



Assess the reporting quality of studies investigating the diagnostic accuracy of Suppression of tumorigenicity 2 (ST2) in the diagnosis of Heart Failure published from 2005 to 2019 using the STARD statement

Αξιολόγηση της ποιότητας αναφοράς μελετών δημοσιευμένων από το 2005 έως το 2019 που ερευνούν τη διαγνωστική ακρίβεια του Suppression of tumorigenicity 2 (ST2) στη διάγνωση της καρδιακής ανεπάρκειας, σύμφωνα με τα κριτήρια STARD

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

ΠΜΣ «Μεθοδολογία Βιοϊατρικής Έρευνας,
Βιοστατιστική και Κλινική Βιοπληροφορική»

Τριμελής Συμβουλευτική Επιτροπή: Δοξάνη Χρυσούλα
Στεφανίδης Ιωάννης
Ζιντζαράς Ηλίας

Χρυσοχοΐδης-Τράντας Θωμάς
Έτος υποβολής: 2019

ΠΕΡΙΛΗΨΗ

Εισαγωγή: Ένας αυξανόμενος αριθμός μελετών εκτιμάει τον ρόλο του διαλυτού υποδοχέα ιντερλευκίνης ST2 στην διάγνωση και στην πρόγνωση της καρδιακής ανεπάρκειας. Παρόλα αυτά η ανεπαρκής αναφορά όλων των απαραίτητων σημείων στις μελέτες διαγνωστικής ακρίβειας αυξάνει την πιθανότητα σφάλματος και μπορεί να οδηγήσει σε λανθασμένη εφαρμογή των αποτελεσμάτων τους.

Στόχοι: Ο στόχος της παρούσας μελέτης είναι να αξιολογήσει την ποιότητα ερευνών που μελετούν την διαγνωστική ακρίβεια του δείκτη ST2, όσων αφορά την πληρότητα του περιεχομένου τους.

Μέθοδοι: Η μηχανή αναζήτησης PubMed χρησιμοποιήθηκε για να βρεθούν μελέτες διαγνωστικής ακρίβειας του ST2. Εκτιμήθηκαν οι μελέτες σύμφωνα με τον κατάλογο κριτηρίων STARD 2015 statement. Υπολογίστηκε το συνολικό ποσοστό κριτηρίων που εκπληρώθηκαν και ταξινομήθηκαν οι μελέτες σε υψηλής και χαμηλής ποιότητας βάσει του STARD statement. Διενεργήθηκαν συγκρίσεις μεταξύ των ποσοστών των δύο ομάδων όπως επίσης και μεταξύ των άρθρων που δημοσιεύτηκαν πριν και μετά το έτος έκδοσης του STARD 2015 statement.

Αποτελέσματα: Παρόλο που ένα ικανοποιητικό ποσοστό των κριτηρίων STARD εκπληρώθηκε από τις μελέτες, σοβαρές ελλείψεις διαπιστώθηκαν στο περιεχόμενο των μεθόδων και των αποτελεσμάτων. Συγκρίνοντας τις μελέτες υψηλής και χαμηλής ποιότητας όπως και τις μελέτες πριν και μετά το 2015 δεν διαπιστώθηκαν στατιστικά σημαντικές διαφορές.

Συμπέρασμα: Η ποιότητα των μελετών διαγνωστικής ακρίβειας επιδέχεται περαιτέρω βελτίωσης ώστε να εξασφαλίσουν πληρέστερο περιεχόμενο και μικρότερο σφάλμα.

Λέξεις κλειδιά: καρδιακή ανεπάρκεια, διαγνωστικός ρόλος, βιοδείκτης ST2, ποιότητα μελέτης.

ABSTRACT

Introduction: An increasing number of studies appreciate the role of soluble interleukin receptor ST2 in diagnosis and prognosis of heart failure. However, insufficient reporting of all the necessary points in the diagnostic accuracy studies increases the likelihood of bias and may result in incorrect application of their results.

Objectives: The aim of this study is to evaluate the reporting quality of diagnostic accuracy studies of ST2 in heart failure.

Methods: PubMed was used to search for diagnostic accuracy studies of ST2 in heart failure. Studies were evaluated according to STARD 2015 statement. The overall percentage of fulfilled STARD criteria was measured and the studies were characterized as high- and low-quality based on STARD statement. Comparisons were made between the percentages of the two groups as well as between the articles published before and after the year of announcement of the STARD 2015 statement.

Results: Although a satisfactory percentage of the STARD criteria was fulfilled by the studies, serious deficiencies were found in the content of the methods and results. Comparing high-

and low-quality studies as well as studies before and after 2015 no statistically significant differences were found.

Conclusion: The reporting quality of diagnostic accuracy studies needs further improvement in order to ensure small bias.

Key words: Heart failure, diagnostic role, soluble ST2, reporting quality

ΕΙΣΑΓΩΓΗ

Η καρδιακή ανεπάρκεια (ΚΑ) είναι μια απειλητική για τη ζωή νόσος με εκτιμώμενη επίπτωση 1-2% στον πληθυσμό ενώ αποτελεί την κύρια αιτία θνητότητας και θνησιμότητας στην βιομηχανικές χώρες. Το οικονομικό κόστος της ΚΑ αγγίζει τα 108 δισεκατομμύρια δολάρια ετησίως^[1]. Η κλινική συμπτωματολογία της καρδιακής ανεπάρκειας οφείλεται στην αναδιαμόρφωση της καρδιάς, ως αποτέλεσμα μιας σειράς αλλαγών που αφορούν τα κύτταρα του μυοκαρδίου και του εξωκυττάριου ιστού και περιλαμβάνουν την κυτταρική υπερτροφία, την υπερβολική παραγωγή ινώδους ιστού καθώς και την νέκρωση μυοκαρδιακών κυττάρων.

Παρόλο που η ενημέρωση της κοινότητας για την πολυπαραγοντική αυτή νόσο λαμβάνει όλο και μεγαλύτερες διαστάσεις, αποτελεί ακόμα δυσκολία η έγκαιρη διάγνωση στην καθημερινή κλινική πράξη καθώς η συμπτωματολογία της δεν είναι ειδική. Τα τελευταία χρόνια προστέθηκε στην διαγνωστική σκέψη για την ΚΑ και ο ρόλος δύο βιοδεικτών στο αίμα των ασθενών, του B-type natriuretic peptide (BNP) και του N-terminal pro-B-type natriuretic peptide (NTproBNP)^[2] ενώ τις τελευταίες δύο δεκαετίες ερευνάται και ο ρόλος άλλων ανιχνεύσιμων στον ορό του αίματος βιοδεικτών.^[3] Ανάμεσα τους όλο και περισσότερο ενδιαφέρον αποκτάει ο διαλυτός στον ορό των ασθενών υποδοχέας ιντερλευκίνης, Suppression of Tumorigenicity-2 (soluble ST2, sST2).

