View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education 2 (2016) 18–23

www.elsevier.com/locate/hpe

The Effect of Patients' Appearance on Doctors' Diagnostic Decision Making: Do Poor People Get Poorer Medical Care?

Fathia Mohamed[a,b], Silvia Mamede[a,b,*], Manahil Mohamedani[a,b], Ibrahim Al Alwan[a,b], Mohi Eldin M. Magzoub[a,b], Henk G. Schmidt[a,b]

[a]*Department of Medical Education, King Saud bin Abdulaziz University, Riyadh, Saudi Arabia*

[b]*Institute of Medical Education Research Rotterdam, Erasmus University, Rotterdam, The Netherlands*

Received 31 January 2016; accepted 31 January 2016

Available online 4 May 2016

Abstract

Purpose: Health inequalities are of great concern to health providers. Studies on the influence of social class on diagnostic performance are rare and inconclusive. We investigated whether patients' appearance (poor versus rich) affects physicians thinking and their ability to reach a correct diagnosis.

Method: Forty-six internal medicine residents participated in this purposely designed computerized study. Every participant solved four case scenarios with one of two versions of a patient's picture for each scenario. In this study simulated patients' pictures were used to play the role of poor and dirty patients in one condition and rich and clean in another condition. The basic analysis was aimed at diagnostic accuracy. Time needed to reach a decision and participants' ratings of how extensively they had processed the case, the latter composed by ratings of confidence in the diagnosis, case complexity and mental effort required to diagnose the cases, were measured for each participant and used as indications of the extent to which participants diagnosed the case analytically.

Results: There were no significant differences between the two conditions in terms of diagnostic accuracy and time spent in diagnosing the cases. However, even if the cases were exactly the same, participants reported to have processed the cases more extensively when the patient appeared poor than in rich-looking patients ($p = .04$).

Discussion: Social class seems to influence how extensively doctors think about the patient's problem during the diagnostic process but does not influence diagnostic accuracy. Given our findings, it may be worthwhile to replicate the study with a larger number of cases and larger differences in experience between groups of physicians.

© 2016 King Saud bin AbdulAziz University for Health Sciences. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Health inequalities; Diagnostic accuracy; Time for diagnosis

1. Introduction

Physicians generally believe that the diagnostic method is a more or less "objective" approach to solving patient problems, in which physicians use only the complaints, signs, and symptoms presented by the patient to arrive at a diagnosis. This belief is grounded in the assumption that medicine is a natural science and hence the application of knowledge from that science

*Correspondence to: Institute of Medical Education Research Rotterdam, Erasmus MC – Ae Building, Room 242, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands. Tel.: +31 10 7038270; fax: +31 10 7044752.

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region.

http://dx.doi.org/10.1016/j.hpe.2016.01.011

2452-3011/© 2016 King Saud bin AbdulAziz University for Health Sciences. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

does not leave room for non-medical factors to interfere with the process.

However, there are reasons to believe that clinical reasoning is a less rational endeavor than many seem to think. For instance, patients' social characteristics seem to play a role in the treatment of coronary heart disease. Although coronary heart disease is the main cause of death for both men and women, twice as many women as men aged 45–64 have undetected myocardial infarctions, suggesting later coronary heart disease diagnosis among women, [1] possibly mirroring lay opinions shared by doctors about coronary heart disease as a primarily male disease. [2] Race, social class, and gender were also shown to be influencing the diagnosis of other diseases, including psychiatric disorders. [3, 4] In a recent study by our own group we presented physicians with one of two versions of the same patient: a difficult patient version, exhibiting aggressive or non-compliant behaviors, or a neutral version in which these behaviors were absent. The physicians made significantly more diagnostic errors when they were presented with a "difficult" version than when presented with a neutral version of the same patient. Unlike most studies discussed here, that were correlational in nature, this effect of a patient characteristic on the accuracy of medical diagnosis was demonstrated in a tightly controlled experiment.

