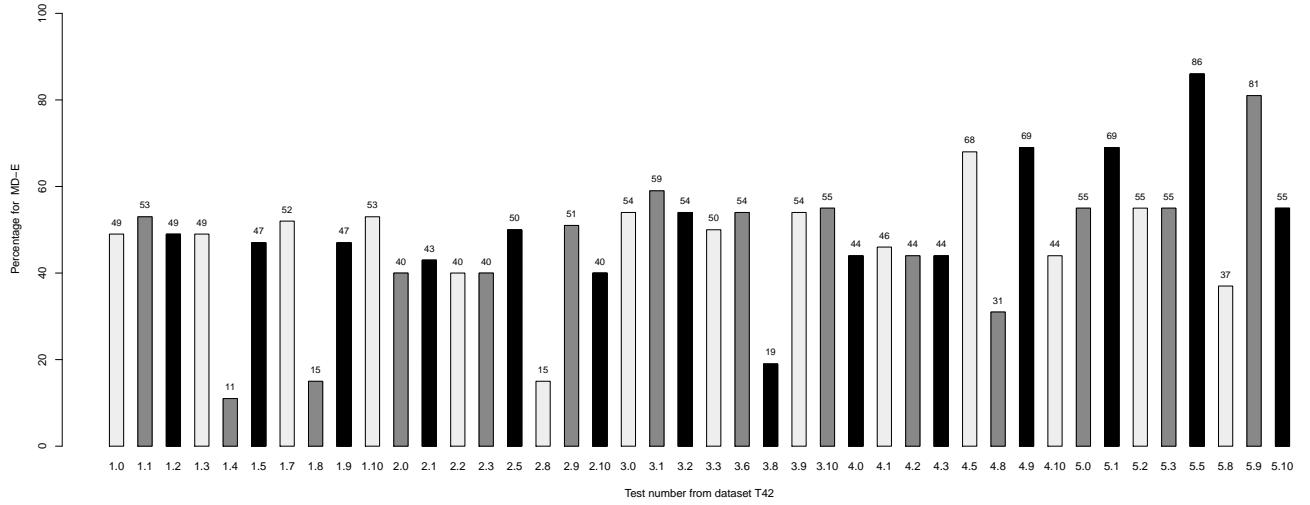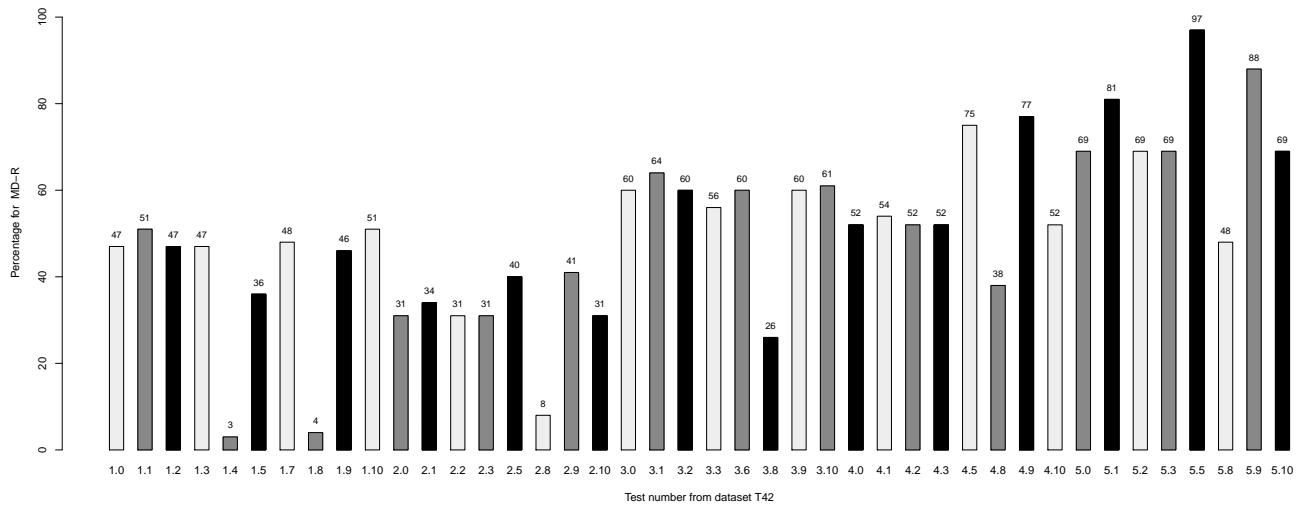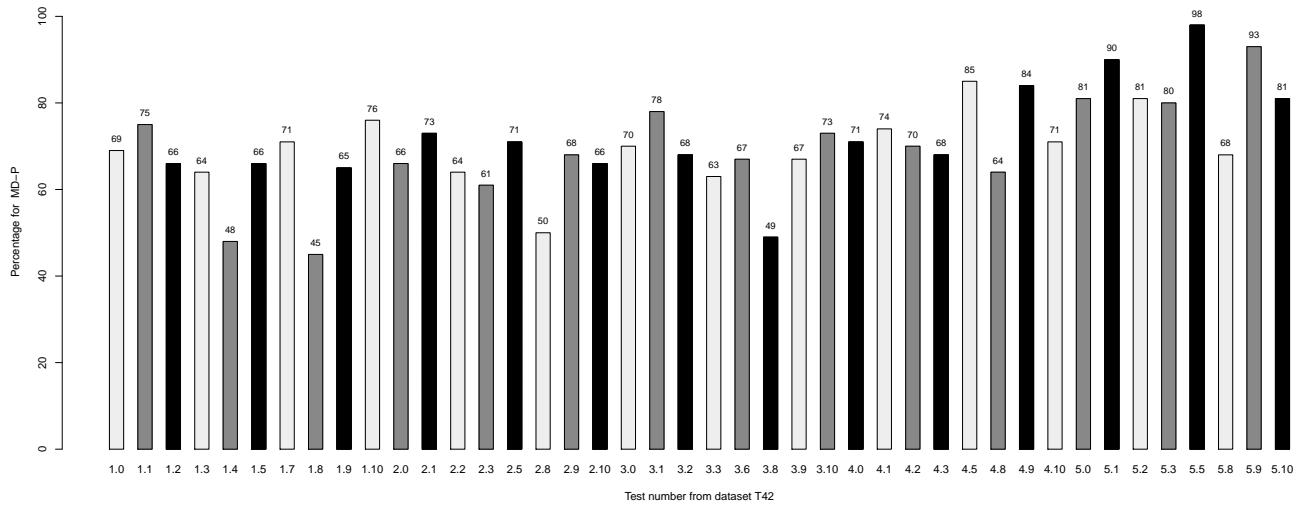
Fig. 37 Quality results for Variation/VFRBR with dataset T42 and metric MD-E

