

SPICED-ACS: Study of the Potential Impact of a Computer-generated ECG Diagnostic Algorithmic Certainty Index in STEMI diagnosis: towards transparent AI

Keywords: algorithm, automated decision-making, automation bias, electrocardiogram, ischaemic heart disease, myocardial infarction

Authors

C.R. Knoery, MBChB^{1,4}, R. Bond, PhD², A. Iftikhar, MSc², K. Rjoob, MSc², V. McGilligan, PhD³, A. Peace, MD, PhD^{3,4}, J. Heaton, PhD¹, S. J. Leslie, FRCP, PhD^{1,5}

Affiliations

- 1. Division of Rural Health and Wellbeing, University of Highlands and Islands, Inverness, UK, IV2 3JH*
- 2. Ulster University, Jordanstown Campus, Shore Rd, Newtownabbey, Northern Ireland, UK, BT37 0QB*
- 3. Centre for Personalised Medicine, Ulster University, Londonderry, Northern Ireland, UK, BT47 6SB*
- 4. Cardiology department, Altnagelvin hospital, Londonderry, Northern Ireland, UK, BT47 6SB*
- 5. Cardiac Unit, Raigmore Hospital, NHS Highland, Inverness, UK, IV2 3UJ*

Correspondence:

Charles Knoery. Division of Rural Health and Wellbeing, University of Highlands and Islands, Centre for Health Science, Old Perth Road, Inverness, UK, IV2 3JH

Email: charles.knoery@uhi.ac.uk

Phone number: +44 1463 279872

Word count: 2,989

Abstract

Background

Computerised electrocardiogram (ECG) interpretation diagnostic algorithms have been developed to guide clinical decisions like with ST segment elevation myocardial infarction (STEMI) where time in decision making is critical. These computer-generated diagnoses have been proven to strongly influence the final ECG diagnosis by the clinician; often called automation bias. However, the computerised diagnosis may be inaccurate and could result in a wrong or delayed treatment harm to the patient. We hypothesise that an algorithmic certainty index alongside a computer-generated diagnosis might mitigate automation bias. The impact of reporting a certainty index on the final diagnosis is not known.

Purpose

To ascertain whether knowledge of the computer-generated ECG algorithmic certainty index influences operator diagnostic accuracy.

Methodology

Clinicians who regularly analyse ECGs such as cardiology or acute care doctors, cardiac nurses and ambulance staff were invited to complete an online anonymous survey between March and April 2019. The survey had 36 ECGs with a clinical vignette of a typical chest pain and which were either a STEMI, normal, or borderline (but do not fit the STEMI criteria) along with an artificially created certainty index that was either high, medium, low

or none. Participants were asked whether the ECG showed a STEMI and their confidence in the diagnosis. The primary outcomes were whether a computer-generated certainty index influenced interpreter's diagnostic decisions and improved their diagnostic accuracy. Secondary outcomes were influence of certainty index between different types of clinicians and influence of certainty index on user's own-diagnostic confidence.

Results

A total of 91 participants undertook the survey and submitted 3,262 ECG interpretations of which 75% of ECG interpretations were correct. Presence of a certainty index significantly increased the odds ratio of a correct ECG interpretation (OR 1.063, 95% CI 1.022-1.106, $p=0.004$) but there was no significant difference between correct certainty index and incorrect certainty index (OR 1.028, 95% CI 0.923-1.145, $p=0.615$). There was a trend for low certainty index to increase odds ratio compared to no certainty index (OR 1.153, 95% CI 0.898 – 1.482, $p=0.264$) but a high certainty index significantly decreased the odds ratio of a correct ECG interpretation (OR 0.492, 95% CI 0.391 – 0.619, $p<0.001$). There was no impact of presence of a certainty index ($p=0.528$) or correct certainty index ($p=0.812$) on interpreters' confidence in their ECG interpretation.

Conclusions

Our results show that the presence of an ECG certainty index improves the users ECG interpretation accuracy. This effect is not seen with differing levels of confidence within a certainty index, with reduced ECG interpretation success with a high certainty index compared with a trend for increased success with a low certainty index. This suggests that a certainty index improves interpretation when there is an increased element of doubt, possibly forcing the ECG user to spend more time and effort analysing the ECG. Further research is

needed looking at time spent analysing differing certainty indices with alternate ECG diagnoses.

