



Whiting, P. F., Davenport, C., Jameson, C., Burke, M., Sterne, J. A. C., Hyde, C., & Ben-Shlomo, Y. (2015). How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open*, 5(7), [e008155]. <https://doi.org/10.1136/bmjopen-2015-008155>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1136/bmjopen-2015-008155](https://doi.org/10.1136/bmjopen-2015-008155)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via BMJ at <http://bmjopen.bmj.com/content/5/7/e008155>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

## Web Appendix 2: Individual Study details

### a. Self-rating of understanding

| Study                                   | Study design | Participants           | Measures of accuracy assessed                          | How was the diagnostic information presented? | Information provided   | How was understanding assessed? | Type of scenario | Results  |
|---|--------------|------------------------|--|---|--|---------------------------------|------------------|--|
| Reid (1998) <sup>41</sup><br>USA        | Single group | 300 practicing doctors | Sensitivity<br>Specificity<br>LR+<br>LR-<br>ROC curves | None  | Questioned regarding use and understanding of various measures | Telephone interview             | None             | 8 (3%) used the recommended formal Bayesian calculations, 3 used ROC curves, and 2 used likelihood ratios. The main reasons cited for non-use included impracticality of the Bayesian method (74%), and non-familiarity with ROC curves and likelihood ratios (97%).<br>246 (82%) used sensitivity and specificity but only 174 (58%) physicians used them when interpreting test results. |
| Young (2002) <sup>45</sup><br>Australia | Single group | 50 GPs                 | Sensitivity<br>Specificity<br>PPV                      | No information                                | Asked to self-rate understanding of diagnostic terms.          | Telephone interview             | None             | 13 of 50 indicated that “‘I understand this and could explain to others’ the above answer” for the 3 diagnostic terms.<br><br>Participants self ratings of their understanding differed from an objective, criterion based assessment.   |

## b. Accuracy Definition

| Study  | Study design | Participants   | Measures of accuracy assessed                   | How was the diagnostic information presented? | Information provided  | How was understanding assessed?   | Type of scenario  | Results   |
|--|--------------|--|---|---|---|---|---|---|
| Argimon-Pallas (2011) <sup>21</sup><br>Spain | Single group | 152 family medicine residents in their second year of the Family Medicine training programme | Sensitivity<br>Specificity<br>PPV<br>NPV<br>LR+ | Population based scenario                     | Information provided on total number of patients with target condition and number with and without condition testing positive | Questionnaire asked to calculate accuracy measures from raw data in scenario<br><br>Administered before and after educational intervention (intensive and interactive four half-day sessions) | Unclear   | Before task number of doctors correctly calculating figures were:<br>Sensitivity: 42%<br>Specificity: 34%<br>PPV: 33%<br>NPV: 26%<br>LR+: 8%<br><br>After intervention numbers more than doubled for all accuracy measures.<br>Sensitivity: 82%<br>Specificity: 79%<br>PPV: 82%<br>NPV: 80%<br>LR+: 48% |
| Bergus(2004) <sup>23</sup><br>USA            | Single group | 43 medical students and residents (psychiatry and Internal Medicine)                         | Sensitivity<br>Specificity                      | Extract from research study                   | Asked to identify sensitivity and specificity from report   | Questionnaire (open ended)  | Real life (major depression and panic disorder, congestive heart failure) | 88% correctly identified the specificity and sensitivity of the test from the paper.  |
| Berwick ( 1981) <sup>24</sup><br>USA         | Single group | 36 medical students, 45 interns and residents, 49 research doctors, 151                      | Sensitivity<br>Specificity<br>FPR               | 2x2 table                                     | Asked to identify definitions based on 2x2 table (a, b, c, d used rather than numbers)  | Questionnaire (MC)  | Hypothetical (Disease K)  | Practicing physicians were less able to correctly define sensitivity and specificity than medical students and research doctors. Exact values not reported  |

