Multi-omics in research: epidemiology, methodology, and advanced data analysis

Faquih, T.O.

Part I
Methodological Challenges in Proteomics and Metabolomics

Chapter 2

Agreement of aptamer proteomics with standard methods for measuring venous thrombosis biomarkers

Authors: Tariq Faquih, M.S.[1], Dennis Mook-Kanamori, MD PhD.[2,3], Frits Rosendaal, MD PhD.[1], Trevor Baglin, MD PhD.[4], Ko Willems van Dijk, MD PhD.[5,6,7], Astrid van Hylckama Vlieg, PhD.[2]

1. Department of Clinical Epidemiology, Leiden University Medical Center, Postal Zone C7-P, PO Box 9600, 2300 RC, Leiden, The Netherlands. Email: t.o.faquih@lumc.nl, Phone Number: +31 626 696 712
2. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.
3. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands.
4. Medicxi Ventures LLP, 25 Great Pulteney Street, W1F 9LT, London, UK
5. Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands.
6. Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands.
7. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Authors' contributions:
T. Faquih designed the study, analysed and drafted the manuscript. A. van Hylckama Vlieg conceived and designed the study, acquired the data, obtained funding, supervised the study. K. Willems van Dijk and D. O. Mook-Kanamori conceived, supervised, and designed the study. T. P. Baglin acquired the data supervised the study. F. R. Rosendaal supervised the study. All authors reviewed the manuscript.

Acknowledgements:
The authors wish to thank the director of the anticoagulation clinic of Leiden, the Netherlands (F. J. M. van der Meer), who made the recruitment of patients in Leiden possible. L. Willems of Brilman is thanked for performing interviews and blood draws. I. de Jonge is thanked for her administrative support and data management. We are grateful to P. Noordijk, L. Mahic and A. Hoenderdos for sample processing and laboratory analysis.

Conflict of Interests:
T. Baglin has received honoraria for consultancy from Organon Teknika and bioMerieux. D. O. Mook-Kanamori is a part-time clinical research consultant for Metabolon, Inc. All other authors have nothing to disclose.
Summary word count = 220/250; Body word count = 2293/2000; references = 24/30; 1 Table; 1 Figure
Keywords: Venous Thrombosis; Proteomics; Biomarkers; Aptamers, Nucleotide; Immunoassay; Blood Coagulation Factors

Res Pract Thromb Haemost. **2021.** 5:e12526 doi:10.1002/rth2.12526, PMID: 34013156

Essentials

- Measurement agreement of aptamer proteomics for venous thromboembolism (VTE) markers is unknown.
- We selected 27 cases with unprovoked VTE and 27 controls from the THE-VTE study.
- Agreement between the aptamers and the laboratory methods for the VTE biomarkers was poor.
- Currently the usage of aptamer proteomics for VTE biomarkers should be considered with caution.

1 ABSTRACT

Background

Venous thromboembolism (VTE) is a complex disease with an incidence rate of about 1/1000 per year. Despite the availability of validated biomarkers for VTE, unprovoked events account for 50% of first events. Therefore, emerging high-throughput proteomics are promising methods for the expansion of VTE biomarkers. One such promising high-throughput platform is SomaScan, which utilizes a large library of synthetic oligonucleotide ligands known as aptamers to measure thousands of proteins.

Objective

The aim of this study was to evaluate the viability of the aptamer-based SomaScan platform for VTE studies by examining its agreement with standard laboratory methods.

Methods

We examined the agreement between eight established VTE biomarkers measured by SomaScan and standard laboratory immunoassay and viscosity-based instruments in 54 individuals (27 cases and 27 controls) from the THE-VTE study. We preformed the agreement analysis by using a regression model and predicting the estimates and the 95% prediction interval (PI) of the laboratory instruments values using SomaScan values.

Results

SomaScan measurements exhibited overall poor agreement, particularly for D-dimer (average fit [95% PI]: 492.7 ng/mL [110.0-1998.2]) and fibrinogen (average fit [95% PI]: 3.3g/L [2.0-4.7]).

Conclusion

Our results indicate that SomaScan measurement had poor agreement with the standard laboratory measurements. These results may explain why some genome wide association studies with VTE proteins measured by SomaScan did not confirm previously identified loci. Therefore, SomaScan should be considered with caution in VTE studies.

