Measures of Agreement

MRM515 - Biomedical Imaging Study Design & Research Methods

Marcio Augusto Diniz, Ph.D.
Biostatistics and Binformatics Research Center
Cedars Sinai Medical Center

July 5, 2022

Zhang *et al. Journal of Cardiovascular Magnetic Resonance* (2018) 20:39
https://doi.org/10.1186/s12968-018-0453-z

Journal of Cardiovascular Magnetic Resonance

RESEARCH Open Access

3D whole-brain vessel wall cardiovascular magnetic resonance imaging: a study on the reliability in the quantification of intracranial vessel dimensions

Na Zhang[1,2,3], Fan Zhang[2], Zixin Deng[2,4], Qi Yang[2], Marcio A. Diniz[5], Shlee S. Song[6], Konrad H. Schlick[6], M. Marcel Maya[7], Nestor Gonzalez[8], Debiao Li[2,4,9], Hairong Zheng[1,3], Xin Liu[1,3*] and Zhaoyang Fan[2,9*]

3D intracranial vessel wall

Summary

- One of the potentially important applications of three-dimensional (3D) intracranial vessel wall (IVW) cardiovascular magnetic resonance (CMR) is to monitor disease progression and regression via quantitative measurement of IVW morphology during medical management or drug development;
- Lumen volume, Wall volume, Normalized wall index, Mean wall thickness, Maximum wall thickness;
- Five different regions: ACA, BA, ICA, MCA, VA;
- 24 healthy subjects and 10 patients;

Variability

- Any biological variable in a number of individuals or repeatedly within an individual always exhibit a range of values;

Measurement variability

- The variability is associated to external conditions under which the biological variable is being measured.

Error variability

- The variability is associated to the instrument used to measure the biological variable.
  - Random: The observed values may be sometimes higher or lower than the true values, but on average it is the true value.
  - Systematic: The observed values have a tendency to be high or low than the true values, such that on average it is biased.

Reliability

Repeatability

- Will a second measurement in the same subject by the same observer under identical conditions be the same?
- It is also know as intra-observer variability;

Reproducibility

- Will two measurements in the same subject by two different observers under identical conditions be the same?
- It is also know as inter-observer variability;
- Two observers can be two different machines, techniques (including gold standard) or operators.
- It only evaluated if there is repeatability;

Reliability

3D intracranial vessel wall

- Imaging protocol: A patient can be scanned to have either a 3D or 2D cardiovascular image resonance (CMR) and a reader performs above vessel wall and lumen measurements: Lumen volume, Wall volume, Normalized wall index, Mean wall thickness, Maximum wall thickness;

What are the possible causes of variability?

- Technique (3D or 2D CMR): Inter-technique;
- Scan: Inter-scan;
- Reader: Inter-observer, intra-observer.

Statistical methods

- Most of the methods can be applied to evaluate intra- and inter-variability;
- The choice depends on the nature of the biological variable.

Numerical

- Bland and Altman diagram;
- Intra-class correlation;
- Lin's concordance correlation;
- Paired t-test.

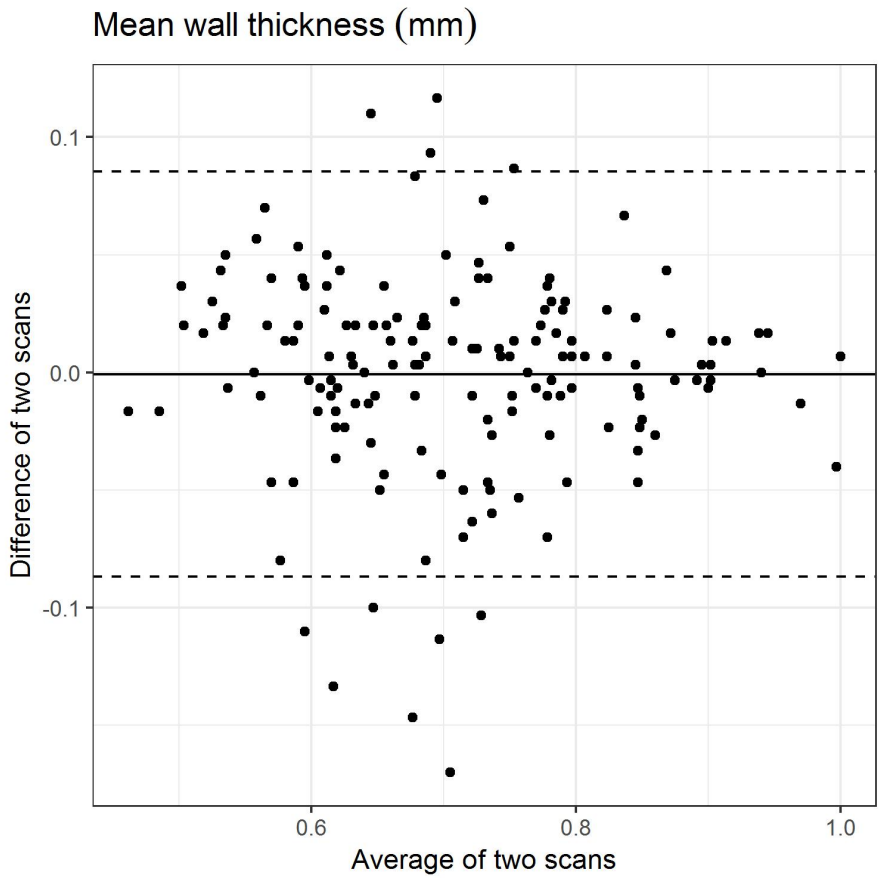
Statistical methods

- Most of the methods can be applied to evaluate intra- and inter-variability;
- The choice depends on the nature of the biological variable.

Categorical

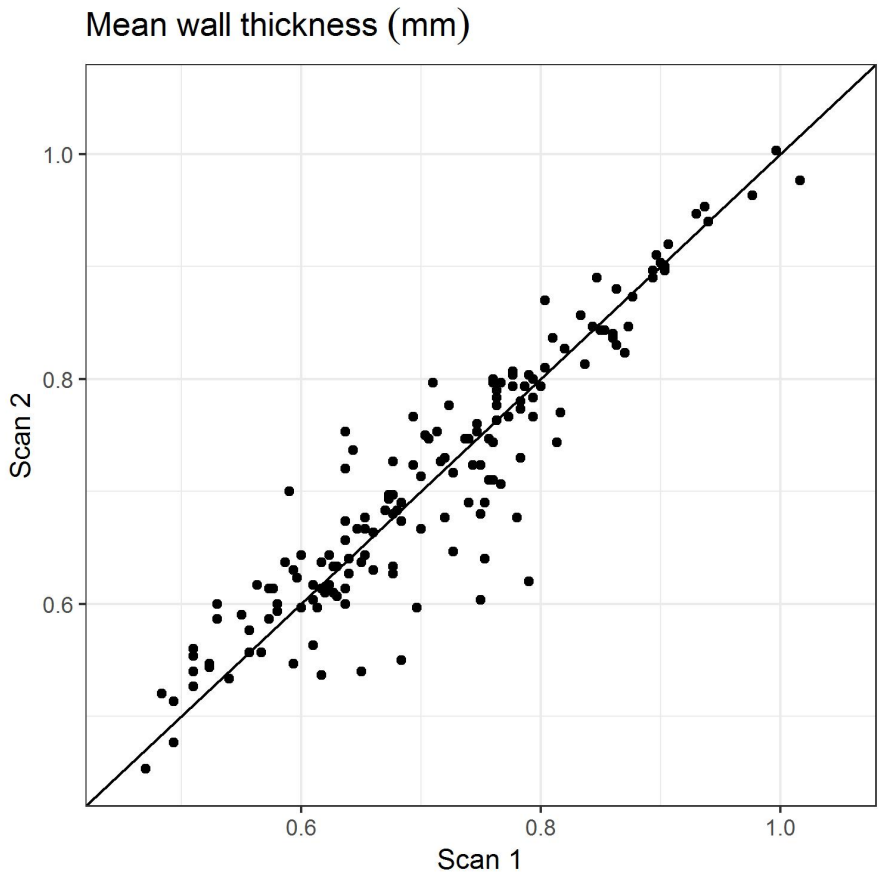
- Kappa statistic;
- McNemar test;
- If one of the techniques is the gold standard:
  - Sensitivity and Specificity;
  - Positive and Negative Predictive Values.

