Summary of empirical finds

Description of contributors

Four radiologist participated in the study (A through D). Additionally, automated readings by the ABDA software was added to the dataset, referred to as radiologist Z.

The manual readings (done by radiologists A through D) are based on heuristics and the value for the denser breast is picked as the breast density for both breasts – for those cases where they differ. We therefore have manual readings per patient taking density of both breast into account. This makes sense clinically, since the risk associated with having denser breasts (any breast) affects the individual regardless of breast.

The automatic reading (done by software) on the other hand, are based on a combination of the two projections per breast (MLO & CC) and often have different values for the left and right breast. For reasons of being able to compare these automatic readings with the manual readings, the denser value is chosen for the patient.

The data used in the study is available at xxx and contains the individual readings – specified for each breast individually, as it were – together with the value used for statistics (which would be the denser value).

Each of the radiologist (A through D and Z) has contributed:

Radiologist	Number of readings made by radiologist
A	8872
B	9502
C	7365
D	4902
Z (software)	15498

The number of manual readings is 30641, which together with the 15498 automatic readings gives a total number of 46139 readings.

Description of data cleansing

The study is based on mammography screening of 15504 patients, but some readings (of the total 46139 mentioned above) are excluded from the study.

Less than two readings for the same patient

We only have one single reading for (anonymized) patients c4a09c29-daf2-428a-8747-d6377c51b5e9 and bf0dc9c3-ab98-4912-9efe-286ed2ab411d made by radiologist D on 2014-03-11 and radiologist A on 2014-03-19, respectively.

These two readings will only be used in statistics for breast density per age class and not in the calculation of Kappa.

Less than two readings for the same screen

We only have one single reading for (anonymized) patients 1fe10012-4462-477b-803f-ccbb50ca8f74 and ce11cef0-0246-4de6-81f9-60f12bd3612f in same screen, assessed by radiologist A on 2014-09-03 and radiologist C on 2014-02-27, respectively. There may be other screens with multiple readings for these patients, but these cases where only one radiologist have made a reading are excluded.

These two assessments will only be used in statistics for breast density per age class and not in the calculation of Kappa.

Duplicate assessments for same patient and same screen

We have 64 duplicate readings for same patient and same screen:

Patient	Screening date	Pair of radiologists
47aa3c45-1735-4df2-ab9f-343a2f95a378	2014-03-19	Z:B
47aa3c45-1735-4df2-ab9f-343a2f95a378	2014-03-19	A:B
47aa3c45-1735-4df2-ab9f-343a2f95a378	2014-03-19	B:C
c588d2c7-99d2-47b4-b762-dd291a9a3f9f	2014-03-19	Z:B
c588d2c7-99d2-47b4-b762-dd291a9a3f9f	2014-03-19	Z:C
c588d2c7-99d2-47b4-b762-dd291a9a3f9f	2014-03-19	B:C
c588d2c7-99d2-47b4-b762-dd291a9a3f9f	2014-03-19	B:C
c588d2c7-99d2-47b4-b762-dd291a9a3f9f	2014-03-19	B:C
173a4d44-bc62-45c2-9d31-97f48689deb3	2014-10-29	Z:B
173a4d44-bc62-45c2-9d31-97f48689deb3	2014-10-29	Z:C
173a4d44-bc62-45c2-9d31-97f48689deb3	2014-10-29	B:C
173a4d44-bc62-45c2-9d31-97f48689deb3	2014-10-29	B:C
173a4d44-bc62-45c2-9d31-97f48689deb3	2014-10-29	B:C
f0b2fbb1-1257-4679-8de8-3a1c9fd8a4d9	2014-09-25	Z:A
f0b2fbb1-1257-4679-8de8-3a1c9fd8a4d9	2014-09-25	A:B
f0b2fbb1-1257-4679-8de8-3a1c9fd8a4d9	2014-09-25	A:D
7f971e96-9e14-4020-8237-94ffdc4d65e2	2014-01-21	Z:B
7f971e96-9e14-4020-8237-94ffdc4d65e2	2014-01-21	Z:D
7f971e96-9e14-4020-8237-94ffdc4d65e2	2014-01-21	B:D
7f971e96-9e14-4020-8237-94ffdc4d65e2	2014-01-21	B:D
7f971e96-9e14-4020-8237-94ffdc4d65e2	2014-01-21	B:D
22a2861a-624c-4abb-abc7-b40361b008e1	2014-11-12	Z:A
22a2861a-624c-4abb-abc7-b40361b008e1	2014-11-12	A:B
82057225-da7a-4893-8d68-43e0e937e863	2014-03-19	Z:B
82057225-da7a-4893-8d68-43e0e937e863	2014-03-19	A:B
82057225-da7a-4893-8d68-43e0e937e863	2014-03-19	B:C