Introduction

Electrocardiogram (ECG) interpretation is the cornerstone of diagnosis of a ST-elevation myocardial infarction (STEMI). There are significant risks in delayed STEMI treatment[1] or misdiagnosis with risk of dangerous and unnecessary interventions such as coronary angiography or thrombolysis with a mimicking diagnosis[2]. To reduce delays and improve diagnosis, it has been shown that a prehospital ECG can help reduce time-to-treatment for STEMI and improve mortality[3].

However, ECG interpretation is acknowledged to be difficult due to the influence of background noise, subtle abnormalities and complex interpretation with artefact, temporal, spatial and vectoral concepts requiring increased cognitive workload[4]. Furthermore, human factors like impulsivity can lead to biases like anchoring (over-reliance on initial impressions) and confirmation (favouring information to confirm previously held belief)[5] which can further hinder judgement. Thus the accuracy of ECG interpretation can be as low as 40-50%[6].

Computerised diagnostic algorithms can assist clinical decision making by generating a automated diagnosis. Algorithms have been developed further to integrate with human interpretation to provide a differential diagnosis[7]. These computer-generated diagnoses have been proven to strongly influence the clinicians' final ECG diagnosis[8,9]. However, the computerised diagnosis may be inaccurate and often be unable to adapt to the clinical scenario[10]. Thus, a wrong computer diagnosis can negatively influence clinical decision making[11]. The negative influence of an automation process on human decision making is termed automation bias, leading to potential misguided clinical decisions[11].

The computer diagnosis is usually presented in a standard manner with little option for describing the computers confidence in its own diagnosis. The computers confidence in its own automated interpretation could be presented as a graded scale known as a ‘certainty index’. If this ‘certainty index’ were presented this might mitigate against automation bias leading to the wrong diagnosis. Currently, modern ECGs already have diagnostic modifiers such as “possible”, “probable” and “consider” to compensate for diagnostic inaccuracy that might arise from the algorithms that may use set-values, such as with q-wave depth or QRS duration[12]. However, these modifiers convey only a crude sense of scale of the computers confidence and the impact of these diagnostic modifiers is unknown. In aviation, the influence of certainty index on aeroplane diagnostic computerised systems has been shown to be beneficial in preventing pilots stalling in flight-simulations[13]. However, the impact of algorithmic certainty index on ECG diagnosis is not known. The aim of this study was to ascertain whether computer-generated ECG certainty indexes can improve operator diagnostic accuracy.

Methodology

Study protocol

This was a feasibility controlled study and was undertaken by interpreters using an online survey tool[14]. Basic demographic details were recorded, such as age, gender, clinical role, time in clinical role, time since qualification and self-reported ECG skills (years of experience). A pilot trial of the survey was run beforehand to identify any potential problems or errors with the survey and it was refined accordingly.

In order to answer the specific research question, interpreters were shown a series of 36 ECGs in random order as determined by a computerised random number generator.

Anonymised ECGs had any patient identifying data along with the prior computer diagnostic results removed. Anonymised ECGs were then assessed by an experienced cardiologist, with a background of research in ECG analysis, to determine whether they had changes consistent with STEMI, normal sinus or features consistent with ischaemia but not fulfilling STEMI criteria (“borderline”). There were 12 normal ECG with normal sinus rhythm, 12 ECGs with degrees of change but not classifiable as a STEMI (“borderline”) according to the ESC STEMI guidelines[15] and 12 ECGs with voltage criteria fulfilling STEMI classification (figure 1). Left bundle branch block ECGs were not included as often a prior ECG is needed to clarify whether the left bundle branch block is a new or old. Each ECG had a computerised diagnosis with an accompanying statement of ‘Computer certainty index for a STEMI is ...’ followed by the percentage algorithmic certainty index conveying the computer’s certainty in its own diagnosis and a text scale. The certainty index was randomly generated within the designated range with the low range between 0 and 20%, medium between 30% to 60% and high between 80% to 100%. ECGs without a percentage algorithmic certainty index acted as the controls.