Bland-Altman diagram



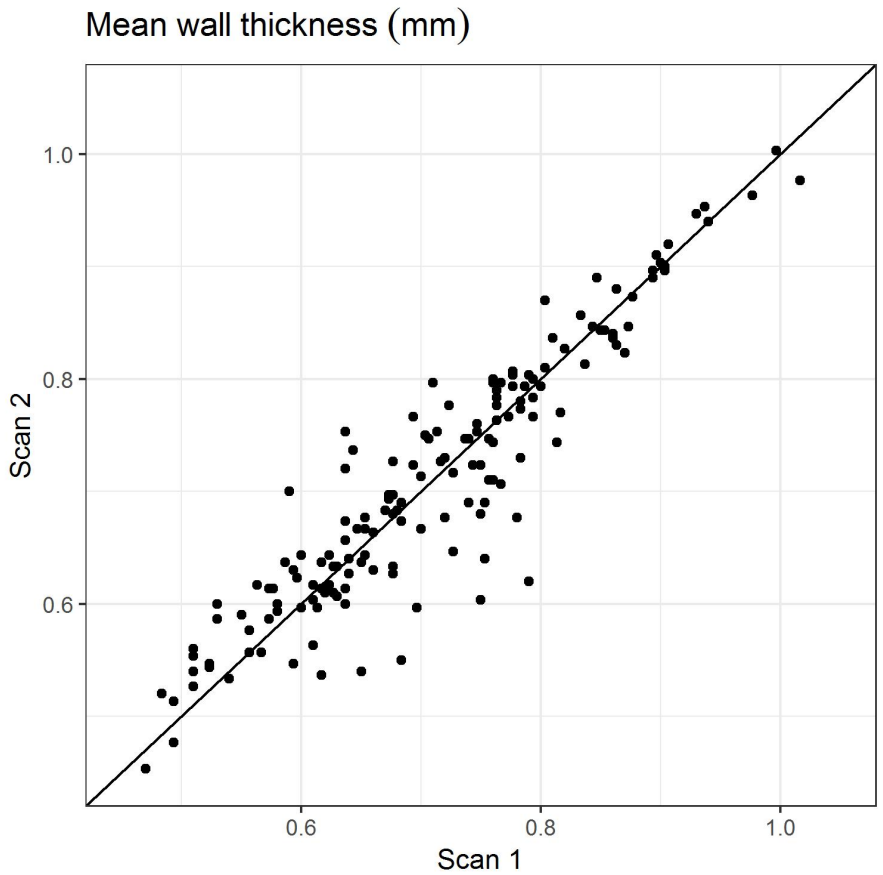
- It helps to identify systematic differences between the measurements as function of the true measure;
- The average of measures is a good estimate of the true measure, which is unknown;

Intra-class correlation



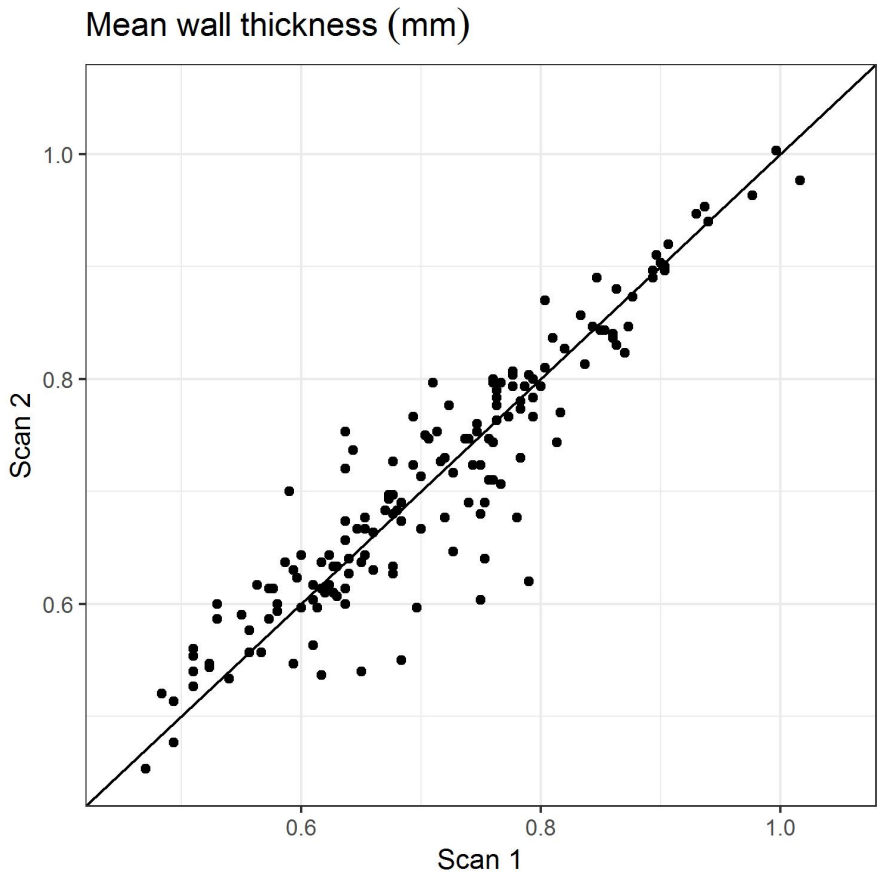
- It quantifies the resemblance between two measures in the same patient.
- It ranges from 0 to 1.

Intra-class correlation



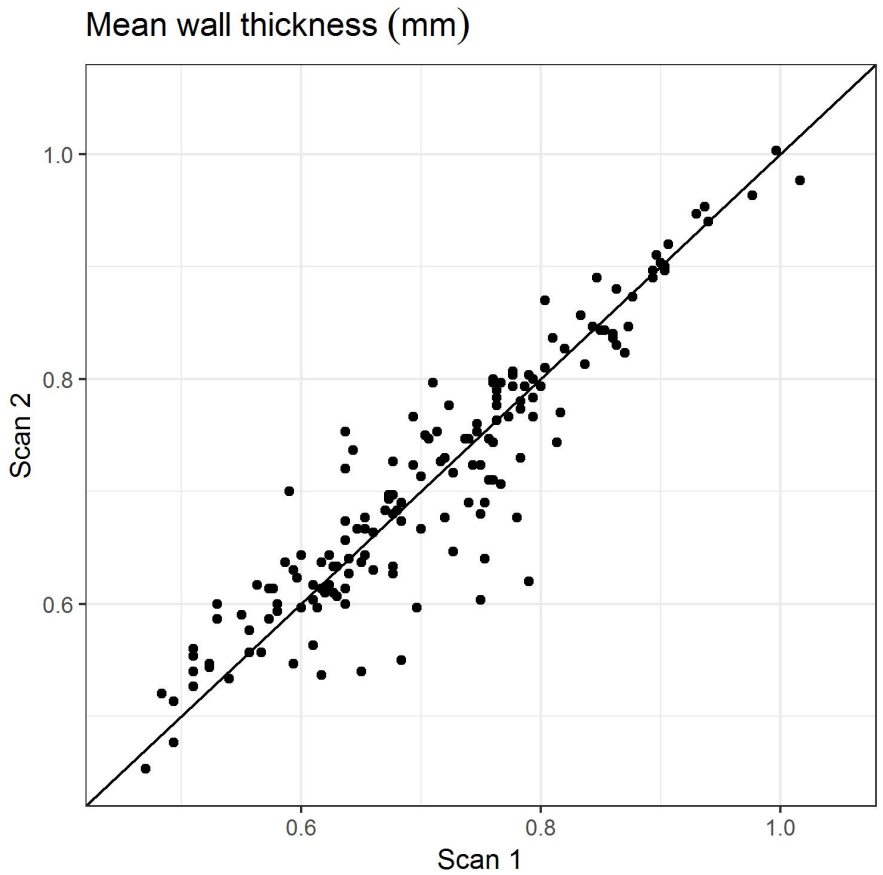
- It can be calculated in three different ways depending on the experiment,
  - One-way random: each patient is measured by a different set randomly selected observers;