Patient	Screening date	Pair of radiologists
f0f1049b-b614-4369-a95b-1e8c86ff1b91	2014-02-28	Z:A
f0f1049b-b614-4369-a95b-1e8c86ff1b91	2014-02-28	Z:B
f0f1049b-b614-4369-a95b-1e8c86ff1b91	2014-02-28	A:B
f0f1049b-b614-4369-a95b-1e8c86ff1b91	2014-02-28	A:B
f0f1049b-b614-4369-a95b-1e8c86ff1b91	2014-02-28	A:B
5cd04a6a-6f35-41dc-9975-88b8dcf16b08	2014-10-02	Z:A
5cd04a6a-6f35-41dc-9975-88b8dcf16b08	2014-10-02	Z:D
5cd04a6a-6f35-41dc-9975-88b8dcf16b08	2014-10-02	A:D
5cd04a6a-6f35-41dc-9975-88b8dcf16b08	2014-10-02	A:D
5cd04a6a-6f35-41dc-9975-88b8dcf16b08	2014-10-02	A:D
dd9f6f8b-9489-4ff1-8a0d-2c6220ac9a55	2014-03-19	Z:B
dd9f6f8b-9489-4ff1-8a0d-2c6220ac9a55	2014-03-19	Z:C
dd9f6f8b-9489-4ff1-8a0d-2c6220ac9a55	2014-03-19	B:C
dd9f6f8b-9489-4ff1-8a0d-2c6220ac9a55	2014-03-19	B:C
dd9f6f8b-9489-4ff1-8a0d-2c6220ac9a55	2014-03-19	B:C
5f4b8800-3972-4271-97b3-b23e26179243	2014-03-20	A:D
9a159673-bebd-4896-a613-5e27ee3116b9	2014-12-10	Z:A
9a159673-bebd-4896-a613-5e27ee3116b9	2014-12-10	Z:D
9a159673-bebd-4896-a613-5e27ee3116b9	2014-12-10	A:D
9a159673-bebd-4896-a613-5e27ee3116b9	2014-12-10	A:D
9a159673-bebd-4896-a613-5e27ee3116b9	2014-12-10	A:D
0fa8f49c-37bd-4988-817b-82c8d1ffd9bb	2014-03-27	Z:D
0fa8f49c-37bd-4988-817b-82c8d1ffd9bb	2014-03-27	A:D
0fa8f49c-37bd-4988-817b-82c8d1ffd9bb	2014-03-27	B:D
63b3ad43-edfa-414c-a587-79fa0e6bfea5	2014-09-02	Z:D
63b3ad43-edfa-414c-a587-79fa0e6bfea5	2014-09-02	B:D
0a055682-0614-4726-a016-9541babe50f0	2014-03-19	Z:B
0a055682-0614-4726-a016-9541babe50f0	2014-03-19	A:B
0a055682-0614-4726-a016-9541babe50f0	2014-03-19	B:C

Patient	Screening date	Pair of radiologists
375c4300-27ac-4814-b367-17f961aeede0	2014-02-27	Z:D
375c4300-27ac-4814-b367-17f961aeede0	2014-02-27	C:D
03698956-5d0c-4e65-89ff-65eb073a6f2d	2014-02-26	Z:C
03698956-5d0c-4e65-89ff-65eb073a6f2d	2014-02-26	B:C
7c00fb42-3b4b-4192-ba6f-4bbd0e62aefa	2014-11-19	Z:A
7c00fb42-3b4b-4192-ba6f-4bbd0e62aefa	2014-11-19	Z:C
7c00fb42-3b4b-4192-ba6f-4bbd0e62aefa	2014-11-19	A:C
7c00fb42-3b4b-4192-ba6f-4bbd0e62aefa	2014-11-19	A:C
7c00fb42-3b4b-4192-ba6f-4bbd0e62aefa	2014-11-19	A:C

These 64 paired readings are excluded from the calculation of Kappa.

Calculation of inter and intra reader agreement

When calculating Kappa, we are interested in both *inter reader agreement* as well as *intra reader agreement* (which is based on a separate study):

Pair of radiologists	Number of readings made by pair
A:A (auto paired)	203
A:B	4125
A:C	2635
A:D	1933
B:B (auto paired)	205
B:C	3491
B:D	1875
C:C (auto paired)	202
C:D	1061
D:D (auto paired)	204
Z:A	8861
Z:B	9491
Z:C	7354
Z:D	4889

There are 15120 paired manual readings that are used in the calculation of *inter reader agreement*. The total number of paired automatic and manual readings is 30595.

There are 814 auto-paired manual readings that are used in the calculation of *intra reader agreement*.

Description of differences in datasets

This study contains three datasets; one emanating from manual readings done by radiologists, one emanating from automatic readings done by the ABDA software, and one emanating from repeated readings by same radiologist separated in time by 6 months on an arbitrarily chosen subset of patients.

For most women, only one mammography screen was done during 2014, but some had more than one mammography screen. Because of this, the number of readings will be higher than the number of unique patients participating in the study.