|   |              |  |                                   |                |  |  |  |   |
|---|--------------|--|-----------------------------------|----------------|--|--|--|---|
|   |              | full time doctors                              |                                   |                |  |  |  |   |
| Estellat (2006) <sup>30</sup>   | Single group | Senior doctors research and full time practice | Sensitivity<br>Specificity<br>LR+ | 2x2 table      | 2x2 table and short extract from study report.                   | Questionnaire. (multiple choice, Postal or given directly by one investigator) | Real life (CT for Pulmonary Embolism)                      | 85% selected correct definition for sensitivity, 80% for specificity and 17% for LR+. High rate of 'do not know' for LR's (72%)   |
| Steurer (2002) <sup>20</sup><br><br><i>Related publication:</i><br>Bachmann (2003) <sup>43</sup><br>Switzerland | Single group | 263 GPs  | Sensitivity<br>PPV                | No information | Asked to select correct definition for various accuracy measures | Questionnaire (multiple choice)  | Real life (Transvaginal ultrasound for endometrial cancer) | 76% (95% CI 70-81%) correctly identified the definition of sensitivity, 61% (95% CI 45-67%) correctly identified the definition of PPV  |
| Young (2002) <sup>45</sup><br><br>Australia   | Single group | 13 GPs   | Sensitivity<br>Specificity<br>PPV | No information | Asked for verbal explanations of diagnostic terms                | Interview  | None   | <b>Sensitivity:</b> In interview, 1 met some of the criteria to show that they knew the correct meaning of the term, 7 met none of the criteria and 5 could not or refused to answer or participate.<br><b>Specificity:</b> In interview, 6 met none of the criteria and 7 could not answer or refused to participate.<br><b>PPV:</b> In interview, 1 met all the criteria, 1 met none of the criteria and 11 could not answer or refused to participate. |

### c. Bayesian Reasoning

| Study  | Study design | Participants                                | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided   | How was understanding assessed?  | Type of scenario                                   | Results  |
|--|--------------|---|-------------------------------|---|--|--|--|--|
| Agoritsas(2011) <sup>22</sup><br>Switzerland | RCT          | 1361 physicians of all clinical specialties | Sensitivity<br>Specificity    | Population based scenario                     | Sensitivity and specificity described in words and numerical frequencies (terms not used) for very accurate test (sensitivity and specificity 99%)<br><br>Doctors randomised to receive information on different prevalence (1%, 2%, 10%, 25%, 95%) and no information | Multiple choice<br>Questionnaire:<br>Different categories of post-test probability offered: <60%, 60-79%, 80-94%, 95-99.9%, >99.9% | Screening test for viral disease in primary school | <b>Test result evaluated (positive or negative):</b><br>Positive<br><b>Post-test probability proportion correct: 22%</b><br>Most respondents (66.7% to 80.3%) selected a post-test probability of 95–99.9%, regardless of the prevalence of disease and even when no information on prevalence was provided.<br>We estimated that 9.1% (95% CI 6.0–14.0) of respondents knew how to assess correctly the post-test probability. This proportion did not vary with clinical experience or practice setting. |

| Study                                | Study design | Participants  | Measures of accuracy assessed | How was the diagnostic information presented?     | Information provided  | How was understanding assessed? | Type of scenario  | Results   |
|--------------------------------------|--------------|---|-------------------------------|---|---|---------------------------------|---|---|
| Bergus(2004) <sup>23</sup><br>USA    | Single group | 43 medical students and incoming residents (psychiatry and Internal Medicine)             | Sensitivity<br>Specificity    | Extract from research study and simulated patient | Asked to identify sensitivity and specificity from report and asked to apply these to a patient with a specified pre-test probability | Questionnaire (open ended)      | Real life (major depression and panic disorder, congestive heart failure) | <b>Test result evaluated:</b><br>Unclear<br><b>PPV/NPV proportion correct:</b> 1/28 Med students, 0/15 residents<br><b>PPV proportion over/under:</b> NR  |
| Berwick ( 1981) <sup>24</sup><br>USA | Single group | 36 medical students, 45 interns and residents, 49 research doctors, 151 full time doctors | Sensitivity<br>Specificity    | Population based scenario                         | Sensitivity and specificity described in words (terms not used)   | Questionnaire (MC)              | Hypothetical (Disease K)  | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b> 32%<br><b>PPV proportion over:</b> 68%<br><b>PPV proportion under:</b> 0<br><b>Effect of research:</b> 65% research vs 21% practicing correct |

| Study                            | Study design | Participants   | Measures of accuracy assessed | How was the diagnostic information presented?       | Information provided   | How was understanding assessed? | Type of scenario  | Results   |
|----------------------------------|--------------|--|-------------------------------|---|--|---------------------------------|---|---|
| Borak(1982) <sup>25</sup><br>USA | Single group | 42 practising physicians based in a non-teaching hospital, 43 'statistically sophisticated' community medicine physicians, 43 nurses | Sensitivity<br>Specificity    | 2 population based and 1 simulated patient scenario | Sensitivity and specificity described in words (terms not used) to a population or a patient with a specified pre-test probability also described in words | Questionnaire (open ended)      | Real life (streptococcal sore throat, bowel cancer)<br>Non-medical scenarios also included but not presented here | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>34% statistically sophisticated doctors, <2% of nurses and other doctors<br><b>PPV proportion over/under:</b> NR   |
| Bramwell (2006) <sup>26</sup>    | RCT          | 42 midwives, 41 obstetricians  | Sensitivity<br>FPR            | Population based scenario                           | Sensitivity and FPR described in words; terms not used. Group 1 received information in % format, group 2 in natural frequencies                           | Questionnaire (open ended)      | Real life (Down's screening)  | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b> 0 midwives, 5% obstetricians<br><b>PPV proportion over:</b><br>46% midwives, 76% obstetricians<br><b>PPV proportion under:</b><br>55% midwives, 19% obstetricians |