Ο δείκτης ST2 είναι μέλος της οικογένειας των υποδοχέων της ιντερλευκίνης IL-1 και εμφανίζεται σε δύο μορφές, την διαλυτή που κυκλοφορεί στο αίμα και την διαμεμβρανική που έχει βρεθεί ότι εκφράζεται μεταξύ άλλων και στα μυοκαρδιακά κύτταρα.^[4] Το μόριο που ενσωματώνεται στον υποδοχέα αυτό είναι η IL-33 και προσφέρει προστατευτικές αντιυπερτροφικές αντιινωτικές δράσεις στα μυοκαρδιακά κύτταρα. Όταν η διαλυτή μορφή του ST2 είναι αυξημένη στο αίμα απομακρύνει την IL-33 από την κυκλοφορία επάγοντας επομένως την παθολογική ιστική αναδιαμόρφωση της καρδιάς μέσω ίνωσης και υπερτροφίας.^[5]

Η έκφραση του υποδοχέα αυτού πιθανολογείται ότι επάγεται από παθολογική έκταση του καρδιακού τοιχώματος και βρέθηκε ότι αυξάνεται σε διάφορες παθολογικές καταστάσεις του καρδιαγγειακού συστήματος όπως η ισχαιμία του μυοκαρδίου, η οξεία και η χρόνια καρδιακή ανεπάρκεια, η πνευμονική υπέρταση.^[6-10]

Ένας αυξανόμενος αριθμός μελετών κατέδειξαν ότι τα επίπεδα του κυκλοφορόντος ST2 είναι αυξημένα σε ασθενείς με ΚΑ και επομένως μπορεί να συμβάλει στην διάγνωση και την εκτίμηση της πρόγνωσης της ασθένειας όπως και στην εντόπιση ασθενών που διαθέτουν μεγάλο κίνδυνο να αναπτύξουν σοβαρή καρδιακή ανεπάρκεια που θα οδηγήσει στο θάνατο.^{[11][12]} Παρόλα αυτά, τα αποτελέσματα άλλων μελετών^{[13][14]} για την διάγνωση της ΚΑ μέσω του ST2 οδηγούν σε αντίθετα συμπεράσματα καθώς προτείνουν ότι ο ρόλος του ST2 είναι κατώτερος από τους ήδη διαθέσιμους βιοδείκτες.

Οι μελέτες διαγνωστικής ακρίβειας μιας εξέτασης εγκυμονούν, όπως όλες οι μελέτες, τον κίνδυνο του σφάλματος. Αυτό εξαρτάται από τις μεθοδολογικές ανεπάρκειες, τον σχεδιασμό μελέτης με λίγους συμμετέχοντες, την λανθασμένη συλλογή και ανάλυση δεδομένων και την εξαγωγή συμπερασμάτων που δεν μπορούν να γενικευτούν με ασφάλεια στον πληθυσμό. Τα σφάλματα αυτά μειώνουν την αξιοπιστία, την εγκυρότητα και την γενικευσιμότητα των αποτελεσμάτων. Πλέον υπάρχουν οδηγίες σχετικά με το περιεχόμενο μιας δημοσιευμένης μελέτης έτσι ώστε να συμπεριλαμβάνονται όλες οι απαραίτητες πληροφορίες. Συγκεκριμένα για τις μελέτες που ασχολούνται με την διαγνωστική αξία έχουν προταθεί από το 2003 τα κριτήρια “Standards for Reporting of Diagnostic Accuracy” (STARD statement) (<http://www.equator-network.org/reporting-guidelines/stard/>)^[15] Αυτή η λίστα αποτελείται από 30 κριτήρια που αφορούν τον σχεδιασμό της μελέτης, τα χαρακτηριστικά των ασθενών, τις αναλύσεις που διενεργήθηκαν και τα συμπεράσματα που προέκυψαν. Αναθεωρήθηκαν το 2015 και προσφέρουν ένα χρήσιμο και εύκολο εργαλείο στον ερευνητή και στον αναγνώστη για να αξιολογήσει κατά πόσο μια μελέτη διαγνωστικής ακρίβειας περιέχει λεπτομερώς όλες τις απαραίτητες πληροφορίες.^[16]

Είναι σκόπιμο να αναφέρουμε ότι τα κριτήρια STARD αξιολογούν το κατά πόσο αναφέρονται λεπτομερώς όλα τα απαραίτητα σημεία μιας μελέτης. Δηλαδή αποτελούν εργαλείο ελέγχου της αναφοράς ορισμένων πληροφοριών και όχι αν πληρούνται οι προϋποθέσεις μιας ποιοτικής μελέτης. Το γεγονός αυτό δεν διασφαλίζει απαραίτητα ότι μια μελέτη είναι ποιοτική αλλά το ότι ο ερευνητής παρουσίασε λεπτομερώς το πως η μελέτη του εκπονήθηκε.

Παράλληλα με τα κριτήρια STARD ευρεία εφαρμογή στις μελέτες διαγνωστικής αξίας έχουν και τα κριτήρια QUADAS-2^[17] ενώ παρόμοια χρήση έχουν και τα κριτήρια TRIPOD. Τα τελευταία έχουν προταθεί για μελέτες που επικεντρώνονται κυρίως στην προγνωστική αξία μιας διαγνωστικής εξέτασης και στην διαδικασία κατασκευής ενός προγνωστικού μοντέλου. Τα STARD κριτήρια δεν κατασκευάστηκαν συγκεκριμένα για μελέτες που αφορούν την προγνωστική αξία ενός δείκτη παρόλα αυτά στην επεξήγηση τους αναφέρεται ότι μπορούν να εφαρμοστούν και σε τέτοιου είδους μελέτες^[18].

Οι σκοποί αυτής της μελέτης είναι: 1) να αξιολογηθεί η ποιότητα μελετών που αφορούν την διαγνωστική αξία του δείκτη sST2 στην ΚΑ σύμφωνα με τα κριτήρια STARD, 2) να συγκριθεί η διαφορά της ποιότητας των μελετών πριν και μετά το έτος 2015 κατά το οποίο προτάθηκαν τα ανανεωμένα κριτήρια STARD 2015.

ΜΕΘΟΔΟΙ

Πηγές δεδομένων και όροι αναζήτησης

Η μηχανή αναζήτησης PubMed χρησιμοποιήθηκε για να βρεθούν όλα τα είδη μελετών που είναι δημοσιευμένες από τον Ιανουάριο του 2005 έως τον Σεπτέμβριο του 2019 και οι οποίες αξιολογούσαν τον ρόλο του ST2 στην διάγνωση της καρδιακής ανεπάρκειας. Στην αναζήτηση χρησιμοποιήθηκαν οι παρακάτω όροι (“diagnosis” or “diagnostic”) and (“suppression of tumorigenicity-2” or “ST2” or “soluble ST2” or “sST2” or “biomarkers”) and (“heart failure” or “HF” or “dyspnea”). Επιλέχθηκαν άρθρα μόνο γραμμένα στην αγγλική γλώσσα.