2 INTRODUCTION

Venous thromboembolism (VTE) is a complex disease caused by an imbalance in the coagulation and fibrinolysis pathways. VTE, which encompasses deep vein thrombosis (DVT) and pulmonary embolism (PE), has an incidence rate of 1/1000 per year[1]. Venous thromboembolism is assumed to be caused by both acquired and genetic risk factors[1], but the mechanisms that evoke VTE involve complex interactions of pathways that are still not fully understood[1, 2]. Several genetic variants and proteins and have been identified as risk factors for VTE, such as factor V Leiden (*F5* rs6025), prothrombin 20210A (*FII* rs1799963), low levels of antithrombin, protein S and protein C, and high levels of factor VIII, IX, and XI[1].

The current standard methods to quantify coagulation factors are by viscosity or optical detection based coagulation analysers and immunoassay-based laboratory instruments targeting one or more coagulation factors. For example, elevated levels of D-Dimer are associated with increased risk of DVT, recurrent DVT and mortality[3]. Moreover, D-dimer is measured with >95% diagnostic sensitivity by immunoassay-based instruments and hence broadly used by clinicians for the exclusion of VTE in patients with low or intermediate risk[4].

Further expansion of the number of biomarkers is imperative for studying the aetiology and improving prediction of VTE. Emerging high throughput proteomic platforms are promising tools to identify such novel biomarkers as these platforms are capable of quantifying large numbers of proteins simultaneously from a single sample[5, 6]. The aim of our study was to assess the measurement agreement of one such platform, the aptamer-based SomaScan platform, by comparing its measurements with the current established laboratory methods for eight VTE biomarkers in THE-VTE study.

2.1 SomaScan

SomaScan (SomaLogic, Inc. Boulder, CO, USA) is a high throughput proteomics platform capable of simultaneous measurement of thousands of proteins. Unlike traditional immunoassay instruments, SomaScan utilizes Systematic Evolution of Ligands by Exponential enrichment (SELEX), a biochemical technique used to create a library with a wide range of modified synthetic oligonucleotide ligands (known as aptamers) designed to bind to their respective protein targets. Aptamers provide several benefits over immunoassay methods; They are inexpensive to produce, highly modifiable and are chemically stable. SomaScan has developed a vast library of unique aptamers (SOMAmers) to detect thousands of proteins[7]. This makes the SomaScan platform appealing for researchers and, indeed, several large studies have utilized the platform for various purposes including the identification novel protein markers[8-10].

3 METHODS

3.1 THE-VTE study

For our cases and controls, we used samples from the Thrombophilia, Hypercoagulability and Environmental Risks in Venous Thromboembolism (THE-VTE) study — a multicentre case control study from Leiden (The Netherlands) and Cambridge (UK)[11]. Inclusion took place between March 2003 and December 2008. In total, 626 patients were included, aged 18-75, with a first DVT or PE. Partners of the patients were invited as controls. Subsequent follow-up of the cases was performed to assess recurrence risk. The mean follow-up duration was 4.8 years after discontinuation of oral anticoagulant therapy. Blood samples were taken 2-3 months after discontinuation of anticoagulants. The blood samples were collected into Sarstedt Monovette tubes, in a 0.1 volume of 0.106 m trisodium citrate and separated into plasma by centrifugation then stored at -80 °C. All participants gave written informed consent. The study was approved by the Medical Ethics Committee of the Leiden University Medical Centre (Leiden, Netherlands) and the NHS Research Ethics Committee in Cambridge, UK.