So it seems that background characteristics and particular behaviors of patients can have a negative influence on the quality of medical practice. But what about how the patient looks like? Will physicians be led astray by the sheer appearance of a patient all other things being equal? On first thought this is unlikely to happen unless appearance is directly related to the complaints, signs, and symptoms of that patient. For instance, when a patient looks pale or yellowish or sad, one can suspect a relationship with disease. But why would sheer appearance unrelated to disease interfere with clinical reasoning? We know from psychology that a person's appearance can influence judgment of that person. An early example is a study of Dion, Berscheid, and Walster. [5] They studied the relationship between attractiveness and judgment of personality. Students from University of Minnesota took part in the experiment. Each subject was given three different pictures to examine; one of an attractive individual, one of an individual of average attractiveness, and one of an unattractive individual. The participant judged the pictures' subjects along 27 different personality traits (including altruism, conventionality, self-assertiveness, stability, emotionality, trustworthiness, extraversion, kindness, and sexual promiscuity). Results showed that participants overwhelmingly believed the more attractive subjects to have more socially desirable personality traits than either the averagely attractive or unattractive subjects. [5] Of course, these students had no special knowledge of the persons judged and of how somebody looked like, and their personality. In such cases a first impression may be an important determinant of judgment. [6] Doctors however have deep knowledge about disease and its relationship with appearance and may therefore be less susceptible to the influence of first impressions. In addition, doctors are thought to be analytical in their thinking, weighing signs and symptoms of a case in the light of possible diagnostic hypotheses. Or are they? There is at least one theory of clinical reasoning that suggests that physicians' modal response to a case is governed by its likeliness to diseases seen previously. [7] This process of pattern recognition is considered to be fast, effortless, and not under control of conscious processing. Only if the signs displayed do not spontaneously activate a particular diagnostic hypothesis, physicians tend to engage in analytical reasoning. [8] This theory suggests that not directly relevant features of a patient, such as his appearance, may influence diagnosis, if the physician has seen a similar-looking patient (with a similar but different disease) previously. [9]

In the present study we were in particular interested in one aspect of a patient's appearance, namely whether he or she appears to be rich or poor. Although some studies suggest that lower-class patients are treated differently from higher-class patients, for instance in referral for psychotherapy, or in the diagnosis of breast cancer, [3, 4] we could find only one study in which social class was experimentally manipulated. [10] In this study, primary care doctors viewed a video-vignette of a scripted consultation where the patient presented with standardized symptoms of coronary heart disease. Videotapes were identical apart from varying patients' gender, age, class and race. Gender of patient significantly influenced doctors' diagnostic and management activities. However, there was no influence of social class, neither on the doctor's diagnosis, nor on the management activities undertaken. It is however possible that the two social class roles (teacher versus janitor) enacted in these videotapes were not sufficiently different to allow for an effect.

To test the hypothesis that the appearance of a patient indicating his social class has an effect on diagnostic accuracy, we presented beginning and advanced internal medicine residents with either a picture of a poor, dirty, patient or a well-dressed, clean, patient, before presenting them with the same clinical scenario. To avoid confounding by different persons impersonating the poor and the rich version, pictures were always of the same person (but in different guises). Time needed to arrive at a diagnosis was measured as an indication of the extent to which the physician used analytical reasoning to arrive at a diagnosis.

How fast and accurately the participants recognized findings related to a previously seen case and their ratings of confidence in their diagnosis, case difficulty and mental effort required for the diagnosis were also taken as indications of the degree of reflection involved in diagnosing the cases.

2. Method

2.1. Participants and setting

The participants were forty-six Internal Medicine senior residents (in the third and fourth year of their residency program), mean age = 33.61, SD = 4.70; 20 females,), enrolled in the Sudanese National Medical Specialization Board, who were willing to participate in the study following telephone calls. All participants signed a written consent agreeing to contribute in the study, and they all received financial compensation consisting of 150 Sudanese pounds for their participation. Also a verbal consent was obtained from the participants to keep confidentiality of the cases, so as to avoid contamination. Ethical approval to conduct the study was obtained from the Ethics committee in the Ministry of Health.

2.2. Materials

The materials consisted of four clinical cases with the following diagnoses: Inflammatory bowel disease, Addison's disease, liver cirrhosis, and community-acquired pneumonia. Cases were randomly selected from a set of cases used in a previous study with internal medicine residents. The case descriptions contained information about patient's complaints, findings from physical examinations, labs, and other test results. See Appendix A for an example (the case of liver cirrhosis).