Επιλογή μελετών

Εξετάστηκαν προσεχτικά οι περιλήψεις των άρθρων ώστε να βρεθούν αυτά που περιείχαν μέσα στους υπό εξέταση βιοδείκτες τον ST2. Οι μελέτες επιλέχθηκαν μόνο αν περιλάμβαναν περισσότερους από 10 ασθενείς. Επιλέχθηκαν κλινικές μελέτες ασθενών-μαρτύρων αλλά και κοόρτης. Στις μελέτες ασθενών μαρτύρων η ομάδα των ασθενών αποτελούσε είτε ασθενείς με γνωστή καρδιακή ανεπάρκεια, είτε ασθενείς που επισκέπτονταν το τμήμα επειγόντων περιστατικών με αναφερόμενα συμπτώματα που παραπέμπουν την διαγνωστική σκέψη σε καρδιακή ανεπάρκεια. Η ομάδα των μαρτύρων αποτελούσε υγιείς ανθρώπους. Στις μελέτες κοόρτης συμπεριλήφθηκαν ασθενείς με γνωστή καρδιακή ανεπάρκεια που παρακολουθούνταν τακτικά. Σε αυτούς εξετάστηκε κατά πόσο τα επίπεδα του ST2 την 1^η ημέρα παρατήρησης σχετίζονται με την διάγνωση καρδιακής ανεπάρκειας με κακή πρόγνωση δηλαδή με αυξημένη θνητότητα μετά από ένα χρόνο ή κατά πόσο τα επίπεδα του ST2 είναι διαγνωστικά για την σοβαρότητα των κλινικών εκδηλώσεων των ασθενών.

Με τον όρο καρδιακή ανεπάρκεια συμπεριλήφθη η ΚΑ με οξεία εμφάνιση που οδήγησε τους ασθενείς σε τμήμα επειγόντων περιστατικών, η χρόνια καρδιακή ανεπάρκεια υπό παρακολούθηση, η ΚΑ με διατηρημένο ή με μειωμένο κλάσμα εξώθησης, η οφειλόμενη σε συγγενείς καρδιοπάθειες και η ΚΑ στην ειδική ομάδα των ασθενών με υπέρταση. Στις μελέτες συμπεριλήφθηκε και η μοναδική μελέτη που αφορούσε παιδιατρικό πληθυσμό.

Εκτίμηση δεδομένων

Τα άρθρα που επιλέχθηκαν εξετάστηκε αν αναφέρουν όλες τις απαραίτητες πληροφορίες που πρέπει να αναφέρονται σε μια μελέτη διαγνωστικής ακρίβειας σύμφωνα με τα 30 συστατικά του STARD 2015. Ένα κριτήριο θεωρήθηκε ότι εκπληρώθηκε όταν από τις πληροφορίες του άρθρου αναφέρεται ρητά ότι εκπληρώθηκε και όχι αν από το περιεχόμενο συμπεραίνει ο αναγνώστης ότι πραγματοποιήθηκε στη μελέτη. Λαμβάνοντας υπόψιν την επεξήγηση των κριτηρίων STARD^[18] όλα τα συστατικά του πρέπει να αναφέρονται με επαρκή λεπτομέρεια για να αποφασίσει ο αναγνώστης ότι πληρούνται. Αν από το περιεχόμενο η πληροφορία ήταν ασαφής ή διφορούμενη λαμβανόταν ως αρνητική απάντηση. Όλα τα κριτήρια αποφασίστηκε να υπολογίζονται ισόβαθμα, χωρίς δηλαδή ορισμένα από αυτά να έχουν αυξημένη βαρύτητα στον υπολογισμό του τελικού score. Τα κριτήρια με αριθμό 10,12,13,21 περιλαμβάνουν 2 υποομάδες a και b. Όταν τουλάχιστον η μία από τις 2 υποομάδες ελάμβανε θετική απάντηση αποφασίστηκε ότι όλο το κριτήριο θα θεωρείται ότι πληρείται.

Ανάλυση δεδομένων

Υπολογίστηκε για κάθε μελέτη ο συνολικός αριθμός των θετικών απαντήσεων σύμφωνα με τα κριτήρια STARD και στη συνέχεια ταξινομήθηκαν αυτές σε υψηλής και χαμηλής ποιότητας (high and low quality). Το όριο της βαθμολογίας για την ταξινόμηση αυτή ήταν το 17,2. Η επιλογή του 17,2 ήταν μέσος όρος όλων των βαθμολογιών. Επομένως 7 μελέτες που συγκέντρωσαν βαθμολογία από 18 και άνω χαρακτηρίστηκαν ως υψηλής ποιότητας ενώ όσες συγκέντρωσαν από 17 και κάτω ως χαμηλής ποιότητας. Στη συνέχεια

υπολογίστηκαν τα ποσοστά για κάθε ένα από τα κριτήρια του STARD statement που πληρούνταν συνολικά. Έπειτα υπολογίστηκαν αυτά τα ποσοστά ξεχωριστά για τα υψηλής και χαμηλής ποιότητας άρθρα. Τέλος βρέθηκε η στατιστική σημαντικότητα της διαφοράς των ποσοστών με Fisher's exact test.

Όσον αφορά τη αξιολόγηση των άρθρων ανάλογα με την χρονολογία έκδοσης τους, τα άρθρα ταξινομήθηκαν σε 2 κατηγορίες αναλόγως αν εκδόθηκαν πριν ή μετά το 2015, δηλαδή την χρονιά που προτάθηκε το ανανεωμένο STARD 2015. Παρόλα αυτά υπήρχε από το 2003 η πρώτη έκδοση των κριτηρίων STARD. Αρχικά υπολογίστηκε αν διέφεραν οι συνολικές βαθμολογίες των άρθρων μεταξύ των δύο κατηγοριών. Η κανονικότητα των δεδομένων ελέγχθηκε με το Shapiro-Wilk test και στη συνέχεια εφαρμόστηκε t-test για ανεξάρτητα δεδομένα. Στη συνέχεια υπολογίστηκαν τα ποσοστά που εκπληρώθηκε ή κάθε υποκατηγορία του STARD ξεχωριστά για τα άρθρα πριν και μετά το 2015 και εξετάστηκε αν η διαφορά για το καθένα ήταν στατιστικά σημαντική με Fisher's exact test. Οι υπολογισμοί διενεργήθηκαν στο ηλεκτρονικό πρόγραμμα στατιστικής ανάλυσης SPSS version 25 (SPSS Inc., Chicago)

ΑΠΟΤΕΛΕΣΜΑΤΑ

Χαρακτηριστικά ερευνών

Συνολικά 13 άρθρα, δημοσιευμένα στο διάστημα 2007-2018, αξιολογήθηκαν σύμφωνα με το STARD 2015 statement. 8 από αυτά δημοσιεύτηκαν πριν το 2015 δηλαδή πριν την έκδοση του STARD 2015 statement και 5 από αυτά μετά. Παρόλα αυτά αξίζει να σημειωθεί ότι είχαν προταθεί από το 2003 τα πρώτα κριτήρια STARD 2003 statement που αναθεωρήθηκαν το 2015. Συνολικά από 8 διαφορετικές χώρες προήλθαν τα άρθρα. 4 άρθρα προήλθαν από την επιστημονική κοινότητα της Ευρώπης (2 από Αυστρία, 1 από Γερμανία και 1 από Ισπανία), 4 άρθρα προήλθαν από ΗΠΑ, 4 από Ασία (2 από Κίνα, 1 από Ινδία και 1 από Σιγκαπούρη) και 1 άρθρο από Νέα Ζηλανδία.