Intra-class correlation



- It can be calculated in three different ways depending on the experiment,
  - Two-way random: observers are randomly selected, then, each patient is measured by the same set of observers;

Intra-class correlation



- It can be calculated in three different ways depending on the experiment,
  - Two-way mixed: observers are fixed, then, each patient is measured by the same set of observers.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

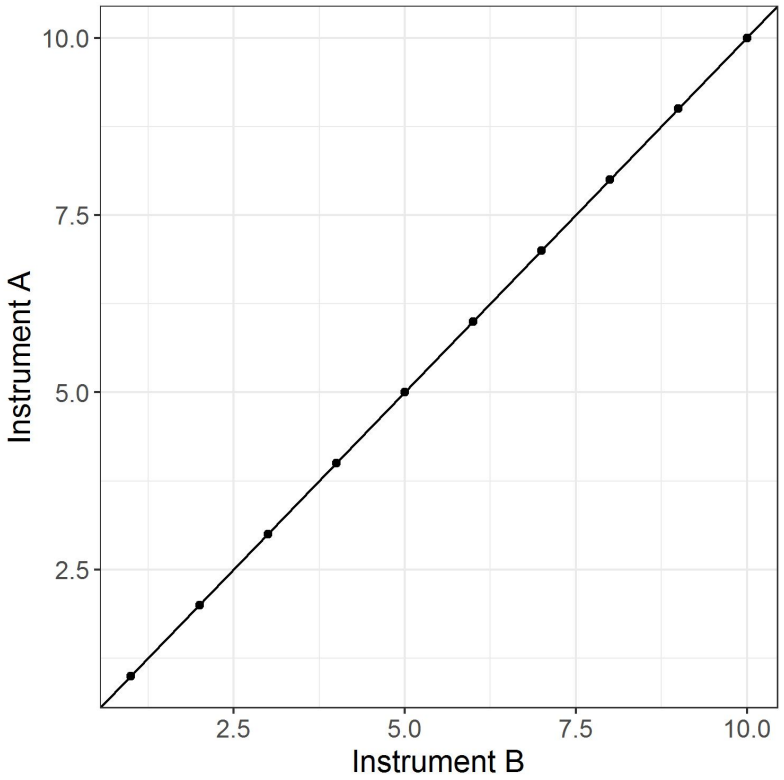


Figure: Pearson correlation = ? and Intra-class correlation = ?

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

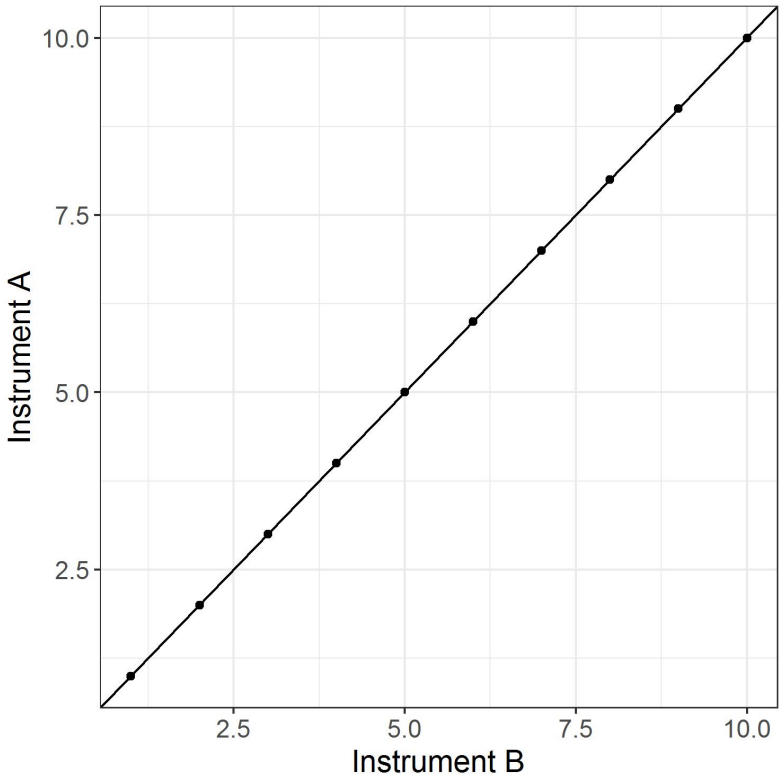


Figure: Pearson correlation = 1 and Intra-class correlation = 1

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

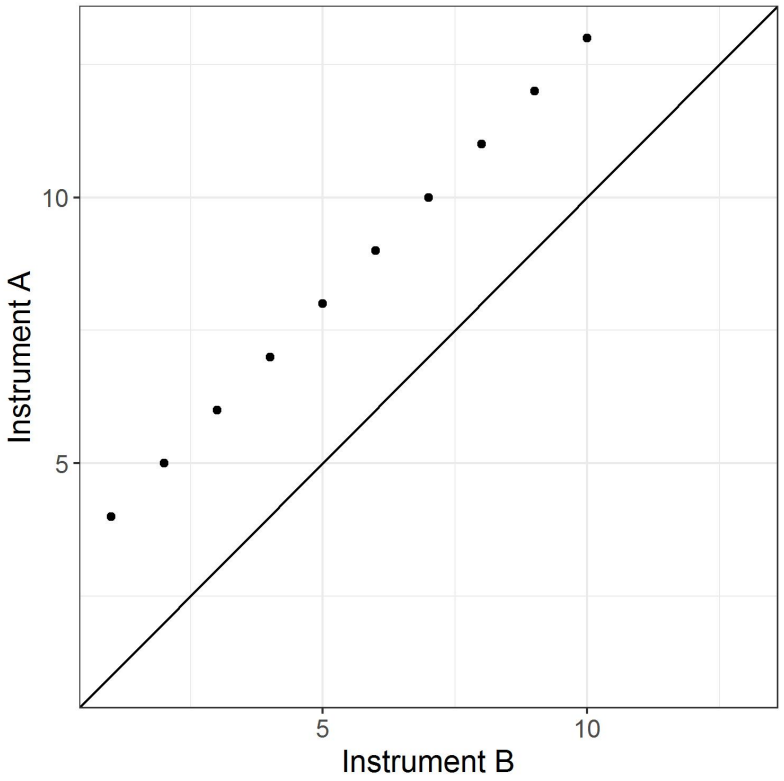


Figure: Pearson correlation = ? and Intra-class correlation = ?