Fig. 38 Quality results for Variation/VFRBR with dataset T42 and metric MD-R

Fig. 39 Quality results for Variation/VFRBR with dataset T42 and metric MD-P
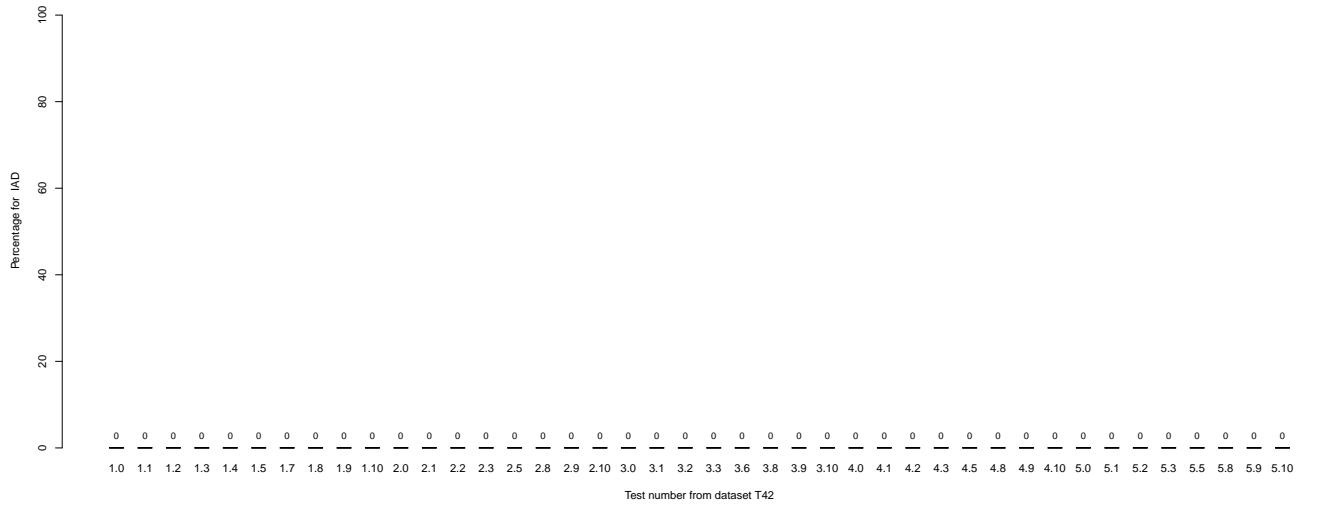
Fig. 40 Quality results for Variation/VFRBR with dataset T42 and metric IAD
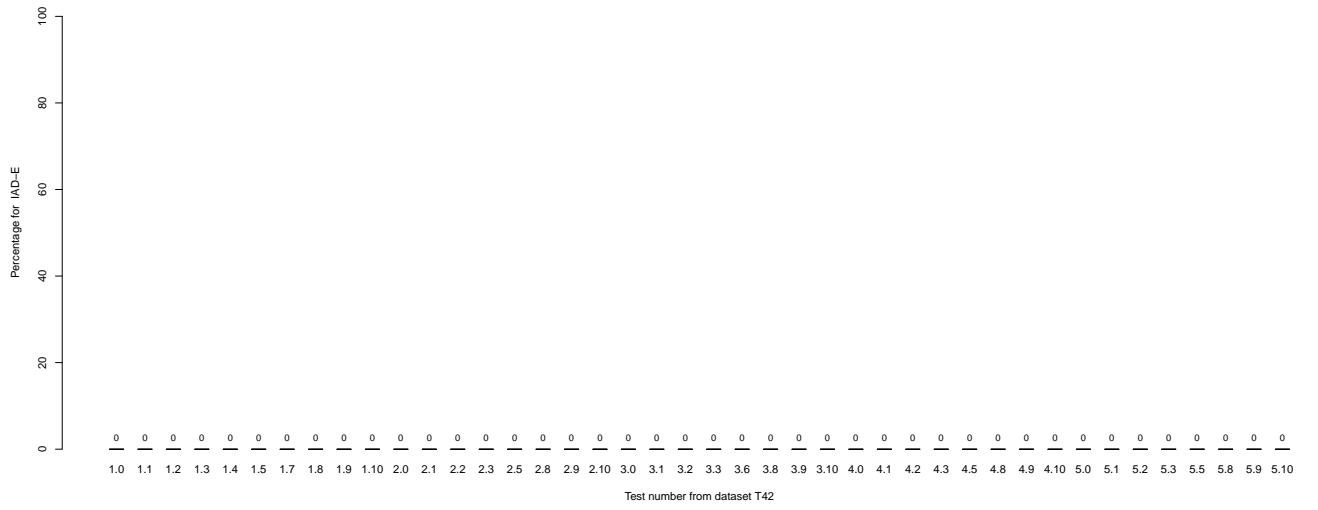
Fig. 41 Quality results for Variation/VFRBR with dataset T42 and metric IAD-E
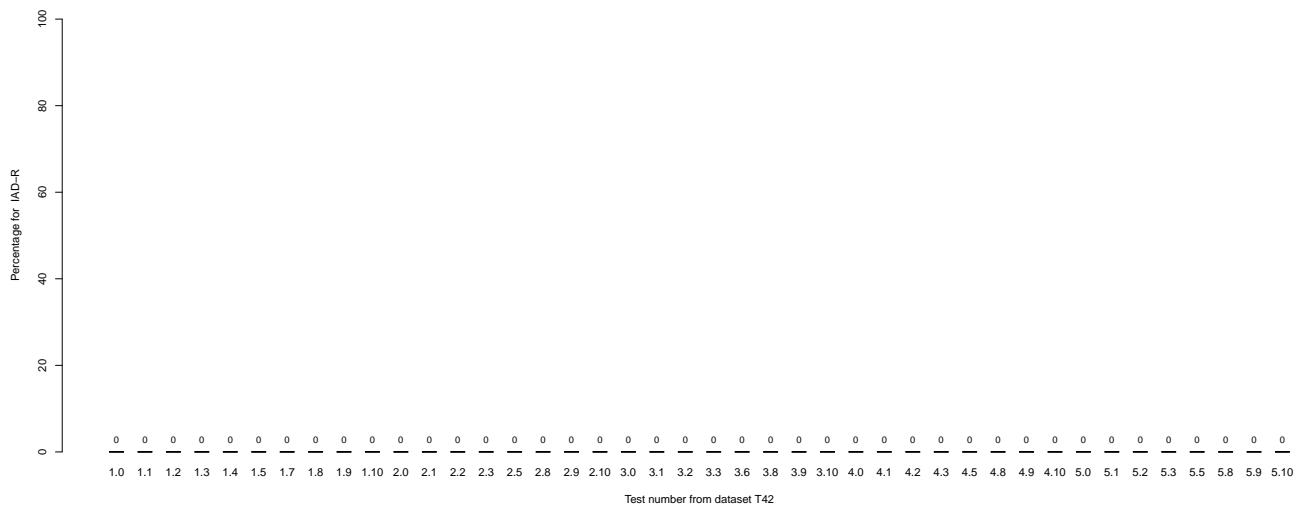
Fig. 42 Quality results for Variation/VFRBR with dataset T42 and metric IAD-R
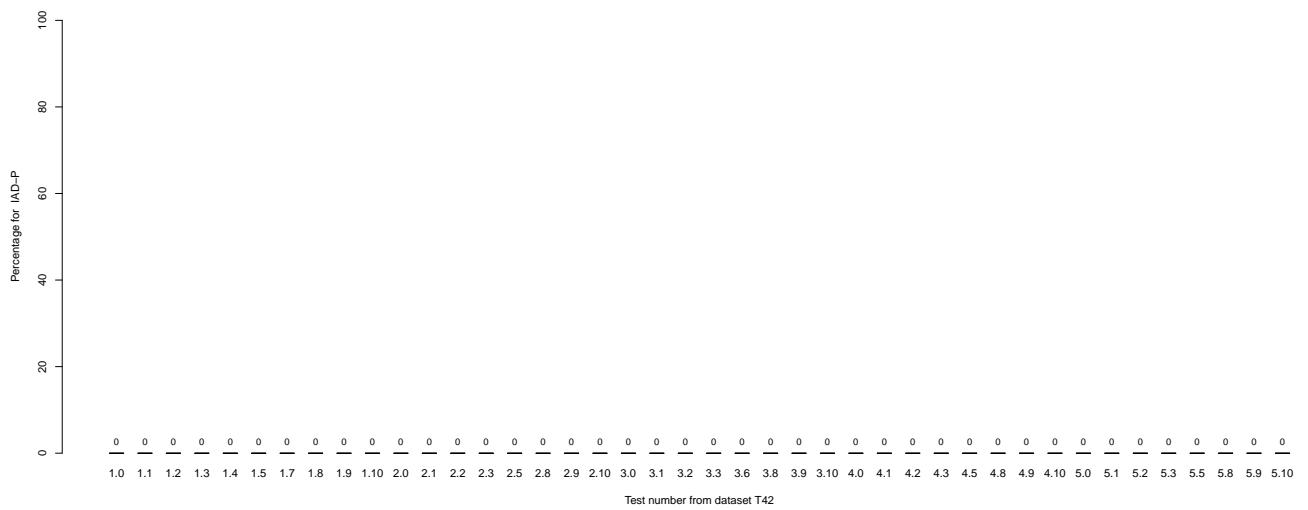
Fig. 43 Quality results for Variation/VFRBR with dataset T42 and metric IAD-P
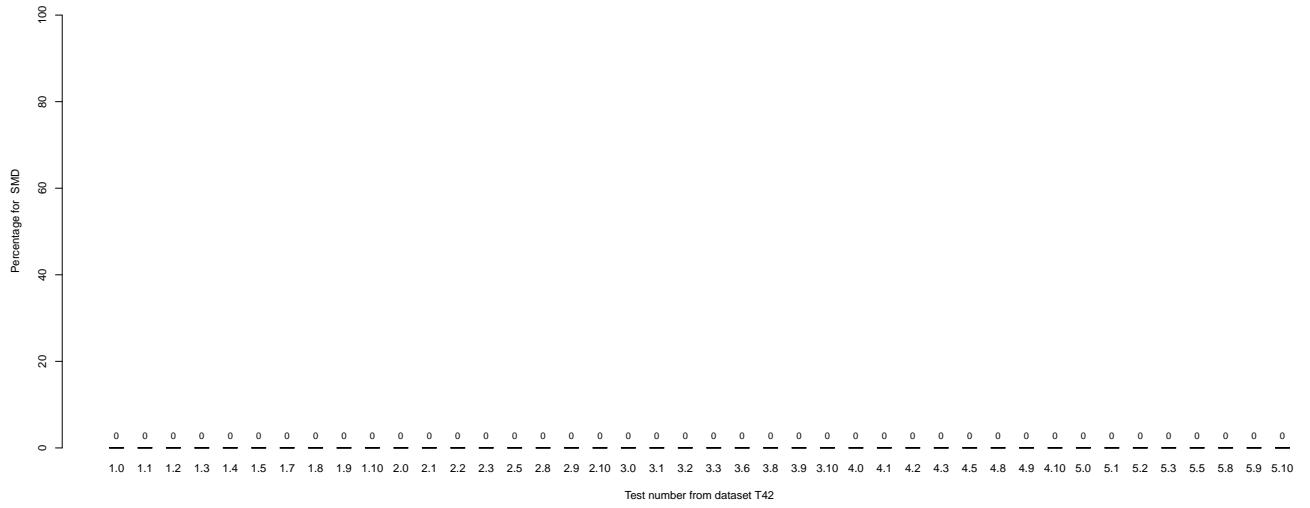