There was a vignette presented at the start, intended to accompany each ECG interpretation. The vignette was ‘A 45 year old man calls 999 as he is describing anterior chest pain, he has no past medical history. He is sweaty and described the pain as sudden onset. This is the ECG taken by the paramedics done on scene’. This vignette is a typical history and has the high potential to be a myocardial infarction or also another cause. A history of potentially ambiguous symptoms was intended to help the interpreters focus on distinguishing ECG features to help establish a diagnosis. Each ECG was accompanied by 2 questions for interpreters to answer. Firstly, “are the above ECG findings consistent with a diagnosis of a ST-elevation myocardial infarction?” answered as either yes or no. Secondly, “how

confident are you in your diagnosis whether the above ECG is or is not a STEMI?”. The question was answered on a scale from 1 to 5.

Research questions

The primary questions of the study were:

1. Does a computer-generated diagnostic algorithmic certainty index influence and improve ECG interpreter’s diagnosis success rate?

The secondary questions of the study were:

1. What is the influence of algorithmic certainty index between differing grades and types of clinicians in diagnostic accuracy?
2. What is the influence of algorithmic certainty index between ECG interpreters and their own confidence in the diagnosis?

Participants and recruitment

Target participants included health care practitioners who would have had training and experience in ECG interpretation of suspected myocardial infarction. A convenience sample were invited to participate the online survey via a web-link, distributed by local advertising, social media, and professional bodies. Invitations were sent out by email to the appropriate organisations with details of the study and web-link to the online survey. Additional letters of invitation were also sent to general practitioner practices with the web-link attached. It was estimated that 100 to 200 participants would respond to the survey.

Participants were divided into 5 groups determined by their clinical roles:

1. Cardiology doctors
2. Non-cardiology doctors

3. Nursing and allied healthcare staff routinely involved in cardiac care
4. Pre-hospital practitioners including paramedics and technicians who both have training and experience in ECG interpretation
5. Nurse practitioners

Statistical analysis.

As this was a feasibility study, there was no set number of participants required to achieve an adequate statistical power. Initial data was entered onto SPSS™ version 24 (IBM, Armonk NY) for analysis. Crosstabs with Pearson Chi-squared testing and logistic regression was used to determine odds ratios, 95% confidence intervals (95% CI) and significance for categorical variables. Independent sample t-test was used to determine statistical significance for continuous variables. A p-value less than 0.05 was considered significant.

To analyse the impact of a correct or incorrect certainty index on ECG interpretation success, ECGs with certainty index were classified as correct or incorrect. A correct certainty index was determined to be an ECG that was a STEMI displaying a high certainty index or an ECG which were normal sinus rhythm or borderline ischaemic features and displayed a low certainty index. ECGs which showed a STEMI but displayed a low or medium certainty index and ECGs with sinus rhythm which displayed a medium or high certainty index were classified as incorrect. ECGs with no certainty index or medium certainty index and borderline ECG were not included in analysis for the impact of correct certainty index compared to incorrect certainty index.

Ethics

NHS Research and Development Management approval was obtained. The initial page of the survey outlined the study purpose, information regarding data storage and handling, privacy statement and further contact details. By clicking proceed, it was judged as a surrogate for providing consent to continue with the survey. Participants were able to contact the study researchers and ask for their answers to be removed retrospectively.

Results

Demographics

The demographics of the respondents are outlined in supplementary table 1. Between March 2019 and April 2019 there were 91 replies, which totalled 3,262 ECG interpretations, with 14 interpretations missing due to no answer entered. The biggest group of respondents were pre-hospital practitioners followed by non-cardiology doctors. Baseline characteristics of participants were 63% male and the average age was 43 years. The average time since qualification was 17 years and the most frequent self-reported ECG ability was ‘good’ with only nurse practitioners reporting their skills as ‘average’ and non-cardiology doctors reporting equal amounts of ‘average’ or ‘good’. Overall, 75% of respondents were correct with non-cardiology doctors having the highest score of 78%.