| Study                                      | Study design | Participants  | Measures of accuracy assessed                 | How was the diagnostic information presented? | Information provided   | How was understanding assessed?        | Type of scenario | Results  |
|--|--------------|---|---|---|--|--|------------------|--|
| Casscells (1978) <sup>17</sup><br>USA      | Single group | 40 doctors<br>20 medical students   | FPR   | Population based scenario                     | Single scenario including prevalence and FPR   | Interview (1 on 1 corridor discussion) | Hypothetical     | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>11/60<br><b>PPV proportion over:</b> not stated; 27/60 said 95% and mean was 56% - correct value was 2%<br><b>PPV proportion under:</b><br>NR<br><b>Effect of experience:</b> No effect   |
| Chernushkin (2012) <sup>27</sup><br>Canada | Single group | 94 Pharmacists;<br>55 completed diagnostics knowledge and skills section (extracted here) | Sensitivity<br>Specificity<br>LR+ (numerical) | Population based scenario                     | Various different knowledge and skills questions related to application of accuracy measures | Online questionnaire                   | Real life        | <b>Test result evaluated (positive or negative):</b><br>Positive and negative<br><b>Post-test probability proportion correct:</b><br>When information on sensitivity was provided 61% were correct, when information on specificity was provided 48% were correct, when information on LR+ was provided 39% were correct. The mean proportion of “don’t know” answers was 13% for sensitivity, 9% for specificity and 49% for LR+. |



| Study                            | Study design                        | Participants  | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided   | How was understanding assessed? | Type of scenario                      | Results   |
|----------------------------------|-------------------------------------|---|-------------------------------|---|--|---------------------------------|---------------------------------------|---|
| Curley 1990 <sup>28</sup><br>USA | Unclear allocation to 1/8 scenarios | 36 fellowship physicians, 29 chief medical residents, 18 medical students.<br><br>208 undergraduates (non-medical) also included but results not presented here | Sensitivity<br>Specificity    | Vignette/Case-study                           | In 6/8 scenarios sensitivity, specificity and prevalence in words (terms not provided). In 2/8 scenarios specificity was purposefully not provided | Questionnaire (open ended)      | Real life (Coronary heart disease)    | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>Most participants revised probability in correct direction but reasonable proportion did not. Between 0% and 69% of participants correctly estimated the magnitude and direction of change in post-test probability following a positive test result (PPV) (on a visual scale from 0-100%).<br><b>Values of sens/Spec:</b><br>Values of sens/spec did not influence proportion correct<br><b>Effect of experience:</b> No significant difference in correct responses between medical students, physicians and undergraduates. |
| Eddy (1982) <sup>29</sup><br>USA | Single group                        | 100 doctors   | FPR                           | Population based scenario                     | Single scenario including prevalence and FPR   | Unclear                         | Real life (mammography breast cancer) | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>95/100 estimated answer as 75% rather than 7.5%  |

| Study   | Study design | Participants   | Measures of accuracy assessed     | How was the diagnostic information presented?                   | Information provided  | How was understanding assessed?                                 | Type of scenario                                  | Results  |
|---|--------------|--|-----------------------------------|---|---|---|---|--|
| Estellat (2006) <sup>30</sup><br>France       | Single group | 130 Senior doctors research and full time practice                                       | Sensitivity<br>Specificity<br>LR+ | Population scenario (different scenarios for sens/spec and LR+) | Sensitivity, specificity, LR+ (in words) and prevalence given   | Questionnaire. (multiple choice for sens/spec and open for LR+) | Hypothetical                                      | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>32% correct, 42% incorrect, 25% do not know based on sens and spec.<br><b>PPV proportion over/under:</b> NR<br><b>LR Effect:</b> 9% correct PPV with LR+, 58% incorrect, 25% did not know |
| Garcia-Retamero (2013) <sup>31</sup><br>Spain | RCT          | 81 GPs with a minimum of 1 year of practice and 81 patients; data only extracted for GPs | Sensitivity<br>FPR                | Population based scenario                                       | Information on sensitivity FPR and prevalence reported in words (terms not used) or as natural frequencies. Half participants received this information depicted with visual aids | Paper questionnaire   | Real life (Breast cancer, colon cancer, diabetes) | <b>Test result evaluated (positive or negative):</b><br>Positive<br><b>Post-test probability proportion correct:</b><br>Probabilities alone: 23%<br>Natural frequencies alone: 48%<br><br>Probabilities with visual aid: 68%<br>Natural frequencies with visual aid: 73% |

