LESS IS MORE

Reassurance After Diagnostic Testing With a Low Pretest Probability of Serious Disease

Systematic Review and Meta-analysis

Alexandra Rolfe, MBChB; Christopher Burton, MD

Importance: Diagnostic tests are often ordered by physicians in patients with a low pretest probability of disease to rule out conditions and reassure the patient.

Objective: To study the effect of diagnostic tests on worry about illness, anxiety, symptom persistence, and subsequent use of health care resources in patients with a low pretest probability of serious illness.

Evidence Acquisition: Systematic review and meta-analysis. We searched MEDLINE, the Cochrane Central Register of Controlled Trials, EMBASE, PsychINFO, CINAHL, and ProQuest Dissertations electronic databases through December 31, 2011, for eligible randomized controlled trials. We independently identified studies for inclusion and extracted the data. Disagreements were resolved by discussion. We performed meta-analysis if heterogeneity was low or moderate ($I^2 < 50\%$).

Results: Fourteen randomized controlled trials that included 3828 patients met the inclusion criteria and were analyzed with outcomes categorized as short term (≤3 months) or long term (>3 months). Three trials showed no overall effect of diagnostic tests on illness worry (odds ratio, 0.87 [95% CI, 0.55-1.39]), and 2 showed no effect on nonspecific anxiety (standardized mean difference, 0.06 [−0.16 to 0.28]). Ten trials showed no overall long-term effect on symptom persistence (odds ratio, 0.99 [95% CI, 0.85-1.15]). Eleven trials measured subsequent primary care visits. We observed a high level of heterogeneity among trials (I^2=80%). Meta-analysis after exclusion of outliers suggested a small reduction in visits after investigation (odds ratio, 0.77 [95% CI, 0.62-0.96]).

Conclusions and Relevance: Diagnostic tests for symptoms with a low risk of serious illness do little to reassure patients, decrease their anxiety, or resolve their symptoms, although the tests may reduce further primary care visits. Further research is needed to maximize reassurance from medically necessary tests and to develop safe strategies for managing patients without testing when an abnormal result is unlikely.

JAMA Intern Med. 2013;173(6):407-416.
Published online February 25, 2013.
doi:10.1001/jamainternmed.2013.2762

Author Affiliations: Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland. Dr Burton is now affiliated with the Centre of Academic Primary Care, University of Aberdeen, Aberdeen, Scotland.

MANY PATIENTS PRESENT in primary care with symptoms that are not caused by serious illness. Approximately one-sixth of primary care visits and more than one-third of referrals from primary to secondary care occur for symptoms for

CME available online at jamanetworkcme.com and questions on page 405

which no organic pathology is apparent.[1-4] Such symptoms with a low probability of disease pose a problem for clinicians and health services in terms of whether or how far to perform diagnostic tests.[5] Although clinical guidelines promote the rational use of diagnostic tests,[6] clinicians often order these tests for patients who do not meet these criteria. Physicians commonly express the belief that patients want these diagnostic tests and that the tests reassure patients[3,7]; consequently, they may propose such tests more often than patients actually seek them.[8,9]

See also page 418

See Invited Commentary at end of article

Psychological models of reassurance suggest the following 2 components: short-term emotional relief and longer-term cognitive assurance.[10] The cognitive component of reassurance is necessary for long-term benefit through alterations in symptom appraisal.[11,12] This change in turn leads to reduced anxiety, less awareness of symptoms, less seeking of medical help, and a change in the belief that the symptoms may represent serious disease.[10]

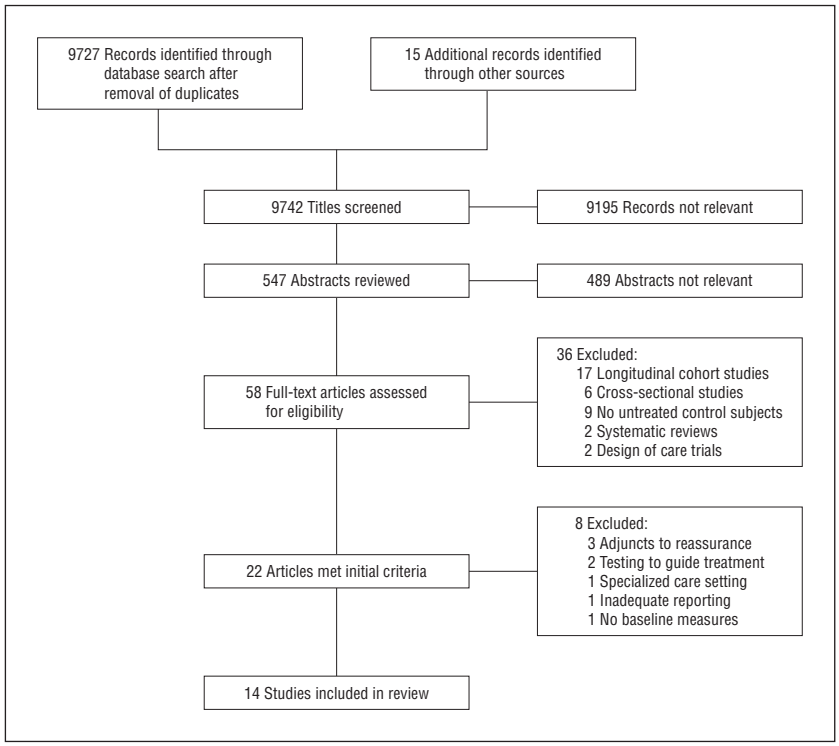

Figure 1. Flowchart indicating selection of studies for inclusion. Reasons for exclusion of trials exceed the numbers of excluded trials owing to more than 1 reason applied to some trials.

We performed a systematic review to measure the effect of diagnostic tests on reassurance in patients with a low probability of serious disease. Because a previous systematic review identified very few trials that directly measured reassurance,[13] we expanded the concept of reassurance to include the following 4 separate components and consequences: the specific concern that symptoms might represent serious illness (illness concern), nonspecific anxiety, persistence of the symptoms, and subsequent primary care visits.

METHODS

We performed a systematic review and meta-analysis of published trials following meta-analysis guidelines[14] and using the preferred reporting items for systematic review and meta-analysis guidelines (the PRISMA statement).[15] The protocol for the review is held by the authors. Because this study did not involve individual patients, ethical approval was not required.

