Accuracy of gestalt perception of acute chest pain in predicting coronary artery disease

Cláudio Marcelo Bittencourt das Virgens, Laudenor Lemos Jr, Márcia Noya-Rabelo, Manuela Campelo Carvalhal, Antônio Maurício dos Santos Cerqueira Junior, Fernanda Oliveira de Andrade Lopes, Nicole Cruz de Sá, Jéssica Gonzalez Sueldieck, Thiago Menezes Barbosa de Souza, Vitor Calixto de Almeida Correia, Gabriella Sant’Ana Sodré, André Barcelos da Silva, Felipe Kalil Beirão Alexandre, Felipe Rodrigues Marques Ferreira, Luís Cláudio Lemos Correia

Data sharing statement: Technical details and statistical methods are available with the corresponding author at luisclcorreia@gmail.com. Participants gave informed consent for data sharing.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Manuscript source: Invited manuscript

Correspondence to: Luís Cláudio Lemos Correia, MD, PhD, Research Coordinator of Hospital São Rafael, Associate Professor of Bahiana School of Medicine and Public Health, Department of Cardiology, Hospital São Rafael, Av. Princesa Leopoldina 19/402, Salvador, Bahia 41253-190, Brazil. luisclcorreia@gmail.com
Telephone: +55-71-99971032

Received: September 14, 2016
Peer-review started: September 18, 2016
First decision: November 14, 2016
Revised: December 15, 2016
Accepted: January 2, 2017
Article in press: January 3, 2017
Published online: March 26, 2017

Abstract

AIM
To test accuracy and reproducibility of gestalt to predict
obstructive coronary artery disease (CAD) in patients with acute chest pain.

METHODS
We studied individuals who were consecutively admitted to our Chest Pain Unit. At admission, investigators performed a standardized interview and recorded 14 chest pain features. Based on these features, a cardiologist who was blind to other clinical characteristics made unstructured judgment of CAD probability, both numerically and categorically. As the reference standard for testing the accuracy of gestalt, angiography was required to rule-in CAD, while either angiography or non-invasive test could be used to rule-out. In order to assess reproducibility, a second cardiologist did the same procedure.

RESULTS
In a sample of 330 patients, the prevalence of obstructive CAD was 48%. Gestalt's numerical probability was associated with CAD, but the area under the curve of 0.61 (95%CI: 0.55-0.67) indicated low level of accuracy. Accordingly, categorical definition of typical chest pain had a sensitivity of 48% (95%CI: 40%-55%) and specificity of 66% (95%CI: 59%-73%), yielding a negligible positive likelihood ratio of 1.4 (95%CI: 0.65-2.0) and negative likelihood ratio of 0.79 (95%CI: 0.62-1.02). Agreement between the two cardiologists was poor in the numerical classification (95% limits of agreement = -71% to 51%) and categorical definition of typical pain (Kappa = 0.29; 95%CI: 0.21-0.37).

CONCLUSION
Clinical judgment based on a combination of chest pain features is neither accurate nor reproducible in predicting obstructive CAD in the acute setting.

Key words: Acute chest pain; Clinical judgment; Gestalt; Coronary artery disease; Acute coronary syndrome

© The Author(s) 2017. Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: In the scenario of acute chest pain, individual features of chest pain presentation are intuitively combined to form physician’s impression, by a process called “gestalt”. Physicians commonly assess probability of disease by unstructured clinical judgment. Although commonly used and presumed to be accurate, diagnostic assessment by gestalt of acute chest pain lacks validation. In the present manuscript, we investigated the accuracy of gestalt in the prediction of coronary artery disease (CAD). Our results indicate that clinical judgment (gestalt) of acute chest pain characteristics has low diagnostic accuracy for obstructive CAD. Thus, physicians should be cautious when relying on chest pain characteristics and investigators should redirect their focus to identify validated predictors.
nature, radiation to left arm, radiation to neck, severe intensity, similarity to previous infarction (if applicable), presence of vagal symptoms, worsening with body movement, worsening with palpation, worsening with arms movement, worsening with deep breath, and relief by nitrate. Characteristics were considered positive if patient’s answer was clearly affirmative. Dubious answers (“maybe”, “sometimes”, “not sure”) were taken as negative. In addition, 3 numeric variables were recorded: Intensity of chest pain from 0 to 10 (defined by the patient according to a visual scale), number of pain episodes at rest and duration of the longest episode in minutes. No additional information regarding demographic or clinical characteristics was recorded on this form.