Comparing ECG with certainty index to without certainty index

Table 1 outlines the scores of respondents with ECGs with a certainty index compared to those without a certainty index. There were 2535 ECGs interpretations with a certainty index and 727 without. From the 2535 with certainty index, 1868 were answered correctly.

Overall, the presence of a certainty index significantly increased the odds ratio of a correct ECG interpretation (OR 1.063, 95% CI 1.022 – 1.106, $p=0.004$) and this finding was also seen with the cardiology doctors (OR 1.156, 95% CI 1.075 – 1.244, $p=0.001$). The other

clinical roles had a trend for increased odds ratio but this was not significant, with only cardiology doctors being significant with a large difference in their p-value compared to other clinical roles. Although the significance effect may have only been seen due to the high number of ECG interpretations in the cardiology doctors group and may have been seen in other groups with larger numbers.

Correct certainty index compared to incorrect certainty index

Table 1 outlines the scores of respondents with ECGs with a correct certainty index compared to an incorrect certainty index. There were 906 ECGs interpretations had a correct certainty index, 1086 had an incorrect certainty index and 1274 were classified as missing (i.e. not incorrect or correct). There was a trend for increased odds ratio for correct ECG interpretations with a correct certainty index although this did not reach statistical significance (OR 1.028, 95% CI 0.923 – 1.145, p=0.615). There was a trend for increased odds ratio for correct ECG interpretation within the clinical roles although this was not significant. The exceptions were with non-cardiology doctors and ambulance staff, who both had a decreased odds ratio for correct ECG interpretation when presented with an ECG with a correct certainty index but this did not reach significance.

Impact of varying certainty index on ECG interpretation success

Table 2 displays the impact of no certainty index and a low, medium and high certainty index on ECG interpretation. Of the 3626 ECGs in the survey, 727 had no certainty index, 815 had a low certainty index, 906 had a medium certainty index and 814 had a high certainty index. Of the 3263 ECG interpretations, 2442 were answered correctly. There was a trend for increased odds ratio for correct interpretation of ECGs displaying a low certainty index compared to no certainty index although this not achieve significance (OR 1.153, 95% CI

0.898 – 1.482, $p = 0.264$). There was a trend for decreased odds ratio for correct interpretation of ECGs displaying a medium certainty index (OR 0.793, 95% CI 0.628 – 1.001, $p = 0.051$) and a high certainty index (OR 0.492, 95% CI 0.391 – 0.619, $p < 0.001$) compared to ECGs with no certainty index, although only ECGs with a high certainty index reached significance.

Influence of certainty index on ECG interpreter's confidence in answers

Figure 2 displays the impact of a certainty index on user confidence with no significant effect ($p = 0.528$) of a certainty index on ECG interpreter's confidence in their own answers. Figure 3 displays the impact a correct certainty index has on user's confidence, with no significant impact seen ($p = 0.812$).

Discussion

As expected, our results show the presence of a certainty index improved the performance of ECG interpretation. However, there was no significant difference in the impact of a correct certainty index on ECG interpretation success. When looking at the influence of no certainty index compared to low, medium or high certainty index, there was a trend for an inverse relationship. A lower certainty index appeared to have an increased odds ratio for correct ECG interpretation compared to higher certainty index but this only significant with no and high certainty index.

The positive influence of a certainty index on automated decision making have already been shown within the aviation industry where an updated certainty index improved flying ability[13]. In addition, this study also suggested that a low or variable (equivalent to medium) certainty index lead to flexible responses, indicating that a higher level of

uncertainty increased appropriate interventions. Our own results similarly found that a low certainty index led to improved ECG interpretation accuracy. This could suggest that when presented with a certainty index which is low, the human reflex was to rely less on the automated decision and therefore, help mitigate the potential for automation bias. The effect of automation bias could possibly be seen with the decreased odds ratio for ECG interpretation success with a high certainty index, which was significant. In addition, when presented with an automated system which has low certainty, it could be speculated that the ECG interpreter is spending more time analysing the ECG or gaining additional help and thus more likely to identify any abnormalities. Further analysis and research could look into analysing the time spent with each ECG with the hypothesis being that interpreters spend more time analysing ECGs with low certainty index.