| Study   | Study design | Participants                                  | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided  | How was understanding assessed?  | Type of scenario  | Results   |
|---|--------------|---|-------------------------------|---|---|--|---|---|
| Hoffrage (1998) <sup>32</sup><br><br><i>Related publications:</i><br>Giggerenzer(1996) <sup>33</sup><br>Giggerenzer (2003) <sup>34</sup><br><br>Germany | Two groups   | 48 Doctors, mixture of full time and research | Sensitivity<br>FPR            | Vignette/Case study                           | Information on sensitivity and specificity reported in words (terms not used) or as natural frequencies                                   | Questionnaire (multiple choice) & interview about reasoning strategies | Real life (Breast cancer, colorectal cancer, Phenylketonuria and Ankylosing Spondylitis.) | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>10% as probabilities, 46% as natural frequencies<br><b>PPV proportion over:</b><br>17/24 for prob, 8/24 for nat freq<br><b>PPV proportion under:</b><br>5/25 for prob, 5/24 for nat freq |
| Hoffrage (2000) <sup>19</sup><br><br><i>Related publication:</i><br>Hoffrage (2004) <sup>35</sup><br><br>Germany  | Single group | 87 medical students, 9 first year interns     | Sensitivity<br>FPR            | Population based scenario                     | 4 different scenarios 2 presented as probabilities (terms defined in words), and two as natural frequencies. Short and long formats used. | Questionnaire  | Real life (colorectal cancer, breast cancer, phynylketonuria, ankylosing spondylitis)     | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b><br>Long prob 18%, long nat 57%, short prob 50%, short nat 68%   |

| Study                             | Study design | Participants  | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided   | How was understanding assessed? | Type of scenario                          | Results   |
|-----------------------------------|--------------|---|-------------------------------|---|--|---------------------------------|---|---|
| Lyman (1993) <sup>36</sup><br>USA | Single group | 29 doctors; 21 nurses and pharmacists                           | Sensitivity<br>Specificity    | Vignette/Case study                           | Asked to estimate prevalence, sensitivity and specificity based on vignette then apply their values to get a post-test probability | Questionnaire (open ended)      | Real life (mammography for breast cancer) | <b>Test result evaluated:</b><br>Positive and negative<br><b>PPV:</b> Consistently overestimated<br><b>NPV:</b> Estimates correct   |
| Lyman (1994) <sup>37</sup><br>USA | Single group | 39 mixed doctors, 15 nurses and pharmacists, 4 medical students | Sensitivity<br>Specificity    | Population based scenario                     | Various different estimates of sensitivity, specificity and prevalence   | Questionnaire (open ended)      | Hypothetical                              | <b>Test result evaluated:</b><br>Positive and negative<br><b>PPV:</b> Physicians and non-physicians overestimate post-test probabilities with increasing error associated with decreasing disease risk. |

| Study                                   | Study design | Participants                                     | Measures of accuracy assessed                                  | How was the diagnostic information presented? | Information provided   | How was understanding assessed?                | Type of scenario                             | Results   |
|---|--------------|--|--|---|--|--|--|---|
| Moreira (2008) <sup>38</sup><br>Belgium | Single group | 50 Doctors attending course on tropical medicine | Sensitivity<br>Specificity<br>Categorical grouping based on LR | Unclear                                       | Sensitivity and specificity values and LRs categorised as: 'quite useless', 'weak', 'good', 'strong', 'very strong'. | Questionnaire (multiple choice and open ended) | Mixed (4 real diseases and 2 dummy diseases) | <p><b>Test result evaluated:</b><br/>Positive</p> <p><b>PPV proportion over:</b><br/>Overestimated for real and dummy diseases.</p> <p><b>PPV not estimate:</b> 40% could not calculate PPV with sensitivity and specificity data</p> <p><b>LR Effect:</b> More accurate results with categorical description of LR compared to numerical presentation of sens and spec</p> |

| Study                                 | Study design | Participants         | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided  | How was understanding assessed? | Type of scenario                                | Results   |
|---------------------------------------|--------------|----------------------|-------------------------------|---|---|---------------------------------|---|---|
| Noguchi (2002) <sup>39</sup><br>Japan | Single group | 224 medical students | Sensitivity<br>Specificity    | Vignette/Case-study                           | Participants provided with 1/3 descriptions of a patients' history representing low, intermediate or high pre-test probability and a diagnostic test result (+ve or -ve) and asked to estimate pre-test probability and PPV and NPV | Questionnaire (open ended)      | Coronary Heart Disease and Exercise Stress Test | <b>Test result evaluated:</b><br>Positive and negative<br><b>PPV:</b> Correct reasoning<br><b>NPV:</b> Poorly estimated |