Subsequently, a cardiology faculty member (CV, with 23 years of experience in the field of acute chest pain) assessed the forms and classified chest pain according to the 14 characteristics. This investigator did not have any contact with the patients, and was completely blind to additional information such as name, gender, age, previous history or additional tests. This method guaranteed that medical judgment was based exclusively on chest pain characteristics. In order to assess reproducibility of medical judgment, the same procedure was independently performed by a second faculty member (LLJ) in all patients and his classification was compared with the first.

Chest pain was classified in four ways: (1) typical or atypical; (2) Non-anginal, Undefined or Anginal Chest Pain; (3) definitely angina, probably angina, probably not or definitely not; and (4) numeric probability of coronary etiology from 0 to 100. No objective predefinition of these classifications was provided to the evaluators of chest pain, enabling the definition to be a result of the physician’s unstructured discretion. This method guaranteed that answers reflected authentic clinical judgment.

Obstructive CAD
Outcome data was collected by 3 other independent investigators (MC, FK, FF) and adjudicated by a fourth investigator (LC). Obstructive CAD was defined by a stenosis \( \geq 70\% \) on angiography. For diagnostic evaluation, patients underwent invasive coronary angiography or a non-invasive test (perfusion magnetic resonance imaging or nuclear single-photon emission computed tomography), at the discretion of the assistant cardiologist. In case of a positive non-invasive test, patients underwent angiography for confirmation. A negative non-invasive test indicated absence of obstructive CAD and no further test was required. In case of a dominant alternative diagnosis as confirmed by imaging (such as pericarditis, pulmonary embolism, aortic dissection or pneumonia), the etiology was defined as non CAD.

Statistical analysis
Frequencies were compared by Pearson’s \( \chi^2 \) test and means by Student’s t test. The accuracy of clinical judgment in predicting CAD was described by point-estimate and 95%CI of sensitivity, specificity, likelihood ratios and predictive values. The accuracy of numerical estimative of CAD probability was described by the area under the ROC curve with 95%CI.

For analysis of reproducibility, the Kappa test was utilized to assess agreement between two observers regarding the different forms of categorical classification. For numeric estimation of CAD probability, the Bland-Altman analysis was used: mean absolute error between the two observers (mean of differences without the signal), mean signed difference (bias) and 95% limits of agreement.

Sample size was calculated based on an expected CAD prevalence of 50%. Thus, a sample size of 300 would provide 150 patients with and 150 patients without CAD. Considering assumptions of 70% sensitivity and specificity, 150 patients would yield a \( \pm 8\% \) precision for the 95%CI.

RESULTS

Sample characteristics
From 2011 to 2013, a sample of 330 patients was studied, 59 ± 15 years old, 58% males, 54% presented ischemic electrocardiographic changes and 48% had positive troponin. All individuals had gestalt evaluation and reference standard performed during the same admission. Obstructive CAD was identified according to study protocol in 48% of the individuals. Baseline characteristics are depicted on Table 1.

Accuracy of clinical judgment
Typical vs atypical chest discomfort: Chest discomfort was classified as typical in 41% of patients. Obstructive CAD was present in 56% of individuals with typical symptoms, compared with 42% of those with atypical symptoms (\( P = 0.02 \)). Among 158 individuals with obstructive CAD, the discomfort was defined as typical in 75, yielding a sensitivity of 48% (95%CI: 40%-55%). Conversely, in 172 individuals free of CAD, 113 had symptoms defined as atypical, leading to a specificity of 66% (95%CI: 59%-73%). Consequently, typical pain had a negligible positive likelihood ratio of 1.4 (95%CI: 0.65-2.0), as well as a negative likelihood ratio of 0.79 (95%CI: 0.62-1.02). The positive predictive value of typical chest pain was only 56% (95%CI: 48%-64%), while the negative predictive value was 58% (95%CI: 51%-65%), Table 2.