Fig. 44 Quality results for Variation/VFRBR with dataset T42 and metric SMD
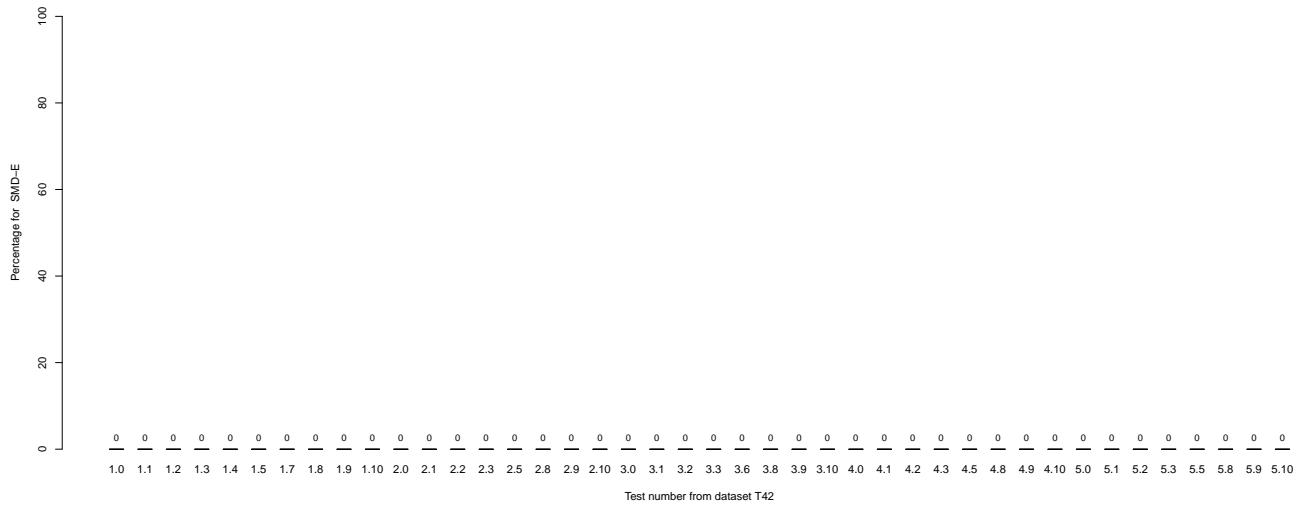
Fig. 45 Quality results for Variation/VFRBR with dataset T42 and metric SMD-E
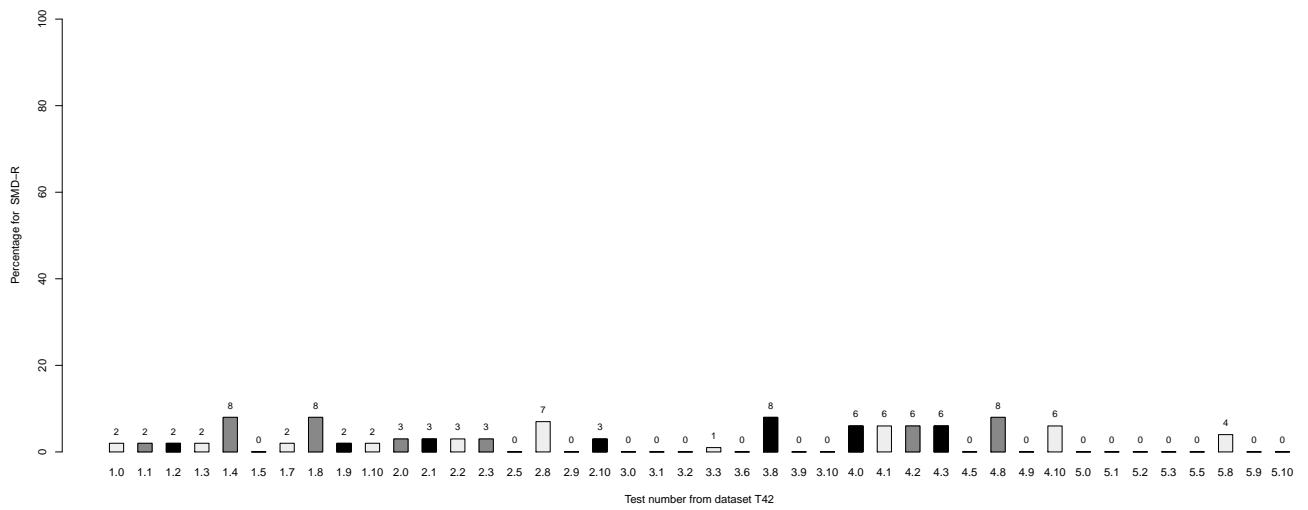
Fig. 46 Quality results for Variation/VFRBR with dataset T42 and metric SMD-R
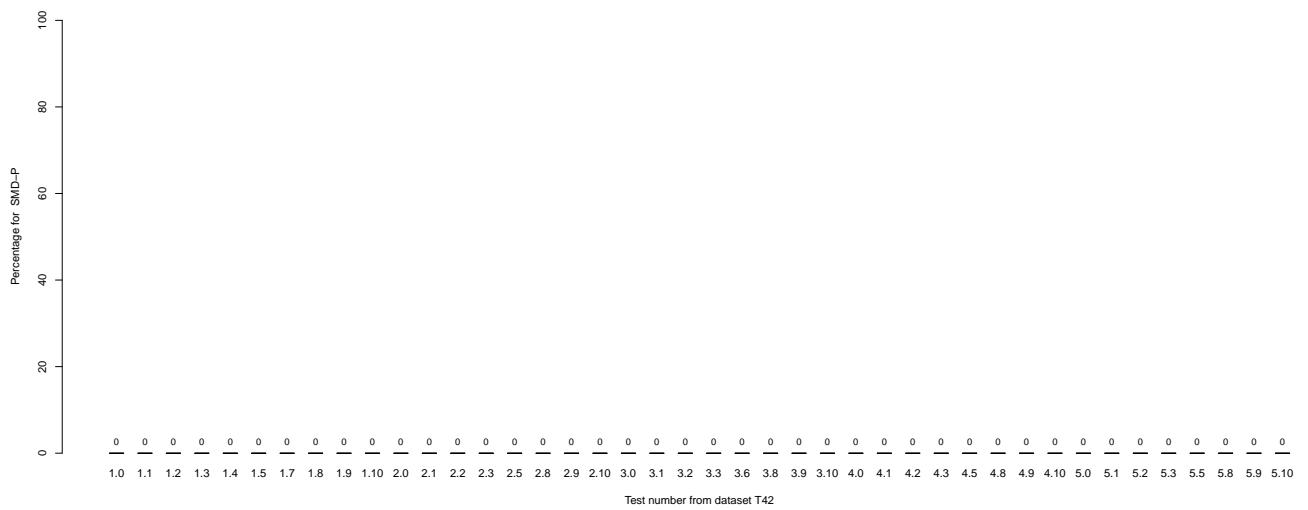
Fig. 47 Quality results for Variation/VFRBR with dataset T42 and metric SMD-P
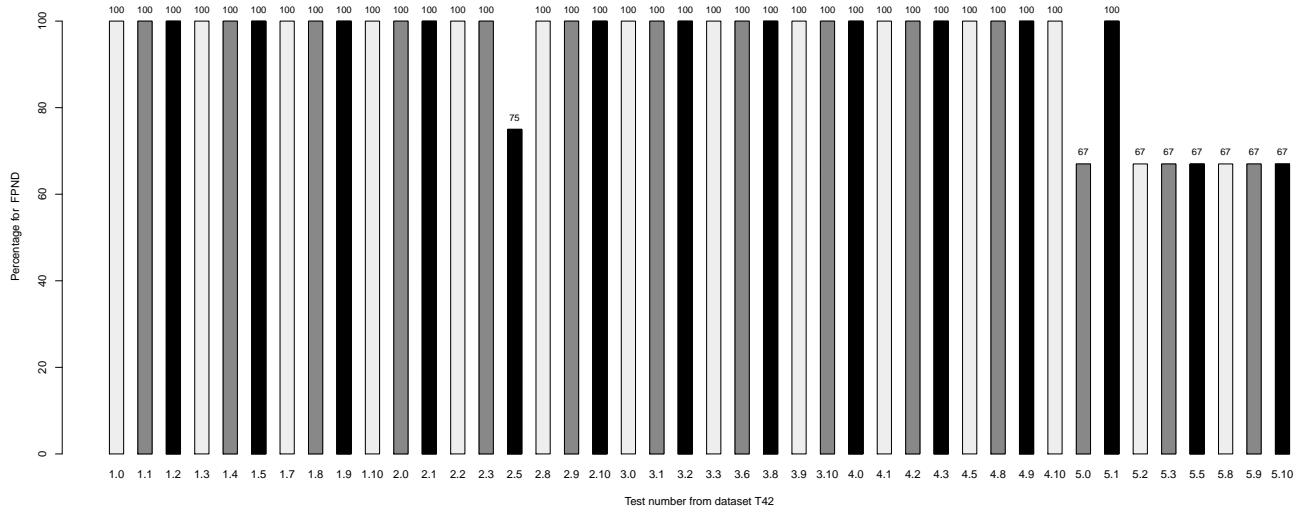
Fig. 48 Quality results for XC with dataset T42 and metric FPND
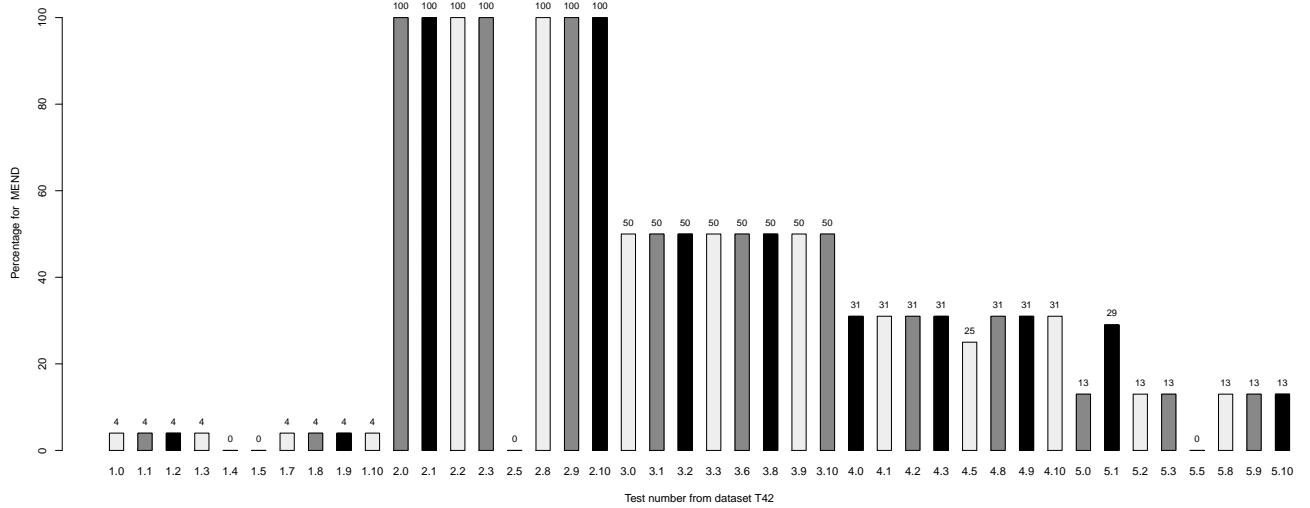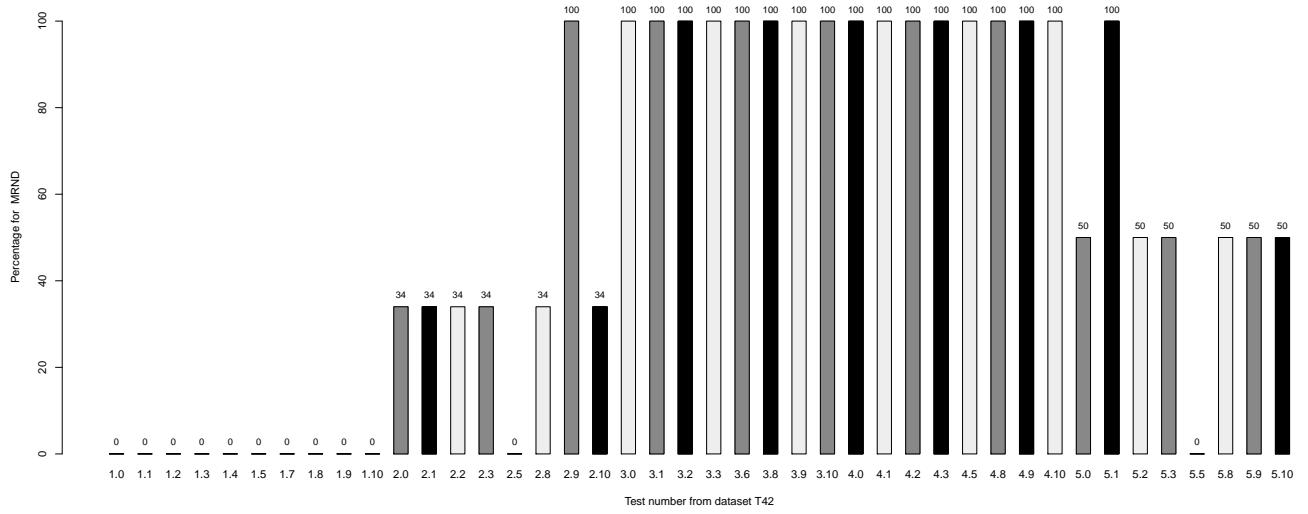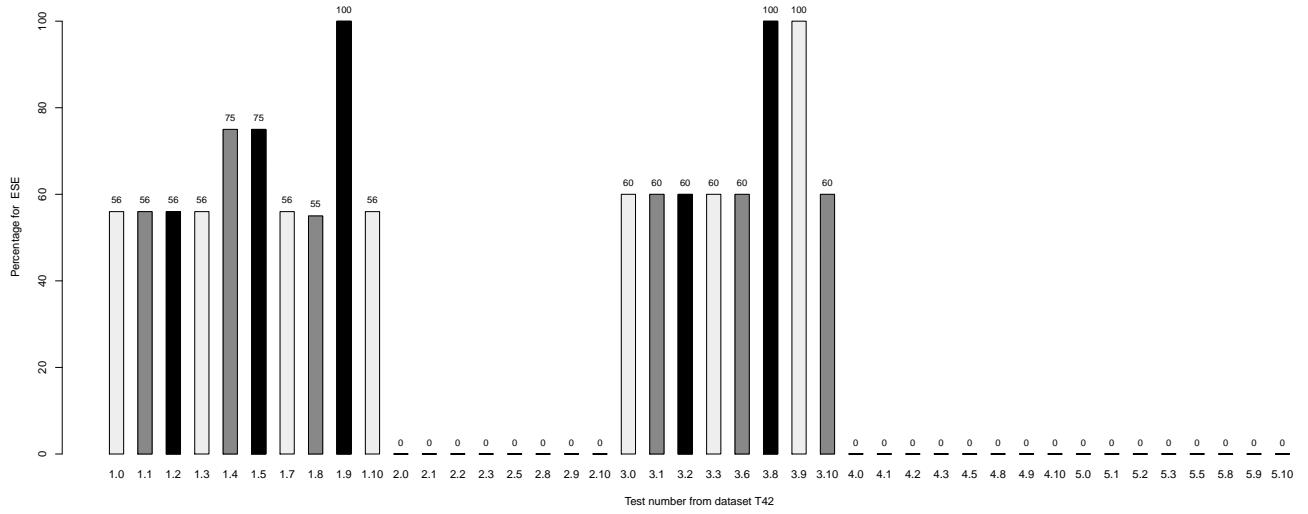
Fig. 49 Quality results for XC with dataset T42 and metric MEND