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

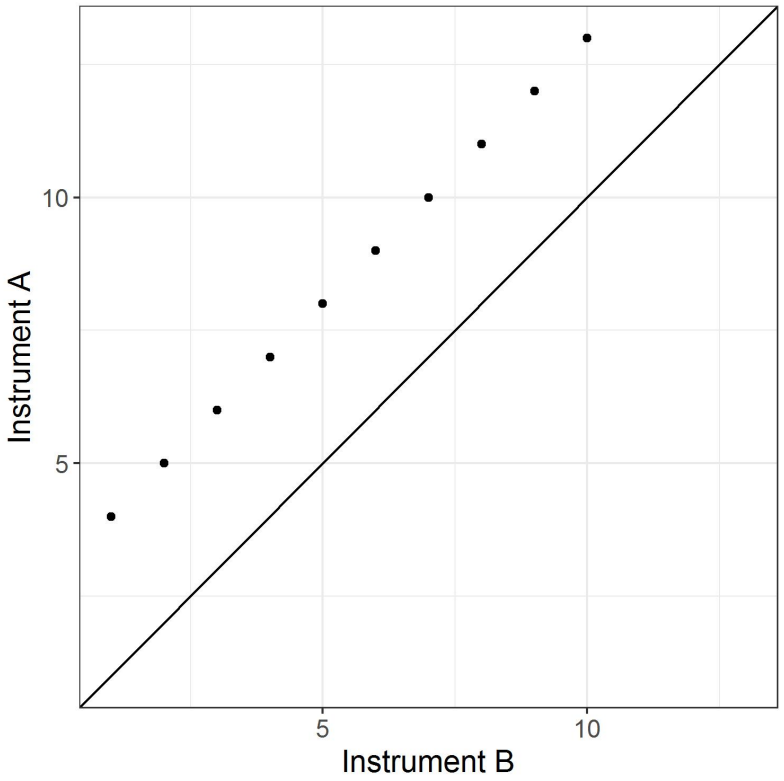


Figure: Pearson correlation = 1 and Intra-class correlation = 0.357

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

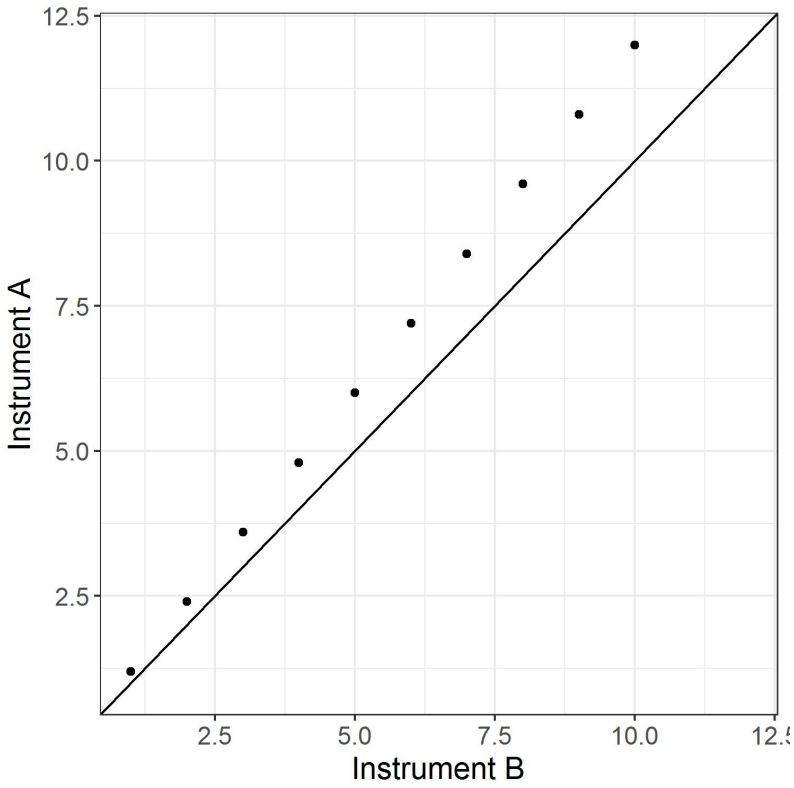


Figure: Pearson correlation = ? and Intra-class correlation = ?

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

Intra-class correlation

What is the difference between Intra-class and Pearson correlation?

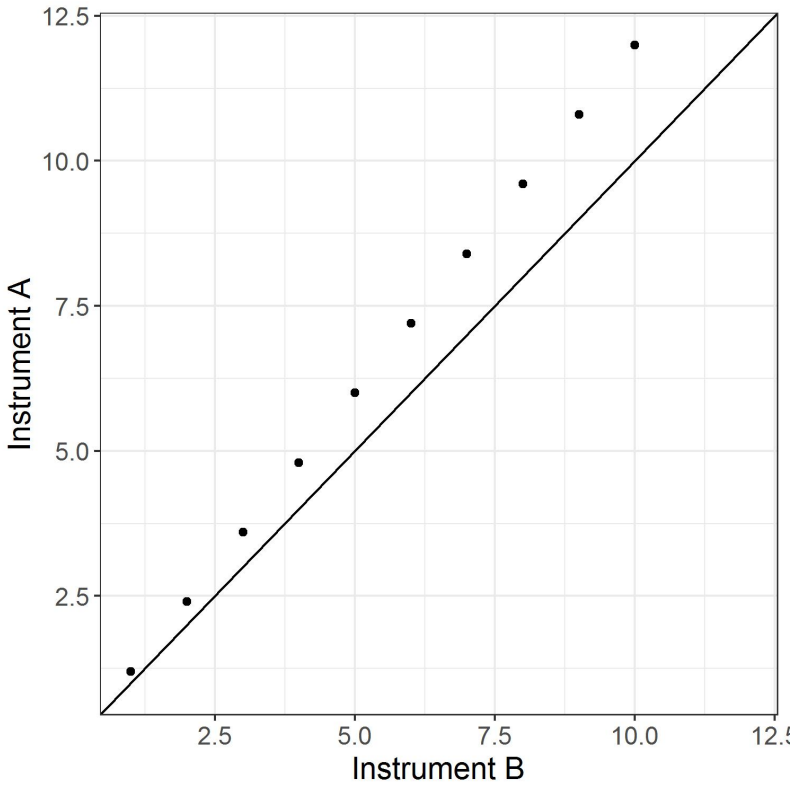
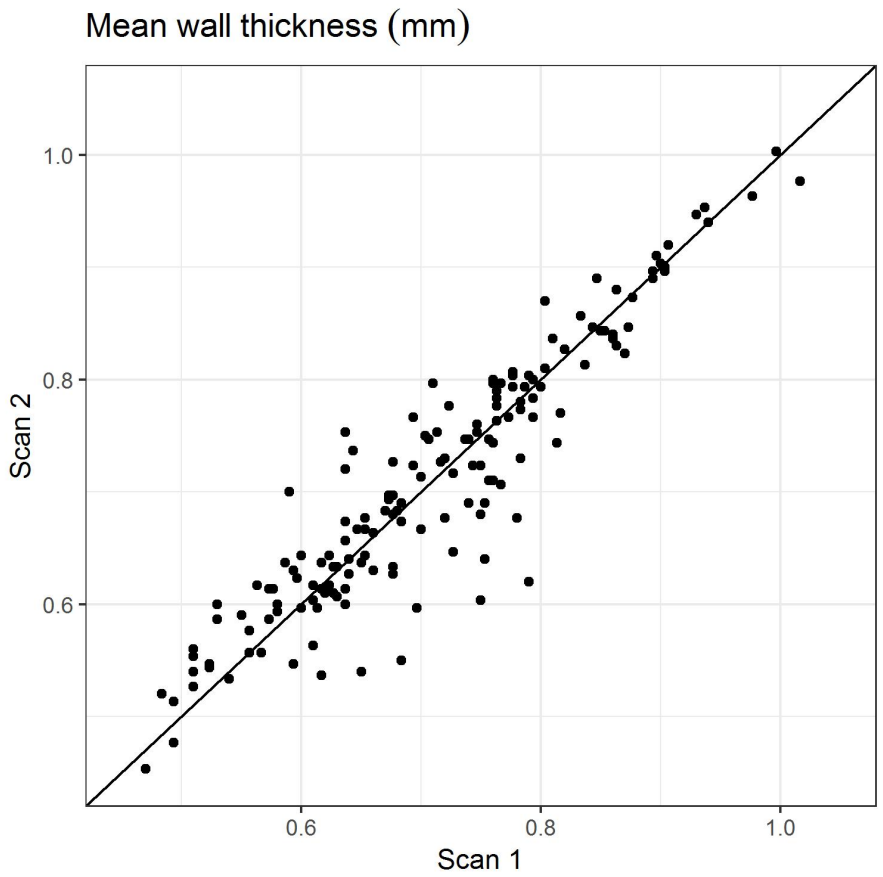


Figure: Pearson correlation = 1 and Intra-class correlation = 0.343

- Pearson correlation measures how strongly pairs of variables are linear related;
- Intra-class correlation measures how strongly pairs of measures in the same sample unit relate to each other.