The selection of patients included in the manual reading dataset overlap to a large extent with the selection included in the automatic reading dataset. Considering that we want to compare readings made by radiologists with those made by the ABDA software, we can only use readings for patients and screens available in both datasets.

Size of datasets

The ABDA software has processed and assessed all patients and all screens, while the individual radiologist only has covered a fraction of patients and screens. There are some discrepancies regarding the size of the datasets that may be a result of filtering done during capure from the PACS – the extraction of manual readings was strictly filtered to only contain those done in 2014 while the accession numbers identifying the data in this study to some small extent have covered examinations done before 2014.

As it were, 984 accession numbers seem to refer to mamography screens done before 2014.

Women having only one breast

Looking at unique patients in the study

In the dataset with automatic readings, 65 women are lacking their left breast and 73 women are lacking their right breast. This corresponds to 138 unique women lacking a breast.

In the dataset with manual readings, 62 women are lacking their left breast and 70 women are lacking their right breast. This corresponds to 132 unique women lacking a breast.

Looking at readings used in Kappa calculations

In this case, only pairs of radiologists (and the ABDA software) and their readings are considered. As mentioned above, only those readings appearing in both datasets can be used. Because of this, the number of patients appearing in the comparison of paired agreement differes from the number of patients in the individual datasets.

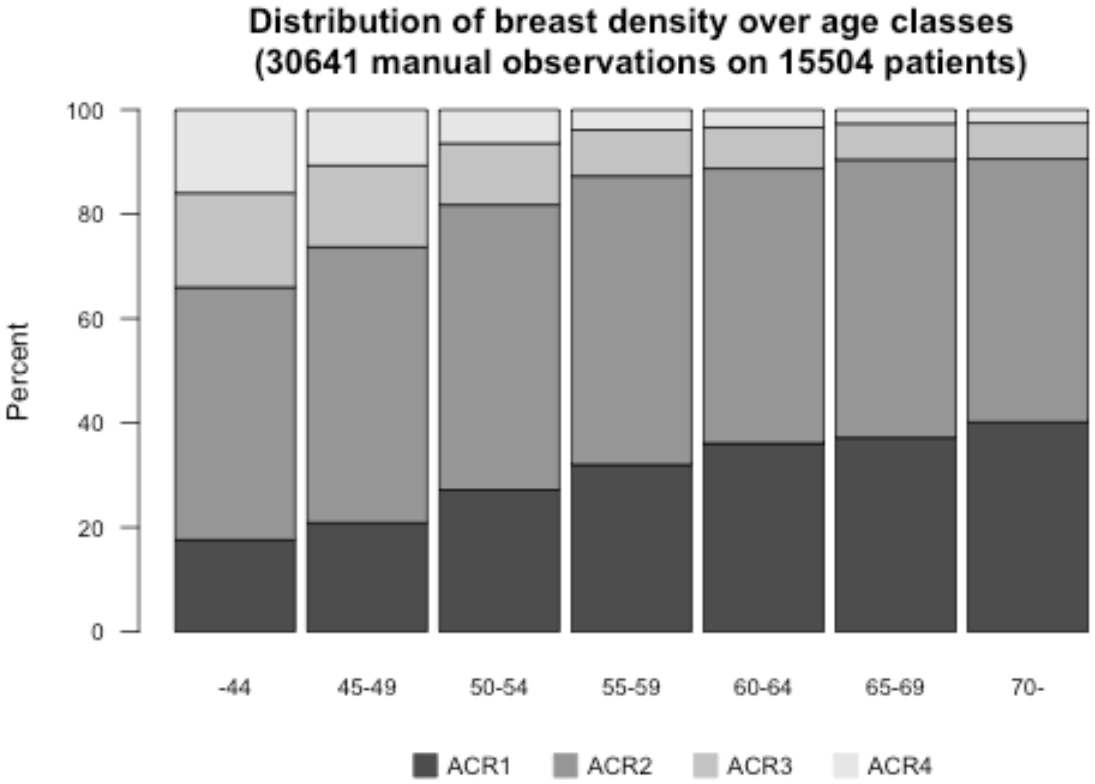
From the data in the dataset with automatic readings, 65 women are lacking their left breast and and 73 women are lacking their right breast – a total of 138 women. Since the readings were made by the same entity (i.e. ABDA software) once for each patient, this corresponds to the 138 unique women mentioned above.

From the dataset with manual readings, multiple readings involve the same women, so that 115 women are found lacking their left breast and 134 women are found lacking their right breast – which produces a total of 249 women, clearly more than the 132 mentioned above. Since at least two radiologists have examined the same patient, this does **not** correspond to unique women but some permutation.