Fig. 50 Quality results for XC with dataset T42 and metric MRND

Fig. 51 Quality results for XC with dataset T42 and metric ESE
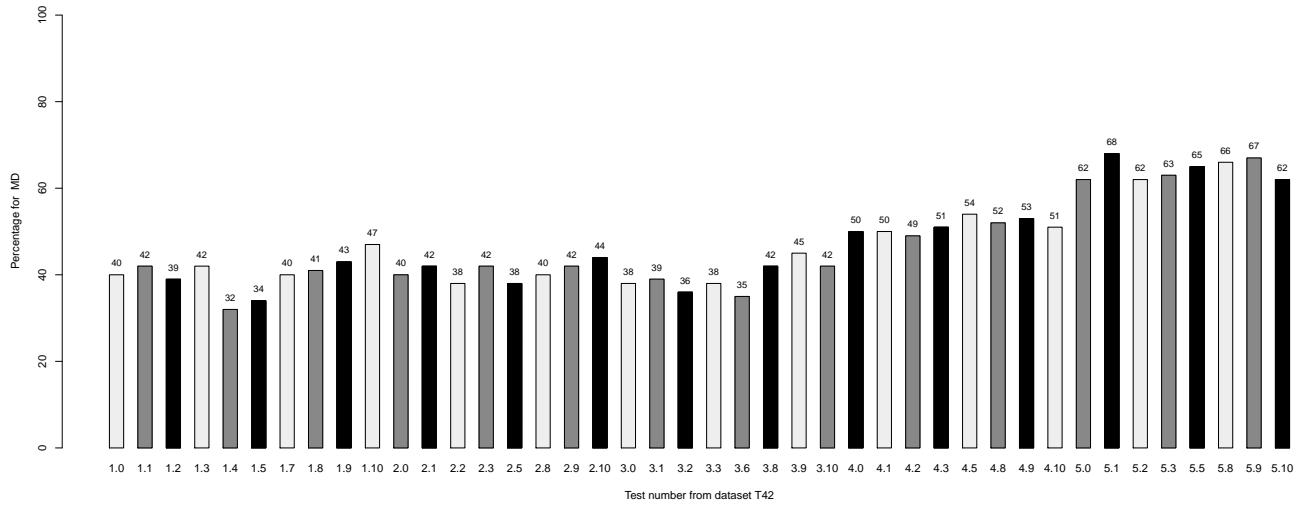
Fig. 52 Quality results for XC with dataset T42 and metric MD
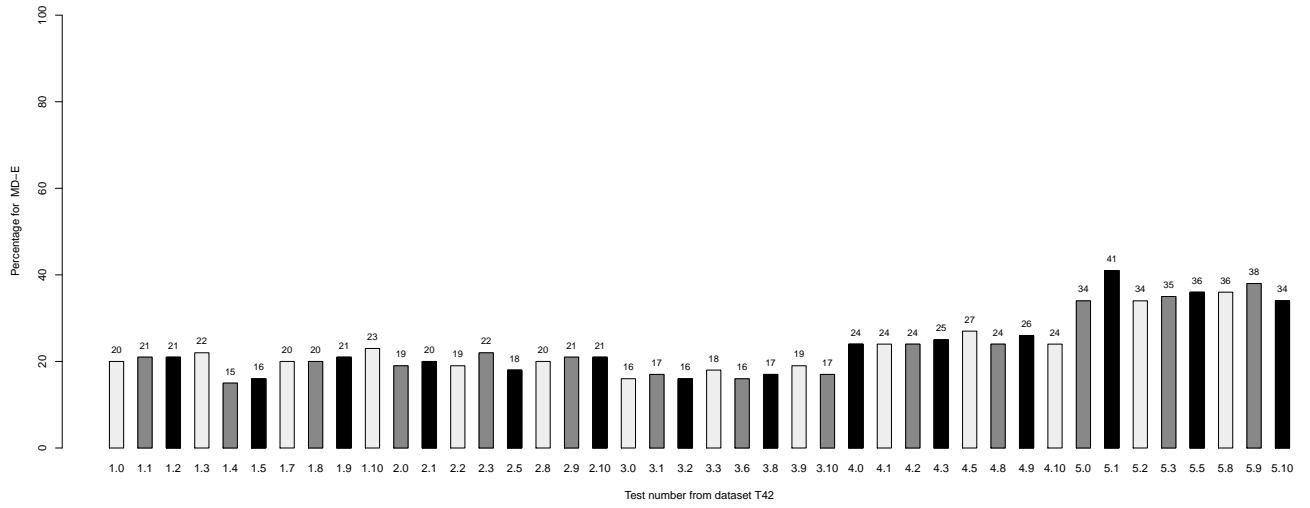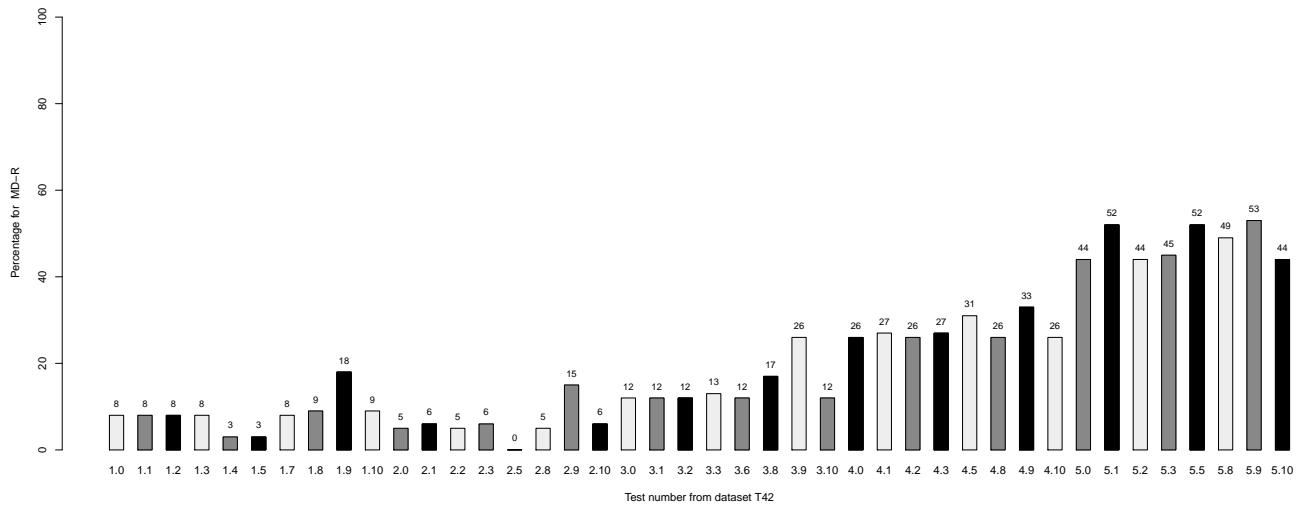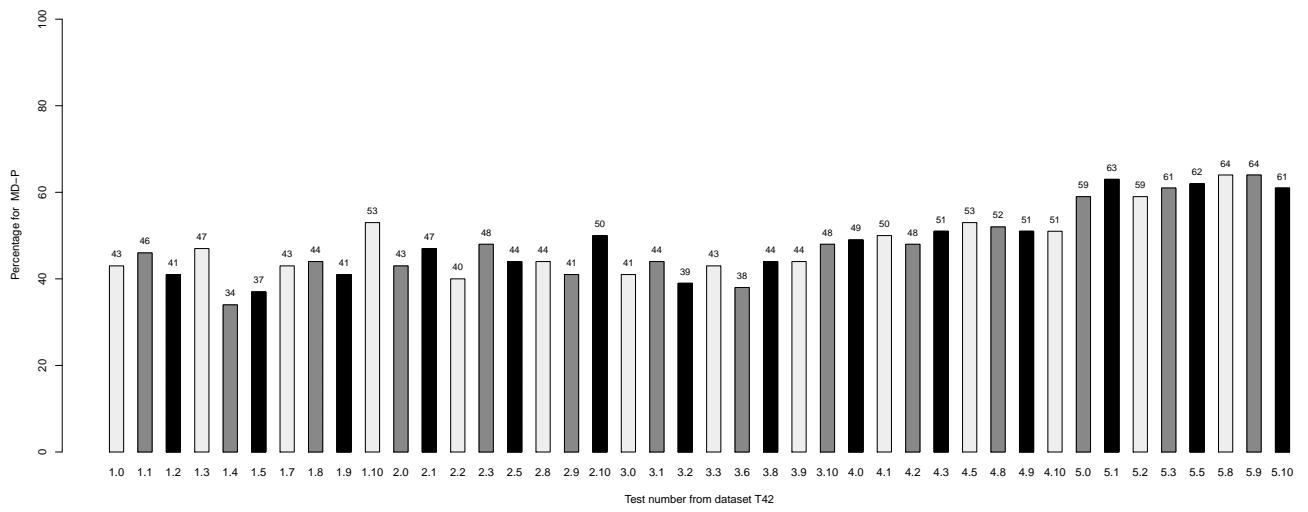
Fig. 53 Quality results for XC with dataset T42 and metric MD-E