SEARCH STRATEGY

We performed a systematic search of the OVID MEDLINE, Cochrane Central Register of Controlled Trials, EMBASE, PsychINFO, CINAHL, and ProQuest Dissertations electronic databases for publications from inception through December 31, 2011. The search was designed to be sensitive rather than specific and identified randomized controlled trials that included the following search terms: *reassur*, anxi*, quality of life,* or *satisfaction* (1 of these); *investigat*, test*, imag*, x-ray, radiography, endoscopy, colonoscopy,* or *scan* (1 of these); and *negativ*, normal,* or *benign* (≥1 of these) (eAppendix; http://www.jamainternalmed.com). We included original articles and checked the reference lists of relevant systematic reviews for trials that met the eligibility criteria. Trials published in languages other than English were eligible for inclusion. We also obtained references cited in other identified publications and considered them for inclusion.

REVIEW OF ELIGIBILITY CRITERIA

We considered trials for inclusion if they met the following 5 criteria:

1. The study design was a randomized controlled trial (including cluster trials).
2. Participants included adults (aged ≥18 years) with symptoms indicating a low probability of serious disease.
3. Interventions consisted of initial diagnostic testing in primary or secondary care for symptoms with a low probability of serious disease based on clinical features. Tests included imaging (radiography, computed tomography, magnetic resonance imaging, and ultrasonography), endoscopy, and cardiac testing. For trials with more than 2 intervention arms, we included the intervention most closely representing testing alone.
4. Comparisons included initial nontesting. Control groups received usual care or empirical treatment. Trials that permitted subsequent diagnostic testing if symptoms persisted were eligible.
5. Outcomes included illness concern, generalized or nonspecific anxiety, change in the original symptoms, and subsequent use of health care resources in primary care defined as subsequent visits to a physician. We defined outcomes as short term (≤3 months) or long term (>3 months); if multiple points after 3 months were available, we used those closest to 1 year.

Trials were excluded if they were not published in a peer-reviewed journal or were undertaken in tertiary care (eg, a spinal surgery center)[16] where the prevalence of serious conditions was expected to be high. In addition, we excluded trials for which outcomes were not reported at baseline and completion.

TRIAL SELECTION AND ANALYSIS

The methods for trial selection and assessment of risk of bias (using the Cochrane risk of bias tool[17]) are described in detail in the eMethods. Eligible data were converted to odds ratios and displayed as forest plots. If statistical heterogeneity was low to moderate ($I^2 < 50\%$), we performed random-effects meta-analysis using methods from the Cochrane Handbook[18]; details of this procedure are available in the eMethods. We converted odds ratios for the use of health care resources to a number needed to investigate.

RESULTS

Figure 1 shows the yield of relevant literature identified from the search strategy and the application of the review eligibility criteria.

Table 1. Characteristics of RCTs

Source[a]	Setting	No. of Patients/ Controls	Mean Age, y	Male Sex, %	Symptom	Dx Test vs Control	Outcomes	Follow-up	Additional Information[b]
Asante et al,[28] 1998; Asante et al,[33] 1999 (1996)	UK Secondary care	78/76	32	56	Dyspepsia	Endoscopy vs no endoscopy	Symptoms Use of health care	6 mo	Patients aged <45 y and seronegative for *Helicobacter pylori* 14 Controls underwent endoscopy No cancer detected in 92 endoscopies
Bytzer et al,[19] 1994 (1989)	Denmark Primary care	208/206	44	43	Dyspepsia	Endoscopy vs empirical treatment	Symptoms Use of health care	12 mo	Dyspepsia for several years, 10% debilitating symptoms, 33% vomiting, 50% night pain 136 Controls underwent endoscopy 4 Cancers detected in 344 endoscopies
Cuddihy et al,[20] 2005 (1998)	US Primary care	13/11	52	29	Dyspepsia	Endoscopy vs empirical treatment	Use of health care	1.5-6 mo	Additional Dx testing uncommon No cancer detected in endoscopies
Delaney et al,[21] 2000 (1995)	UK Primary care	256/186	62	49	Dyspepsia	Endoscopy vs usual care	Symptoms Use of health care	12-18 mo	Patients aged >50 y with new or recently recurrent dyspepsia 75 Controls underwent endoscopy 8 Cancers detected in 363 endoscopies
Delaney et al,[22] 2001	UK Primary care	285/193	37	57	Dyspepsia	*H pylori* testing with selective endoscopy vs usual care	Symptoms Use of health care	12-18 mo	Patients aged <50 y with dyspepsia >4 wk 127 Patients and 48 controls underwent endoscopy No cancers detected in endoscopies
Djais and Kalim,[29] 2005 (2003)	Indonesia Secondary care	51/50	40	55	Back pain	Radiography vs usual care	Symptoms	3 wk	Low back pain <3 mo No serious disease detected
Duggan et al,[23] 2009 (1995)	UK Primary care	187/178	45% > 50	55	Dyspepsia	Endoscopy vs empirical treatment	Symptoms Use of health care	2-12 mo	70 Controls underwent endoscopy No cancer detected

(continued)

STUDY CHARACTERISTICS

We found 14 trials comparing testing with nontesting; these trials included 3828 patients. All the trials were published in English. Nine trials took place in a general practice/family medicine setting,[19-27] and 5, in a general or specialist internal medicine practice setting.[28-32] Eight of the trials involved diagnostic testing for dyspepsia (endoscopy or radiography)[19-23,27,28,30]; 3 involved radiography for back pain.[25,26,29] One each involved blood tests and electrocardiography for chest pain,[32] imaging for headache,[31] and continuous event recorders for palpitations.[24] With the exception of 2 trials,[19,27] studies involved recent rather than persistent symptoms. Several studies specified a minimum duration to exclude trivial conditions for which diagnostic testing may not be clinically necessary. Trials comparing testing with nontesting reported outcome data on illness concern,[25,31,32] nonspecific or general anxiety,[24,26] symptoms,[19,21-23,25-30,32] and subsequent primary care visits.[19-23,25-28,31,32] Ten trials reported short-term data[20,23-27,29-32] and 13, long-term data.[19-28,30-32] Long-term follow-up varied from 4 to 18 months. Substantial variation existed between trials in the measurement tools used. Trial characteristics are summarized in **Table 1** and main outcomes in **Table 2**.

RISK OF BIAS

Figure 2 summarizes the risk of bias for all included trials. All trials were randomized at the level of the individual patient. No study had clearly inadequate randomization procedures or allocation concealment, although details were unclear in 6 trials for each of these criteria.[19,24,27,29,30,32] One study used a modified Zelen preconsent randomization to minimize the risk of patients feeling more anxious about being declined a diagnostic test.[31] Most studies were of moderate quality; outcomes from those with the highest quality[23,27,31,35] did not differ noticeably from the other outcomes.