A picture of the patient was attached to each clinical scenario. These pictures were taken from four simulated patients, 3 females and one male aged 30–50 years. They were chosen to participate in the study according to the patients' description in the clinical scenarios. After explaining to them their roles in the study, verbal consent was obtained from each. A professional photographer was hired to take two pictures from each of the simulated patients: one depicting them as a well-to-do patient and another depicting them as a poor patient. The effect was obtained by different clothing and Photoshop-added dirt on face and clothes (in the case of the poor patient). See for an example of two versions of the same patient (Fig. 1).

Fig. 1. Example of a picture of a patient in two versions.

2.3. Procedure

The study was an experimental study using a within-subjects design. This implies that the participants solved the cases under both experimental conditions. The independent variable was whether the patient was appearing rich or poor. The dependent variables were diagnostic accuracy, response time to make a diagnosis, number of errors made when deciding whether a finding belonged to the previously seen case or not and time needed to make this decision, and participants' rating of confidence in their diagnoses, case complexity and mental effort required to make the diagnoses. In addition, information about the participants' gender, number of years in clinical practice and the current training year were collected.

The experimental materials were included in a computer program, a runtime version of software for running psychology experiments called E-Prime (http://www.pstnet.com/eprime.cfm). The runtime version was downloaded on 25 computers, the number of computers locally available, at the computer lab of The Sudan Medical Council for Health Specialties. So, it required two sessions to complete the experiment.

Participants were first instructed verbally to remain silent through all the experiment and to switch off their mobile phones to avoid any interruption and then they went through written instructions on the computer screen before proceeding with the cases. Subsequently, participants' inquiries about the experiment were answered by the investigators. The first scenario was an example case to ensure familiarity with the procedure. Each case was presented on two screens. On the first screen the patient picture appeared. On the second screen, the scenario was printed with the same patient picture attached on the right upper corner and an empty bar at the bottom of the page to type a diagnosis. All participants diagnosed the same four cases, two of them with the picture displaying the rich version of the patient and two cases with the patient in the poor version. The order in which the cases were presented and the versions of the picture attached to the cases were counterbalanced to ensure that each case was diagnosed equally often in each experimental condition.

Participants were first asked to read the case and type a diagnosis. Their responses and the amount of time needed to arrive at this diagnosis was registered. After diagnosing each case, participants performed a "recognition task". They were presented to a series of concepts that may or may not be related to the case. The concepts appeared one by one on the computer screen, and the participants had to decide whether the concept was related to the case by pressing the key 'c' for YES and the key 'n' for NO. Twenty concepts appeared after each scenario, 10 were findings that were either literally stated in the case description (symptoms and signs) or not literally stated in the case descriptions but were generated on the basis of information presented in the description, and 10 were distractors or fillers that were not related to the case. Each concept was shown on screen for only few seconds. Participants were asked to make their decisions as quickly and as accurately as possible. As soon as one of the keys was pressed, the concept disappeared and the next concept appeared. Response times and the accuracy of the responses were registered automatically. It was assumed that the response time needed to decide whether a concept was or was not related to the case is a measure of the extent to which the case was mentally processed. More extensive processing would lead to shorter recognition times and fewer mistakes. See Appendix B for the concepts presented after the case printed in Appendix A.

In the second phase of the experiment, the cases appeared again, one by one, and each participant had to answer three questions for each case scenario: 1) how difficult the case was, 2) how confident he or she was about the diagnosis, and 3) how much mental effort he or she had to do to diagnose the case. To answer these questions, participants had to choose a percentage between (0% and 100%), where 0% reflected the very easy cases, lowest level of confidence and lowest mental effort, and 100% reflected the opposite. Participants' responses to these three questions were also taken as a measure of how extensively the case was mentally processed: more difficulty, more mental effort and less confidence indicate more processing to diagnosing the case.