Fig. 54 Quality results for XC with dataset T42 and metric MD-R

Fig. 55 Quality results for XC with dataset T42 and metric MD-P
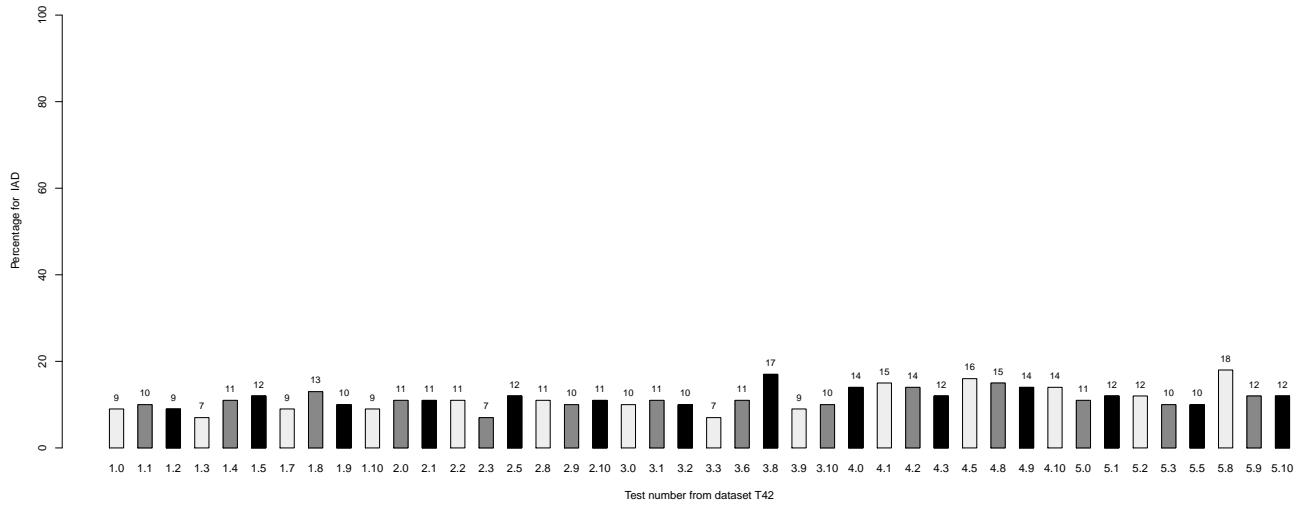
Fig. 56 Quality results for XC with dataset T42 and metric IAD
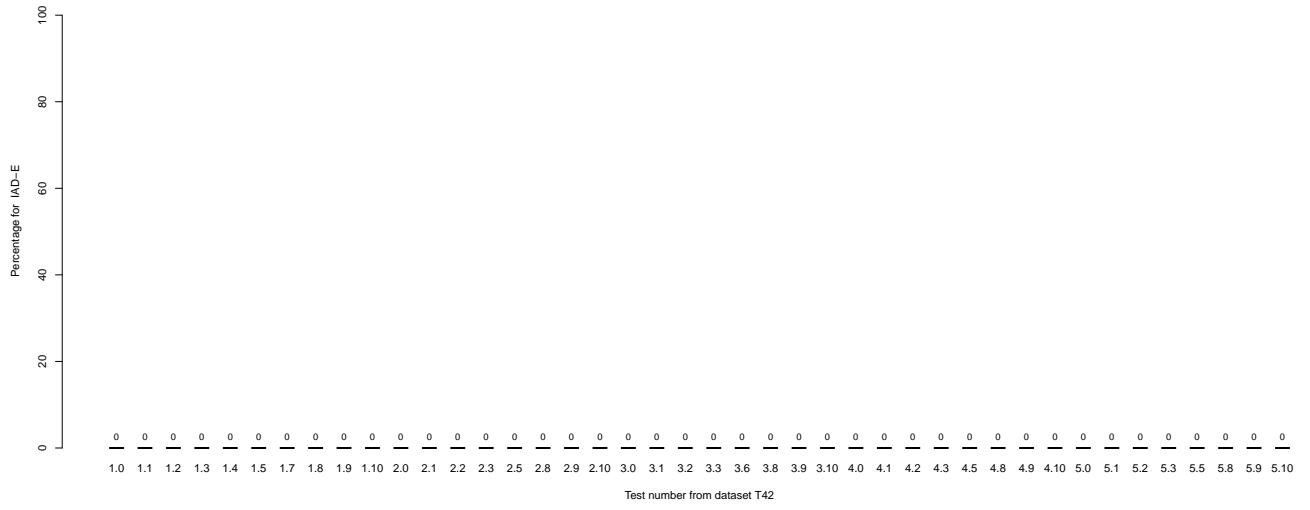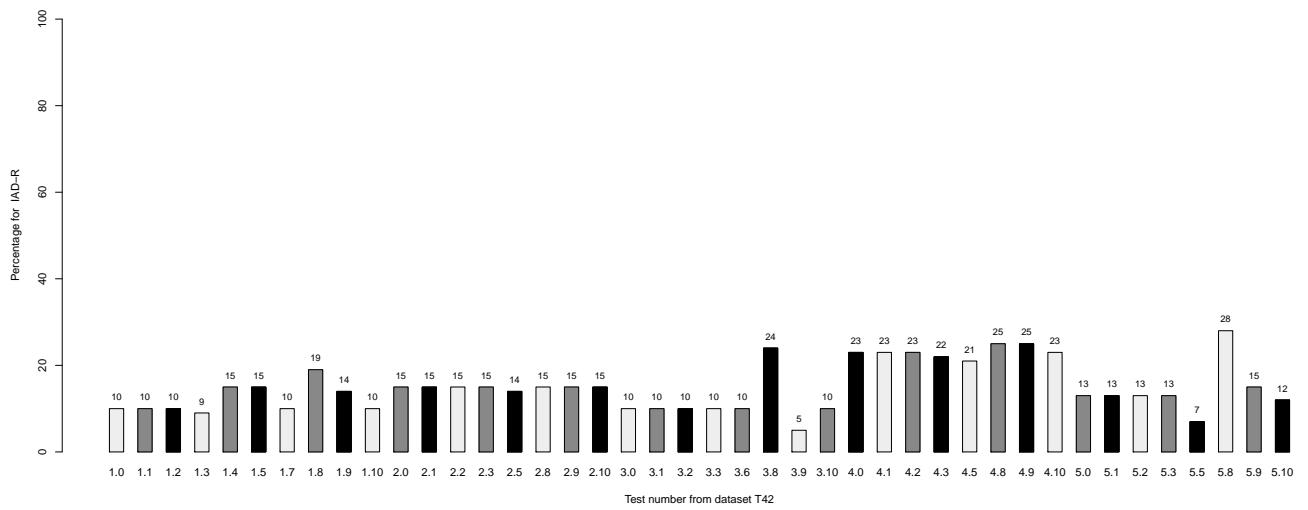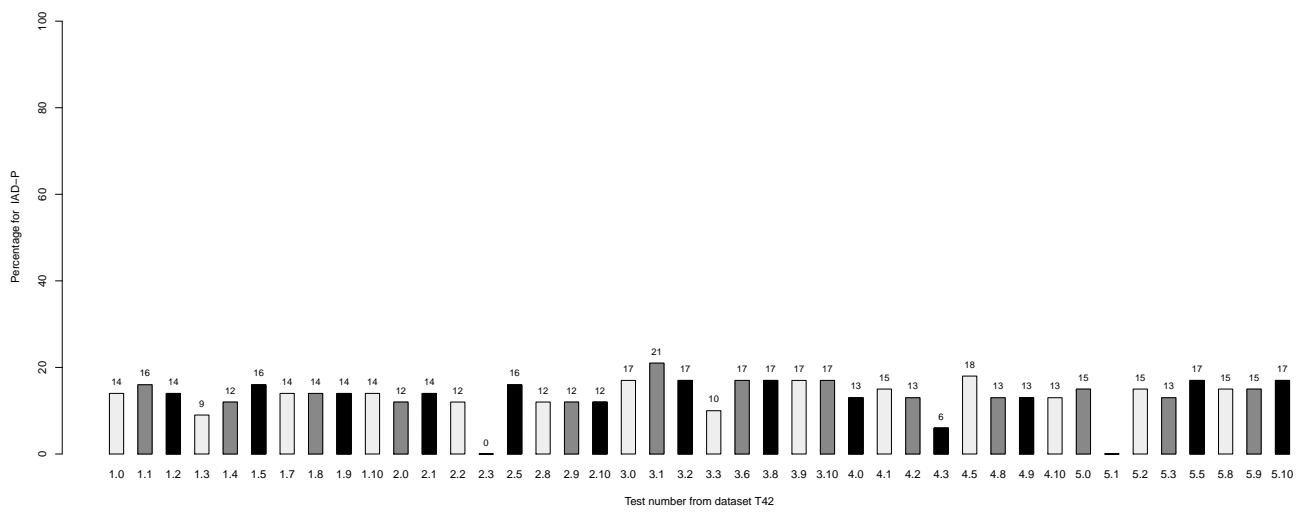
Fig. 57 Quality results for XC with dataset T42 and metric IAD-E