| Study                                     | Study design | Participants                                    | Measures of accuracy assessed                                   | How was the diagnostic information presented? | Information provided   | How was understanding assessed?        | Type of scenario  | Results  |
|---|--------------|---|---|---|--|--|---|--|
| Puhan (2005) <sup>40</sup><br>Switzerland | RCT          | 183 Senior family and internal medicine doctors | Sensitivity<br>Specificity<br>LR+<br>LR-<br>Graphic based on LR | Vignette/Case study                           | Group 1: Sensitivity and specificity<br>Group 2: Positive or negative likelihood ratio defined in words<br>Group 3: simple graphic of 5 circles based on LR. | Questionnaire (open ended, conference) | Pulmonary Embolus, Myocardial Infarction, COPD, Temporal arteritis, flu, heart failure. | <b>Test result evaluated:</b> Positive and negative<br><b>Post-test probability proportion correct:</b> Deviations from correct estimates were similar for all modes of presentation, for some scenarios the graphic produced the closest estimates<br><b>Post-test probability proportion over:</b> Overall post-test probability in wrong direction in 9% of sens/spec group, 4% in LR group, and 4% in LR graphic group |
| Reid (1998) <sup>41</sup><br>USA          | Single group | 300 practicing doctors                          | Sensitivity<br>Specificity                                      | None  | Questioned regarding use and understanding of various measures   | Telephone interview                    | None  | <b>Test result evaluated:</b> No test result defined<br><b>PPV proportion correct:</b> Of the 174 physicians who said they used sensitivity and specificity, 165 (95%) did not do so in the recommended formal manner.   |

| Study                           | Study design | Participants       | Measures of accuracy assessed | How was the diagnostic information presented? | Information provided  | How was understanding assessed?   | Type of scenario              | Results  |
|---------------------------------|--------------|--------------------|-------------------------------|---|---|-----------------------------------|-------------------------------|--|
| Sox (2009) <sup>42</sup><br>USA | RCT          | 653 paediatricians | Sensitivity<br>Specificity    | Vignette/Case study                           | <i>Group 1:</i> No test accuracy info<br><i>Group 2:</i> Sensitivity and specificity (%)<br><i>Group 3:</i> Sensitivity and specificity (natural frequencies) | Questionnaire (open ended postal) | Real life (DFA for pertussis) | <b>Test result evaluated:</b> Negative<br><b>Post-test probability proportion correct: 1% (n=5)</b> (all from group 3) estimated correct value. Proportion nearly correct was 13% (group 1), 20% (group 2) and 19% (group 3)<br><b>Post-test probability proportion over: 56%</b> estimated post test prob higher than pre-test prob, 11% estimated post test probability same as pre-test probability. 32% estimated post-test prob as 50% (same as sensitivity)<br><b>Effect of experience:</b> Greater proportion of residents estimated a nearly correct probability (29%) compared to paediatricians with (15%) or without (15%) an academic affiliation. |



| Study   | Study design | Participants                                    | Measures of accuracy assessed                                   | How was the diagnostic information presented? | Information provided  | How was understanding assessed?                | Type of scenario   | Results  |
|---|--------------|---|---|---|---|--|--|--|
| Steurer (2002) <sup>20</sup><br><br><i>Related publication:</i><br>Bachmann (2003) <sup>43</sup><br><br>Switzerland | RCT          | 263 GPs   | Sensitivity<br>Specificity<br>LR+<br>(described in words)       | Vignette/Case study                           | Generic question based on sensitivity and specificity for population based scenario.<br><br>Group 1: Test positive, no information on accuracy<br>Group 2: sensitivity and specificity<br>Group 3: LR+ defined in words | Questionnaire (multiple choice and open ended) | Real life (Transvaginal ultrasound for endometrial cancer) | <b>Test result evaluated:</b><br><b>Positive</b><br><b>PPV proportion correct:</b> 22%.<br><b>PPV proportion over:</b> 56% selected value close to 100%. PPV overestimated: no test accuracy info > sensitivity & specificity (%) > LR in plain language.  |
| Vermeesch (2010) <sup>44</sup>  | Single group | 117 GPs and 55 specialists in internal medicine | Sensitivity<br>Specificity<br>LR+<br>Probability modifying plot | Population based scenario                     | Three questions with different info:<br>Q 1: Sensitivity, specificity and prevalence<br>Q 2: Prevalence & LR+ described in words (terms not used)<br>Q 3: Prevalence and probability modifying plot                     | Questionnaire (multiple choice, conference)    | Hypothetical   | <b>Test result evaluated:</b><br>Positive<br><b>PPV proportion correct:</b> Q1: 7%, Q2: 27%, Q3: 50%.<br><b>PPV "Don't know":</b> Q1 15%, Q2 22%, Q3 33%<br><b>PPV proportion over:</b> Q1: 73%, Q2: 43%, Q2: 7%<br><b>PPV proportion under:</b> Q1: 6%, Q2: 8%, Q3: 2%<br><b>Effect of experience:</b> Results similar according to age |