2.4. Analysis

Two clinicians (F.M.; M.M.) evaluated, through a consensus model, the diagnoses provided by the participants as "correct", "partially correct" or "incorrect", scored, respectively, as 1.0, 0.5, or 0. A mean diagnostic accuracy score was computed for cases with patients in the rich version and in the poor version. The participants' responses in the recognition task (i.e., when they decided whether a finding presented on the computer screen was related to the previously seen case or not) were automatically scored by the computer program as 1 or 0, when the response was, respectively, correct or incorrect. The mean proportion of correct responses in the recognition task for the related

findings and the mean time required to for these responses were computed for the cases with rich-looking and poor-looking patients. To compute a mean rating of the amount of case processing for these two categories of cases, we averaged the ratings of confidence in diagnosis (reversed), case complexity, and mental effort. For all dependent variables, paired *t*-tests were performed to check for differences between cases solved with the patient presenting in the rich and the poor version. The data were analyzed using the statistical program SPSS for Mac version 20.0. Significance level was set at $p < .05$ for all comparisons.

3. Results

Table 1 presents the mean diagnostic accuracy scores and time required to give a diagnosis for the cases with patients presenting in the rich and the poor version.

Although the residents have made more mistakes in poor-looking patients than in patients who appear to be rich, this difference did not reach significance, $t(45) = 0.64$, $p = .52$. The time spent on diagnosing the cases showed a similar pattern: participants diagnosed the cases faster in patients with a poor appearance than in rich-looking patients but the difference was again not significant, $t(45) = 0.64$, $p = .53$.

The results of the recognition task are presented in Table 2. No significant differences emerged between cases with rich-looking or poor-looking patients. Participants recognized findings that belonged to the case with similar accuracy [$t(45) = 1.06$; $p = .29$)] and in a similar time [$t(45) = 0.41$; $p = .68$] independent of the patient appearance.

Finally, even if the cases were exactly the same, except for the patient's appearance, participants indicated to have processed the case more extensively when the patient appeared to be rich than when the patient looked poor (respectively, mean = 48.08%, standard deviation = 17.28 *vs* mean = 41.21%, standard deviation = 20.32), $t(45) = 2.12$, $p = .04$.

Table 1
Mean diagnostic accuracy score (range, 0–1) and mean time (seconds) required to give a diagnosis as a function of patient presentation, $N = 46$.

	Poor-looking patients Mean (SD)	Rich-looking patient Mean (SD)
Diagnostic accuracy	0.62 (0.40)	0.55 (0.44)
Time on diagnosis	238.93 (120.02)	252.21 (133.69)

Table 2
Mean proportion of accurate responses in recognition of findings related to the previously seen case and mean time (seconds) required to provide the response as a function of patient presentation, $N = 46$.

	Rich-looking patient Mean (SD)	Poor-looking patients Mean (SD)
Proportion of accurate responses	0.83 (0.15)	0.80 (0.14)
Response time	1.34 (0.27)	1.36 (0.25)

4. Discussion

Health inequalities are of great concern to health policy makers and providers of health care. Most of the research carried out operationalized inequalities in term of health care availability and utilization. [1-3] Studies directly attempting to measure the influence of social class on diagnostic performance are however rare. The only experimental study that we were able to unearth, [10] failed to provide evidence of the influence of social class on diagnostic accuracy. We argued that in this study perhaps the differences in social class were not sufficiently apparent and therefore designed an experiment in which this difference would be more extreme. To that end we had simulated patients photographed either as members of the Sudanese well-to-do class or as poor persons. Every physician involved in the study solved four case scenarios with one of these two versions. The basic analysis was aimed at diagnostic accuracy. Time needed to diagnose the case, time needed to reach a decision on items belonging (or not) to the case, and participants' ratings of the amount of case processing involved in diagnosing the case, (the latter composed by ratings of confidence in the diagnosis, case complexity and mental effort required to diagnose the cases), were measured for each participant and taken as indications of the extent to which participants diagnosed the case analytically. We failed to find effects of appearance on our variables, with the exception of the physicians' ratings of how extensively they had processed the case during their diagnoses. Participants processed cases more analytically when the patient looked rich than when the patient appeared to be poor. However, as the differences did not reach significance in the other variables, this finding may be only coincidental.