Fig. 58 Quality results for XC with dataset T42 and metric IAD-R

Fig. 59 Quality results for XC with dataset T42 and metric IAD-P
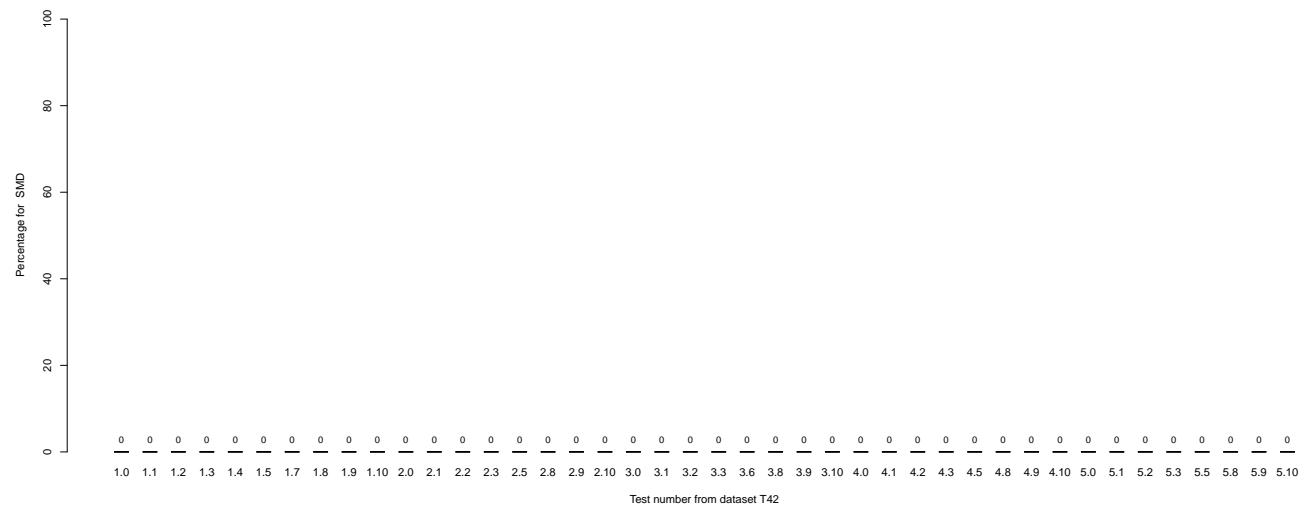
Fig. 60 Quality results for XC with dataset T42 and metric SMD
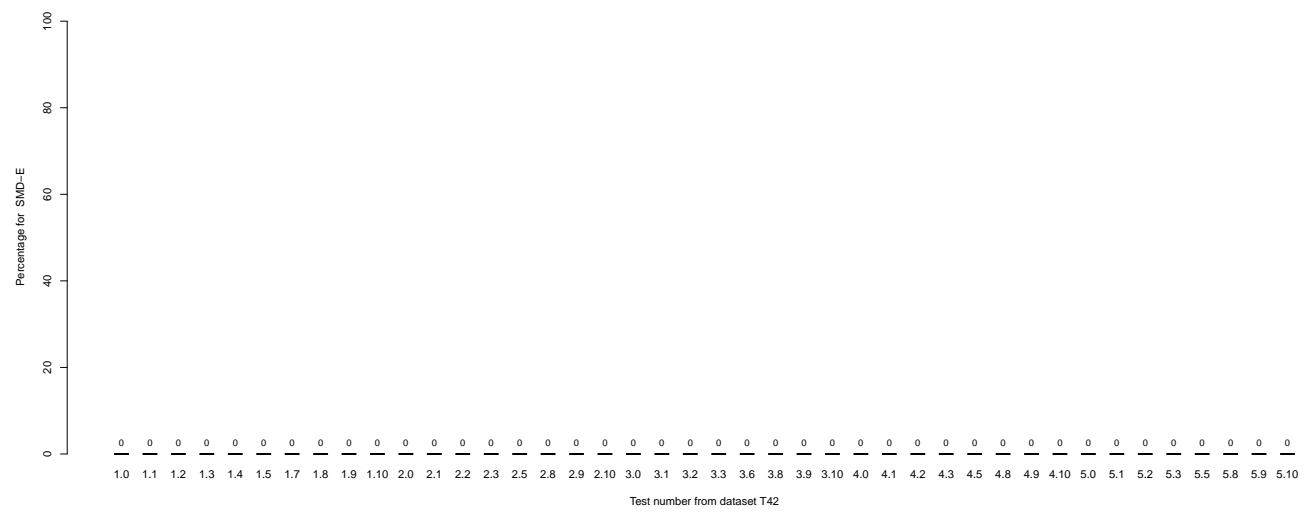
Fig. 61 Quality results for XC with dataset T42 and metric SMD-E
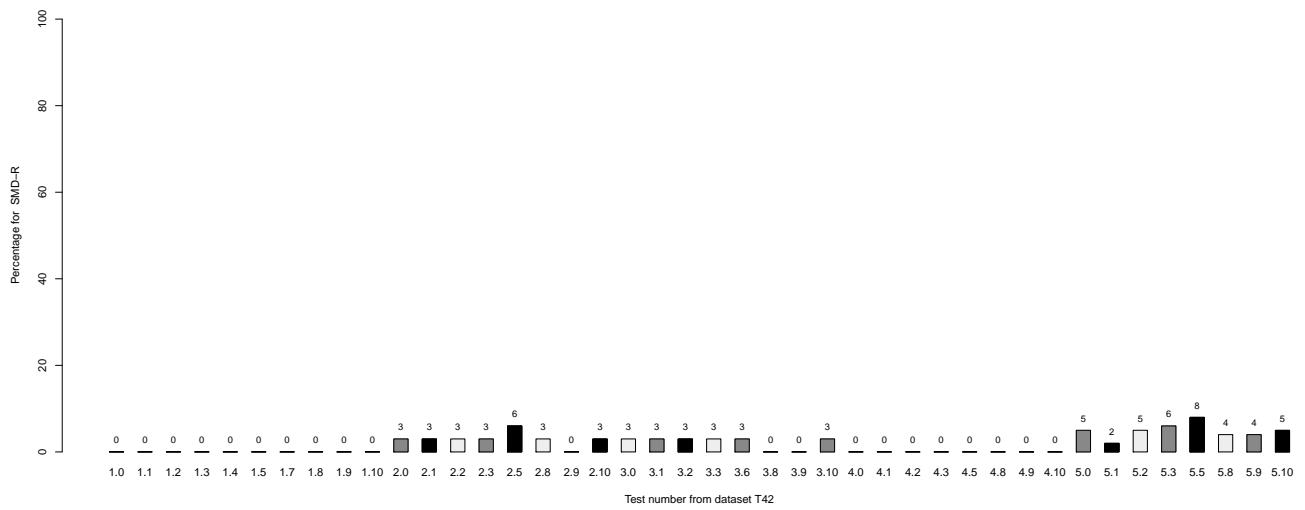
Fig. 62 Quality results for XC with dataset T42 and metric SMD-R
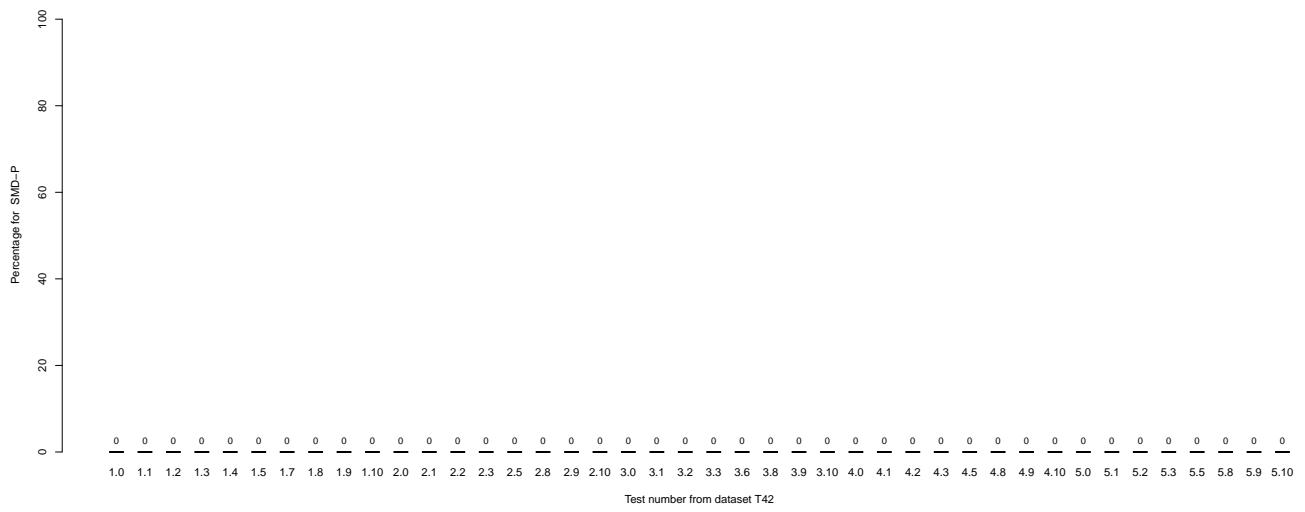
Fig. 63 Quality results for XC with dataset and metric SMD-P