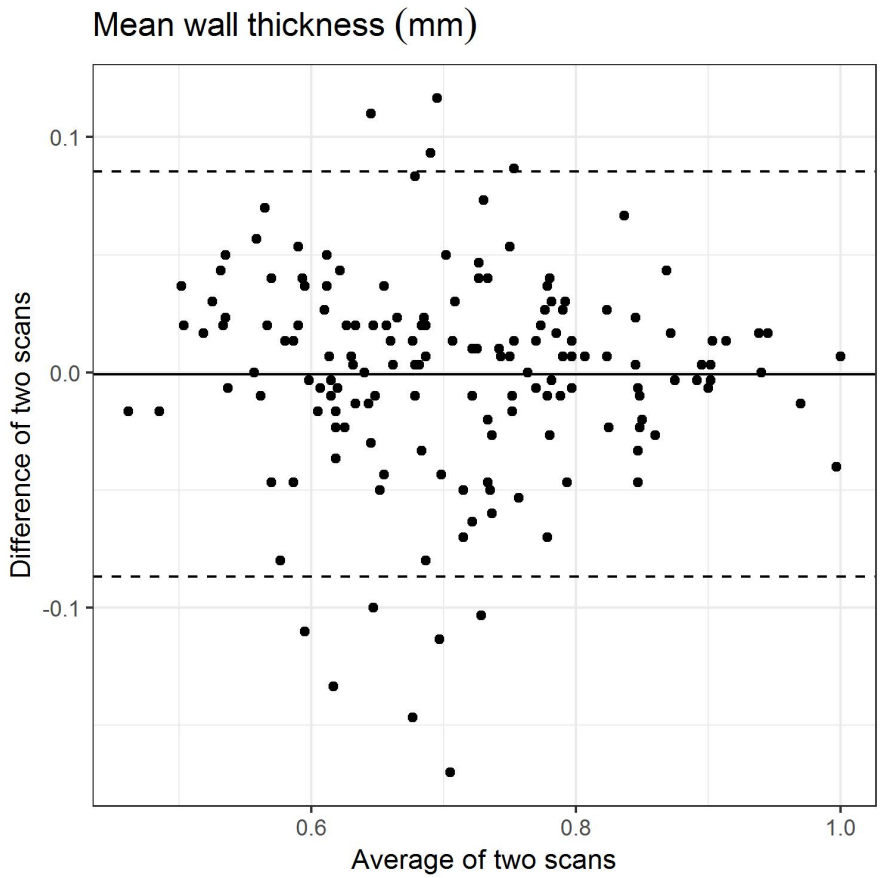
Lin's concordance correlation coefficient



- It ranges from -1 to 1;
- It yields similar values to ICC;
- It can be applied to ordinal and nominal qualitative variables.

Figure: ICC = 0.9795195 and CCC = 0.9795091

Paired t-test



- Paired t-test is equivalent to a one sample t-test for the difference;
- One sample t-test:
  - H_0: mean difference $= 0$
    vs
  - H_1: mean difference $\neq 0$;
  - p-value $= 0.8387$.

Chen *et al. Journal of Cardiovascular Magnetic Resonance* (2018) 20:42
https://doi.org/10.1186/s12968-018-0459-6

Journal of Cardiovascular Magnetic Resonance

RESEARCH Open Access

Cardiovascular magnetic resonance black-blood thrombus imaging for the diagnosis of acute deep vein thrombosis at 1.5 Tesla

Hanwei Chen[1,2†], Xueping He[1,2†], Guoxi Xie[3,4*], Jianke Liang[1], Yufeng Ye[1], Wei Deng[1], Zhuonan He[1], Dexiang Liu[1], Debiao Li[5], Xin Liu[6] and Zhaoyang Fan[5]

Acute deep vein thrombosis

Summary

- Cardiovascular magnetic resonance (CMR) black-blood thrombus imaging (BBTI) technique;
- Diagnosis of acute deep vein thrombosis (DVT);
- 15 healthy subjects and 30 acute DVT patients;
- Two blinded and independent readers;
- Contrast-enhanced CMR (CE-CMRV) as reference (gold standard).

Acute deep vein thrombosis

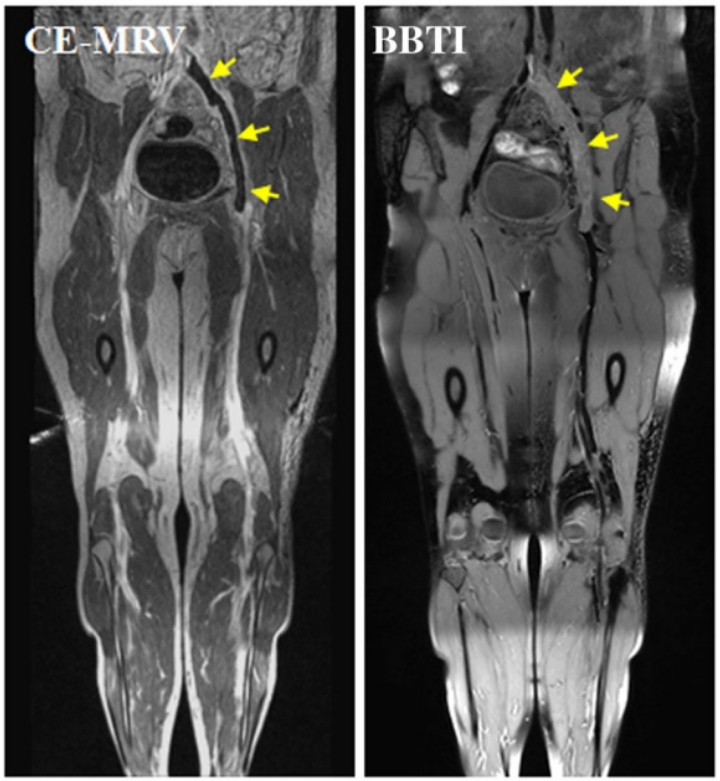


Fig. 1 Representative images obtained by contrast-enhanced cardiovascular magnetic resonance (CE-CMRV) and black blood thrombus imaging (BBTI) from a 55-year-old woman with deep venous thrombosis (DVT) symptom onset at 5 days. The thrombus detected by BBTI showed iso-intense signals within the black-blood venous lumen. The locations and sizes of the thrombi between BBTI and CE-CMRV matched (yellow arrows)

- BBTI: positive or negative;
- CE-CMRV: positive or negative;
- What are the causes of variability?

Acute deep vein thrombosis

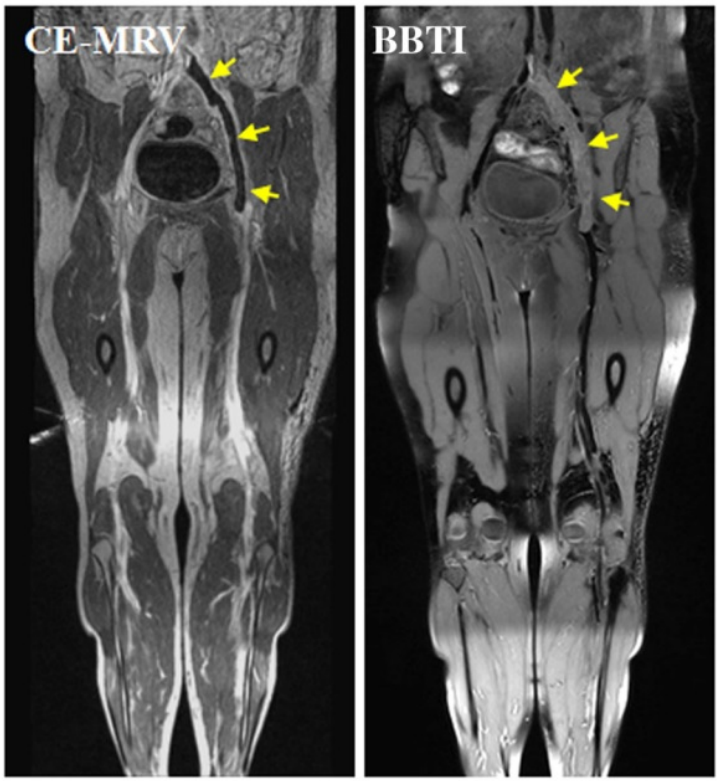


Fig. 1 Representative images obtained by contrast-enhanced cardiovascular magnetic resonance (CE-CMRV) and black blood thrombus imaging (BBTI) from a 55-year-old woman with deep venous thrombosis (DVT) symptom onset at 5 days. The thrombus detected by BBTI showed iso-intense signals within the black-blood venous lumen. The locations and sizes of the thrombi between BBTI and CE-CMRV matched (yellow arrows)

- BBTI: positive or negative;
- CE-CMRV: positive or negative;
- What are the causes of variability?
  - Technique: Inter-technique;