#### d. Presentation Format

| Study                         | Study design | Participants                  | Measures of accuracy assessed   | How was the diagnostic information presented? | Information provided   | How was understanding assessed? | Type of scenario             | Results   |
|-------------------------------|--------------|-------------------------------|---------------------------------|---|--|---------------------------------|------------------------------|---|
| Bramwell (2006) <sup>26</sup> | RCT          | 42 midwives, 41 obstetricians | Sensitivity (1-specificity) FPR | Population based scenario                     | Information on sensitivity and 1-specificity (as FPR) reported in words (terms not used) or as natural frequencies | Questionnaire (open ended)      | Real life (Down's screening) | <p><b>Probability format (sensitivity and FPR as words):</b></p> <ul style="list-style-type: none"> <li>-None of the midwives and 1 (5%) of the obstetricians gave the correct answer.</li> <li>- 46% of midwives and 76% of obstetricians overestimated the PPV</li> <li>- 55% of midwives and 19% of obstetricians underestimated the PPV.</li> </ul> <p><b>Natural frequency format:</b></p> <ul style="list-style-type: none"> <li>- None of the midwives and 13 (65%) of the obstetricians gave the correct answer.</li> <li>-35% of midwives and 15% of obstetricians overestimated the PPV</li> <li>-65% of midwives and 20% of obstetricians underestimated the PPV.</li> </ul> |

|  |              |  |                         |                           |  |   |  |   |
|--|--------------|--|-------------------------|---------------------------|--|---|--|---|
| Garcia-Retamero (2013) <sup>31</sup><br>Spain  | RCT          | 81 GPs with a minimum of 1 year of practice and 81 patients; data only extracted for GPs | Sensitivity FPR         | Population based scenario | Information on sensitivity FPR and prevalence reported in words (terms not used) or as natural frequencies. Half participants received this information depicted with visual aids      | Paper questionnaire                         | Real life (Breast cancer, colon cancer, diabetes)  | <b>Test result evaluated (positive or negative):</b> Positive<br><b>Post-test probability proportion correct:</b><br>Probabilities alone: 23%<br>Natural frequencies alone: 48%<br><br>Probabilities with visual aid: 68%<br>Natural frequencies with visual aid: 73%   |
| Hoffrage (1998) <sup>32</sup><br><br><i>Related publications:</i><br>Giggerenzer (1996) <sup>33</sup><br>Giggerenzer (2003) <sup>34</sup><br>Germany | Two groups   | 48 Doctors, mixture of full time and research  | Sensitivity Specificity | Vignette/Case study       | Information on sensitivity and specificity reported in words (terms not used) or as natural frequencies  | Questionnaire (multiple choice) & interview | Real life (Breast cancer, colorectal cancer, Phenylketonuria and Ankylosing Spondylitis .) | <b>Probability format:</b> Clinicians correct post-test probability only 10%<br><b>Natural frequency format:</b> Clinicians correct post-test probability increased to 46%.<br><br>Doctors spent an average of 25% more time on probability formats than natural frequency formats  |
| Hoffrage (2000) <sup>19</sup><br><br><i>Related publication:</i><br>Hoffrage (2004) <sup>35</sup><br><br>Germany                                     | Single group | 87 medical students, 9 first year interns  | Sensitivity FPR         | Population based scenario | Information on sensitivity and specificity reported in words (terms not used) or as natural frequencies. Four scenarios two for each presentation format using short and long versions | Questionnaire (open ended)                  | Real life (colorectal cancer, breast cancer, phenylketonuria, ankylosing spondylitis )     | <b>LONG FORMAT:</b><br><b>Probability format:</b> Clinicians correct post-test probability only 10% correct<br><b>Natural frequency format:</b> Clinicians correct post-test probability increased to 57%.<br><br><b>SHORT FORMAT:</b><br><b>Probability format:</b> Clinicians correct post-test probability only 50% correct<br><b>Natural frequency format:</b> Clinicians correct post-test probability increased to 68%. |