In fact, the Kappa calculation for pairs consisting of a radiologist and ABDA software includes 135 women who lack either breast, women that appear in both datasets

Distribution of breast density over age classes

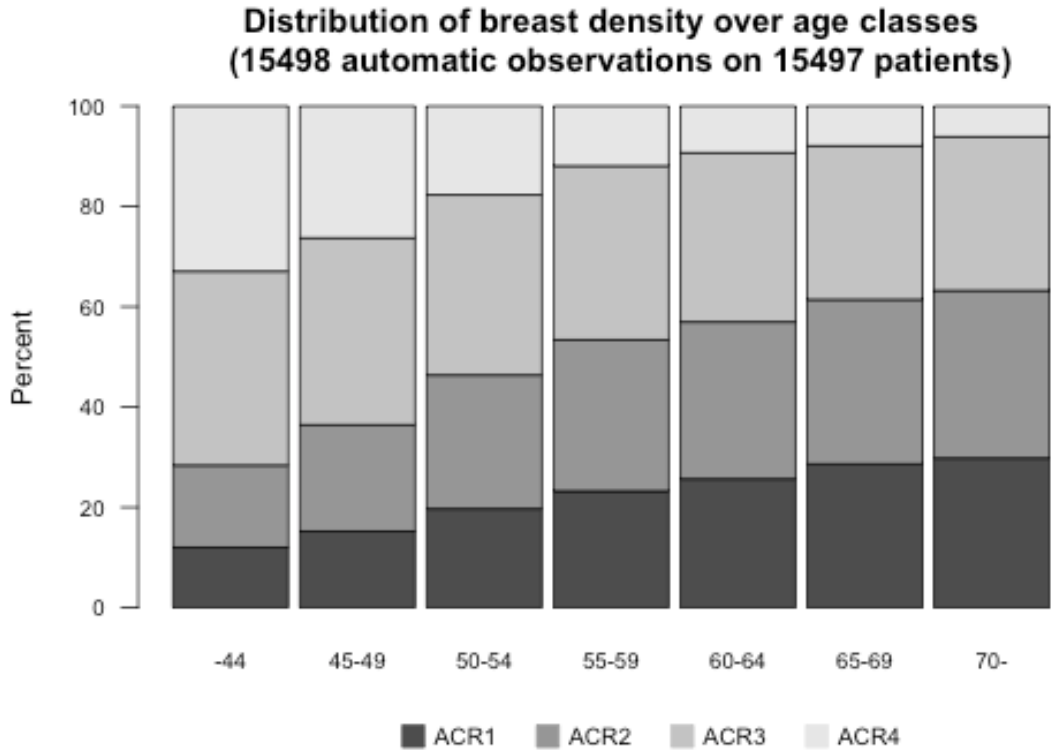
Manual readings made by radiologists



Age class	ACR1%	ACR1#	ACR2%	ACR2#	ACR3%	ACR3#	ACR4%	ACR4#
-44	17,57172131	686	48,41188525	1890	18,05840164	705	15,9579918	623
45-49	20,85255988	949	52,82355526	2404	15,64491321	712	10,67897165	486
50-54	27,14870396	1194	54,63847203	2403	11,77808095	518	6,434743065	283
55-59	32,01522521	1514	55,27595686	2614	8,817931909	417	3,890886022	184
60-64	36,09713411	1650	52,61430759	2405	7,87573835	360	3,412819952	156
65-69	37,18623482	1837	53,2388664	2630	6,923076923	342	2,651821862	131
70-	40,13528749	1424	50,42277339	1789	6,877113867	244	2,564825254	91
	Sum ACR1#	*9254*	*Sum ACR2#*	*16135*	*Sum ACR3#*	*3298*	*Sum ACR4#*	*1954*

Summing percentages by row yields 100%.

Automatic readings made by software



Age class	ACR1%	ACR1#	ACR2%	ACR2#	ACR3%	ACR3#	ACR4%	ACR4#
-44	11,99392713	237	16,39676113	324	38,66396761	764	32,94534413	651
45-49	15,20797227	351	21,2305026	490	37,26169844	860	26,29982669	607
50-54	19,73033708	439	26,65168539	593	35,91011236	799	17,70786517	394
55-59	23,26359833	556	30,12552301	720	34,68619247	829	11,92468619	285
60-64	25,6932409	593	31,28249567	722	33,70883882	778	9,31542461	215
65-69	28,6287089	714	32,79871692	818	30,63352045	764	7,939053729	198
70-	29,82749026	536	33,38898164	600	30,6622148	551	6,1213133	110
	Sum ACR1#	*3426*	*Sum ACR2#*	*4267*	*Sum ACR3#*	*5345*	*Sum ACR4#*	*2460*

Summing percentages by row yields 100%.

Inter reader agreement

The result covers agreement both between pairs of radiologists (A through D) as well as agreement between individual radiologists and software (Z).