Table 1. Characteristics of RCTs (continued)

Source[a]	Setting	No. of Patients/ Controls	Mean Age, y	Male Sex, %	Symptom	Dx Test vs Control	Outcomes	Follow-up	Additional Information[b]
Giannini et al,[30] 2008	Italy Secondary care	303/309	44	57	Dyspepsia (GERD only)	Endoscopy vs empirical treatment	Symptoms	1-6 mo	≥3 mo typical GERD symptoms 42 Controls underwent endoscopy No cancer detected
Hoefman et al,[24] 2007; Hoefman et al,[34] 2005 (1999)	The Netherlands Primary care	127/117	50	26	Palpitations	Continuous event recorder vs usual care	Anxiety	1.5-6 mo	New episodes or lightheadedness Relevant cardiac Dx in 22% of both groups
Howard et al,[31] 2005 (1999)	UK Secondary care	76/74	38	78	Chronic daily headache	Cranial MRI vs not offered scan	Illness concern Use of health care	3-12 mo	Of 33 controls with HADS, 11 underwent MRI No serious disease detected
Kendrick et al,[25,35] 2001 (1995)	UK Primary care	210/211	39	41	Back pain	Radiography vs usual care	Illness concern Symptoms Use of health care	1.5-2 mo	Recruited at first visit for back pain
Kerry et al,[26] 2002; Kerry et al,[36] 2000 (1996)	UK Primary care	73/80	44	50	New-onset back pain	Radiography vs no radiography	Anxiety Symptoms Use of health care	1.5-12 mo	Recruited at first visit for back pain
Laheij et al,[27] 1998 (1995)	The Netherlands Primary care	42/42	43	51	Dyspepsia	Endoscopy vs empirical treatment	Use of health care Symptoms	2.5-12 mo	13 Controls underwent endoscopy 3 Cancers detected in 51 endoscopies
Sox et al,[32] 1981 (1978)	US Secondary care	93/93	Not stated	Not stated	Chest pain	ECG and blood tests vs no tests	Illness concern Symptoms Use of health care	3 wk to 4 mo	Open-access clinic, recent chest pain and low probability of cardiac disease 5 Controls underwent ECG

Abbreviations: Dx, diagnosis; ECG, electrocardiography; GERD, gastroesophageal reflux disease; HADS, Hospital Anxiety and Depression Scale; MRI, magnetic resonance imaging; RCT, randomized controlled trial; UK, United Kingdom; US, United States.

[a] The year of first recruitment is reported in parentheses. If only the year of publication is given, the dates of recruitment were not reported.

[b] Includes patient characteristics, diagnostic testing rate in controls, and serious illness detection rates where available.

No trials blinded the patients or their clinicians to the intervention, and only 1 trial reported adequate blinding of the outcome assessor at follow-up.[31] Most (but not all) patients randomized to diagnostic tests received them. Most trials included the option for patients allocated to no testing to receive subsequent diagnostic testing. The rates of later testing in patients randomized to nontesting varied, from 10% to 66% for endoscopy to 1.5% to 13% for other diagnostic tests. Three trials did not provide clear information about subsequent testing.[20,24,29]

Attrition rates varied between trials and with the duration of follow-up. Short-term outcome data were available for 75% to 100% of randomized patients; longer-term data were available for 71% to 100%. Few trials used statistical techniques to adjust for the effects of loss to follow-up.[22,24,31]

ILLNESS CONCERN

Three trials examined illness concern in relation to magnetic resonance imaging for patients with headache,[31] lumbar spine radiography for patients with back pain,[25] and blood tests and electrocardiography in patients with chest pain.[32] Heterogeneity between studies was low (I^2 = 0%). Investigation was associated with no significant reduction in illness concern in the short (odds ratio, 0.90 [95% CI, 0.51-1.59]) or in the longer term (0.87 [0.55-1.39]), as shown in **Figure 3**.

ANXIETY

Only 2 trials examined nonspecific or generalized anxiety: one in relation to lumbar radiography for patients with back pain[26] and the other in relation to the use of a continuous cardiac event recorder for patients with palpitations.[24] Heterogeneity between studies was low (I^2 = 0%). We observed a statistically insignificant increase in anxiety at longer-term follow-up in patients who underwent investigation, as shown in **Figure 4**.

SYMPTOMS

Eleven trials reported the original symptoms after 1 or more follow-up periods.[19,21-23,25-30,32] Heterogeneity between studies was high in the short term (I^2 = 67%) and low in the longer term (I^2 = 0%). Meta-analysis (**Figure 5**) indicated no effect of diagnostic testing on symptoms in the longer term (odds ratio, 0.99 [95% CI, 0.85-1.15]).