Στα περισσότερα από τα άρθρα αναφέρεται ότι οι συμμετέχοντες ήταν ασθενείς που προσήλθαν σε τμήμα επειγόντων περιστατικών με αναφερόμενα συμπτώματα που παραπέμπουν την διαγνωστική σκέψη σε οξεία καρδιακή ανεπάρκεια (6 άρθρα, 46,1%), 3 άρθρα περιγράφουν ότι οι συμμετέχοντες ασθενείς ήταν ασθενείς που νοσηλεύονταν σε κλινική (23,0%), σε 2 άρθρα οι ερευνητές μελέτησαν ασθενείς που παρακολουθούνται σε τακτική βάση σε εξωτερικά ιατρεία (15,4%), και σε 2 άρθρα οι ασθενείς ήταν και από εξωτερικά ιατρεία και από την κλινική (15,4%). Ο πιο δημοφιλής κατασκευαστής των εργαστηριακών αναλυτών ήταν ο "Critical Diagnostics" που απαντήθηκε σε 6 άρθρα (46,1%) ενώ σε 2 άρθρα απαντήθηκε ο "R&D systems" (15,4%) και σε 2 άρθρα ο "Medical and Biological Laboratories" (15,4%).

Οι συμμετέχοντες στις μελέτες μπορούν να χωριστούν ηλικιακά σε 3 κατηγορίες. 11 άρθρα μελέτησαν άτομα ηλικίας 55 και άνω με μέσο όρο ηλικιακά 64,9 έτη. 1 άρθρο αφορούσε νεότερους ασθενείς με διάγνωση συγγενούς καρδιοπάθειας^[29] με μέσο όρο ηλικίας τα 28 έτη. 1 άρθρο αφορούσε μελέτη σε παιδιατρικό πληθυσμό με ηλικίες 0-2 έτη. Συνολικά επομένως ο μέσος όρος ηλικίας των ενήλικων συμμετεχόντων ήταν 46,5 έτη.

Τα άρθρα της μελέτης προέρχονται από 10 διαφορετικά επιστημονικά περιοδικά. Ο μέσος όρος του impact factor ήταν 6,183 (1,728 – 18,639) για τα 8 περιοδικά καθώς για 2

περιοδικά δεν βρέθηκε impact factor.5 από τα άρθρα συγκέντρωναν impact factor πάνω από το μέσο όρο (38,4%) του συνόλου. Τέλος, μετά τον υπολογισμό του μέσου όρου που ορίστηκε σαν cut-off σημείο, 7 άρθρα (53,8%) συμπεριλήφθηκαν στα άρθρα υψηλής ποιότητας σύμφωνα με τα κριτήρια STARD 2015 δηλαδή με βαθμολογία από 18 και άνω και 6 άρθρα κρίθηκαν ως χαμηλής ποιότητας με βαθμολογία από 17 και κάτω.

Κύρια αποτελέσματα

Ο πίνακας 1 δείχνει το ποσοστό που εκπληρώνεται το κάθε κριτήριο του STARD statement συνολικά στις 13 έρευνες καθώς και ξεχωριστά μετά την ταξινόμηση τους σε υψηλής ή χαμηλής ποιότητας με βάση την βαθμολογία.

Συνολικά 13 πληροφορίες του STARD (2 στον τίτλο και στην περίληψη, 1 στην εισαγωγή, 5 στις μεθόδους, 2 στα αποτελέσματα, 2 στη συζήτηση και 1 στις λουιπές πληροφορίες) αναφέρθηκαν σε ποσοστό άνω του 85% του συνόλου. Αυτά αφορούν 1)την αναγνώριση της μελέτης ως μελέτη διαγνωστικής ακρίβειας 2)την ύπαρξη δομημένης σύνοψης του σχεδιασμού, των μεθόδων, αποτελεσμάτων και συμπερασμάτων της μελέτης, 3)το επιστημονικό και κλινικό υπόβαθρο για την υπό μελέτη εξέταση, 4)τα κριτήρια καταλληλότητας για τους συμμετέχοντες, 5)τον τρόπο που εντοπίστηκαν οι εν δυνάμει συμμετέχοντες (συμπτώματα, μητρώα ασθενών, αποτελέσματα προηγούμενων εξετάσεων), 6)την περιγραφή της υπό μελέτη διαγνωστικής δοκιμασίας ή την συγκρινόμενη με αυτή εξέταση αναφοράς, 7)τον ορισμό του σημείου cut-off για την θετικότητα της υπό μελέτης ή της διαγνωστικής εξέτασης αναφοράς καθώς και τον τρόπο με τον οποίο υπολογίστηκε, 8)τις μεθόδους που υπολογίστηκαν τα μεγέθη διαγνωστικής ακρίβειας, 9) τα βασικά κλινικά και δημογραφικά χαρακτηριστικά των συμμετεχόντων, 10) τον υπολογισμό της διαγνωστικής ακρίβειας με στατιστική μεγέθη, 11) τους περιορισμούς της δοκιμής, τις πηγές πιθανών σφαλμάτων, την ανακρίβεια και πώς επηρεάζεται η πολλαπλότητα των αναλύσεων, 12)προτάσεις για εφαρμογή, συμπεριλαμβανομένου της προβλεπόμενης χρήσης και του κλινικού ρόλου της πειραματικής εξέτασης, 13)τον αριθμό καταχώρησης και το όνομα του μητρώου δοκιμών.

Επιπλέον 4 υποκατηγορίες του STARD (2 στις μεθόδους, 1 στα αποτελέσματα και 1 στις επιπλέον πληροφορίες) απαντήθηκαν θετικά σε ποσοστό από 60-85% του συνόλου. Αυτές ήταν 1)αν η συλλογή δεδομένων έγινε πριν την διεξαγωγή της υπό μελέτη και της διαγνωστικής εξέτασης αναφοράς 2) συνθήκες και τόποι στους οποίους συλλέχθηκαν τα δεδομένα, 3)η κατανομή της σοβαρότητας της πάθησης στους ασθενείς που νοσοούν από αυτή ή η κατανομή των εναλλακτικών διαγνώσεων στα άτομα χωρίς τη συγκεκριμένη νόσο 4)πηγές χρηματοδότησης και άλλης υποστήριξης, ο ρόλος των χορηγών.