Pair of readers: Z & D

Number of pair readings: 4889

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   2346   101 |  ACR1&2
|   1529   913 |  ACR2&3
+--------------+
Kappa = 0,3327977 (Fair agreement)
95% confidence interval: 0.3063526 0.3592427
Z = 23.256, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|    439   622    16     2 |  ACR1
|    379   906    75     8 |  ACR2
|    195  1046   348    73 |  ACR3
|      7   281   324   168 |  ACR4
+--------------------------+
Kappa = 0,1507739 (Slight agreement)
95% confidence interval: 0.1321120 0.1694358
Z = 17.304, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   1001   1121    397 |
+---------------------+
```

Percent of all observations:

ACR1/2	ACR2/3	ACR3/4
20,47	22,93	8,12

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
211	289

Percent of all observations:

ACR1/3	ACR2/4
4,32	5,91

Third degree disagreements (combination of ACR1 with ACR4) (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
9

Percent of all observations:

ACR1/4
0,18

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR3  ACR4
+--------------+
|    348    73 |  ACR3
|    324   168 |  ACR4
+--------------+
Kappa = 0,1611350 (Slight agreement)
95% confidence interval: 0.0991003 0.2231696
Z = 5.051, p-value = 2.198e-07
```

Pair of readers: A & D

Number of pair readings: 1933

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   1492   182 |  ACR1&2
|     31   228 |  ACR2&3
+--------------+
Kappa = 0,6190506 (Substantial agreement)
95% confidence interval: 0.5707921 0.6673091
Z = 17.363, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|    245    40     0     0 |  ACR1
|    153  1054   173     9 |  ACR2
|      1    27   121    36 |  ACR3
|      3     0    22    49 |  ACR4
+-------------------------+
Kappa = 0,5514171 (Moderate agreement)
95% confidence interval: 0.5158353 0.5869988
Z = 26.01, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|    193    200    58 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   9,98  10,35   3,00 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
1	9

Percent of all observations:

ACR1/3	ACR2/4
0,05	0,47

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
3

Percent of all observations:

ACR1/4
0,16

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
121	36	ACR3
22	49	ACR4

Kappa = 0,4372287 (Moderate agreement)
95% confidence interval: 0.3121674 0.5622900
Z = 5.9962, p-value = 1.01e-09

Pair of readers: B & C

Number of pair readings: 3491

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   2694    55 |  ACR1&2
|    309   433 |  ACR2&3
+--------------+
Kappa = 0,6440288 (Substantial agreement)
95% confidence interval: 0.6094188 0.6786388
Z = 24.491, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|    755  1151     1     0 |  ACR1
|     52   736    51     3 |  ACR2
|      0   212    90     1 |  ACR3
|      4    93   228   114 |  ACR4
+-------------------------+
Kappa = 0,2735618 (Fair agreement)
95% confidence interval: 0.2501517 0.2969719
Z = 25.181, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   1203    263    229 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|  34,46   7,53   6,56 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
1	96

Percent of all observations:

ACR1/3	ACR2/4
0,03	2,75

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
4

Percent of all observations:

ACR1/4
0,11

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
90	1	ACR3
228	114	ACR4

Kappa = 0,1682925 (Slight agreement)
95% confidence interval: 0.09435375 0.24223125
Z = 4.6278, p-value = 1.848e-06

Pair of readers: Z & A

Number of pair readings: 8861

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   4407    55 |  ACR1&2
|   3276  1123 |  ACR2&3
+--------------+
Kappa = 0,2442214 (Fair agreement)
95% confidence interval: 0.2239457 0.2644971
Z = 22.87, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|    532  1454     5     4 |  ACR1
|    459  1962    38     8 |  ACR2
|    266  2303   377    86 |  ACR3
|     15   692   425   235 |  ACR4
+--------------------------+
Kappa = 0,1076433 (Slight agreement)
95% confidence interval: 0.09399361 0.12129302
Z = 16.57, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|   1913   2341    511 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|  21,59  26,42   5,77 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
271	700

Percent of all observations:

ACR1/3	ACR2/4
3,06	7,90

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
19

Percent of all observations:

ACR1/4
0,21

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
377	86	ACR3
425	235	ACR4

Kappa = 0,1535380 (Slight agreement)
95 percent confidence interval: 0.09935899 0.20771698
Z = 5.5475, p-value = 1.449e-08

Pair of readers: C & D

Number of pair readings: 1061

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|    831   108 |  ACR1&2
|     16   106 |  ACR2&3
+--------------+
Kappa = 0,5676219 (Moderate agreement)
95% confidence interval: 0.4961043 0.6391395
Z = 11.253, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|    179    69     0     0 |  ACR1
|     22   561   101     7 |  ACR2
|      0    15    53    23 |  ACR3
|      0     1     6    24 |  ACR4
+-------------------------+
Kappa = 0,5778512 (Moderate agreement)
95% confidence interval: 0.5313706 0.6243319
Z = 20.59, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|     91    116     29 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   8,58  10,93   2,73 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	8

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,75

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
0

Percent of all observations:

ACR1/4
0,00

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
53	23	ACR3
6	24	ACR4

Kappa = 0,4245601 (Moderate agreement)
95% confidence interval: 0.2460587 0.6030614
Z = 4.1614, p-value = 1.582e-05