Table 2. Individual Trial Outcome Measures and Summary Results

Source	Trial Type	Measure	Diagnostic Test Group		Control Group		SMD (95% CI)	OR (95% CI)[a]
			Value	Total No. of Patients	Value	Total No. of Patients		
Illness Concern >3 mo								
Howard et al,[31] 2005	C	VAS, mean (SD)	−21 (49)	54	−17 (49)	42	−0.08 (−0.48 to 0.32)	0.86 (0.42 to 1.79)
Kendrick et al,[25,35] 2001	D	Statement agreement, No. of patients	17	40	17	44		117 (0.49 to 2.81)
Sox et al,[32] 1981	D	Statement agreement, No. of patients	12	74	16	72		0.68 (0.30 to 1.56)
Illness Concern ≤3 mo								
Kendrick et al,[25,35] 2001	D	Statement agreement, No. of patients	39	104	26	58		0.74 (0.38 to 1.42)
Sox et al,[32] 1981	D	Statement agreement, No. of patients	17	84	17	87		1.04 (0.49 to 2.21)
Anxiety >3 mo								
Hoefman et al,[24] 2007; Hoefman et al,[34] 2005	C	STAI score, mean (SD)[b]	−1.1 (9.94)	103	−4.0 (10.1)	91	0.29 (0.01 to 0.57)	
Kerry et al,[26] 2002; Kerry et al,[36] 2000	C	HADS score, mean (SD)	−1.1 (5.5)	50	−1.5 (5.5)	58	0.07 (−0.31 to 0.45)	
Anxiety ≤3 mo								
Hoefman et al,[24] 2007; Hoefman et al,[34] 2005	C	STAI score, mean (SD)[b]	−1.6 (8.3)	108	−2.6 (8.3)	82	0.12 (−0.17 to 0.41)	
Kerry et al,[26] 2002; Kerry et al,[36] 2000	C	HADS score, mean (SD)	−0.6 (4.2)	59	−0.5 (4.2)	67	−0.02 (−0.37 to 0.33)	
Symptoms >3 mo								
Bytzer et al,[19] 1994[c]	C	3-Point Likert scale, mean (SD)	−20 (22.6)	187	−20 (23.2)	186	0.01 (−0.19 to 0.21)	*1.02 (0.71 to 1.46)*
Delaney et al,[21] 2000	C	BDSS, mean (SD)	−4.7 (5.0)	190	−3.5 (5.0)	135	−0.24 (−0.46 to −0.02)	*1.36 (0.91 to 2.03)*
Delaney et al,[22] 2001	C	BDSS, mean (SD)	−3.8 (4.8)	183	−3.5 (4.5)	107	−0.06 (−0.30 to 0.17)	*0.90 (0.59 to 1.37)*
Giannini et al,[30] 2008[c]	C	4-Point Likert scale, mean (SD)	−7 (7.6)	209	−7.7 (7.6)	222	0.09 (−0.10 to 0.28)	*1.24 (0.88 to 1.75)*
Kerry et al,[26] 2002; Kerry et al,[36] 2000	C	SF-36 bodily pain score, mean (SD)	−5.7 (5.57)	50	−6.6 (5.31)	58	0.16 (−0.21 to 0.54)	1.11 (0.56 to 2.22)
Asante et al,[28] 1998; Asante et al,[33] 1999	D	Still has symptom, No. of patients	34	60	30	57		118 (0.57 to 2.44)
Bytzer et al,[19] 1994	D	Still has symptom, No. of patients	50	187	50	186		0.99 (0.63 to 1.57)
Duggan et al,[23] 2009	D	Still has symptom, No. of patients	64	143	69	137		0.80 (0.50 to 1.28)
Giannini et al,[30] 2008	D	Still has symptom, No. of patients	55	262	46	268		1.28 (0.83 to 1.98)
Kendrick et al,[25,35] 2001	D	Still has symptom, No. of patients	126	195	133	199		0.91 (0.60 to 1.37)
Laheij et al,[27] 1998	D	No. of days without treatment	96	255	100	266		1.00 (0.70 to 1.43)
Sox et al,[32] 1981	D	Still has symptom, No. of patients	14	74	14	72		0.97 (0.42 to 2.20)
Symptoms ≤3 mo								
Djais and Kalim,[29] 2005[c]	C	VAS, mean (SD)	−43 (5.44)	59	−4 (5.92)	67	−0.05 (−0.04 to 0.30)	*1.82 (0.80 to 4.12)*
Kerry et al,[26] 2002; Kerry et al,[36] 2000	C	SF-36 bodily pain score, No. of events	−2 (2.56)	38	−3 (2.08)	38	0.42 (−0.03 to 0.88)	*0.91 (0.48 to 1.72)*
Djais and Kalim,[29] 2005	D		10	36	7	33		1.43 (0.47 to 4.33)
Duggan et al,[23] 2009	D	Still has symptom, No. of patients	40	154	61	137		0.44 (0.27 to 0.72)
Giannini et al,[30] 2008	D	Still has symptom, No. of patients	30	395	31	298		0.98 (0.57 to 1.66)
Kendrick et al,[25,35] 2001	D	Still has symptom, No. of patients	148	199	132	203		1.56 (1.02 to 2.40)
Sox et al,[32] 1981	D	Still has symptom, No. of patients	19	84	23	87		0.81 (0.40 to 1.64)

(continued)

Table 2. Individual Trial Outcome Measures and Summary Results (continued)

Source	Trial Type	Measure	Diagnostic Test Group Value	Diagnostic Test Group Total No. of Patients	Control Group Value	Control Group Total No. of Patients	SMD (95% CI)	OR (95% CI)[a]
Use of Health Care Resources								
Asante et al,[28] 1998; Asante et al,[33] 1999	C	Case notes review, mean (SD)	−0.8 (3.5)	78	−1.2 (4.4)	76	0.10 (−0.22 to 0.42)	*1.20 (0.67 to 2.14)*
Cuddihy et al,[20] 2005	C	Case notes review, mean (SD)	238 (139)	13	133 (79)	11	0.88 (0.03 to 1.72)	*4.93 (1.07 to 22.85)*
Delaney et al,[21] 2000	C	Case notes review, mean (SD)	3.46 (2.61)	254	3.95 (3.36)	184	−0.17 (−0.36 to 0.02)	*0.73 (0.52 to 1.04)*
Delaney et al,[22] 2001	C	Case notes review, mean (SD)	3.26 (2.73)	284	3.30 (2.67)	191	−001 (−0.20 to 0.17)	*0.98 (0.70 to 1.37)*
Howard et al,[31] 2005	C	Case notes review, mean (SD)	124.8 (86.4)	76	148.9 (132.4)	74	−0.22 (−0.54 to 0.11)	*1.35 (0.45 to 4.08)*
Kerry et al,[26] 2002; Kerry et al,[36] 2000	C	Case notes review, mean (SD)	1.0 (1.6)	50	1.6 (2.1)	58	−0.32 (−0.70 to 0.06)	*0.56 (0.28 to 1.11)*
Laheij et al,[27] 1998	C	Case notes review, mean (SD)	3.0 (2.4)	38	3.9 (2.5)	42	−0.36 (−0.81 to 0.08)	*0.52 (0.23 to 1.17)*
Bytzer et al,[19] 1994	D	Self-reported, No. of events	47	187	114	186		0.21 (0.14 to 0.33)
Duggan et al,[23] 2009	D	Case notes review, No. of events	80	186	108	177		0.48 (0.32 to 0.73)
Kendrick et al,[25,35] 2001	D	Self-reported, No. of events	42	195	47	199		0.89 (0.55 to 1.42)
Sox et al,[32] 1981	D	Questionnaire, No. of events	10	84	13	87		0.77 (0.32 to 1.86)

Abbreviations: BDSS, Birmingham Dyspepsia Symptoms Score; C, continuous measure; D, dichotomous; HADS, Hospital Anxiety and Depression Scale; OR, odds ratio; SF-36, 36-Item Short Form Health Survey; SMD, standardized mean difference; STAI, State-Trait Anxiety Inventory; VAS, visual analog scale.

[a] Values in italics refer to estimated ORs converted from SMD for use in meta-analysis combining continuous and dichotomous variables.

[b] Only the state anxiety component was used in the analysis.

[c] Continuous data are shown for completeness; the comparison uses the OR from dichotomous data, which was also reported.