Acute deep vein thrombosis

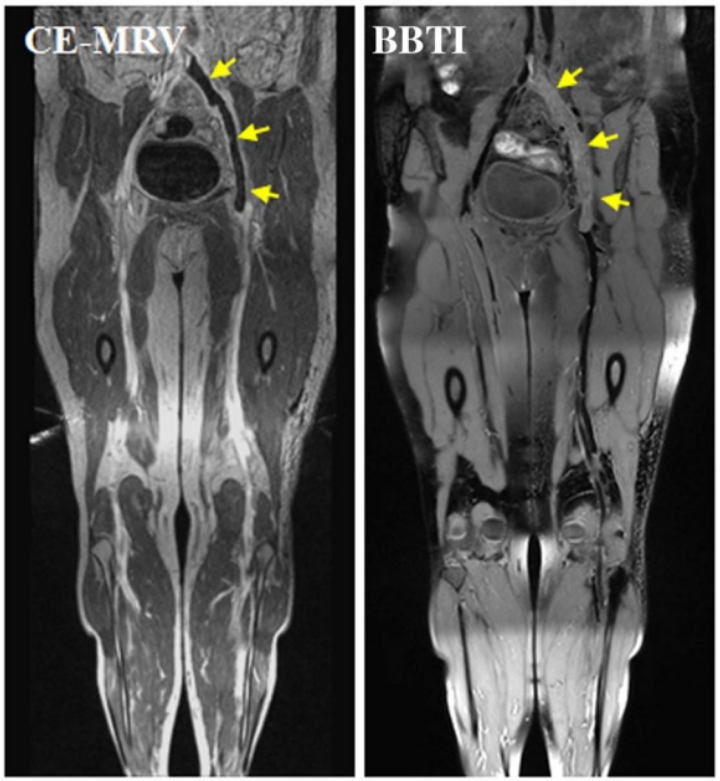


Fig. 1 Representative images obtained by contrast-enhanced cardiovascular magnetic resonance (CE-CMRV) and black blood thrombus imaging (BBTI) from a 55-year-old woman with deep venous thrombosis (DVT) symptom onset at 5 days. The thrombus detected by BBTI showed iso-intense signals within the black-blood venous lumen. The locations and sizes of the thrombi between BBTI and CE-CMRV matched (yellow arrows)

- BBTI: positive or negative;
- CE-CMRV: positive or negative;
- What are the causes of variability?
  - Technique: Inter-technique;
  - Reader: Inter-observer, intra-observer.

Kappa-statistic

- It measures the agreement between readers that is not because of pure chance;
- $\kappa = 1 - \frac{1 - p_o}{1 - p_e}$;
- p_o is the observed agreement;
- p_e is the agreement due to pure chance.

Kappa-statistic

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

- $p_o = \frac{43 + 14}{60} = 0.95;$

Kappa-statistic

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

- $p_o = \frac{43 + 14}{60} = 0.95;$

Reader 2	Reader 1		
	Negative	Positive	Total
Negative			0.76
Positive			0.24
Total	0.71	0.29	1

Table: Observed marginal proportions BBTI Reading - Tibiofibular trunk vein

Kappa-statistic

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

- $p_o = \frac{43 + 14}{60} = 0.95;$

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	0.5396	0.2204	0.76
Positive	0.1704	0.0696	0.24
Total	0.71	0.29	1

Table: Expected proportions BBTI Reading assuming pure chance - Tibiofibular trunk vein

Kappa-statistic

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	32.376	13.224	46
Positive	10.224	4.176	14
Total	43	17	60

Table: Expected BBTI Reading - Tibiofibular trunk vein

- $p_o = \frac{43 + 14}{60} = 0.95;$
- $p_e = \frac{32.376 + 4.176}{60} = 0.6092$

Kappa-statistic

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	32.376	13.224	46
Positive	10.224	4.176	14
Total	43	17	60

Table: Expected BBTI Reading - Tibiofibular trunk vein

- $p_o = \frac{43 + 14}{60} = 0.95;$
- $p_e = \frac{32.376 + 4.176}{60} = 0.6092$
- $\kappa = 1 - \frac{1 - 0.95}{1 - 0.6092} = 0.87$

Kappa-statistic

What is a meaningful value of Kappa?

Kappa	Agreement
0 - 0.2	None
0.21 - 0.4	Slight
0.41 - 0.6	Moderate
0.61 - 0.8	Substantial
0.81 - 1.0	Almost Perfect

Table: Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics. 33 (1): 159-174. doi:10.2307/2529310. JSTOR 2529310. PMID 843571

Kappa-statistic

What is a meaningful value of Kappa?

Kappa	Agreement
0 - 0.4	Poor
0.41 - 0.75	Fair to good
0.75 - 1	Excellent

Table: Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley. ISBN 0-471-26370-2.

Test of Hypothesis

- $H_0 : \kappa = 0$ (Disagreement) vs $H_1 : \kappa > 0$ (Agreement)

Test of Hypothesis

- $H_0 : \kappa = 0$ (Disagreement) vs $H_1 : \kappa > 0$ (Agreement)
- p-value < 0.001

Test of Hypothesis

- $H_0 : \kappa = 0$ (Disagreement) vs $H_1 : \kappa > 0$ (Agreement)
- p-value < 0.001
- 95% CI: [0.72 ; 1]

Kappa-statistic

What is a meaningful value of Kappa?

Test of Hypothesis

- $H_0 : \kappa = 0$ (Disagreement) vs $H_1 : \kappa > 0$ (Agreement)
- p-value < 0.001
- 95% CI: [0.72 ; 1]

Kappa < 0

- It is possible when the agreement in the observed data is lower than the agreement due pure chance;
- It should be considered as zero.

McNemar test

- $H_0 : p_{1.} = p_{.1}$ and $p_{2.} = p_{.2}$ (Agreement);
- $H_1 : p_{1.} \neq p_{.1}$ or $p_{2.} \neq p_{.2}$ (Disagreement);
- The rejection of null hypothesis indicates disagreement, which it is the opposite of Kappa.

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	p_{11}	p_{12}	$p_{1.}$
Positive	p_{21}	p_{22}	$p_{2.}$
Total	$p_{.1}$	$p_{.2}$	1

Table: Proportion table

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	43	3	46
Positive	0	14	14
Total	43	17	60

Table: Observed BBTI Reading - Tibiofibular trunk vein

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	0.71	0.05	0.76
Positive	0	0.24	0.24
Total	0.71	0.29	1

Table: Observed proportions BBTI Reading - Tibiofibular trunk vein

Reader 2	Reader 1		
	Negative	Positive	Total
Negative	0.71	0.05	0.76
Positive	0	0.24	0.24
Total	0.71	0.29	1

Table: p-value = 0.2482

Diagnostic Measures

- Statistical measures of performance for a binary diagnostic test assuming that there is a gold standard or reference;
- Sensitivity (SE) and Specificity (SP) characterize the test for any prevalence;
- Positive Predictive Value (PPV) and Negative Predictive Value (NPV) are specific for a given prevalence.