Pair of readers: B & D

Number of pair readings: 1875

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   1379    72 |  ACR1&2
|    112   312 |  ACR2&3
+--------------+
Kappa = 0,7099300 (Substantial agreement)
95% confidence interval: 0.6701272 0.7497327
Z = 21.981, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|    392   645     1     0 |  ACR1
|     20   322    69     2 |  ACR2
|      0    89   105    11 |  ACR3
|      2    21   106    90 |  ACR4
+--------------------------+
Kappa = 0,2927374 (Fair agreement)
95% confidence interval: 0.2616831 0.3237917
Z = 20.761, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|    665    158    117 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|  35,47   8,43   6,24 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
1	23

Percent of all observations:

ACR1/3	ACR2/4
0,05	1,23

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
2

Percent of all observations:

ACR1/4
0,11

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
105	11	ACR3
106	90	ACR4

Kappa = 0,3121797 (Fair agreement)
95% confidence interval: 0.2136493 0.4107101
Z = 6.0374, p-value = 7.831e-10

Pair of readers: Z & B

Number of pair readings: 9491

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   4505   162 |  ACR1&2
|   2962  1862 |  ACR2&3
+--------------+
Kappa = 0,3478786 (Fair agreement)
95% confidence interval: 0.3291488 0.3666084
Z = 34.214, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|   1749   294    31     6 |  ACR1
|   1894   568    81    44 |  ACR2
|   1398  1086   468   346 |  ACR3
|    168   310   306   742 |  ACR4
+--------------------------+
Kappa = 0,1765961 (Slight agreement)
95% confidence interval: 0.1638570 0.1893352
Z = 30.882, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|   2188   1167    652 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|  23,05  12,30   6,87 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
1429	354

Percent of all observations:

ACR1/3	ACR2/4
15,06	3,73

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
174

Percent of all observations:

ACR1/4
1,83

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
468	346	ACR3
306	742	ACR4

Kappa = 0,2845147 (Fair agreement)
95% confidence interval: 0.2402428 0.3287866
Z = 12.02, p-value < 2.2e-16

Pair of readers: Z & C

Number of pair readings: 7354

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   3568    55 |  ACR1&2
|   2758   973 |  ACR2&3
+--------------+
Kappa = 0,2429835 (Fair agreement)
95% confidence interval: 0.2210007 0.2649662
Z = 21.059, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|    683   940     6     2 |  ACR1
|    641  1304    44     3 |  ACR2
|    393  1784   332    46 |  ACR3
|     23   558   419   176 |  ACR4
+--------------------------+
Kappa = 0,1020610 (Slight agreement)
95% confidence interval: 0.08735495 0.11676698
Z = 14.607, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|   1581   1828    465 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|  21,50  24,86   6,32 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
399	561

Percent of all observations:

ACR1/3	ACR2/4
5,43	7,63

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
25

Percent of all observations:

ACR1/4
0,34

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
332	46	ACR3
419	176	ACR4

Kappa = 0,1475542 (Slight agreement)
95% confidence interval: 0.09157018 0.20353816
Z = 5.1991, p-value = 1.001e-07

Pair of readers: A & B

Number of pair readings: 4125

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   3208   360 |  ACR1&2
|     58   499 |  ACR2&3
+--------------+
Kappa = 0,6469648 (Substantial agreement)
95% confidence interval: 0.6148816 0.6790480
Z = 26.365, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|    582    38     1     3 |  ACR1
|   1681   907   240   116 |  ACR2
|      2    49   125   201 |  ACR3
|      1     6    12   161 |  ACR4
+-------------------------+
Kappa = 0,2200421 (Fair agreement)
95% confidence interval: 0.1993563 0.2407279
Z = 23.263, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   1719    289    213 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|  41,67   7,01   5,16 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
3	122

Percent of all observations:

ACR1/3	ACR2/4
0,07	2,96

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
4

Percent of all observations:

ACR1/4
0,10

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
125	201	ACR3
12	161	ACR4

Kappa = 0,2499841 (Fair agreement)
95% confidence interval: 0.1737303 0.3262379
Z = 6.4179, p-value = 6.909e-11