|                          |     |                    |                            |                     |  |                                   |                               |  |
|--------------------------|-----|--------------------|----------------------------|---------------------|--|-----------------------------------|-------------------------------|--|
| Sox (2009) <sup>42</sup> | RCT | 635 paediatricians | Sensitivity<br>Specificity | Vignette/Case study | Group 1: No test accuracy info<br>Group 2: Sensitivity and specificity<br>Group 3: Sensitivity and specificity (natural frequencies) | Questionnaire (open ended postal) | Real life (DFA for pertussis) | <p>18 % correctly estimated post-test probability.</p> <p>There was no difference (p=0.16) in the mean post-test probability between groups 1 and 2 (38% and 41%). Group 3 (45%) had a significantly higher mean post-test probability than group 1 (p=0.007).</p> <p>Even though test result was negative 56% of participants gave a higher post-test probability than the pre-test probability and 11% estimated a post-test probability of 30% (same as pre-test probability). Five participants (all in group 3) correctly estimated the post-test probability. There was no significant difference in the proportion of doctors who nearly estimated the correct post-test probability (defined as within range 13% to 23%) - 13% in group 1, 20% in group 2, and 19% in group 3 - p=0.06 comparing groups 1 and 2, p=0.08 and comparing groups 3 and 1</p> |
|--------------------------|-----|--------------------|----------------------------|---------------------|--|-----------------------------------|-------------------------------|--|

### References (same as main document)

1. Kostopoulou O, Oudhoff J, Nath R, et al. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. *Medical Decision Making* 2008;**28**(5):668-80.
2. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *BMJ* 2009;**338**:b946.
3. Eddy D, Clanton C. The art of diagnosis: solving and clinicopathological exercise. In: Dowie J, Elstein A, eds. *Professional Judgment: A Reader in Clinical Decision Making*. Cambridge: Cambridge University Press, 1988:200-11.
4. Falk G, Fahey T. Clinical prediction rules. *BMJ* 2009;**339**:b2899.
5. Knottnerus JA. Interpretation of diagnostic data: an unexplored field in general practice. *The Journal of the Royal College of General Practitioners* 1985;**35**(275):270-4.
6. Stengel D, Bauwens K, Sehouli J, et al. A likelihood ratio approach to meta-analysis of diagnostic studies. *Journal of medical screening* 2003;**10**(1):47-51.
7. Moons KG, Harrell FE. Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. *Academic radiology* 2003;**10**(6):670-2.
8. Sackett DL, Straus S. On some clinically useful measures of the accuracy of diagnostic tests. *ACP journal club* 1998;**129**(2):A17-9.
9. Dujardin B, Van den Ende J, Van Gompel A, et al. Likelihood ratios: a real improvement for clinical decision making? *European journal of epidemiology* 1994;**10**(1):29-36.
10. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;**365**(9469):1500-5.
11. Hayward RS, Wilson MC, Tunis SR, et al. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *Jama* 1995;**274**(7):570-4.
12. Wilson MC, Hayward RS, Tunis SR, et al. Users' guides to the Medical Literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients? The Evidence-Based Medicine Working Group. *Jama* 1995;**274**(20):1630-2.
13. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *BMJ* 2005;**330**(7499):1080-3.
14. Cochrane AJ. *Effectiveness and Efficiency: Random Reflections on Health Services*. The Nuffield Provincial Hospitals Trust. London: The Royal Society of Medicine Press Ltd., 1972.
15. Knottnerus JA. *Evidence Base of Clinical Diagnosis*: Wiley, 2002.
16. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care* [Internet]. York: University of York, 2009 [accessed 23.3.11].
17. Casscells W, Schoenberger A, Graboyes TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med* 1978;**299**(18):999-1001.
18. Gigerenzer G, Hoffrage U. How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review* 1995;**102**(4):684-704.
19. Hoffrage U, Lindsey S, Hertwig R, et al. Medicine. Communicating statistical information. *Science* 2000;**290**(5500):2261-62.
20. Steurer J, Fischer JE, Bachmann LM, et al. Communicating accuracy of tests to general practitioners: a controlled study.[Erratum appears in *BMJ* 2002 Jun 8;324(7350):1391]. *BMJ* 2002;**324**(7341):824-26.