3 Comparing tools in real-world context

The objective of this second experiment is to compare FRBRization tools in a real-world context using both FRBRization and post-FRBRization metrics. All tools rely on their basic set of rules (no tuning). Contrary to the dataset T42, all records from the dataset BIB-RCAT come from various institutions. This means that these real-world records can be rather simple or may contain several bibliographic patterns and issues. Note that the FRBRization metrics ETC, ETD and NRT could not be presented, because the tools did not implement these features. Instead, we provide the overall execution time for FRBRizing the dataset BIB-RCAT. The post-FRBRization metric DLE is also not given, since there exists many authority files or knowledge bases (e.g., Linked Open Data) and the expert FRBRized collection cannot include a link for each of these sources.

In the following plots, the results are organized by tool.
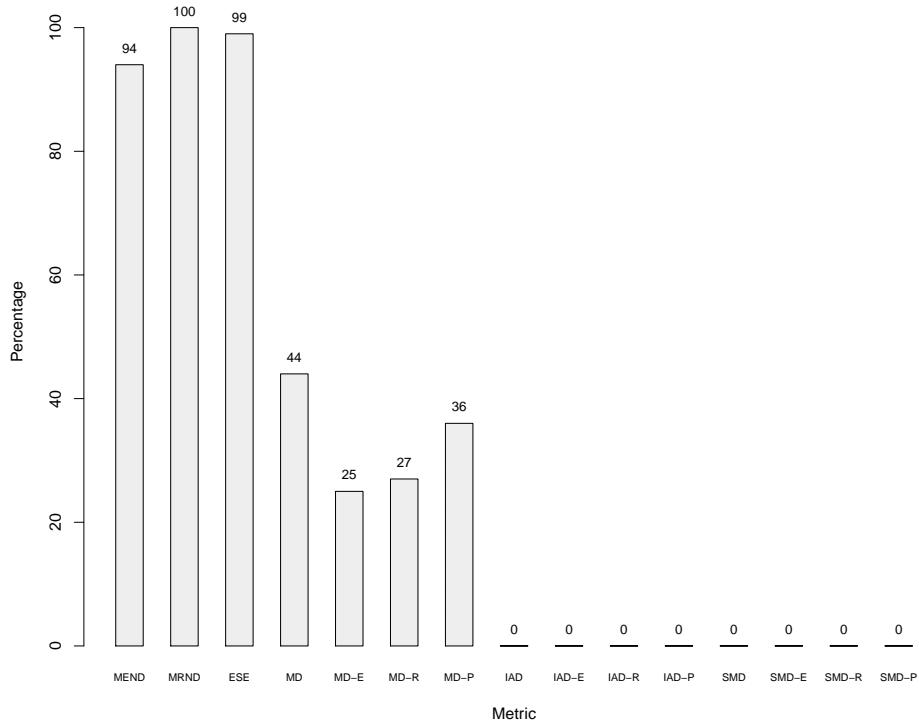
Fig. 64 Quality results for FRBR-ML basic (no tuning) on dataset BIB-RCAT
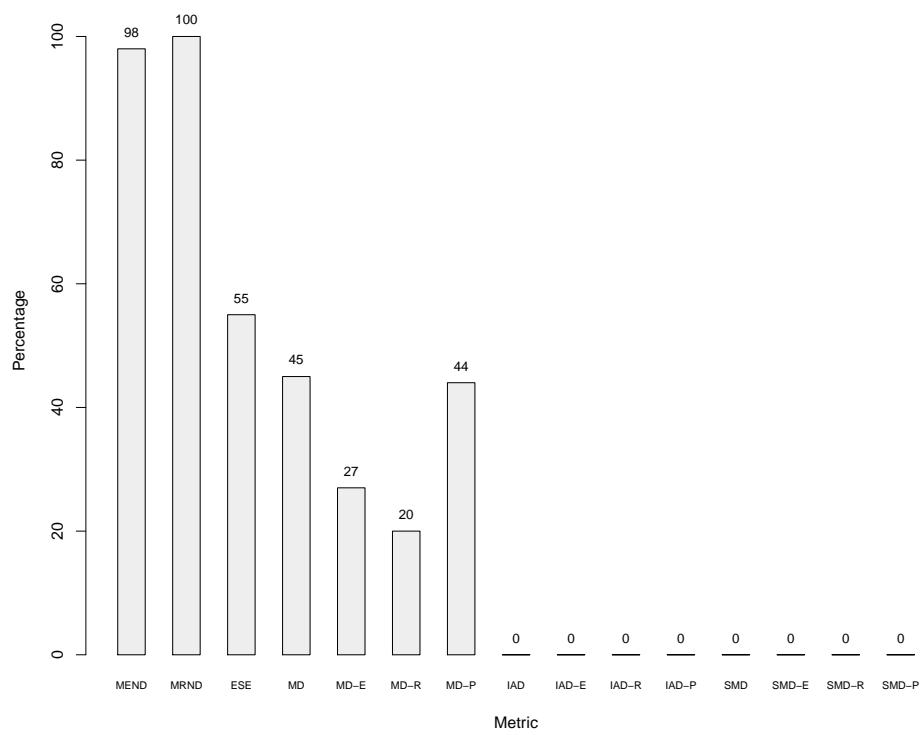
Fig. 65 Quality results for VFRBR on dataset BIB-RCAT
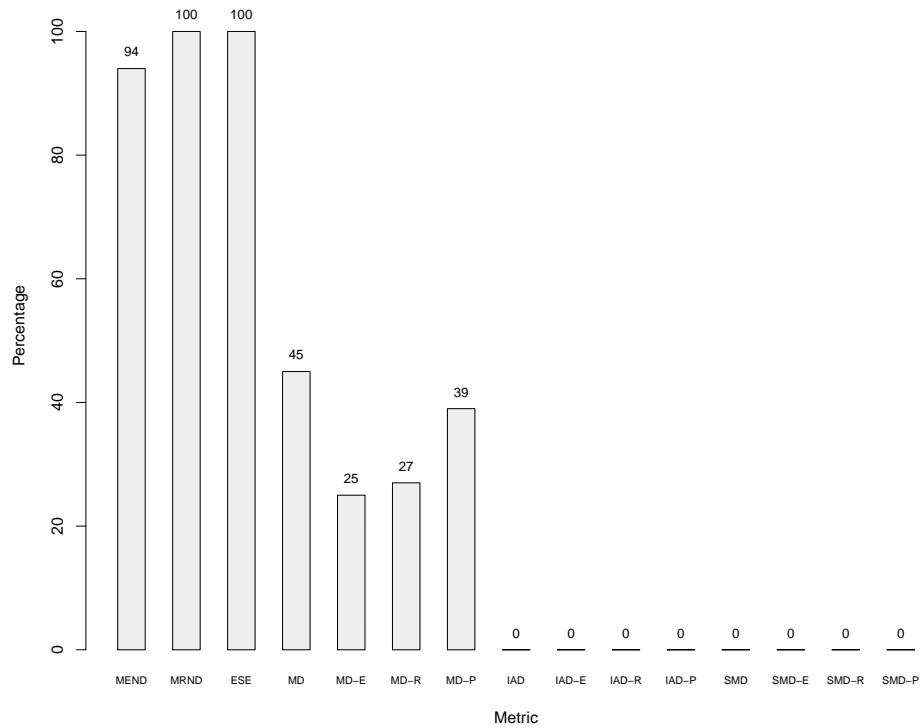
Fig. 66 Quality results for XC on dataset BIB-RCAT
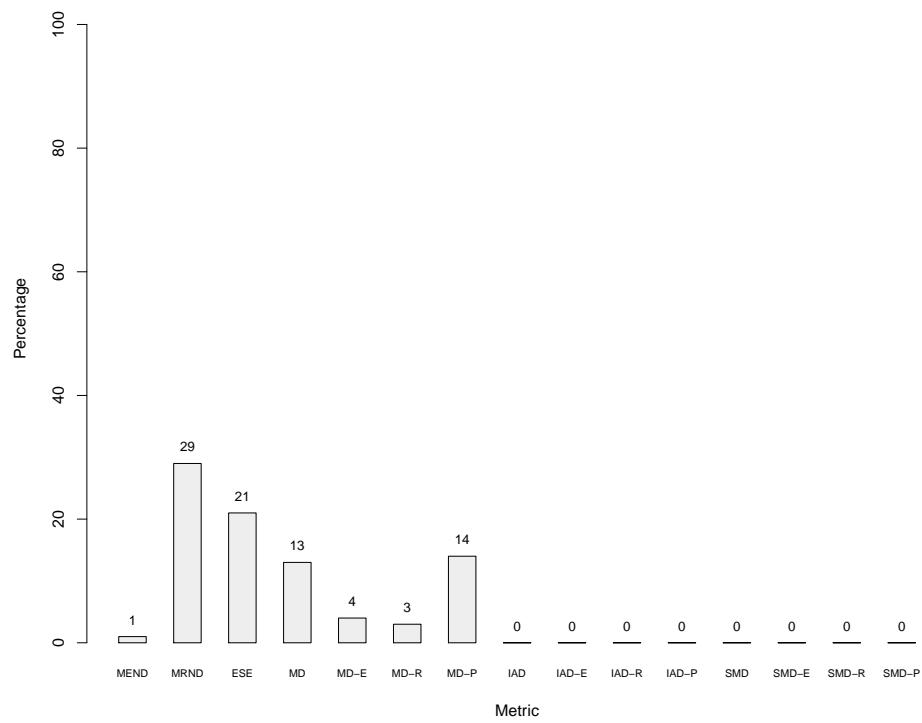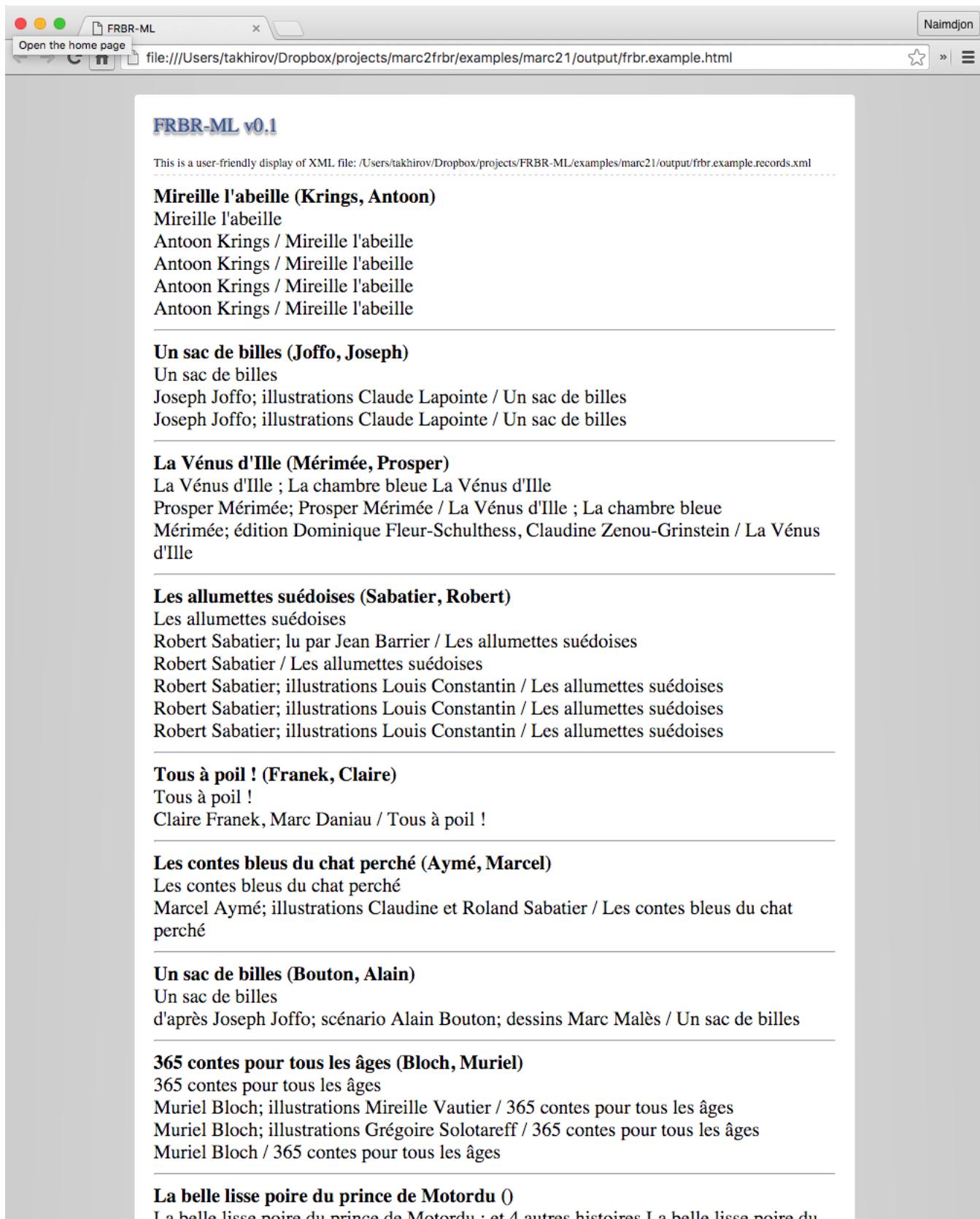
Fig. 67 Quality results for FRBR-ML tuned (enhanced set of rules) on dataset BIB-RCAT