Pair of readers: A & C

Number of pair readings: 2635

Grouped observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|   2141   156 |  ACR1&2
|     90   248 |  ACR2&3
+--------------+
Kappa = 0,6146346 (Substantial agreement)
95% confidence interval: 0.5687813 0.6604880
Z = 17.84, p-value < 2.2e-16
```

All discrete observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|    277    60     0     0 |  ACR1
|    371  1433   141    15 |  ACR2
|      0    85   158    22 |  ACR3
|      1     4    31    37 |  ACR4
+-------------------------+
Kappa = 0,4552301 (Moderate agreement)
95% confidence interval: 0.4216287 0.4888314
Z = 23.771, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|    431    226     53 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|  16,36   8,58   2,01 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	19

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,72

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
1

Percent of all observations:

ACR1/4
0,04

Selected observations of readings for two readers; reader 1 row-wise, reader 2 column-wise

ACR3	ACR4	
158	22	ACR3
31	37	ACR4

Kappa = 0,4400136 (Moderate agreement)
95% confidence interval: 0.3063297 0.5736975
Z = 5.4437, p-value = 2.61e-08

Intra reader agreement

The result covers agreement between two manual readings done by the same radiologist (A through D) separated in time by 6 months.

Reader A on two occasions

Number of readings: 203

Grouped observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|    133     5 |  ACR1&2
|     12    53 |  ACR2&3
+--------------+
Kappa = 0,8019967 (Almost perfect agreement)
95% confidence interval: 0.7119009 0.8920925
Z = 9.7825, p-value < 2.2e-16
```

All discrete observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|     20     4     0     0 |  ACR1
|      3   106     5     0 |  ACR2
|      0    12    40     3 |  ACR3
|      0     0     3     7 |  ACR4
+--------------------------+
Kappa = 0,7463451 (Substantial agreement)
95% confidence interval: 0.6625525 0.8301377
Z = 12.563, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|      7     17      6 |
+----------------------+
```

Percent of all observations:

ACR1/2	ACR2/3	ACR3/4
3,45	8,37	2,96

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	0

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,00

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
0

Percent of all observations:

ACR1/4
0,00

Selected observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
    ACR3  ACR4
+--------------+
|     40     3 |  ACR3
|      3     7 |  ACR4
+--------------+
Kappa = 0,6302326 (Substantial agreement)
Z = 3.0478, p-value = 0.001153
95% confidence interval: 0.3516127 0.9088525
```

Reader B on two occasions

Number of readings: 205

Grouped observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|    137     3 |  ACR1&2
|      8    57 |  ACR2&3
+--------------+
Kappa = 0,8734923 (Almost perfect agreement)
95% confidence interval: 0.8007658 0.9462188
Z = 10.734, p-value < 2.2e-16
```

All discrete observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|     62    20     0     0 |  ACR1
|      8    47     3     0 |  ACR2
|      0     8    30     0 |  ACR3
|      0     0     4    23 |  ACR4
+--------------------------+
Kappa = 0,7052628 (Substantial agreement)
95% confidence interval: 0.6269506 0.7835750
Z = 15.864, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|     28     11      4 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|  13,66   5,37   1,95 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	0

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,00

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
0

Percent of all observations:

ACR1/4
0,00

Selected observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

ACR3	ACR4	
30	0	ACR3
4	23	ACR4

Kappa = 0,8582090 (Almost perfect agreement)
95% confidence interval: 0.7242205 0.9921974
Z = 6.4139, p-value = 7.094e-11

Reader C on two occasions

Number of readings: 202

Grouped observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|    140     7 |  ACR1&2
|      9    46 |  ACR2&3
+--------------+
Kappa = 0,7978231 (Substantial agreement)
95% confidence interval: 0.7027625 0.8928837
Z = 9.1006, p-value < 2.2e-16
```

All discrete observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+--------------------------+
|     66     8     0     0 |  ACR1
|      6    60     7     0 |  ACR2
|      0     9    32     3 |  ACR3
|      0     0     5     6 |  ACR4
+--------------------------+
Kappa = 0,7240834 (Substantial agreement)
95% confidence interval: 0.6450373 0.8031295
Z = 15.064, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|     14     16      8 |
+----------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+----------------------+
|   6,93   7,92   3,96 |
+----------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	0

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,00

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
0

Percent of all observations:

ACR1/4
0,00

Selected observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

ACR3	ACR4	
32	3	ACR3
5	6	ACR4

Kappa = 0,4903047 (Moderate agreement)
95% confidence interval: 0.1692888 0.8113207
Z = 2.3932, p-value = 0.008351

Reader D on two occasions

Number of readings: 204

Grouped observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
  ACR1&2 ACR3&4
+--------------+
|    149     6 |  ACR1&2
|      3    46 |  ACR2&3
+--------------+
Kappa = 0,8816095 (Almost perfect agreement)
95% confidence interval: 0.8059879 0.9572311
Z = 9.7047, p-value < 2.2e-16
```

All discrete observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

```
    ACR1  ACR2  ACR3  ACR4
+-------------------------+
|     28     4     0     0 |  ACR1
|      5   112     6     0 |  ACR2
|      0     3    38     1 |  ACR3
|      0     0     3     4 |  ACR4
+-------------------------+
Kappa = 0,8123275 (Almost perfect agreement)
95% confidence interval: 0.7382548 0.8864002
Z = 13.485, p-value < 2.2e-16
```

First degree disagreements (combination of ACR1 with ACR2, ACR2 with ACR3, ACR3 with ACR4 (i.e. the pair-wise sum of the off-by-one diagonals)

Observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|      9      9      4 |
+---------------------+
```