21. Argimon-Pallas JM, Flores-Mateo G, Jimenez-Villa J, et al. Effectiveness of a short-course in improving knowledge and skills on evidence-based practice. *BMC Family Practice* 2011;**12**:64.
22. Agoritsas T, Courvoisier DS, Combescure C, et al. Does prevalence matter to physicians in estimating post-test probability of disease? A randomized trial. *Journal of General Internal Medicine* 2011;**26**(4):373-8.
23. Bergus G, Vogelgesang S, Tansey J, et al. Appraising and applying evidence about a diagnostic test during a performance-based assessment. *BMC Medical Education* 2004;**4**:20.
24. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med* 1981;**71**(6):991-98.
25. Borak J, Veilleux S. Errors of Intuitive Logic Among Physicians. *Social Science & Medicine* 1982;**16**(22):1939-44.
26. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal* 2006;**333**(7562):284-86A.
27. Chernushkin K, Loewen P, De Lemos J, et al. Diagnostic reasoning by hospital pharmacists: Assessment of attitudes, knowledge, and skills. *Canadian Journal of Hospital Pharmacy* 2012;**65**(4):258-64.
28. Curley SP, Yates JF, Young MJ. Seeking and applying diagnostic information in a health care setting. *Acta Psychol (Amst)* 1990;**73**(3):211-23.
29. Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A, eds. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982:249-67.
30. Estellat C, Faisy C, Colombet I, et al. French academic physicians had a poor knowledge of terms used in clinical epidemiology. *Journal of Clinical Epidemiology* 2006;**59**(9):1009-14.
31. Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science & Medicine* 2013;**83**:27-33.
32. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Academic Medicine* 1998;**73**(5):538-40.
33. Gigerenzer G. The psychology of good judgment: Frequency formats and simple algorithms. *Medical Decision Making* 1996;**16**(3):273-80.
34. Gigerenzer G. *Reckoning with Risk: Learning to live with uncertainty*. UK: Penguin, 2003.
35. Hoffrage U, Gigerenzer GE-MA, Hoffrage Uhm-bmd. How to Improve the Diagnostic Inferences of Medical Experts. [References]. Kurz-Milcke, Elke [Ed]; Gigerenzer, Gerd [Ed] 2004;:(2004):314.
36. Lyman GH, Balducci L. Overestimation of test effects in clinical judgment. *Journal of Cancer Education* 1993;**8**(4):297-307.
37. Lyman GH, Balducci L. The effect of changing disease risk on clinical reasoning. *Journal of General Internal Medicine* 1994;**9**(9):488-95.
38. Moreira J, Bisoffi Z, Narvaez A, et al. Bayesian clinical reasoning: does intuitive estimation of likelihood ratios on an ordinal scale outperform estimation of sensitivities and specificities? *Journal of Evaluation in Clinical Practice* 2008;**14**(5):934-40.
39. Noguchi Y, Matsui K, Imura H, et al. Quantitative evaluation of the diagnostic thinking process in medical students. *Journal of General Internal Medicine* 2002;**17**(11):848-53.

40. Puhan MA, Steurer J, Bachmann LM, et al. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Annals of Internal Medicine* 2005;**143**(3):184-89.
41. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *American Journal of Medicine* 1998;**104**(4):374-80.
42. Sox CM, Doctor JN, Koepsell TD, et al. The influence of types of decision support on physicians' decision making. *Archives of Disease in Childhood* 2009;**94**(3):185-90.
43. Bachmann LM, Steurer J, ter RG. Simple presentation of test accuracy may lead to inflated disease probabilities. *BMJ* 2003;**326**(7385):393.
44. Vermeersch P, Bossuyt X. Comparative Analysis of Different Approaches to Report Diagnostic Accuracy. *Archives of Internal Medicine* 2010;**170**(8):734-35.
45. Young JM, Glasziou P, Ward JE. General practitioners' self ratings of skills in evidence based medicine: validation study. *BMJ* 2002;**324**(7343):950-51.
46. Sassi F, McKee M. Do clinicians always maximize patient outcomes? A conjoint analysis of preferences for carotid artery testing. *J Health Serv Res Policy* 2008;**13**(2):61-66.
47. Gigerenzer G. *What are natural frequencies?*, 2011.
48. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;**327**(7417):741-44.
49. Hoffrage U, Gigerenzer G, Krauss S, et al. Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition* 2002;**84**(3):343-52.
50. Edwards W. 25. Conservatism in human information processing. In: Kahneman D, Slovic P, Tversky A, eds. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982:359-69.
51. Zhelev Z, Garside R, Hyde C. A qualitative study into the difficulties experienced by healthcare decision makers when reading a Cochrane diagnostic test accuracy review. *Systematic reviews* 2013;**2**:32.
52. Cochrane Diagnostic Test Accuracy Working Group. Handbook for DTA Reviews [Internet]: The Cochrane Collaboration, 2013 [accessed 13.10.14].
53. GRADE working group [Internet]. Secondary GRADE working group [Internet] 2014 [accessed 27.3.2014].  
<http://www.gradeworkinggroup.org/index.htm>.