The current pilot study was originally designed to explore the biomarker measurements differences between unprovoked VTE and controls in a case-control design using SomaScan. Before proceeding, we checked the general measurement agreement between SomaScan and standard laboratory immunoassay and viscosity-based instruments in both cases and controls to assess the agreement over the whole range of coagulation factors levels. Unprovoked events were defined as individuals who did not have surgery, trauma or long-term immobilisation 3 months prior to the event. Moreover, patients were excluded if they had an active malignancy, abnormal levels of proteins C, protein S and antithrombin (<80 U/dL), used hormone replacement therapy or hormonal contraceptives at the time of the event. Patients with factor V Leiden or prothrombin (PT20210A) mutations were also excluded. We selected a sample of 16 cases with unprovoked VTE and further added eleven patients who experienced a recurrent venous thrombosis during follow-up, resulting in the inclusion of 27 VTE cases. By including unprovoked VTE cases as well as controls we covered a wide range of VTE biomarker values. Finally, we randomly selected 27 participants without VTE as controls. The frozen samples of the selected cases and controls were sent and thawed for analysis by SomaScan in 2016. No thawing or refreezing was performed during the interim period between 2011-2016. VTE biomarkers were measured by validated immunoassay-based, viscosity-based detection instruments, henceforth referred to as laboratory instruments. D-dimer total concertation (ng/mL) was measured by the Vidas D-dimer immunoassay (BioMérieux, Basingstoke, UK). The activity (international units per millilitre; IU/ml) of protein C, protein S, antithrombin (using chromogenic assay), prothrombin, coagulation factor IX, and coagulation factor XI were measured by STA-R coagulation analyser (Diagnostica Stago, S.A.S, Asnières sur Seine, France). Fibrinogen total concertation (g/L) was measured by STA-R coagulation analyser (Diagnostica Stago, S.A.S, Asnières sur Seine, France)[11].

Samples were sent to SomaLogic (Boulder, CO, USA) and measured by the SomaScan platform. The instrument measured 1310 total proteins of which 24 proteins failed the quality check and were flagged. We selected eight VTE biomarkers that were measured by laboratory instruments and successfully measured by SomaScan: D-dimer, prothrombin, protein C, protein S, antithrombin, fibrinogen, coagulation Factors IX, and XI. One control sample failed quality control was excluded.

To compare the agreement and interchangeability of the different measures we used the 95% agreement statistical method[12]. Since SomaScan uses relative fluoresces units (RFU) as measures for protein concentration, and the laboratory instruments measure absolute protein

concentrations or activity (IU/ml), we applied an alternative method to assess agreement if measurements are on different units, as described by Bland & Altman[12]. First, we performed a linear regression per biomarker with laboratory instruments measures as the outcome and SomaScan measures as the independent variable. Second, we used the regression models to predict estimates of the laboratory instruments values using SomaScan values. After checking the normality of the residuals, we log transformed the D-Dimer measurements as their distributions were very skewed. Finally, we calculated the 95% prediction interval (PI) to represent the equivalent of 95% limits of agreement[12]. This method is equivalent to comparing the mean differences of the two measurement methods. If the bias is consistent and the mean difference is close to 0 the result would show narrow prediction intervals. Consequently, the two methods would be interchangeable and in good agreement[13]. It is difficult to define hard cut-off points for the intervals. Therefore, judging the agreement is considered a clinical question rather than a statistical one[14].