Diagnostic Measures

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	43	3	46
Positive	1	13	14
Total	44	16	60

Table: Observed BBTI Reading for Reader 1 and CE-CMRV - Tibiofibular trunk vein

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	n_{11}	n_{12}	$n_{1.}$
Positive	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Table: Theoretical counts

- Sensitivity = True Positive:

$$\begin{aligned} SE &= P(\text{BBTI} + | \text{CE-CMRV} +) \\ &= \frac{n_{22}}{n_{.2}} \\ &= \frac{13}{16} \\ &= 0.81 \end{aligned}$$

Diagnostic Measures

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	43	3	46
Positive	1	13	14
Total	44	16	60

Table: Observed BBTI Reading for Reader 1 and CE-CMRV - Tibiofibular trunk vein

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	n_{11}	n_{12}	$n_{1.}$
Positive	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Table: Theoretical counts

- Specificity = True Negative:

$$\begin{aligned} SP &= P(\text{BBTI -}|\text{CE-CMRV -}) \\ &= \frac{n_{11}}{n_{.1}} \\ &= \frac{43}{44} \\ &= 0.98 \end{aligned}$$

Diagnostic Measures

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	43	3	46
Positive	1	13	14
Total	44	16	60

Table: Observed BBTI Reading for Reader 1 and CE-CMRV - Tibiofibular trunk vein

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	n_{11}	n_{12}	$n_{1.}$
Positive	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Table: Theoretical counts

- Positive Predictive Value:

$$\begin{aligned} PPV &= P(\text{CE-CMRV+}|\text{BBTI+}) \\ &= \frac{n_{22}}{n_{2.}} \\ &= \frac{13}{14} \\ &= 0.92 \end{aligned}$$

- Prevalence = 26.6%

Diagnostic Measures

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	43	3	46
Positive	1	13	14
Total	44	16	60

Table: Observed BBTI Reading for Reader 1 and CE-CMRV - Tibiofibular trunk vein

BBTI	CE-CMRV		
	Negative	Positive	Total
Negative	n_{11}	n_{12}	$n_{1.}$
Positive	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Table: Theoretical counts

- Negative Predictive Value:

$$\begin{aligned} NPV &= P(\text{CE-CMRV-}|\text{BBTI-}) \\ &= \frac{n_{11}}{n_{1.}} \\ &= \frac{43}{46} \\ &= 0.93 \end{aligned}$$

- Prevalence = 26.6%

Diagnostic Measures

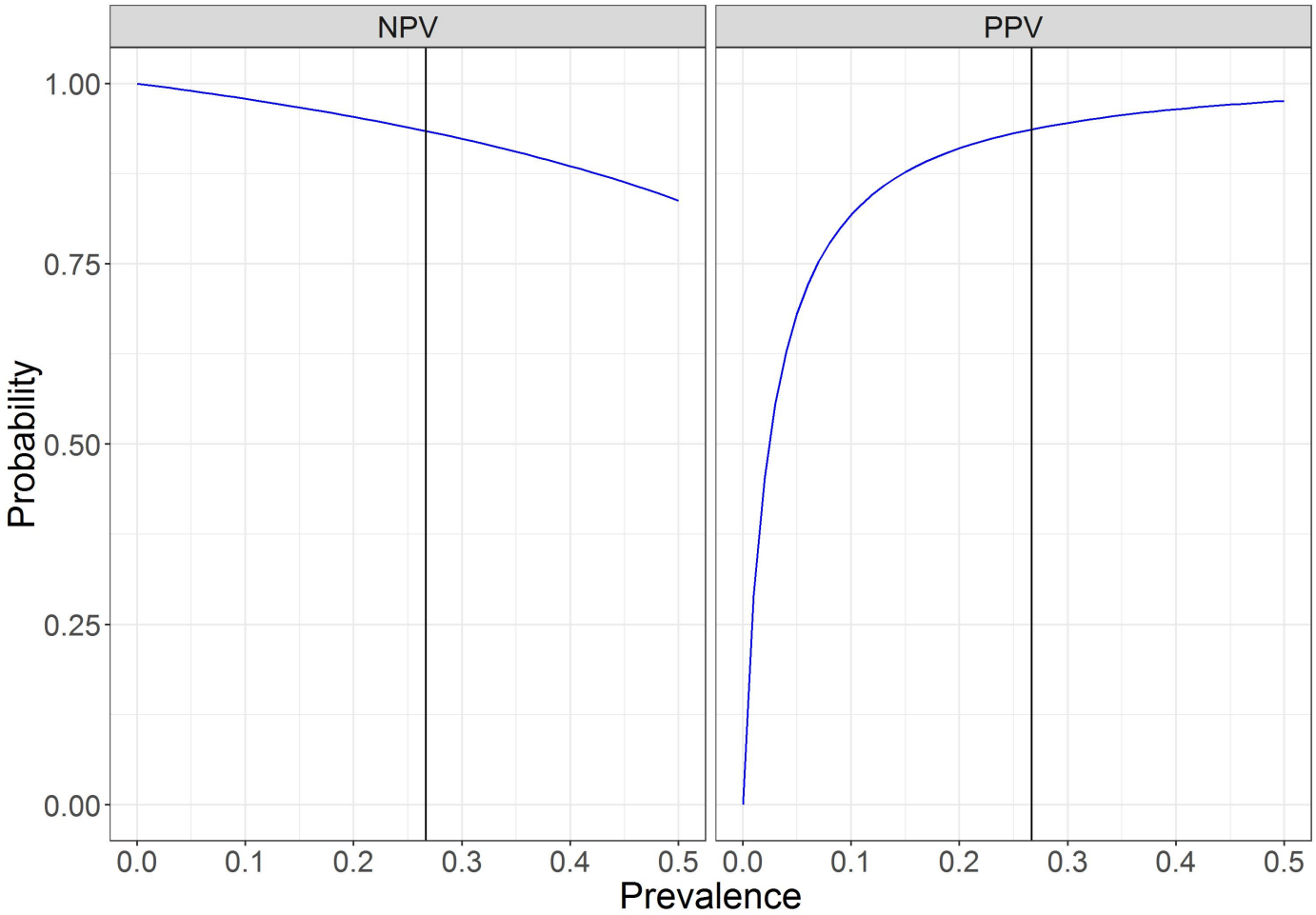


Figure: PPV and NPV values for SE = 0.81 and SP = 0.98

Diagnostic Measures

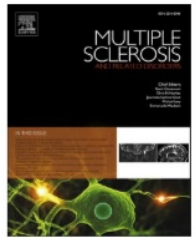
95% CI

- Sensitivity: 81% 95%CI (54% ; 96%)
- Specificity: 98% 95%CI (88% ; 100%)
- PPV: 92% 95%CI (66% ; 100%) for prevalence = 26%;
- NPV: 93% 95%CI (82% ; 99%) for prevalence = 26%;

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

journal homepage: www.elsevier.com/locate/msard



Original article

Preventing multiple sclerosis misdiagnosis using the "central vein sign": A real-world study

Marwa Kaisey [a,*], Andrew J. Solomon [b], Brooke L. Guerrero [a], Brian Renner [a], Zhaoyang Fan [c], Natalie Ayala [a], Michael Luu [d], Marcio A. Diniz [d], Pascal Sati [a], Nancy L. Sicotte [a]

[a] Cedars-Sinai Medical Center Department of Neurology, 127 S. San Vicente Blvd, Suite A6600, Los Angeles, CA 90048, USA
[b] Larner College of Medicine at the University of Vermont, Department of Neurological Sciences, 1 South Prospect Street, Arnold, Level 2, Burlington, Vermont 05401, USA
[c] Cedars-Sinai Biomedical Imaging Research Institute, 116 N Robertson Blvd, Los Angeles, CA 90048, USA
[d] Cedars-Sinai Biostatistics and Bioinformatics Research Center, 8700 Beverly Blvd North Tower, Los Angeles, CA 90048, USA

Multiple Sclerosis - Central Vein Sign

Summary

- Misdiagnosis of multiple sclerosis (MS) is common and often occurs due to misattribution of non-MS magnetic resonance imaging (MRI) lesion;
- A new MRI biomarker, the central vein sign, has been demonstrated high specificity for MS lesions and may prevent misdiagnosis;
- 15 non-MS and 15 MS patients;
- Two blinded and independent readers;
- Goal: Identify a cutoff that discriminate between MS and non-MS patients.