Αντιθέτως 8 υποκατηγορίες απαντήθηκαν σε μικρότερο ποσοστό από το σύνολο των άρθρων. Σε ποσοστό 53,8% απαντήθηκε αν η εκτίμηση των αποτελεσμάτων της πειραματικής και της διαγνωστικής εξέτασης αναφοράς έγινε χωρίς να είναι γνωστά τα αποτελέσματα της εξέτασης αναφοράς ή της υπό εξέτασης αντίστοιχα καθώς επίσης και αν ήταν γνωστά τα κλινικά χαρακτηριστικά των συμμετεχόντων. Σε ποσοστό 23% αναφέρθηκαν οι σκοποί (objectives) και οι υποθέσεις των μελετών καθώς και η λογική που επιλέχθηκε ο τρόπος διάγνωσης της υπό μελέτης νόσου στους συμμετέχοντες. Σε ποσοστό 15,3% αναφέρθηκε αν οι συμμετέχοντες μελετήθηκαν με μια συνεχή ή τυχαία ή βολική για

STARD items	Overall % of reporting item(n=13)	% of reporting item in lower quality articles (n≤17)	% of reporting item in higher quality articles (n>17)	P-value*
TITLE OR ABSTRACT				
1. Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	100	100	100	1
ABSTRACT				
2. Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	100	100	100	1
INTRODUCTION				
3. Scientific and clinical background, including the intended use and clinical role of the index test	100	100	100	1
4. Study objectives and hypotheses	23	0	42,8	0,39
METHODS				
5. Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	69,2	50	85,7	0,164
6. Eligibility criteria	92,3	8,3	100	0,26
7. On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	100	100	100	1
8. Where and when potentially eligible participants were identified (setting, location and dates)	61,5	66,6	57,1	0,72
9. Whether participants formed a consecutive, random or convenience series	15,3	0	28,5	0,154
10. a) Index test, in sufficient detail to allow replication b) Reference standard, in sufficient detail to allow replication	100	100	100	1
11. Rationale for choosing the reference standard (if alternatives exist)	23	16,6	28,5	0,514
12. a) Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory b) Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	92,3	83,3	100	0,26
13. a) Whether clinical information and reference standard results were available to the performers/readers of the index test b) Whether clinical information and index test results were available to the assessors of the reference standard	53,8	16,6	85,7	0,012
14. Methods for estimating or comparing measures of diagnostic accuracy	100	100	100	1
15. How indeterminate index test or reference standard results were handled	7,7	0	14,2	0,335
16. How missing data on the index test and reference standard were handled	15,3	0	28,5	0,154
17. Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	15,3	0	28,5	0,154
18. Intended sample size and how it was determined	0	0	0	1
RESULTS				
19. Flow of participants, using a diagram	7,7	0	14,2	0,335
20. Baseline demographic and clinical characteristics of participants	100	100	100	1
21. a) Distribution of severity of disease in those with the target condition b) Distribution of alternative diagnoses in those without the target condition	61,5	83,3	42,8	0,134
22. Time interval and any clinical interventions between index test and reference standard	0	0	0	1
23. Cross tabulation of the index test results (or their distribution) by the results of the reference standard	0	0	0	1
24. Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	100	100	100	1
25. Any adverse events from performing the index test or the reference standard	0	0	0	1
DISCUSSION				
26. Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	92,3	100	85,7	0,335
27. Implications for practice, including the intended use and clinical role of the index test	100	100	100	1
OTHER INFORMATION				
28. Registration number and name of registry	100	100	100	1
29. Where the full study protocol can be accessed	0	0	0	1
30. Sources of funding and other support; role of funders	84,6	66,6	100	0,096

τους ερευνητές σειρά, το πως διαχειρίστηκαν τα δεδομένα που έλλειπαν από τη μελέτη καθώς και αν έγινε περεταίρω ανάλυση της μεταβλητότητας της διαγνωστικής ακρίβειας. Σε ποσοστό 7,7% αναφέρθηκε πώς οι ερευνητές διαχειρίστηκαν τα απροσδιόριστα/ασαφή δεδομένα και στο ίδιο ποσοστό ήταν διαθέσιμο ένα διάγραμμα ροής των συμμετεχόντων.

5 πληροφορίες (1 στις μεθόδους, 3 στα αποτελέσματα και 1 στις λοιπές πληροφορίες) δεν αναφέρθηκαν σε καμία έρευνα. Αυτές ήταν 1) το πώς ορίστηκε το αναμενόμενο μέγεθος του δείγματος, 2) το χρονικό διάστημα και αν έγιναν κλινικές παρεμβάσεις ανάμεσα στην πειραματική διαγνωστική εξέταση και στην εξέταση αναφοράς, 3) η παρουσίαση των αποτελεσμάτων με μέθοδο cross tabulation, 4) αν υπήρξαν ανεπιθύμητες ενέργειες από την διενέργεια των διαγνωστικών εξετάσεων, 5) πού μπορεί να βρεθεί διαθέσιμο το πλήρες πρωτόκολλο της μελέτης.

Συγκρίνοντας τις δύο κατηγορίες άρθρων (υψηλής και χαμηλής ποιότητας) με βάση τη συνολική βαθμολογία STARD που συγκέντρωσαν παρατηρούμε ότι μόνο σε ένα 1 κριτήριο παρουσιάζεται στατιστικά σημαντική διαφορά ($P\text{-value}<0,05$). Αυτό είναι το 13: «αν η εκτίμηση των αποτελεσμάτων της υπό εξέτασης και της διαγνωστικής εξέτασης αναφοράς έγινε χωρίς να είναι γνωστά τα αποτελέσματα της εξέτασης αναφοράς ή της υπό μελέτης εξέτασης αντίστοιχα στους εκτιμητές, καθώς επίσης και αν ήταν γνωστά τα κλινικά χαρακτηριστικά των συμμετεχόντων» Παρατηρούμε ότι στις υπόλοιπες υποκατηγορίες του STARD statement παρόλο που οι μελέτες υψηλής ποιότητας συγκέντρωσαν μεγαλύτερα ποσοστά σε σχέση με τις χαμηλής ποιότητας δεν βρέθηκε στατιστικά σημαντική διαφορά ($P\text{-value}>0,05$).

Κύρια αποτελέσματα μετά την χρονολογική ταξινόμηση των άρθρων

Τα άρθρα χωρίστηκαν σε δύο κατηγορίες ανάλογα με το αν εκδόθηκαν πριν ή μετά το 2015 που ήταν το έτος που προτάθηκαν τα ανανεωμένα κριτήρια STARD 2015. 8 άρθρα είχαν ημερομηνία έκδοσης πριν το 2015 ενώ 5 άρθρα μετά το 2015. Στον πίνακα 2 φαίνονται τα συνολικά score κατά STARD που συγκέντρωσαν οι έρευνες καθώς και το αποτέλεσμα της σύγκρισης τους. Διαπιστώθηκε ότι δεν παρουσιάζουν στατιστικά σημαντική διαφορά οι δύο κατηγορίες ($p\text{-value}>0.05$).