4 DISCUSSION AND RESULTS

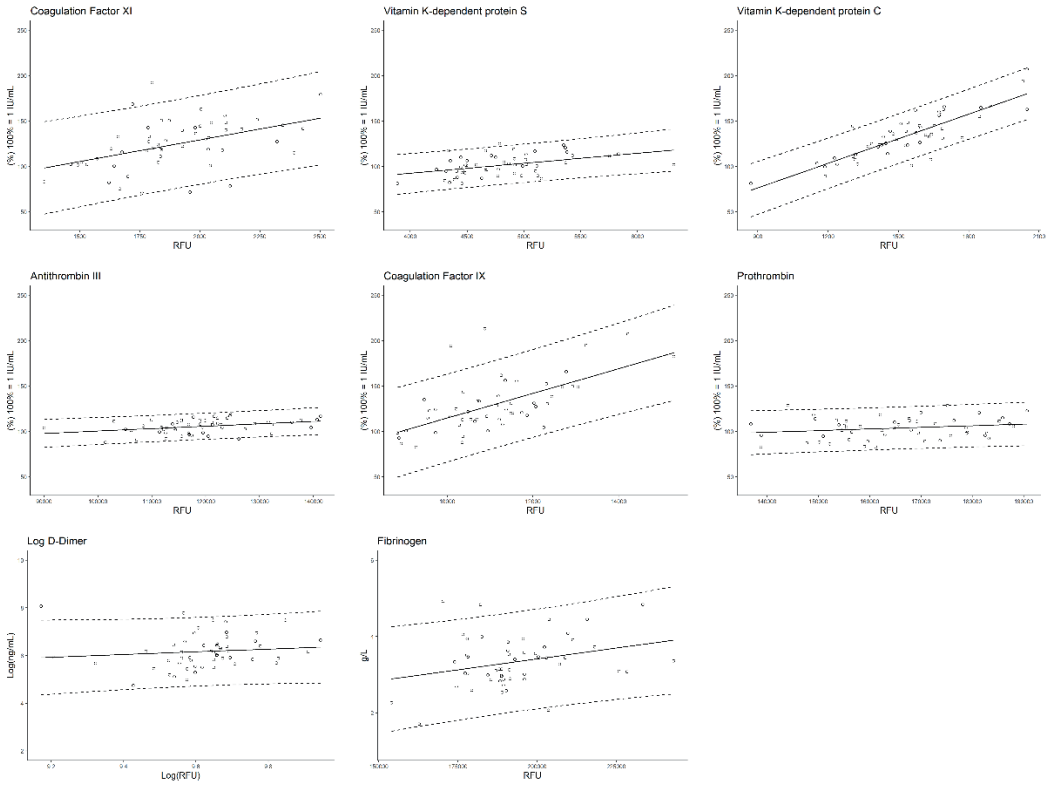
We examined the agreement of SomaScan measurements of VTE biomarkers with the laboratory instruments. The table and figures for the results are shown in Table 1 and Figure 1. Although a particular biomarker may seem to be in good agreement due to oblique slopes, the appropriate indicator for agreement is the width of the prediction interval around the average fit of the regression line[12].

Table 1: Average fit and average prediction intervals for SomaScan and laboratory instruments for the coagulation factors.

Protein Name	Units	Average Fit [Average 95% PIs]	Average Width
Coagulation Factor XI	(%) 100% = 1 IU/mL	124.3 [75.0-173.5]	98.5
Protein S	(%) 100% = 1 IU/mL	101.4 [80.2-122.7]	42.6
Protein C	(%) 100% = 1 IU/mL	133.9 [106.4-161.4]	55.1
Antithrombin	(%) 100% = 1 IU/mL	105.5 [91.0-120.1]	29.1
Coagulation Factor IX	(%) 100% = 1 IU/mL	129.6 [81.0-178.2]	97.2
Prothrombin	(%) 100% = 1 IU/mL	103.7 [80.3-127.1]	46.9
D-Dimer			
Log	Log(ng/mL*)	6.2 [4.7-7.6]	2.9
Back-transformed	ng/mL*	492.7 [110.0-1998.2]	1888.2
Fibrinogen	g/L	3.3 [2.0-4.7]	2.6

Activity of coagulation factor XI, protein S, protein C, antithrombin, coagulation factor IX, and prothrombin were measured by the same instrument and use IU/mL units. D-dimer and fibrinogen total concentrations were measured by immunoassay instruments. *D-dimer was assayed using the Vidas D-dimer assay. Unit type used was FEU = Fibrinogen equivalent unit (500 ng FEU/mL = 250 ng D-dimer/mL). Abbreviations: PIs: Prediction Interval; IU/mL: international units per millilitre; Average Width: the average difference between the lower and upper limits of the prediction interval.

Figure 1: 95% Limits of Agreements plots for each VTE biomarker.



Narrow prediction intervals indicate higher agreement between the SomaScan (x-axis) and laboratory instruments (y-axis). Antithrombin had the narrowest interval and, therefore, the best agreement. D-dimer had the poorest agreement as indicated by the wide interval in the log transformed plot. Abbreviations: RFU: relative fluoresces units; IU/mL: international units per millilitre.