In search of factors that may explain why the differences found in the measurements of diagnostic accuracy and time for diagnosis, though all suggesting a deeper processing in rich-looking patients than in

patients who appeared poor, did not reach significance, we checked whether participants' clinical experience influenced their performance. We divided our sample in more experienced and less experienced physicians, using the median as a cutoff point, and conducted a repeated-measures ANOVA with patient appearance as a within-subjects factor (rich-looking *vs* poor-looking) and clinical practice (less experienced *vs* more experienced) as a between-subjects factor. Although there seems to be a tendency for an interaction between patient appearance and clinical experience, with more experienced doctors performing better in rich-looking patients and the reverse occurring for less experienced doctors, the interaction effect did not reach significance, $F(1,44) = 1.36$, $p = .25$.

We are left with a number of tantalizing possibilities neither of which materialized in this study. Given our findings however, it may be worthwhile to replicate the study with a larger number of cases and larger differences in experience between groups of physicians. In the mean time, our conclusion however must be that although social class seems to influence access to health care, it does not influence diagnostic accuracy.

Appendix A. An example of a case, related to the pictures in Fig. 1

A 52-year-old lawyer presented with persistent upper abdominal pain that started some months ago. The patient believes the pain is stress-related due to a falling number of clients and the death of his wife two years ago. He has had other relationships, but proved to suffer from an erectile dysfunction on the last sexual encounter. There are no indications of food intolerance. He has smoked around 40 cigarettes a day for many years and has consumed considerable amounts of alcohol. Medical history: he underwent a surgery 5 years ago for prostate cancer

Physical examination

Spider angiomas on the thorax. Abdomen: slightly distended, without shifting dullness; liver is palpable with an irregular surface; spleen not enlarged. Ankles with 1+edema. Testes: greatly reduced in size.

Lab tests

ESR: 44 mm/h; Hb: 8.0 g/dL; sodium: 138 mEq/L; potassium: 3.6 mEq/L; ALT: 120 U/L; AST: 84 U/L; LDH: 800 U/L; y-GT 250 U/L; alkaline phosphatase: 200 U/L; bilirubin total: 2.0 mg/dl.

Appendix B. Items related or not directly related to the case presented in Appendix A

Items related to the case	Items non-related to the case
liver with irregular surface	headache in the morning
upper abdominal pain	**ulcerative gingivitis**
alcohol consumption	reduced hearing
spider angiomas	tingling in the arms
reduced testes	low back pain
palpable liver	dizziness
ankles with edema	restlessness
elevated liver enzymes	excessive sleeping
prostate cancer 5 years ago	nasal obstruction
no food intolerance	perianal pain

References

1. McKinlay JB. Some contributions from the social system to gender inequalities in heart disease. *J Health Soc Behav.* 1996;37 (2):1. U1-U. PubMed PMID: WOS:A1996UQ27900001.
2. Emslie C, Hunt K, Watt G. Invisible women? The importance of gender in lay beliefs about heart problems. *Soc Health Illn.* 2001;23(2):203–233. PubMed PMID: WOS:000167366200004.
3. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *Am J Public Health.* 1991;81(5):646–649.
4. Garb HN. Race bias, social class bias, and gender bias in clinical judgment. *Clin Psychol Sci Pract.* 1997;4(2): 99–120. PubMed PMID: WOS:A1997WW06800002.
5. Dion K, Berscheid E, Walster E. What is beautiful is good. *J Personal Soc Psychol.* 1972;24(3):285–290.
6. Willis J, Todorov A. First impressions: making up your mind after a 100-ms exposure to a face. *Psychol Sci.* 2006;17(7):592–598.
7. Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol: Gen.* 1991;120(3):278–287.
8. Evans JSBT. Dual-processing accounts of reasoning, judgment, and social cognition. *Ann Rev Psychol.* 2008;59:255–278.
9. Allen SW, Brooks LR, Norman GR, Rosenthal D. Effect of prior examples on rule-based diagnostic performance. In: *Proceedings of the Annual Conference Research in Medical Education.* Research in Medical Education; 1988; 27: 9–14. PubMed PMID: MEDLINE:3218881.
10. Arber S, McKinlay J, Adams A, Marceau L, Link C, O'Donnell A. Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: a video-simulation experiment. *Soc Sci Med.* 2006;62(1):103–115.