Στη συνέχεια υπολογίστηκε το ποσοστό θετικής απάντησης για την κάθε μία από τις υποκατηγορίες του STARD 2015 statement ξεχωριστά για κάθε κατηγορία άρθρων. Στον πίνακα 3 φαίνονται τα ποσοστά που υπολογίστηκαν καθώς και η τιμή $P\text{-value}$ της σύγκρισης ανάμεσα στα άρθρα που εκδόθηκαν πριν και μετά το 2015. Σε 7 υποκατηγορίες (6 στις μεθόδους, 1 στην συζήτηση) του STARD παρατηρήθηκε αύξηση του ποσοστού αναφοράς στα άρθρα που εκδόθηκαν μετά το 2015. Παρόλα αυτά δεν παρατηρήθηκε σε κανένα κριτήριο στατιστικά σημαντική διαφορά.

Πίνακας 2 Σύγκριση των άρθρων ταξινομημένων σε 2 κατηγορίες ανάλογα με το έτος έκδοσης τους.

Article	number of reporting items in articles before 2015	Article	number of reporting items in articles after 2015	P-value*
1	14	1	16	
2	17	2	15	
3	15	3	19	
4	18	4	18	
5	17	5	18	
6	19			
7	19			
8	18			
Mean	16,2		17,2	
Std. Error	0,735		0,735	0,941

*Η τιμή του P-value υπολογίστηκε με t-test για ανεξάρτητα δεδομένα αφού ελέγχθηκε η κανονικότητα των δεδομένων κατά Shapiro-Wilk.

STARD items	% of reporting item in articles before 2015 (n=8)	% of reporting item in articles after 2015 (n=5)	P-value*
TITLE OR ABSTRACT			
1. Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	100	100	1
ABSTRACT			
2. Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	100	100	1
INTRODUCTION			
3. Scientific and clinical background, including the intended use and clinical role of the index test	100	100	1
4. Study objectives and hypotheses	37,5	0	0,23
METHODS			
5. Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	62,5	80	1
6. Eligibility criteria	87,5	100	1
7. On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	100	100	1
8. Where and when potentially eligible participants were identified (setting, location and dates)	37,5	100	0,075
9. Whether participants formed a consecutive, random or convenience series	12,5	20	1
10. a) Index test, in sufficient detail to allow replication b) Reference standard, in sufficient detail to allow replication	100	100	1
11. Rationale for choosing the reference standard (if alternatives exist)	25	20	1
12. a) Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory b) Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	100	80	0,384
13. a) Whether clinical information and reference standard results were available to the performers/readers of the index test b) Whether clinical information and index test results were available to the assessors of the reference standard	62,5	40	0,592
14. Methods for estimating or comparing measures of diagnostic accuracy	100	100	1
15. How indeterminate index test or reference standard results were handled	12,5	0	1
16. How missing data on the index test and reference standard were handled	12,5	20	1
17. Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	12,5	20	1
18. Intended sample size and how it was determined	0	0	1
RESULTS			
19. Flow of participants, using a diagram	12,5	0	1
20. Baseline demographic and clinical characteristics of participants	100	100	1
21. a) Distribution of severity of disease in those with the target condition b) Distribution of alternative diagnoses in those without the target condition	62,5	60	1
22. Time interval and any clinical interventions between index test and reference standard	0	0	1
23. Cross tabulation of the index test results (or their distribution) by the results of the reference standard	0	0	1
24. Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	100	100	1
25. Any adverse events from performing the index test or the reference standard	0	0	1
DISCUSSION			
26. Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	87,5	100	1
27. Implications for practice, including the intended use and clinical role of the index test	100	100	1
OTHER INFORMATION			
28. Registration number and name of registry	100	100	1
29. Where the full study protocol can be accessed	0	0	1
30. Sources of funding and other support; role of funders	87,5	80	1

*P-values υπολογισμένα με Fisher's exact test

ΣΥΜΠΕΡΑΣΜΑ

Στη μελέτη αυτή αξιολογήθηκε η πληρότητα του περιεχομένου άρθρων, που εξετάζουν τον διαγνωστικό ρόλο του ST2 στην καρδιακή ανεπάρκεια, σύμφωνα με τον κατάλογο κριτηρίων STARD 2015. Από τα άρθρα που επιλέχθηκαν, τα 8 (62% του συνόλου) δημοσιεύτηκαν πριν το έτος έκδοσης του STARD 2015 statement ενώ 5 άρθρα (38% του συνόλου) μετά. Υπολογίστηκε το ποσοστό που πληρούταν κάθε μία από τις υποκατηγορίες που συνιστούν τον κατάλογο STARD. Στη συνέχεια έγινε σύγκριση για κάθε ένα κριτήριο ξεχωριστά ανάμεσα σε άρθρα υψηλής και χαμηλής ποιότητας αναφοράς. Η ποιότητα αυτή χαρακτηρίζει το κατά πόσο αναφέρθηκαν όλες οι απαραίτητες πληροφορίες που συστήνει ο κατάλογος STARD έτσι ώστε μια μελέτη διαγνωστικής αξίας να θεωρείται ολοκληρωμένη ως προς το περιεχόμενο. Τέλος έγινε σύγκριση της ποιότητας αυτής μεταξύ των άρθρων που δημοσιεύτηκαν πριν και μετά το 2015.

Συνολικά 14 από τις 30 υποκατηγορίες του STARD statement απαντήθηκαν σε ποσοστό 80% και άνω στις επιλεγμένες μελέτες. Μπορούμε να πούμε λοιπόν ότι τα άρθρα χαρακτηρίζονται ως ικανοποιητικής ποιότητας αναφοράς. Παρόλα αυτά σημαντικές μεθοδολογικές πληροφορίες όπως οι συνθήκες και ο τρόπος που εντοπίστηκαν οι συμμετέχοντες, η λογική πίσω από την επιλογή της διαγνωστικής εξέτασης αναφοράς, η διαθεσιμότητα των κλινικών χαρακτηριστικών των ασθενών και τα αποτελέσματα των εξετάσεων τους στους εκτιμητές της εκάστοτε διαγνωστικής εξέτασης, αν έγινε αναδρομικά η προοπτικά η συλλογή δεδομένων όπως επίσης και ο τρόπος διαχείρισης των ασαφών ή των χαμένων δεδομένων αναφέρονται σε ποσοστό κάτω του 70%. Προβληματικός φαίνεται να είναι και ο τρόπος παρουσίασης των αποτελεσμάτων καθώς 4 υποκατηγορίες του STARD statement απαντήθηκαν σε ποσοστό κάτω του 65%. Το γεγονός αυτό καθιστά δύσκολη την εκτίμηση της γενικευσιμότητας και της αξιοπιστίας των συμπερασμάτων.

Συγκρίνοντας ξεχωριστά την κάθε υποκατηγορία του STARD statement ανάμεσα στα άρθρα υψηλότερης και χαμηλότερης ποιότητας διαπιστώθηκε στατιστικά σημαντική διαφορά μόνο σε ένα κριτήριο. Για τα υπόλοιπα, η διαφορά δεν έφτανε το όριο στατιστικής σημασίας. Αποδείχθηκε λοιπόν πως για να χαρακτηριστεί τελικά υψηλής ποιότητας μια μελέτη και να διαφέρει σημαντικά από τις χαμηλότερης ποιότητας χρειάζεται να πληροί περισσότερα κριτήρια STARD.

