Combining qualitative and quantitative methods in a study of computer-assisted medical decisions

Key points of talk at the workshop on "Mixed methods: identifying the issues", Manchester 26 - 27 October 2005

Mark Hartswood
School of Informatics
University of Edinburgh

Lorenzo Strigini
Centre for Software Reliability
City University, London

Introduction

This talk will summarise studies and analyses about the use of a Computer Aided Detection (CAD) tool in mammography, performed in DIRC (http://www.dirc.org.uk), a collaborative, multidisciplinary project dealing with the dependability of computer based systems: systems encompassing computer software and hardware together with their human users and the social context of computer use.

Computer Aided Detection (CAD) in mammography

Breast cancer is a common illness (more than 40,000 diagnosed per year in Britain) and screening programmes have been created for early detection; the U.K. national screening programme, for instance, relies on periodic mammography (radiography of the breasts) for women in the 50 to 64 age range. Interpreting screening mammograms is a highly skilled, specialized task; it involves weighing up the risks of missing possible cancers against the threat of overwhelming the programme with the unnecessary recall of healthy women for further tests. The risk of errors and relative scarcity of skilled radiologists[1] have prompted the development of computerised tools for analysing mammograms. These can identify the most common signs of cancer with high probability, achieved at the cost of allowing many false alarms. In CAD, these tools are used to complement the skills of the radiologist. The computer analyses each mammogram and marks 'prompts' on a copy of it to indicate areas on the image that might be worth a second look. The radiologist is required to examine the films in the usual way, but to check the computer prompts, before reaching a final decision.

CAD tools could improve the chances of early cancer diagnosis and/or alleviate the scarcity of radiologists for screening programmes. However, studies of their effectiveness have produced contrasting results: from sizeable projected improvements in the percentage of properly recognised cancers, to a lack of any statistically significant effect in some studies. This is a perplexing situation for any healthcare organisation considering investing in CAD, or for those designing the CAD tools or the procedures for their use. Direct controlled experiments present various difficulties, including the rarity of cancers, the delay with which missed cancers are recognised, and the rapid evolution of the CAD tools.

A multidisciplinary analysis of CAD

The work we will describe was performed in collaboration between computer scientists, clinicians, psychologists, reliability engineers and sociologists. The starting point was a series of previous studies run by UCL researchers on the effectiveness of a specific CAD tool [Taylor 2004], among the largest studies to date, which indicated no statistically significant effect on the average performance (sensitivity and specificity) of radiologists.

DIRC researchers performed additional analyses and studies, including ethnographic work, probabilistic modelling – as used in reliability engineering - of the effects of computer support, supplementary, exploratory statistical analyses on the data collected and additional small experiments. Probabilistic models, describing the effects of error probabilities of the CAD tool and of the radiologists on the overall rates of errors in cancer detection, suggested directions for data analysis and additional experiments. Ethnographic observation of users at work in real and experimental settings provided checks on the realism of experiments, and of modelling assumptions, and suggested possible phenomena and causal mechanisms to be studied via modelling and data analysis. Exploratory statistical analyses and experiments pointed at possible effects deserving further investigation.

Our collaboration revealed aspects of the problem that would elude more conventional approaches. For instance radiologists were not ignoring the computer's indications; rather, the negligible *average* effects hid systematic non-zero effects, beneficial for some subsets of decision types and harmful for others. CAD appeared actually to *damage* the performance of some radiologists on a specific subset of cases, mostly the *more skilled* ones. If these effects were to be present in clinical use, with different mixes of cases and readers from those in the trial, the overall effect might be positive, negative or null, with some possible transfer of risk between categories of patients. Ethnography corroborated some of the possible explanations for the observed statistical effects, for example, evidence of readers not always adhering to the trial protocol of using the tool as an attention cue and, instead, using it as a decision aid.

---

1 By "radiologist" here we mean any person performing the task of interpreting mammograms, although in healthcare practice the term designates a narrower professional category.

Acknowledgments

This work by City, Edinburgh and Lancaster Universities was mostly funded by EPSRC via DIRC, the Interdisciplinary Research Collaboration on the Dependability of computer-based systems. The DIRC team included Eugenio Alberdi, Peter Ayton, Andrey Povyakalo, Rob Procter, Mark Rouncefeld and Roger Slack. The work was made possible by Paul Taylor and Jo Champness (from UCL) who granted us access to their data, facilitated the follow-up studies and helped run them. R2 Technologies provided materials for the experiments conducted. Collaboration between Edinburgh and UCL was also funded by EPSRC grant GR/R24517/01

References

Note: A summary of the case study is on the Web at
http://www.dirc.org.uk/research/DIRC-Results/MammographyCity.html and a bibliography at
http://www.csr.city.ac.uk/people/eugenio/dirc-mammo.html

[Heathfield 1993] Heathfield, H., Wyatt, J. Philosophies for the design and development of clinical decision support systems, Methods of Information in Medicine 1993; 32(1): 1-8.

[Taylor 2004] Taylor, PM, Champness, J, Given-Wilson, RM, Potts, HWE et al. An Evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. Brit J Radiol, 77, pp. 21-27, 2004.