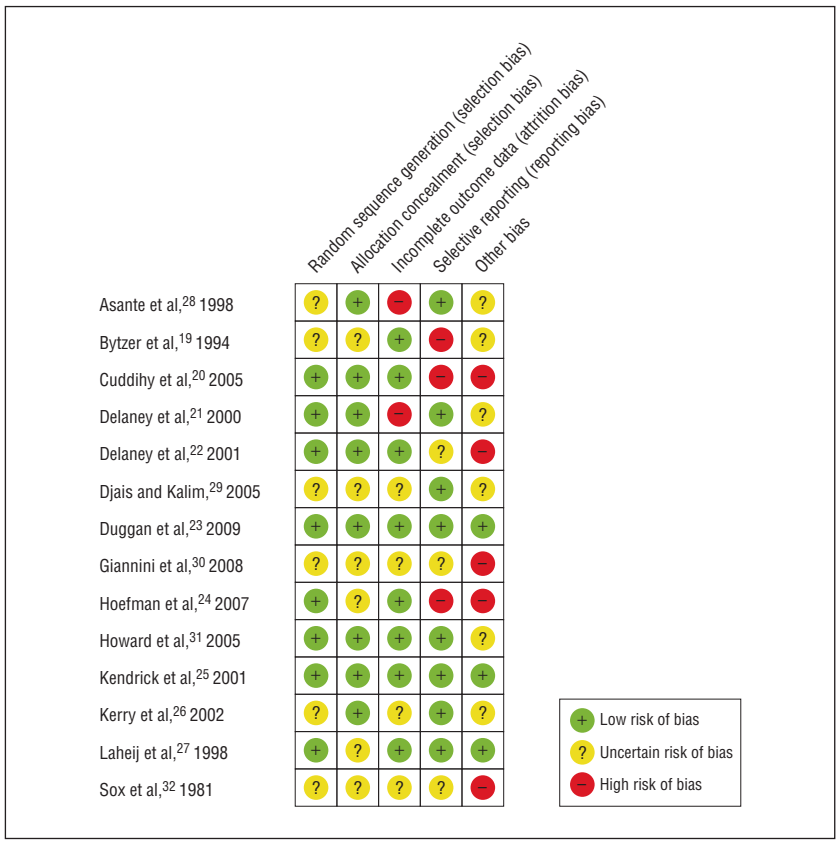

Figure 2. Risk of bias summary.

SUBSEQUENT PRIMARY CARE VISITS

Eleven trials examined primary care visit rates after the intervention.[19-23,25-28,31,32] Eight trials used case note reviews[20-23,26-28,31] and 3 used patient self-report.[19,25,32] Individual and overall effects are shown in **Figure 6**, which indicates marked heterogeneity between trials (I^2 = 80%). Most of this heterogeneity arose from 2 trials: one with a very small number of participants[20] and the other an older trial that predated *Helicobacter pylori* eradication for peptic ulcer disease.[19] When these 2 trials were excluded, heterogeneity was reduced (I^2 = 33%), and the overall effect was a reduction in subsequent visit rates of 0.77 (0.62-0.96). The rates of repeated visits for control patients were 60% in dyspepsia trials, with a number needed to investigate of 16 (95% CI, 8-100), and 20% in back pain trials, with a number needed to investigate of 26 (95% CI, 15-155).

MAIN FINDINGS

This systematic review indicates that patients' illness concern, health anxiety, and symptoms are not reduced by diagnostic testing in the short or the long term. Subsequent use of health care resources may be reduced by diagnostic testing, although the number of patients needed to investigate and avoid 1 subsequent visit varied from 16 to 26 depending on the symptom. In the context of widespread belief that diagnostic testing reassures patients, these findings suggest that physicians overestimate the value of testing when the probability of serious disease is low.

STRENGTHS AND LIMITATIONS

Although previous systematic reviews have used a narrow definition of reassurance[13] or have been limited to 1 specific clinical problem,[37,38] we included a broader assessment of reassurance by including the reduction of illness concern and the expected consequences of reassurance; we also included a wide range of clinical problems. This approach risks comparing trials that are too dissimilar and for which meta-analysis may be inappropriate; however, we took the view that models of symptom appraisal and reassurance are consistent across contexts[10-12] and that all trials centered on the decision to perform diagnostic testing or not. Substantial heterogeneity was seen only for the use of health care resources and, because the number of trials was small, we did not perform a formal subgroup comparison. Outcome measures varied from the well validated (such as the 36-Item Short Form Health Survey) to the ad hoc; the poorly validated measures might have been insufficiently sensitive to change.

The studies in this review were conducted in different places and times during which the practice of medicine and the expectations of patients changed. The study that showed the greatest influence of diagnostic testing on reassurance[19] was for peptic ulcer disease in the 1980s, when surgical treatment was common and before the recognition of *H pylori*. Most of the eligible studies were conducted in European health care systems, where access to diagnostic testing may be more constrained than in US health care. We examined reassurance for patients only; we did not examine the reassurance (including the reassurance that they were less likely to be sued) that diagnostic tests provided for physicians. A health economic analysis was beyond the scope of this review; however, because the cost of a primary care consultation is less than the cost of most diagnostic tests and because several tests were required to avert 1 consultation, the balance would not favor testing.

We did not examine differential effects of anxiety at baseline on subsequent reassurance. One study reported a prespecified comparison between more and less anxious patients and found that illness concern was reduced more by diagnostic testing in patients with high anxiety levels, but these data were not available from other trials.[31] Although our analysis highlights the limited value of diagnostic testing in terms of reassurance, it does not address the wider role of investigations in identifying disease or allowing the physician to rule out a particular differential diagnosis. Although the prevalence of serious disease such as cancer in the eligible trials varied from less than 0.5% to 3%, our findings do not address what pretest probability of disease constitutes an appropriate threshold for investigation to obtain a diagnosis. In addition, our find-

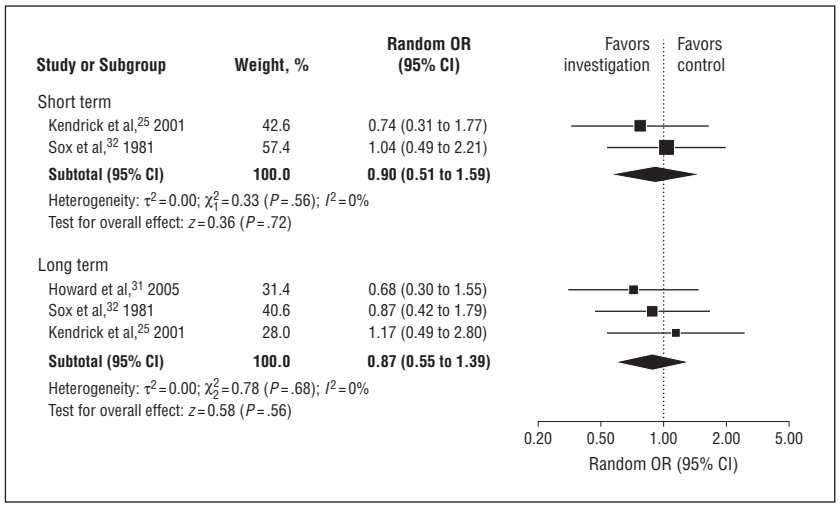

Figure 3. Effect of diagnostic testing on reduction of illness concern. The size of the data marker corresponds to the relative weight assigned in the pooled analysis using random-effects models. OR indicates odds ratio.