Overall, the results indicate poor agreement of SomaScan with the validated laboratory instruments. The narrowest prediction interval, and thus the best agreement, were observed for antithrombin (average fit [95% PI]: 105.52 IU/mL [90.98-120.06]) followed by protein S (average fit [95% PI]: 101.44 IU/mL [80.15-122.73]), Prothrombin (average fit [95% PI]: 103.7 IU/mL [80.27-127.13]) and protein C (average fit: [95% PI] 133.87 IU/mL [106.35-161.4]). Factor IX (average fit [95% PI]: 129.62 IU/mL [81.03-178.21]) and factor XI (average fit [95% PI]: 124.26 IU/mL [75-173.53]) had a wide mean prediction interval and low agreement. The prediction interval for fibrinogen also had a wide interval (average fit [95% PI]: 3.3 g/L [2.0-4.7]). This result indicates that the predicted value of fibrinogen was within a prediction range (average width) of 2.6 g/L (~80%) of the laboratory measurement. Considering the wide prediction interval width and the fact that the normal range of fibrinogen is between 2 and 5 g/L[15], we concluded that the SomaScan measurements of fibrinogen are in poor agreement with the laboratory instrument. Finally, agreement between SomaScan and laboratory instruments for D-dimer had the widest average interval (average fit [95% PI]: 6.2 [4.7-7.6]) among the measured markers as shown in Figure 1. The values and width of the interval for the log D-dimer plot may seem normal compared

to the plots of the other measurements. However, unlike the other biomarkers, D-dimer was measured in ng/mL units and was log transformed in order to fulfil the requirement of normality for the analysis. After back-transforming the values, the agreement was very poor as indicated by the extremely wide prediction interval (average fit [95% PI]: 492.7 ng/mL [110.0-1998.2]).

The current study demonstrated poor agreement of SomaScan VTE measures with the laboratory instruments, particularly for D-dimer, which is a particularly important VTE biomarker[1, 4, 6], despite passing quality control. This poor agreement could explain the reported lack of association between SomaScan D-dimer measurement and the risk of DVT[16].

Despite the advantages of SomaScan for high throughput measurements, the observed disagreement could be due to some of the platform's shortcomings[17]. Some factors that can affect binding affinity are aptamer cross reactivity, genetic variations, post-translational modifications, and the complexity and stability of the target protein structure. Moreover, SomaScan measurements are quantitative and not qualitative and would not be able to detect qualitative defects in the analysis. It is important to note that our assessment was only of the eight VTE biomarkers available in our study and cannot be generalized to the agreement of the remaining proteins measured by SomaScan for our dataset.

Two previous studies assessed the association between SomaScan measurements and VTE SNPs in the SomaScan protein genome wide association study (pGWAS)[8, 10]. Since several genetic loci associated with VTE biomarkers have previously been identified, multiple hits were expected. However only factor XI and protein C measured by SomaScan associated with three loci and one locus, respectively. The lack of genetic correlations with the SomaScan measures for the biomarkers further supports our findings of poor agreement.

Possible limitations of the study are the usage of activity measurements for biomarkers versus the relative concentration reported by SomaScan. However, both D-dimer and fibrinogen showed poor agreement despite being measured as concentration measures. Moreover, viscosity-based activity measurements, such as the STA-R analyser used here, are considered the standard for VTE studies[18-20]. Furthermore, the recommended the sample size for Bland-Altman methods is usually N>100[21]. Our small size may affect the accuracy of the width of the 95% agreement intervals. However, we found that the agreement is very poor for some of the biomarkers, such as D-dimer, which cannot be fully explained by the sample size. Finally, it is unlikely the storage time of the plasma samples before the SomaScan analysis caused major degradation. Since the blood was collected, the samples were stored in -80 °C and the sampled aliquots were used for the primary analysis. Afterwards the samples were not thawed until the analysis by SomaScan five years later. Several studies have shown that these conditions were optimal for the storage of plasma samples and maintain minimal degradation[22-24]. Therefore, storage time and conditions are an unlikely cause to the disagreement in our results. Nevertheless, comparing the agreement of SomaScan with total concentrations for the other biomarkers and in larger studies may provide further insight.

5 CONCLUSION

The 95% limits agreement is a simple and effective statistical method for comparing measurements by different methods. We believe it is important to apply this type of analysis to compare the measurements of exciting novel high throughput platforms with current established measurements; thereby limiting measurement errors from affecting the results and conclusions based on such platforms.

In conclusion, despite the promising applications of aptamers for proteomics studies, we found that the applied SomaScan platform is not interchangeable with validated laboratory instruments for the VTE markers in our study. Therefore, caution is needed when applying SomaScan measurements for hypothesis driven VTE studies using these markers. Whether this is also true for other biomarkers for VTE remains to be determined. It is clear that more studies of agreement with larger sample size and additional markers are needed.