Multiple Sclerosis - Central Vein Sign

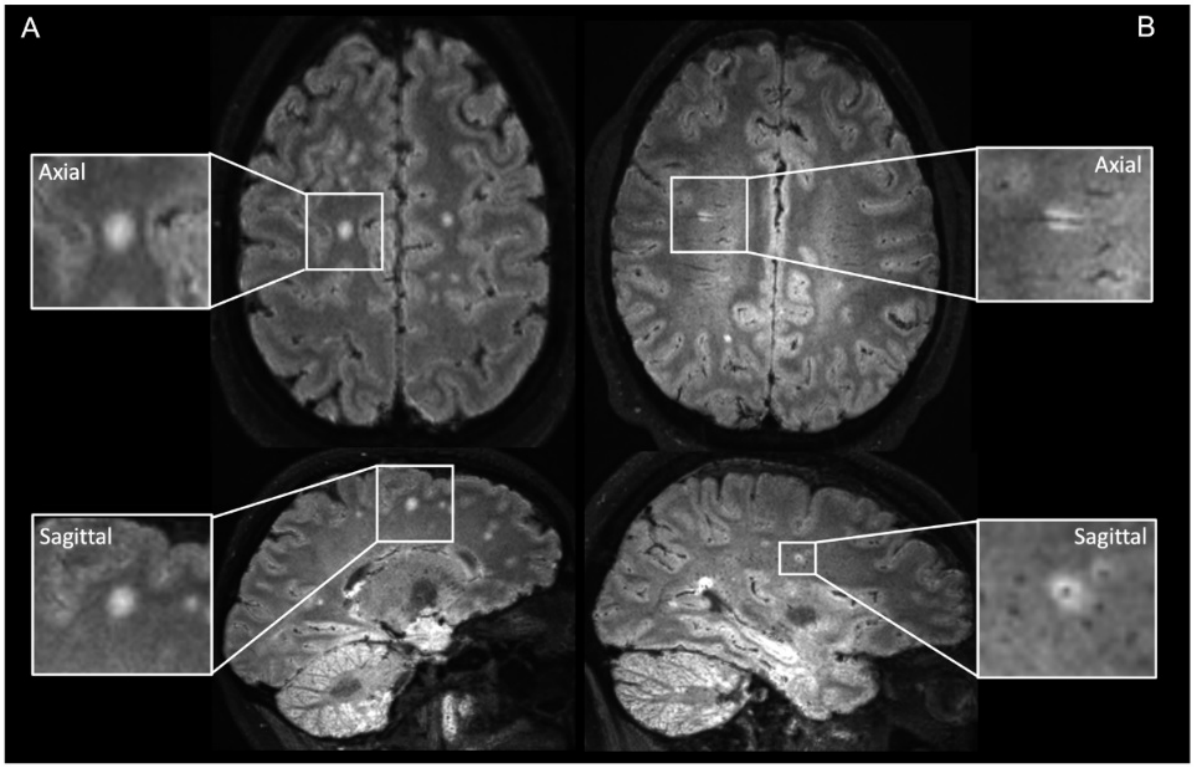
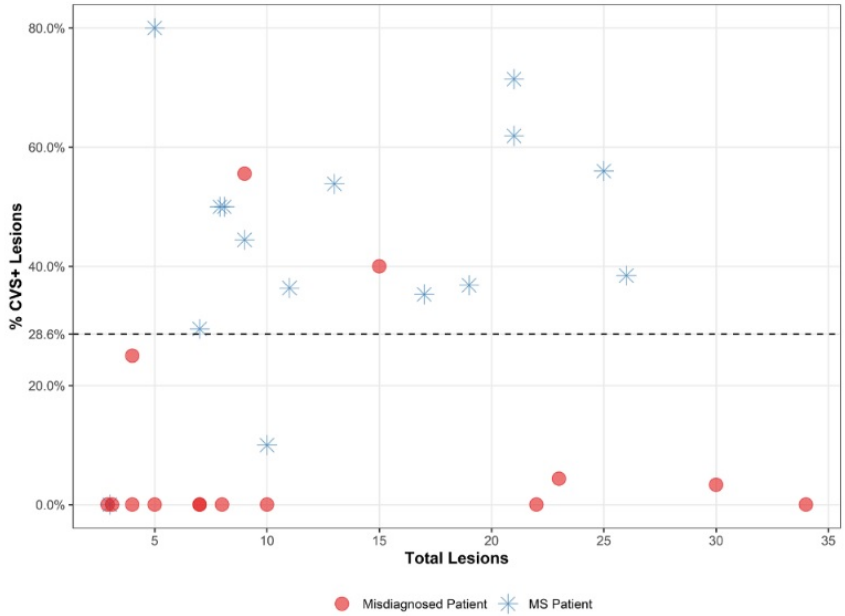


Figure 1. (Larger view file uploaded separately): Axial and sagittal views of 3T FLAIR* brain MRI in two patients. A) Non-inflammatory lesion without CVS in a patient misdiagnosed with MS (final diagnosis: migraine and cervical disc degeneration). B) Inflammatory demyelinating lesion with CVS in a patient with MS. The hypointense vein running centrally through the hyperintense focal lesion can be seen in both views. CVS: central vein sign.

Multiple Sclerosis - Central Vein Sign



- For cutoff = 28.6%,
  - Sensitivity = 86%;
  - Specificity = 86%.

Figure: How do we choose the cut-off that best discriminate between MS and non-MS patients?

Multiple Sclerosis - Central Vein Sign

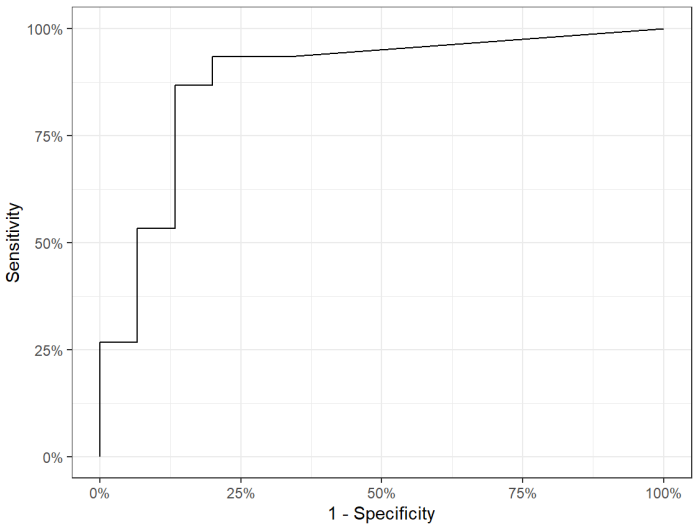


Figure: AUC: 0.88 95%(0.746 ; 1)

- We can calculate Sensitivity and Specificity for every possible cutoff;
- Then, we can plot the Receive Operating Characteristic (ROC) Curve;
- The Area Under the Curve (AUC) indicates the discrimination ability of a biomarker.

Multiple Sclerosis - Central Vein Sign

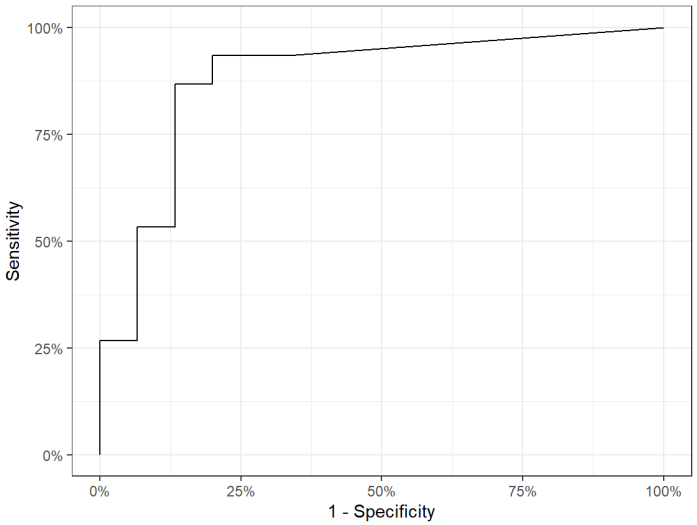


Figure: AUC: 0.88 95%CI (0.746 ; 1)

- One of most common criterion to identify a cutoff based on ROC Curve is maximizing the Youden Index = Specificity + Sensitivity - 1.

Multiple Sclerosis - Central Vein Sign

- When we calculate sensitivity and specificity in the same sample used to identify cut-off, the estimates for sensitivity and specificity will be optimistic;
- Therefore, validation studies should be performed;
- When external validation samples are not available, bootstrap or cross-validation methods can be applied.
  - Sensitivity = 77% 95%CI (39% - 96%);
  - Specificity = 82% 95%CI (36% - 93%).

References

- Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ: British Medical Journal. 1994 Jun 11;308(6943):1552.
- Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. Bmj. 1994 Jul 9;309(6947):102.
- Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. Theriogenology. 2010 Jun 1;73(9):1167-79.