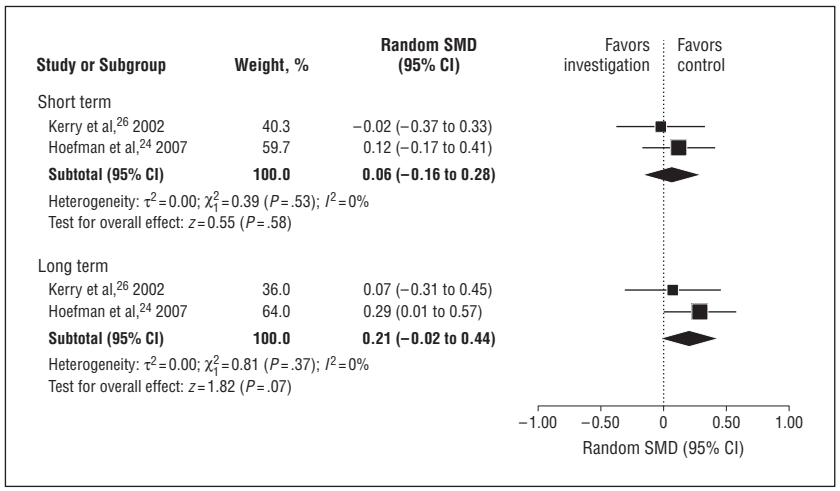

Figure 4. Effect of diagnostic testing on anxiety. The size of the data marker corresponds to the relative weight assigned in the pooled analysis using random-effects models. SMD indicates standardized mean difference.

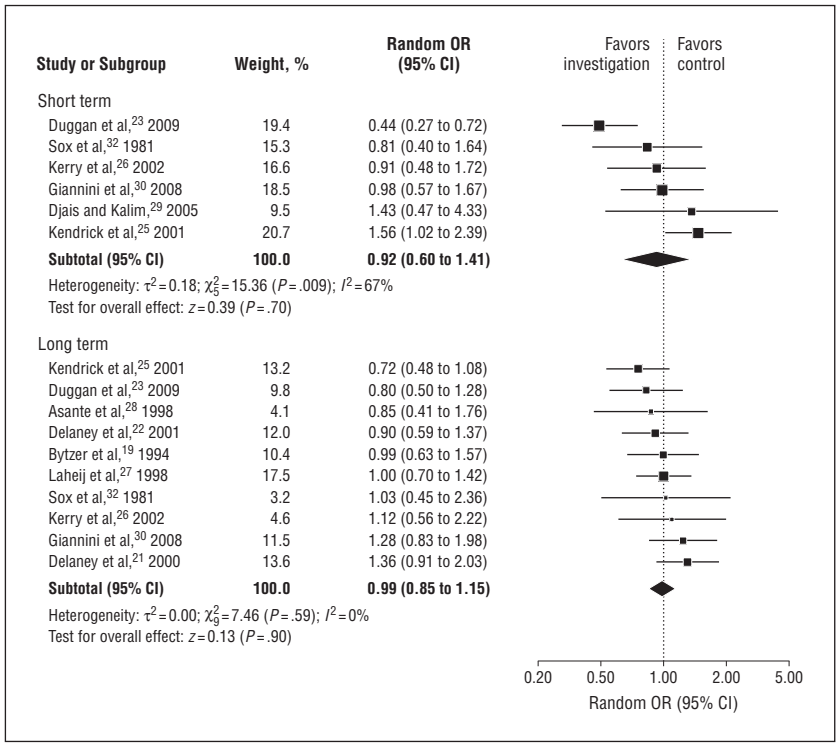

Figure 5. Effect of diagnostic testing on presenting symptoms. The size of the data marker corresponds to the relative weight assigned in the pooled analysis using random-effects models. OR indicates odds ratio.

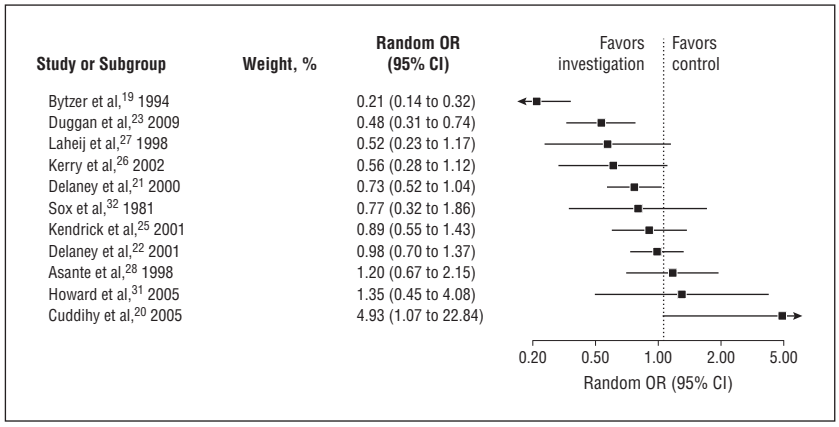

Figure 6. Effect of diagnostic testing on primary care visits. OR indicates odds ratio.

ings do not contradict guidelines for rational requesting of tests that balance benefits against harms.

We did not find studies of complex or chronic symptoms. However, persistent symptoms with negative test results are associated with frustration and dissatisfaction.[39]

INTERPRETATION

We found the use of diagnostic testing did little to reassure patients, and this finding is inconsistent with beliefs expressed by physicians.[3,7] One explanation is that the reassurance obtained by patients from negative diagnostic test results is transient. Observational studies suggest that illness concerns reappear within hours of receiving a normal (negative) test result,[40,41] whereas the trials in our review measured effects after weeks or months. The mechanism of transient reassurance appears to be predominantly emotional—a fleeting sense of relief—in contrast to a more sustained cognitive reassurance.[10] One trial included in the review attempted to place a value on reassurance and found that patients were willing to pay for the reassurance of normal findings on spine radiography, although no discernible effect was observed on measures of concern.[25] Thus, patients and physicians may value the immediate relief of reassurance, although the benefits are not sustained. We found a small reduction in subsequent primary care visits after diagnostic testing, but this reduction required several patients to undergo testing to prevent 1 visit.