Όσον αφορά την σύγκριση ανάλογα με το έτος έκδοσης δεν παρατηρήθηκε στατιστικά σημαντική διαφορά μεταξύ των άρθρων που δημοσιεύτηκαν πριν και μετά το έτος έκδοσης του STARD 2015 statement. Αυτή η έλλειψη στατιστικής διαφοράς επιβεβαιώθηκε και αφού υπολογίστηκαν τα ποσοστά απάντησης για κάθε υποκατηγορία ξεχωριστά. Συμπεραίνουμε επομένως ότι η παρουσία του STARD statement δεν επηρέασε τον τρόπο παρουσίασης των μελετών που ασχολούνται με την διαγνωστική αξία του ST2 στην καρδιακή ανεπάρκεια.

Χρήσιμο είναι να τονιστεί πως τα κριτήρια STARD αξιολογούν κατά πόσο συγκεκριμένες πληροφορίες αναφέρθηκαν ρητά στο περιεχόμενο του άρθρου. Δεν αξιολογούν άμεσα την ποιότητα μιας μελέτης αλλά το πόσο πλήρης είναι η αναφορά όλων των σημείων που χρειάζονται για να εκτιμηθεί σωστά ο σχεδιασμός της. Επομένως υπάρχει η πιθανότητα μια μελέτη παρόλο που σχεδιάστηκε σωστά να χαρακτηριστεί ότι περιέχει σφάλματα από τον ανεπαρκή τρόπο παρουσίασης της στον αναγνώστη. Το αντίθετο μπορεί να συμβεί επίσης, δηλαδή μια μελέτη με σφάλματα στον σχεδιασμό της, να αναφέρει

επαρκώς όλες τις απαραίτητες πληροφορίες και να χαρακτηριστεί ως υψηλής ποιότητας κατά το STARD statement.

Η χρήση του STARD statement για την αξιολόγηση μελετών διαγνωστικής ακρίβειας έχει εφαρμοστεί ξανά σε ποικίλα επιστημονικά πεδία όπως στην χρήση βιοδεικτών για τη διάγνωση της νεοπλασίας του προστάτη^[19], στην υπερηχογραφία για την διάγνωση παθήσεων του περινεού^[20], στην εκτίμηση της ίνωσης σε χρόνιες παθήσεις του ήπατος^[21], στη διάγνωση του γλαυκώματος^[22] και σε άλλες παθήσεις. Όσον αφορά τις καρδιαγγειακές παθήσεις έχει εφαρμοστεί στην εκτίμηση της διαγνωστικής ακρίβειας της υπολογιστικής τομογραφίας στις παθήσεις των στεφανιαίων αγγείων^[23].

Η παρούσα μελέτη παρουσιάζει ορισμένους περιορισμούς. Αρχικά η αναζήτηση της βιβλιογραφίας διενεργήθηκε μόνο μέσω της μηχανής αναζήτησης PubMed. Επιπλέον η γλώσσα αναζήτησης περιορίστηκε στην αγγλική. Στη διεθνή βιβλιογραφία εντοπίστηκαν άρθρα γραμμένα στη κινεζική γλώσσα αλλά απορρίφθηκαν λόγω αδυναμίας μετάφρασης. Επιπλέον στη μελέτη συμπεριλήφθηκαν άρθρα που εξέταζαν τον ρόλο της προγνωστικής αξίας του δείκτη ST2 αλλά ανέφεραν και λιγότερες πληροφορίες για την διαγνωστική αξία του. Παρόλο που στην επεξήγηση του STARD statement αναφέρεται πως μπορεί να εφαρμοστεί και σε μελέτες προγνωστικής ακρίβειας^[18] η χρήση παρόμοιων κριτηρίων με το όνομα TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement κρίνεται καταλληλότερη. Επομένως ορισμένες μελέτες μπορεί να έλαβαν χαμηλή βαθμολογία κατά τα κριτήρια STARD αλλά να κρίνονταν υψηλότερης ποιότητας με τα κριτήρια TRIPOD όσον αφορά το σκέλος που ασχολείται με την προγνωστική ακρίβεια. Τέλος, όλες οι μελέτες αξιολογήθηκαν από έναν αναγνώστη αυξάνοντας έτσι την πιθανότητα σφάλματος πληροφορίας.

Συμπερασματικά, παρόλο που ένας ικανός αριθμός υποκατηγοριών του STARD statement αναφέρθηκε σε ένα υψηλό ποσοστό του συνόλου των άρθρων, διαπιστώθηκαν ανεπάρκειες στην αναφορά σημαντικών πληροφοριών ιδιαίτερα στις μεθόδους και στα αποτελέσματα. Μεγάλη προσπάθεια πρέπει ακόμη να καταβληθεί ώστε να βελτιωθεί ο τρόπος συγγραφής ενός άρθρου για να εξασφαλίζεται η ποιότητα της μελέτης πριν εφαρμοστούν τα αποτελέσματά της στην κλινική πράξη. Ο κατάλογος του STARD statement μπορεί να συμβάλει ουσιαστικά σε αυτό το εγχείρημα.

ΑΝΑΦΟΡΕΣ

1. Cook C, Cole G, Asaris P, et al⁴⁹ The Annual Global Economic Burden of Heart FailureHeart 2014;100:A28-A29.
2. Gaggin HK, Januzzi JL Jr. Biomarkers and diagnostics in heart failure. Biochim Biophys Acta 2013;1832:2442- 50. doi: 10.1016/j. bbadis.2012.12.014
3. van Kimmenade RR, Januzzi JL Jr. Emerging biomarkers in heart failure. Clin Chem 2012;58:127- 38. doi: 10.1373/clinchem.2011.165720
4. Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK, et al. IL-33, an interleukin-1- like cytokine that signals via the IL-1 receptor- related protein ST2 and induces T Helper type 2-associated cytokines. Immunity 2005;23:479– 90.