6 FUNDING

This project has been sponsored by the Leiden University Fund / Nypels van der Zee fonds. THE-VTE study was supported by the Netherlands Organization for Scientific Research (NWO) – grant number 916.56.157 (VENI).]. Tariq Faquih was supported by the King Abdullah Scholarship Program and King Faisal Specialist Hospital & Research Centre [No. 1012879283].

7 REFERENCES

1. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015;12(8):464-74.
2. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. J Thromb Haemost. 2007;5(4):692-9.
3. Halaby R, Popma CJ, Cohen A, Chi G, Zacarkim MR, Romero G, et al. D-Dimer elevation and adverse outcomes. Journal of thrombosis and thrombolysis. 2015;39(1):55-9.
4. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). European Heart Journal. 2019;41(4):543-603.
5. Cristea IM, Gaskell SJ, Whetton AD. Proteomics techniques and their application to hematology. Blood. 2004;103(10):3624-34.
6. Pabinger I, Ay C. Biomarkers and Venous Thromboembolism. 2009;29(3):332-6.
7. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One. 2010;5(12):e15004.
8. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8:14357.
9. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, et al. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. JAMA. 2016;315(23):2532-41.
10. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.
11. van Hylckama Vlieg A, Baglin CA, Luddington R, MacDonald S, Rosendaal FR, Baglin TP. The risk of a first and a recurrent venous thrombosis associated with an elevated D-dimer level and an elevated thrombin potential: results of the THE-VTE study. J Thromb Haemost. 2015;13(9):1642-52.
12. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. 2003;22(1):85-93.
13. Giavarina D. Understanding Bland Altman analysis. Biochemia medica. 2015;25(2):141-51.
14. Bland JM. Interpreting the limits of agreement: do I have good or bad agreement? 2009 [updated 2009 March 202021 January 11]. Available from: https://www-users.york.ac.uk/~mb55/meas/interlim.htm.
15. Kattula S, Byrnes JR, Wolberg AS. Fibrinogen and Fibrin in Hemostasis and Thrombosis. Arteriosclerosis, thrombosis, and vascular biology. 2017;37(3):e13-e21.
16. Tala JA, Polikoff LA, Pinto MG, Li S, Trakas E, Miksa M, et al. Protein biomarkers for incident deep venous thrombosis in critically ill adolescents: An exploratory study. 2020;67(4):e28159.
17. Joshi A, Mayr M. In Aptamers They Trust: The Caveats of the SOMAscan Biomarker Discovery Platform from SomaLogic. Circulation. 2018;138(22):2482-5.
18. Justine B, Thomas S, Bruno P, Marc GB, Anne-Françoise S-S, Aurélien L. Évaluation des performances de l'automate STA R Max®(Stago) pour les paramètres d'hémostase de routine. Annales de Biologie Clinique. 2018;76(2):143-9.
19. Flanders MM, Crist R, Safapour S, Rodgers GM. Evaluation and Performance Characteristics of the STA-R Coagulation Analyzer. Clinical Chemistry. 2002;48(9):1622-4.
20. Cupaiolo R, Govaerts D, Blauwaert M, Cauchie P. Performance evaluation of a new Stago(®) automated haemostasis analyser: The STA R Max(®) 2. Int J Lab Hematol. 2019;41(6):731-7.
21. Bland JM. How can I decide the sample size for a study of agreement between two methods of measurement? 2004 [updated 2004 January 122021 January 1]. Available from: https://www-users.york.ac.uk/~mb55/meas/sizemeth.htm.
22. Hubel A, Spindler R, Skubitz AP. Storage of human biospecimens: selection of the optimal storage temperature. Biopreservation and biobanking. 2014;12(3):165-75.

23. Tworoger SS, Hankinson SE. Collection, processing, and storage of biological samples in epidemiologic studies: sex hormones, carotenoids, inflammatory markers, and proteomics as examples. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006;15(9):1578-81.
24. Wagner-Golbs A, Neuber S, Kamlage B, Christiansen N, Bethan B, Rennefahrt U, et al. Effects of Long-Term Storage at -80 °C on the Human Plasma Metabolome. Metabolites. 2019;9(5).