RECOMMENDATIONS FOR RESEARCH AND PRACTICE

Because the number of trials in this review is relatively small, further trials may demonstrate an effect of diagnostic testing on reassurance among patients with symptoms indicating a low probability of disease. However, the small effect sizes are in keeping with the postulated psychological and behavioral mechanisms underlying symptom appraisal[11] and reassurance.[10,12] Thus, concentration of future research on the following 2 questions may be more important: (1) how to maximize the reassurance value of diagnostic tests and (2) whether reassurance should be targeted to particular patients. Three trials have reported that brief interventions to increase the acceptability of negative test results lead to improved reassurance,[42-44] and theoretical work supports this finding.[45] Targeting interventions (including reassurance and cognitive-based rehabilitation) to patients at higher risk of persistent symptoms is effective in patients with low back pain,[46] and diagnostic testing accompanied by enhanced explanation can be an appropriate strategy. Meanwhile, physicians and health care organizations should be aware of the limitations of the transient reassurance provided by negative diagnostic test results and should limit tests to those that influence clinical management. In summary, commonly used diagnostic tests have little effect on several aspects of reassurance in patients whose symptoms indicate a low pretest probability of serious illness.

Accepted for Publication: October 28, 2012.
Published Online: February 25, 2013. doi:10.1001/jamainternmed.2013.2762

Correspondence: Christopher Burton, MD, Centre of Academic Primary Care, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, Scotland (c.burton@abdn.ac.uk).

Author Contributions: Drs Rolfe and Burton are joint guarantors of this study. *Study concept and design:* Rolfe and Burton. *Acquisition of data:* Rolfe and Burton. *Analysis and interpretation of data:* Rolfe and Burton. *Drafting of the manuscript:* Rolfe and Burton. *Critical revision of the manuscript for important intellectual content:* Burton. *Statistical analysis:* Rolfe and Burton. *Administrative, technical, and material support:* Rolfe.

Conflict of Interest Disclosures: None reported.

Funding/Support: Dr Burton was supported by a Research Career Award from the Chief Scientist Office of the Scottish Government.

Previous Presentation: This study was presented at the Society for Academic Primary Care Annual Scientific Meeting; July 6, 2011; Bristol, England.

Online-Only Material: The eAppendix and eMethods are available at http://www.jamainternalmed.com.

Additional Contributions: Fay Crawford, PhD, and Kurt Kroenke, MD, provided comments on earlier versions of this manuscript.

REFERENCES

1. Burton C. Beyond somatisation: a review of the understanding and treatment of medically unexplained physical symptoms (MUPS). *Br J Gen Pract*. 2003;53(488):231-239.
2. Hatcher S, Arroll B. Assessment and management of medically unexplained symptoms. *BMJ*. 2008;336(7653):1124-1128.
3. Van der Weijden T, Van Velson M, Dinant GJ, Van Hasselt CM, Grol R. Unexplained complaints in general practice: prevalence, patients' expectations and professionals' test-ordering behaviour. *Med Decis Making*. 2003;23:226-231.
4. Barsky AJ, Orav EJ, Bates DW. Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Arch Gen Psychiatry*. 2005;62(8):903-910.
5. Salmon P, Humphris GM, Ring A, Davies JC, Dowrick CF. Primary care consultations about medically unexplained symptoms: patient presentations and doctor responses that influence the probability of somatic intervention. *Psychosom Med*. 2007;69(6):571-577.
6. Royal College of Radiologists. *Making the Best Use of Clinical Radiology Services (MBUR)*. 6th ed. London, England: Royal College of Radiologists; 2007.
7. Ring A, Dowrick C, Humphris G, Salmon P. Do patients with unexplained physical symptoms pressurise general practitioners for somatic treatment? a qualitative study. *BMJ*. 2004;328(7447):157. doi:10.1136/bmj.38057.622639.EE.
8. Salmon P, Ring A, Dowrick CF, Humphris GM. What do general practice patients want when they present medically unexplained symptoms, and why do their doctors feel pressurized? *J Psychosom Res*. 2005;59(4):255-262.
9. Little P, Dorward M, Warner G, Stephens K, Senior J, Moore M. Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *BMJ*. 2004;328(7437):444-446.
10. Coia P, Morley S. Medical reassurance and patients' responses. *J Psychosom Res*. 1998;45(5):377-386.
11. Leventhal H, Meyer D, Nerenz D. *The Commonsense Representation of Illness Danger*. Oxford, England: Pergamon; 1980.
12. Rief W, Broadbent E. Explaining medically unexplained symptoms-models and mechanisms. *Clin Psychol Rev*. 2007;27(7):821-841.
13. van Ravesteijn H, van Dijk I, Darmon D, et al. The reassuring value of diagnostic tests: a systematic review. *Patient Educ Couns*. 2012;86(1):3-8.
14. Centre for Reviews and Dissemination. *Systematic Review: CRD's Guidance for Undertaking Reviews in Health Care*. York, England: Centre for Reviews and Dissemination, University of York; 2009.
15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700. doi:10.1136/bmj.b2700.
16. Gilbert FJ, Grant AM, Gillan MGC, et al. Low back pain: influence of early MR imaging or CT on treatment and outcome: multicentre randomized trial. *Radiology*. 2004;231(2):343-351.
17. *Review Manager (RevMan), Version 5.1* [computer program]. København, Denmark: Nordic Cochrane Centre; 2011.
18. Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0*. The Cochrane Collaboration website. http://www.cochrane-handbook.org. Updated March 2011. Accessed January 18, 2013.
19. Bytzer P, Hansen JM, Schaffalitzky de Muckadell OB. Empirical H_2-blocker therapy or prompt endoscopy in management of dyspepsia. *Lancet*. 1994;343(8901):811-816.
20. Cuddihy MT, Locke GR III, Wahner-Roedler D, et al. Dyspepsia management in primary care: a management trial. *Int J Clin Pract*. 2005;59(2):194-201.
21. Delaney BC, Wilson S, Roalfe A, et al. Cost effectiveness of initial endoscopy for dyspepsia in patients over age 50 years: a randomised controlled trial in primary care. *Lancet*. 2000;356(9246):1965-1969.
22. Delaney BC, Wilson S, Roalfe A, et al. Randomised controlled trial of *Helicobacter pylori* testing and endoscopy for dyspepsia in primary care. *BMJ*. 2001;322(7291):898-901.
23. Duggan AE, Elliot CA, Miller P, Hawkey CJ, Logan RF. A clinical trial: a randomized trial of early endoscopy, *Helicobacter pylori* testing and empirical therapy for the management of dyspepsia in primary care. *Aliment Pharmacol Ther*. 2009;29(1):55-68.
24. Hoefman E, Boer KR, van Weert HCPM, Reitsma JB, Koster RW, Bindels PJP. Continuous event recorders did not affect anxiety or quality of life in patients with palpitations. *J Clin Epidemiol*. 2007;60(10):1060-1066.
25. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *BMJ*. 2001;322(7283):400-405.
26. Kerry S, Hilton S, Dundas D, Rink E, Oakeshott P. Radiography for low back pain: a randomised controlled trial and observational study in primary care. *Br J Gen Pract*. 2002;52(479):469-474.
27. Laheij RJF, Severens JL, Van de Lisdonk EH, Verbeek ALM, Jansen JBMJ. Randomized controlled trial of omeprazole or endoscopy in patients with persistent dyspepsia: a cost-effectiveness analysis. *Aliment Pharmacol Ther*. 1998;12(12):1249-1256.
28. Asante MA, Mendall M, Patel P, Ballam L, Northfield TC. A randomized trial of endoscopy vs no endoscopy in the management of seronegative *Helicobacter pylori* dyspepsia. *Eur J Gastroenterol Hepatol*. 1998;10(12):983-989.
29. Djais N, Kalim H. The role of lumbar spine radiography in the outcomes of patients with simple acute low back pain. *APLAR J Rheumatol*. 2005;8(1):45-50.
30. Giannini EG, Zentilin P, Dulbecco P, Vigneri S, Scarlata P, Saverino V. Management strategy for patients with gastroesophageal reflux disease: a comparison between empirical treatment with esomeprazole and endoscopy-orientated treatment. *Am J Gastroenterol*. 2008;103(2):267-275.
31. Howard L, Wessely S, Leese M, et al. Are investigations anxiolytic or anxiogenic? a randomised controlled trial of neuroimaging to provide reassurance in chronic daily headache. *J Neurol Neurosurg Psychiatry*. 2005;76(11):1558-1564.
32. Sox HC Jr, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. *Ann Intern Med*. 1981;95(6):680-685.
33. Asante MA, Lord J, Mendall M, Northfield T. Endoscopy for *Helicobacter pylori* sero-negative young dyspeptic patients: an economic evaluation based on a randomized trial. *Eur J Gastroenterol Hepatol*. 1999;11(8):851-856.
34. Hoefman E, van Weert HC, Reitsma JB, Koster RW, Bindels PJ. Diagnostic yield of patient-activated loop recorders for detecting heart rhythm abnormalities in general practice: a randomised clinical trial. *Fam Pract*. 2005;22(5):478-484.
35. Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M. The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health Technol Assess*. 2001;5(30):1-69.
36. Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J. Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography? *Health Technol Assess*. 2000;4(20):i-iv, 1-119.
37. Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet*. 2009;373(9662):463-472.
38. Delaney B, Ford AC, Forman D, Moayeddi P, Qume M. Initial management strategies for dyspepsia. *Cochrane Database Syst Rev*. 2005;19(4):CD001961.
39. Speckens AEM, Spinhoven P, Van Hemert AM, Bolk JH, Bolk JH. The Reassurance Questionnaire (RQ): psychometric properties of a self-report questionnaire to assess reassurability. *Psychol Med*. 2000;30(4):841-847.
40. Lucock MP, Morley S, White C, Peake MD. Responses of consecutive patients to reassurance after gastroscopy: results of self administered questionnaire survey. *BMJ*. 1997;315(7108):572-575.
41. Wiklund I, Glise H, Jerndal P, Carlsson J, Talley