When comparing clinical roles, the highest scores were achieved by non-cardiology doctors, with cardiology doctors with the second highest score although the difference is 1% and most likely falls within the standard error of the results. However, when looking at the influence of the presence of certainty index on ECG interpretation on individual clinical roles, the cardiology doctors were the only role that significantly improved with the presence of a certainty index compared to no certainty index (although there is a trend for improvement with all roles). This could be due to the potential increased time analysing the ECG with a low or conflicting certainty index, which would lead the more experienced ECG interpreters, such as cardiology doctors, to notice nuances in the ECGs. There was no significant influence between a correct certainty index and incorrect certainty index on ECG interpretation overall and between roles. It appears that the influence of a certainty index is most prominent when presenting differing certainty rather than a correct or incorrect judgement. There was no significant impact of a certainty index on ECG interpreter's confidence in their own answers,

potentially indicating the mitigating effect of a certainty index on automation bias. A further explanation could be that the ECG interpreter's over-reliance in the computer interpretation could be counter-balanced with encouraged use of the interpreters own diagnostic skill and the computer diagnosis acting as an aid rather than a definitive tool.

Currently, there is no pre-defined algorithm for ECG certainty index. However, there would be the potential to develop an algorithm based on factors such as disease prevalence, known accuracy for suggested ECG findings, background noise and lead misplacement. Such an algorithm could be implemented into a computerised ECG machine and a certainty index could be displayed alongside the suggested diagnosis. Other ways to help counter automation bias could include the use of touch-screen technology to provide a more user-friendly interface and provide prompts alongside a clinical decision tool to augment decision making[16].

Limitations

This study had a number of limitations. Firstly, we can only assume that the displayed certainty index had an influential effect as there is a possibility it was ignored and the interpreter concentrated solely on the ECG. However, the significance of results when judging the impact between ECG interpretation with and without certainty index displayed suggests that a certainty index had an impact on the decision making. Secondly, the ability to interpret the ECG online is very different from doing so in a clinical environment where a varying amount of additional information and distractions would be present. Therefore, these results must be taken with caution when comparing to a clinical environment. Fourthly, the reference standard only had one cardiologist to validate that the ECGs fulfilled the STEMI criteria. To help compensate for this, the ECG analysis was done in a methodical process,

following the ECS defined criteria for STEMI[15] alongside the principal researcher. In addition, the ECGs and their correct interpretation were analysed by another consultant cardiologist and a senior lecturer who specialises in ECG analysis during the pilot trial. Any ECGs which were thought to be incorrectly labelled as a STEMI were removed and replaced with another until there was concordance. Finally, there was a limited number of respondents to the study and coupled with a high amount of negative results suggests that there is a high chance of type 2 error with the findings.

Future research

The premise of this study was as a feasibility study and therefore it could be repeated with greater numbers with increased distribution. Development of a clinical decision support system and software to aid in the management of myocardial infarction is of growing interest[17]. Along with point-of-care biomarkers[18], computerised ECG interpretation has the potential for significant improvement in outcomes[19]. Further work could look at the use of digital eyewear[20] to analyse the time spent looking at the ECG computerised diagnosis and certainty index. This study was also just limited to STEMI, however it could be used for a variety of ECG findings.

Conclusion

This study illustrates the potential impact that a certainty index has on computerised ECG diagnosis interpretation and its benefits in helping to mitigate against automation bias. Further work is required to develop a potential algorithm for certainty index utilisation on a variety of ECGs and their computerised diagnosis to help aid ECG interpretation and diagnosis.

Declarations

C Knoery, R Bond, A Iftikhar, K Rjoob, A Peace, V McGilligan and S Leslie are supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB). Project: Centre for Personalised Medicine – Clinical Decision Making and Patient Safety.