Non-anginal, undefined or anginal chest pain: Patients were equally distributed among the 3 classifications, 36% defined as non-anginal, 34% as undefined and 30% as anginal. Prevalence of CAD was respectively 38%, 49% and 55% (\( P = 0.04 \)). Among 158 individuals with CAD, only 66 had anginal pain, leading to a sensitivity of 42% (95%CI: 34%-50%), positive likelihood ratio of 1.35 (95%CI: 0.89-2.1) and...
Table 1  Clinical characteristic n (%)  

| Variable                              | Description |
|---------------------------------------|-------------|
| Sample size                           | 330         |
| Age (yr)                              | 59 ± 15     |
| Male gender                           | 192 (58)    |
| History of coronary disease           | 112 (34)    |
| Diabetes                              | 104 (32)    |
| Ischemia on EKG                       | 179 (54)    |
| Positive troponin                     | 157 (48)    |
| Signs of left ventricular failure     | 28 (8.5)    |
| Final diagnosis                       |             |
| Unstable angina                       | 52 (16)     |
| Myocardial infarction                 | 142 (43)    |
| No CAD, but undefined diagnosis       | 103 (31)    |
| Gastrointestinal disorder             | 5 (1.5)     |
| Osteo-muscular disorder               | 1 (0.3)     |
| Pericarditis                          | 12 (3.6)    |
| Pulmonary embolism                    | 2 (0.6)     |
| Aortic dissection                     | 2 (0.6)     |
| Pneumonia                             | 2 (0.6)     |
| Other                                 | 9 (2.7)     |

CAD: Coronary artery disease.

positive predictive value of 56% (95%CI: 47%-65%). Conversely, in 172 individuals free of CAD, 62 had symptoms defined as non-anginal, leading to a specificity of 36% (95%CI: 29%-43%), negative likelihood ratio of 0.67 (95%CI: 0.40-1.1) and negative predictive value of 62% (95%CI: 53%-62%) (Table 2).

Definitely angina, probably angina, probably not and definitely not: Patients were equally distributed among the 4 categories, with 25% classified as definitely angina, 32% as probably angina, 23% as probably not and 20% as definitely not. Prevalence of CAD was similar among the first 3 groups, respectively 49%, 56% and 51%, while patients classified as definitive no-angina had a lower prevalence of 30%, which was responsible for the statistical difference among the 4 groups (P = 0.008). Thus, the threshold of definitely not was not utilized for accuracy. Among 158 individuals with CAD, 138 were classified as definitely not, yielding 27% specificity (95%CI: 20%-34%). Thus, the negative likelihood ratio of definitely not was a negligible 0.63 (95%CI: 0.32-1.15), with a negative predictive value of 70% (95%CI: 59%-81%) (Table 2).

Subjective estimation of CAD probability: Probability of CAD had a mean of 59% ± 34%, with a median of 70% (interquartile range = 30%-90%). Individuals with CAD had a median probability of 80% (interquartile range = 50%-95%), compared with 60% in patients free of disease (interquartile range = 10%-90%) - P < 0.001. The diagnostic area under the ROC for numeric probability was 0.61 (95%CI: 0.55-0.67) (Figure 1).

Reproducibility of clinical judgment
The two observers agreed in 62% of the patients regarding typical vs atypical chest pain, yielding a weak Kappa of 0.29 (95%CI: 0.21-0.37; P < 0.001). For non-anginal, undefined or anginal chest pain, agreement was 53% (Kappa = 0.28; 95%CI: 0.20-0.36; P < 0.001). For definitely angina, probably angina, probably not and definitely not, agreement was 42%, leading to a weak Kappa of 0.21 (95%CI: 0.14-0.28; P < 0.001).

Regarding numeric estimation of probability, mean absolute error was 23% ± 23%, with a mean signed difference (bias) of - 9.7% ± 31%, with 95% limits of agreement from - 71% to + 51%. The Bland-Altman plot showed a diamond pattern with reasonable agreement in very low (< 20%) or very high (> 80%) ranges of probability, with increasing disagreement as probability becomes more intermediate (Figure 2).

DISCUSSION
The present study indicates that clinical judgment (gestalt) of acute chest pain has low diagnostic accuracy for obstructive CAD. In addition, there was poor agreement between the gestalt of two physicians, indicating low precision of intuitive interpretation of chest pain features. These findings confront the common belief that physicians should take into account the typicality of symptoms when evaluating patients with acute chest pain.

Our primary interest was to assess the role of chest pain features on clinical evaluation. Thus, our methods were designed to evaluate accuracy of clinical judgment that comes specifically from chest pain characteristics, as opposed to the entire clinical presentation. In order to do this, we blinded the physician to demographics, clinical characteristics or patient’s appearance. Secondly, we tested physician’s intuitive judgment that comes from the combination of all features, instead of the accuracy of specific symptom characteristics. Thus, there was not an a priori criterion for classifying chest pain, allowing the physician to use his own intuition (unstructured clinical judgment).