Percent of all observations:

```
  ACR1/2 ACR2/3 ACR3/4
+---------------------+
|   4,41   4,41   1,96 |
+---------------------+
```

Second degree disagreements (combination of ACR1 with ACR3, ACR2 with ACR4 (i.e. the pair-wise sum of the off-by-two diagonals)

Observations:

ACR1/3	ACR2/4
0	0

Percent of all observations:

ACR1/3	ACR2/4
0,00	0,00

Third degree disagreements (combination of ACR1 with ACR4 (i.e. the pair-wise sum of the off-by-three diagonals)

Observations:

ACR1/4
0

Percent of all observations:

ACR1/4
0,00

Selected observations of readings at two occasions; occasion 1 row-wise, occasion 2 column-wise

ACR3	ACR4	
38	1	ACR3
3	4	ACR4

Kappa = 0,6182573 (Substantial agreement)
95% confidence interval: 0.2607914 0.9757232
Z = 2.2774, p-value = 0.01138

Appendix 1: Software used in study

Empirical data was extracted from DICOM archive and processed into datasets using software developed for this purpose.

Calculation of statistics for age classification was done using the statistics package of `Apache Commons Math3 version 3.6.1` (*org.apache.commons:commons-math3:3.6.1*) running on `JRE 1.8.0_111` (*build 1.8.0_111-b14*).

Calculation of paired radiologist agreement (Kappa) was done using `R version 3.3.2` (*2016-10-31*) and the `fmsb` package, installable from CRAN and running on an `Apple MacBook Pro` (*MacBookPro11,3*) with `OS x86_64-apple-darwin15.0.0`.

The `fmsb` package does an estimation of Cohen's kappa statistics and test the null hypothesis that the extent of agreement is same as random (kappa=0).

The Kappa calculations have been manually verified using information from these articles; Article [1] discusses the binomial case with two outcomes during reading), while article [2] discusses the more general case with ACR1 - ACR4 outcomes.

[1] Viera, A.J. and Garrett, J.M., 2005. *Understanding interobserver agreement: the kappa statistic*. Fam Med, 37(5), pp.360-363.

[2] Haley, S.M. and Osberg, J.S., 1989. *Kappa coefficient calculation using multiple ratings per subject: a special communication*. Physical Therapy, 69(11), pp.970-974.

Appendix 2: Extraction of characteristics

Manual readings dataset

Number of patients not having the left breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE left_breast_density IS
NULL AND radiologist <> 'Z' AND (study_date >= '2014-01-01' AND study_date <=
'2014-12-31')
--62
```

Number of patients not having the right breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE right_breast_density IS
NULL AND radiologist <> 'Z' AND (study_date >= '2014-01-01' AND study_date <=
'2014-12-31')
--70
```

Number of patients without either breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE (right_breast_density IS
NULL OR left_breast_density IS NULL) AND radiologist <> 'Z' AND (study_date
>= '2014-01-01' AND study_date <= '2014-12-31')
--132
```

Number of assessments on mamography studies done earlier than 2014

```
SELECT COUNT(*) FROM datum WHERE radiologist <> 'Z' AND study_date < '2014-
01-01'
--655
```

Automatic readings dataset

Number of patients not having the left

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE left_breast_density IS
NULL AND radiologist = 'Z' AND (study_date >= '2014-01-01' AND study_date <=
'2014-12-31')
--65
```

Number of patients not having the right breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE right_breast_density IS
NULL AND radiologist = 'Z' AND (study_date >= '2014-01-01' AND study_date <=
'2014-12-31')
--73
```

Number of patients without either breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE (right_breast_density IS
NULL OR left_breast_density IS NULL) AND radiologist = 'Z' AND (study_date >=
'2014-01-01' AND study_date <= '2014-12-31')
--138
```

Number of assessments on mamography studies done earlier than 2014

```
SELECT COUNT (*) FROM datum WHERE radiologist = 'Z' AND study_date < '2014-
01-01'
--329
```

Either dataset

Number of patients not having the left breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE left_breast_density IS
NULL AND (study_date >= '2014-01-01' AND study_date <= '2014-12-31')
--65
```

Number of patients not having the right breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE right_breast_density IS
NULL AND (study_date >= '2014-01-01' AND study_date <= '2014-12-31')
--73
```

Number of patients without either breast

```
SELECT COUNT(DISTINCT patient_id) FROM datum WHERE (right_breast_density IS
NULL OR left_breast_density IS NULL) AND (study_date >= '2014-01-01' AND
study_date <= '2014-12-31')
--138
```

Number of assessments on mamography studies done earlier than 2014

```
SELECT COUNT(*) FROM datum WHERE study_date < '2014-01-01'
--984
```

Number of assessments on mamography studies done after 2014

```
SELECT COUNT(*) FROM datum WHERE study_date > '2014-12-31'
--0
```