References

- [1] De Luca G, Suryapranata H, Ottervanger JP, Antman EM. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. *Circulation* 2004;109:1223–5. doi:10.1161/01.CIR.0000121424.76486.20.
- [2] Means G, End C, Kaul P. Management of Percutaneous Coronary Intervention Complications. *Curr Treat Options Cardiovasc Med* 2017;19:25. doi:10.1007/s11936-017-0526-6.
- [3] Brown JP, Mahmud E, Dunford JV, Ben-Yehuda O. Effect of Prehospital 12-Lead Electrocardiogram on Activation of the Cardiac Catheterization Laboratory and Door-to-Balloon Time in ST-Segment Elevation Acute Myocardial Infarction. *The American Journal of Cardiology* 2008;101:158–61. doi:10.1016/j.amjcard.2007.07.082.
- [4] Sibbald M, de Bruin ABH, van Merrienboer JJG. Checklists improve experts' diagnostic decisions. *Med Educ* 2013;47:301–8. doi:10.1111/medu.12080.
- [5] Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate? *Behav Brain Sci* 2000;23:645–65; discussion 665-726.
- [6] Fisch C. Clinical competence in electrocardiography. A statement for physicians from the ACP/ACC/AHA Task Force on clinical privileges in cardiology. *J Am Coll Cardiol* 1995;25:1465–9.
- [7] Cairns AW, Bond RR, Finlay DD, Guldenring D, Badilini F, Libretti G, et al. A decision support system and rule-based algorithm to augment the human interpretation of the 12-lead electrocardiogram. *J Electrocardiol* 2017;50:781–6. doi:10.1016/j.jelectrocard.2017.08.007.
- [8] Tsai TL, Fridsma DB, Gatti G. Computer Decision Support as a Source of Interpretation Error: The Case of Electrocardiograms. *J Am Med Inform Assoc* 2003;10:478–83. doi:10.1197/jamia.M1279.
- [9] Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 2012;19:121–7. doi:10.1136/amiajnl-2011-000089.
- [10] Estes NAM. Computerized interpretation of ECGs: supplement not a substitute. *Circ Arrhythm Electrophysiol* 2013;6:2–4. doi:10.1161/CIRCEP.111.000097.

- [11] Bogun F, Anh D, Kalahasty G, Wissner E, Bou Serhal C, Bazzi R, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med* 2004;117:636–42. doi:10.1016/j.amjmed.2004.06.024.
- [12] Smulyan H. The Computerized ECG: Friend and Foe. *Am J Med* 2019;132:153–60. doi:10.1016/j.amjmed.2018.08.025.
- [13] McGuirol JM, Sarter NB. Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information. *Hum Factors* 2006;48:656–65. doi:10.1518/001872006779166334.
- [14] Jisc. Online surveys n.d. <https://www.onlinesurveys.ac.uk/> (accessed November 20, 2018).
- [15] Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). *Eur Heart J* 2018. doi:10.1093/eurheartj/ehy462.
- [16] Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, et al. Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. *J Electrocardiol* 2018;51:S6–11. doi:10.1016/j.jelectrocard.2018.08.007.
- [17] Nehme Z, Boyle MJ, Brown T. Diagnostic accuracy of prehospital clinical prediction models to identify short-term outcomes in patients with acute coronary syndromes: a systematic review. *J Emerg Med* 2013;44:946-954.e6. doi:10.1016/j.jemermed.2012.07.078.
- [18] van Dongen DN, Tolsma RT, Fokkert MJ, Badings EA, van der Sluis A, Slingerland RJ, et al. Pre-hospital risk assessment in suspected non-ST-elevation acute coronary syndrome: A prospective observational study. *Eur Heart J Acute Cardiovasc Care* 2018;2048872618813846. doi:10.1177/2048872618813846.
- [19] Bhalla MC, Mencl F, Gist MA, Wilber S, Zalewski J. Prehospital electrocardiographic computer identification of ST-segment elevation myocardial infarction. *Prehosp Emerg Care* 2013;17:211–6. doi:10.3109/10903127.2012.722176.
- [20] Stetler J, Resendes E, Martinez-Parachini JR, Patel K, Amsavelu S, Tarar MNJ, et al. Hands-free zoom and pan technology improves the accuracy of remote electrocardiogram interpretation using Google Glass. *Int J Cardiol* 2016;204:147–8. doi:10.1016/j.ijcard.2015.11.144.