4 Additional information about experiments

In this section, we provide screenshots of the tools that have been used during experiments.

Figure 68 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.

Fig. 68 Extract of the FRBRized BIB-RCAT dataset in FRBR-ML

Figure 69 depicts (an extract of) the relationships between FRBR entities in the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.
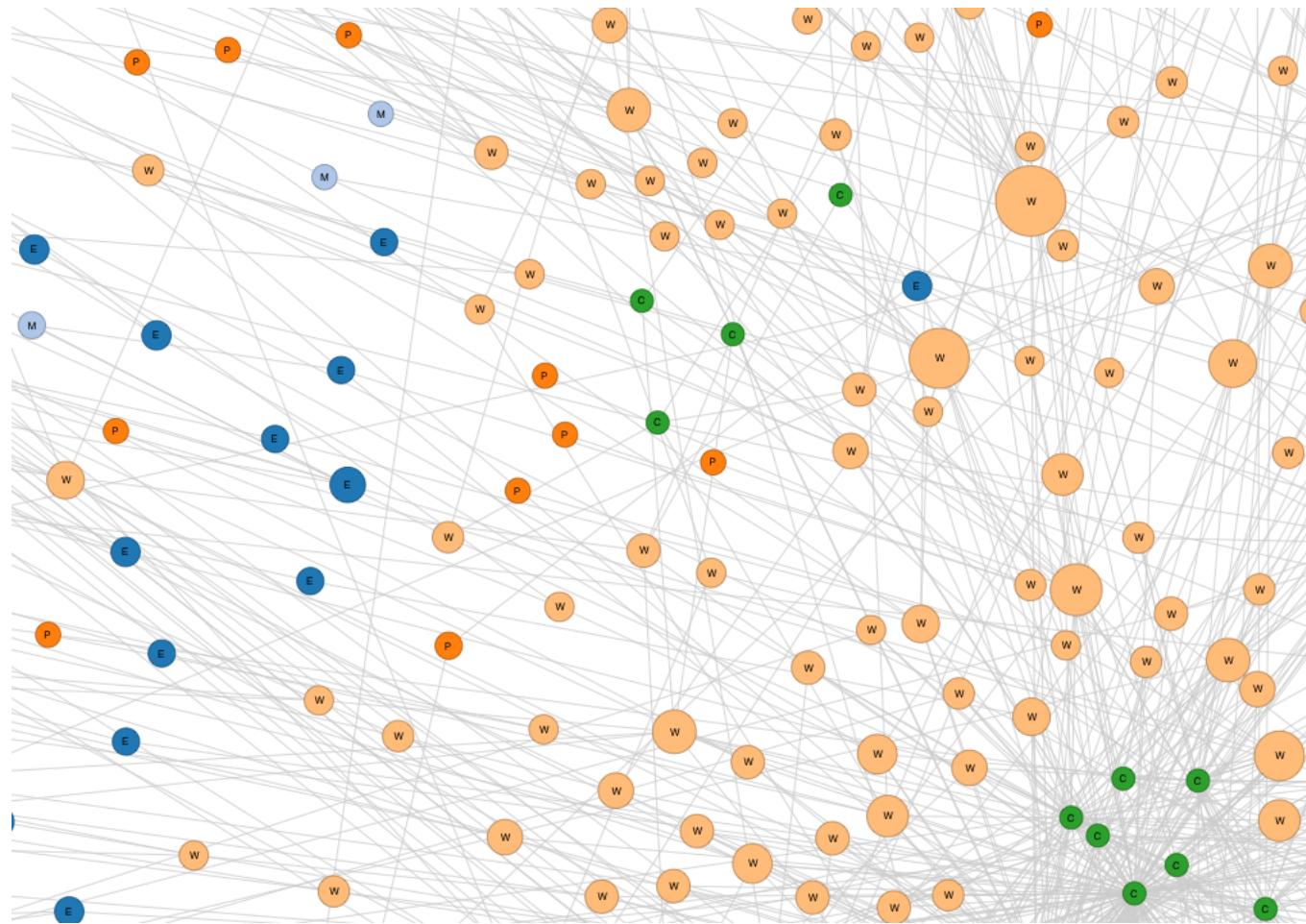
Fig. 69 Extract of a graph-based visualization of the FRBRized BIB-RCAT dataset in FRBR-ML

Figure 70 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using Variations (Scherzo).

Fig. 70 Extract of the FRBRized BIB-RCAT dataset in Scherzo (Variations/VFRBR)

Below we present the interpretation of a MARC record in FRBR using XC. Figure 71 depicts the initial MARC record. Figures 72, 73 and 74 respectively illustrate the FRBRization into a Work, an Expression and a Manifestation.

Fig. 71 A MARC record displayed in XC web interface

Fig. 72 A FRBR Work entity displayed in XC web interface

Fig. 73 A FRBR Expression entity displayed in XC web interface

Fig. 74 A FRBR Manifestation entity displayed in XC web interface

5 Acknowledgments

References

1. Decourselle, J., Duchateau, F., Aalberg, T., Takhirov, N., Lumineau, N.: BIB-R: a Benchmark for the Interpretation of Bibliographic Records. In: Theory and Practice of Digital Libraries (TPDL). Hannover, Germany (2016). URL `https://hal.archives-ouvertes.fr/hal-01324529`