5. Sanada S, Hakuno D, Higgins LJ, Schreiter ER, McKenzie AN, Lee RT. IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system. *J Clin Invest* 2007;117: 1538–49
6. S.U. Rehman, T. Mueller, J.L. Januzzi Jr., Characteristics of the novel interleukin family biomarker ST2 in patients with acute heart failure, *J. Am. Coll. Cardiol.* 52 (2008) 1458–1465.
7. E.O. Weinberg, M. Shimpo, G.W. De Keulenaer, et al., Expression and regulation of ST2, an interleukin-1 receptor family member, in cardiomyocytes and myocardial infarction, *Circulation* 106 (2002) 2961–2966.
8. G. Carlomagno, G. Messalli, R.M. Melillo, et al., Serum soluble ST2 and interleukin-33 levels in patients with pulmonary arterial hypertension, *Int. J. Cardiol.* 168 (2013)1545–1547.
9. B. Dieplinger, M. Egger, M. Haltmayer, et al., Increased soluble ST2 predicts longterm mortality in patients with stable coronary artery disease: results from the Ludwigshafen risk and cardiovascular health study, *Clin. Chem.* 60 (2014) 530–540.
10. K.M. Eggers, P.W. Armstrong, R.M. Califf, et al., ST2 and mortality in non-ST segment elevation acute coronary syndrome, *Am. Heart J.* 159 (2010) 788–794.
11. Aldous SJ, Richards AM, Troughton R, Than M. ST2 has diagnostic and prognostic utility for all- cause mortality and heart failure in patients presenting to the emergency department with chest pain. *J Card Fail* 2012;18:304- 10. doi: 10.1016/j.cardfail.2012.01.008.
12. Wang YC, Yu CC, Chiu FC, Tsai CT, Lai LP, Hwang JJ, et al. Soluble ST2 as a biomarker for detecting stable heart failure with a normal ejection fraction in hypertensive patients. *J Card Fail* 2013;19:163- 8. doi: 10.1016/j.cardfail.2013.01.010.
13. Henry- Okafor Q, Collins SP, Jenkins CA, Miller KF, Maron DJ, Naftilan AJ, et al. Soluble ST2 as a diagnostic and prognostic marker for acute heart failure syndromes. *Open Biomark J* 2012;2012:1- 8. doi: 10.2174/1875318301205010001.
14. Januzzi JL Jr., Peacock WF, Maisel AS, Chae CU, Jesse RL, Baggish AL, et al. Measurement of the interleukin family member ST2 in patients with acute dyspnea: Results from the PRIDE (Pro- brain natriuretic peptide investigation of dyspnea in the emergency department) study. *J Am Coll Cardiol* 2007;50:607- 13. doi: 10.1016/j.jacc.2007.05.014
15. Bossuyt, P. M., & Reitsma, J. B. (2003). The STARD initiative [1]. *Lancet*. [https://doi.org/10.1016/S0140-6736\(03\)12122-8](https://doi.org/10.1016/S0140-6736(03)12122-8)
16. Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., ... Bossuyt, P. M. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open*. <https://doi.org/10.1136/bmjopen-2016-012799>
17. Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, Vol. 155, pp. 529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
18. <http://www.equator-network.org/wp-content/uploads/2015/03/STARD-2015-checklist.pdf> (Explanation)
19. Narayan, V. M. (2019). A critical appraisal of biomarkers in prostate cancer. *World Journal of Urology*. <https://doi.org/10.1007/s00345-019-02759-x>

20. Grob, A. T. M., van der Vaart, L. R., Withagen, M. I. J., & van der Vaart, C. H. (2017). Quality of reporting of diagnostic accuracy studies on pelvic floor three-dimensional transperineal ultrasound: a systematic review. *Ultrasound in Obstetrics & Gynecology : The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 50(4), 451–457. <https://doi.org/10.1002/uog.17390>
21. Boursier, J., de Ledinghen, V., Poynard, T., Guéchet, J., Carrat, F., Leroy, V., ... AFEF. (2015). An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. *Journal of Hepatology*, 62(4), 807–815. <https://doi.org/10.1016/j.jhep.2014.10.042>
22. Michelessi, M., Lucenteforte, E., Miele, A., Oddone, F., Crescioli, G., Fameli, V., ... Virgili, G. (2017). Diagnostic accuracy research in glaucoma is still incompletely reported: An application of Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015. *PloS One*, 12(12), e0189716. <https://doi.org/10.1371/journal.pone.0189716>
23. Walther, S., Schueler, S., Tackmann, R., Schuetz, G. M., Schlattmann, P., & Dewey, M. (2014). Compliance with STARD checklist among studies of coronary CT angiography: systematic review. *Radiology*, 271(1), 74–86. <https://doi.org/10.1148/radiol.13121720>
24. Bahuleyan, C. G., Alummoottil, G. K., Abdullakutty, J., Lordson, A. J., Babu, S., Krishnakumar, V. V., ... Dilip, M. N. (2018). Prognostic value of soluble ST2 biomarker in heart failure patients with reduced ejection fraction – A multicenter study. *Indian Heart Journal*, 70, S79–S84. <https://doi.org/10.1016/j.ihj.2017.09.010>
25. Jin, X. L., Huang, N., Shang, H., Zhou, M. C., Hong, Y., Cai, W. Z., & Huang, J. (2018). Diagnosis of chronic heart failure by the soluble suppression of tumorigenicity 2 and N-terminal pro-brain natriuretic peptide. *Journal of Clinical Laboratory Analysis*, 32(3), 1–7. <https://doi.org/10.1002/jcla.22295>
26. Mueller, T., Gegenhuber, A., Leitner, I., Poelz, W., Haltmayer, M., & Dieplinger, B. (2016). Diagnostic and prognostic accuracy of galectin-3 and soluble ST2 for acute heart failure. *Clinica Chimica Acta*, 463, 158–164. <https://doi.org/10.1016/j.cca.2016.10.034>
27. Hauser, J. A., Demyanets, S., Rusai, K., Goritschan, C., Weber, M., Panesar, D., ... Michel-Behnke, I. (2016). Diagnostic performance and reference values of novel biomarkers of paediatric heart failure. *Heart*, 102(20), 1633–1639. <https://doi.org/10.1136/heartjnl-2016-309460>
28. Pascual-Figal, D. A., Manzano-Fernández, S., Boronat, M., Casas, T., Garrido, I. P., Bonaque, J. C., ... Januzzi, J. L. (2011). Soluble ST2, high-sensitivity troponin T- and N-terminal pro-B-type natriuretic peptide: Complementary role for risk stratification in acutely decompensated heart failure. *European Journal of Heart Failure*, 13(7), 718–725. <https://doi.org/10.1093/eurjhf/hfr047>
29. Laqqan, M., Schwaighofer, C., Graeber, S., & Raedle-Hurst, T. (2018). Predictive value of soluble ST2 in adolescent and adult patients with complex congenital heart disease. *PLoS ONE*, 13(8), 1–13. <https://doi.org/10.1371/journal.pone.0202406>
30. Aldous, S. J., Richards, A. M., Troughton, R., & Than, M. (2012). ST2 has diagnostic and prognostic utility for all-cause mortality and heart failure in patients presenting to the emergency department with chest pain. *Journal of Cardiac Failure*, 18(4), 304–310. <https://doi.org/10.1016/j.cardfail.2012.01.008>
31. Santhanakrishnan, R., Chong, J. P. C., Ng, T. P., Ling, L. H., Sim, D., Toh G. Leong, K., ... Lam, C. S. P. (2012). Growth differentiation factor 15, ST2, high-sensitivity troponin T, and N-terminal pro brain natriuretic peptide in heart failure with preserved vs.

reduced ejection fraction. *European Journal of Heart Failure*, 14(12), 1338–1347.
<https://doi.org/10.1093/eurjhf/hfs130>

32. Shah, K. B., Kop, W. J., Christenson, R. H., Diercks, D. B., Henderson, S., Hanson, K., ... DeFilippi, C. R. (2011). Prognostic utility of ST2 in patients with acute dyspnea and preserved left ventricular ejection fraction. *Clinical Chemistry*, 57(6), 874–882.
<https://doi.org/10.1373/clinchem.2010.159277>