Physicians commonly assess probability of disease by unstructured clinical judgment. Although medical doctors normally put confidence into this type of judgment, it tends to be inaccurate. As described by Nobel Prize laureates and psychologists Kahneman and Tversky, judgment under uncertainty is vulnerable to cognitive bias, due to heuristics utilized in the process of intuitive thinking[6]. A common example of heuristics is “representativeness”: If A resembles B, when A is present we think B is highly probable to be present. Oppressive chest pain resembles angina. Hence, a physician may jump to conclude that a patient with oppressive chest pain has a high probability of CAD. However, the likelihood ratio of oppressive chest pain is very low. These cognitive biases that are present in intuitive thinking explain why mechanical models are
usually better predictors than medical judgment. For instance, a systematic review of several medical and non-medical situations consistently showed better predictability of probabilistic models, in comparison with specialist’ decision[1]. Therefore, in order to avoid heuristics when evaluating a chest pain scenario, physicians should increase awareness of the low diagnostic value of chest pain characteristics or invest in probabilistic models able to predict obstructive disease more precisely.

The lack of reproducibility between two independent cardiologists also deserves attention. While lack of accuracy promotes diagnostic errors, lack of agreement impairs consensus regarding medical management. Thus, relying too much on chest pain characteristics does not only promote probabilistic errors, but also promote differences in clinical impressions, leading to confusion and discordance among the medical team.

Although an experienced physician made clinical judgment, we cannot guarantee that his analysis is similar to most physicians. In fact, this would be unlikely, considering the low level of reproducibility found in our head-to-head comparisons. Nevertheless, the concept of accuracy is somewhat independent of agreement. Accuracy depends on the proportion of correct predictions. Two models can have the same proportion of correct predictions and not be related to the same patients. Indeed, we should not expect different people to have the same intuition regarding diagnosis. This rationale is the basis for testing the concept of accuracy of physician judgment by using one specific professional as a proxy of the average physician. Nevertheless, we recognize that further studies are needed to validate our findings, extending it to different populations of

| Classification | Sensitivity | Specificity | Positive LR | Negative LR | Positive PV | Negative PV |
|----------------|-------------|------------|-------------|-------------|-------------|-------------|
| Typical (vs atypical) | 48% (40%-55%) | 66% (59%-73%) | 1.4 (0.65-2.0) | 0.79 (0.62-1.02) | 56% (48%-64%) | 58% (51%-65%) |
| Classification 2 (3-level)
  Angina (vs undefined/no-angina) | 42% (34%-50%) | 69% (62%-76%) | 1.35 (0.89-2.1) | 56% (47%-65%) |
  No-angina (vs undefined/no-angina) | 76% (69%-83%) | 36% (29%-43%) | 0.67 (0.40-1.1) | 62% (53%-72%) |
| Classification 3 (4-level)
  Definitely no-angina (vs the other 3 groups) | 27% (20%-34%) | 0.63 (0.32-1.15) | 70% (59%-81%) |

Numbers in parenthesis are 95%CI. 1Because only no-angina distinguished itself from the other 3 classifications, only specificity, negative LR and negative PV were calculated. LR: Likelihood ratio; PV: Predictive value.

Figure 1  Categorical and numerical accuracy of gestalt, regarding the prediction of obstructive coronary artery disease. CAD: Coronary artery disease.
March 2, Nagurney JT. Value and limitations of chest pain on chest pain characteristics and investigators should of CAD. Physicians should be cautious when relying acuity and reproducibility in estimating the probability sentativeness heuristics. Therefore, the use of probabilistic models is supposed to be a more effective way to avoid repre- of cases. Therefore, the use of probabilistic models is of CAD and mismanagement in under or overdiagnosis of CAD and the current categorization may be misleading, resulting by these subjective classifications. The overvaluing of differences on predicting CAD probability has important significance. Firstly, our findings do not undermine the value of the history as a whole, because our analysis only refers to chest pain characteristics. These classifications merely to chest pain characteristic and this isolated aspect has little or no influence on probability of CAD. A sense of surprise regarding our results may arise from the traditional belief that a careful history is important. Secondly, our data is in line with chest pain characteristics being consistently demonstrated to be inaccurate. The novelty of our study is the gestalt evaluation of these characteristics taken together. And the main application of our results prompts us to reconsider how much value we should assign to classifications such as typical or atypical chest pain, as these have little or no influence on probability of CAD. The fact that typicality of pain did not show significant differences on predicting CAD probability has important practical implications, since decision-making during the clinical management of patients can be initially guided by these subjective classifications. The overvaluing of the current categorization may be misleading, resulting in under or overdiagnosis of CAD and mismanagement of cases. Therefore, the use of probabilistic models is supposed to be a more effective way to avoid repre- sentativeness heuristics.

In conclusion, our findings indicate that physician’s gestalt based on acute chest pain features lacks accuracy and reproducibility in estimating the probability of CAD. Physicians should be cautious when relying on chest pain characteristics and investigators should redirect their focus to identify validated predictors.

COMMENTS

Background
Traditionally, physicians tend to strongly rely on chest pain characteristics to define whether a patient has low or high probability of having coronary artery disease, through a process called gestalt. This kind of clinical judgment, however, does not seem to have good diagnostic accuracy in predicting coronary artery disease (CAD) etiology.

Research frontiers
Many authors have compared the accuracy of scores vs medical judgment in acute coronary syndromes. However, the study intends to clarify a current non-scientific trend of the physician community to make cardiovascular inferences directly from chest pain characteristics. This research establishes a new perspective for chest pain analysis, and reinforces the need to identify strong diagnostic clinical predictors of obstructive disease and then develop a multivariate model to help the emergency physician to assess this condition, instead of intuitive univariate diagnostic association currently applied.

Innovations and breakthroughs
The main idea presented was the evaluation of the diagnostic accuracy of chest pain characteristics in predicting the probability of CAD. This was performed by using only pain characteristics and with no further information. The novelty of the study is the gestalt evaluation of these characteristics taken together, while previous others had analyzed the diagnostic probability of each symptom separately. Additionally, previous literature has tested medical gestalt vs probabilistic scores, while the authors have tested medical gestalt vs real diagnosis.

Applications
The main application of the results relies on avoiding putting too much value of classifications such as typical or atypical chest pain. These classifications merely refer to chest pain characteristic and this isolated aspect has little or no influence on probability of CAD.

Terminology
Clinical gestalt refers to the theory that physicians and healthcare professionals organize clinical perceptions into “unified wholes”. This means that physicians can make clinical decisions without necessarily having complete information, posteriorly using this information to create solutions that can be generalized from one situation to another. Clinical gestalt represents an overall analysis, cultivated mainly by personal experience, history and examination.

Peer-review
The present study essentially supports that the elements of the chest pain history are only a little bit associated with increasing accuracy of diagnosis with CAD. Furthermore, it is very interested that there were poor agreement between the two cardologists. The methods are sound, and the used statistics seem also sound.

REFERENCES

1. Goodacre S, Locker T, Morris F, Campbell S. How useful are clinical features in the diagnosis of acute, undifferentiated chest pain? Acad Emerg Med 2002; 9: 203-208 [PMID: 11874776 DOI: 10.1017/CEB09781107415324.004]
2. Khan NA, Daskalopoulou SS, Karp J, Eisenberg MJ, Pelletier R, Tsadok MA, Dasgupta K, Norris CM, Pilote L. Sex differences in acute coronary syndrome symptom presentation in young patients. JAMA Intern Med 2013; 173: 1863-1871 [PMID: 24043208 DOI: 10.1001/jamainternmed.2013.1049]
3. Swap CJ, Nagurney JT. Value and limitations of chest pain history in the evaluation of patients with suspected acute coronary syndromes. JAMA 2005; 294: 2623-2629 [PMID: 16304077 DOI: 10.1001/jama.294.20.2623]
4. Body R, Carley S, McDowell G, Ferguson J, Mackway-Jones K.
Can a modified thrombolysis in myocardial infarction risk score outperform the original for risk stratifying emergency department patients with chest pain? *Emerg Med J* 2009; 26: 95-99 [PMID: 19164616 DOI: 10.1136/emermed.2008.058495]

5 **Body R**, Cook G, Burrows G, Carley S, Lewis PS. Can emergency physicians ‘rule in’ and ‘rule out’ acute myocardial infarction with clinical judgement? *Emerg Med J* 2014; 31: 872-876 [PMID: 25016388 DOI: 10.1136/emermed-2014-203832]

6 **Tversky A**, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974; 185: 1124-1131 [PMID: 17835457 DOI: 10.1126/science.185.4157.1124]

7 **Grove WM**, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess* 2000; 12: 19-30 [PMID: 10752360 DOI: 10.1037//1040-3590.12.1.19]

**P- Reviewer**: Coccheri S, Chang ST, Deng B, Okumura K

**S- Editor**: Ji FF

**L- Editor**: A

**E- Editor**: Wu HL