NJ. Does endoscopy have a positive impact on quality of life in dyspepsia? *Gastrointest Endosc*. 1998;47(6):449-454.
42. Arnold J, Goodacre S, Bath P, Price J. Information sheets for patients with acute chest pain: randomised controlled trial. *BMJ*. 2009;338:b541. doi: 10.1136/bmj.b541.
43. Petrie KJ, Müller JT, Schirmbeck F, et al. Effect of providing information about normal test results on patients' reassurance: randomised controlled trial. *BMJ*. 2007;334(7589):352 doi:http://dx.doi.org/10.1136/bmj.39093.464190.55.
44. Serinken M, Zencir M, Karcioglu O, Sener S, Turkcuer I. Value of the timing of informing the emergency department patients on cardiac test results: a randomized controlled study. *Eur J Emerg Med*. 2009;16(2):74-79.
45. Salmon P. Conflict, collusion or collaboration in consultations about medically unexplained symptoms: the need for a curriculum of medical explanation. *Patient Educ Couns*. 2007;67(3):246-254.
46. Hill JC, Whitehurst DG, Lewis M, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*. 2011;378(9802):1560-1571.

INVITED COMMENTARY

Diagnostic Testing and the Illusory Reassurance of Normal Results

Diagnostic testing is enticing to patients and clinicians. It appears more objective and less pedestrian than a simple clinical interview and physical examination. Medical certainty is seldom solidified until all the tests results are in. Patients anxiously await the telephone call or letter announcing "your tests are all normal." Indeed, the grander the technology, the more alluring. However, the testing imperative can become addictive. As noted in a 1991 cautionary essay:

> Technology pounds upon the shore, but the danger is the undertow. The effacement of sand castles we abide; the relentless tug is another matter, sucking us deeper. Like systole and diastole, there is faint pause, endless indications. Imaging fits the metaphor, wave after wave: radionuclide scanning, computerized tomography, magnetic resonance imaging, duplex sonography. The very names captivate our diagnostic instincts, and yet excess appears inevitable. . . . Endoscopy is equally irrepressible. . . . To witness a cause transcends the more banal concerns of costs and therapeutic outcome. Follow-up is inconvenient. To wait and see whether growing suspicions will justify exploration or whether signs and symptoms prove merely transitory cannot compete with immediate visualization.[1]

Like many of our treatments, however, diagnostic testing is not without its adverse effects. Increased health care costs are the most obvious: wide geographic variations in the use of expensive tests persist more than 30 years after such inexplicable variation was first exposed.[2] Still more insidious consequences lurk. One is the problem of false-positive results. The prevalence of detecting a serious condition may be as low as 0.5% to 3.0% when diagnostic tests are ordered in patients with a low probability of disease,[3] meaning that a diagnostic test with a 90% sensitivity and 90% specificity would yield 4 to 19 false-positive results for every true-positive result in patients for whom the test is ordered simply to rule out a disease for which the clinical suspicion is already low. This disproportionately high false-positive rate may then cascade into additional and sometimes invasive procedures, not to mention considerable patient anxiety that may persist months after a negative finding of a workup cancels out the initial test results. One might consider this PTSD (post-test stress disorder) an iatrogenic variant of the traditional PTSD (post-traumatic stress disorder).

False-negative results can also be a concern. For example, the high diagnostic accuracy of abdominal computed tomography for appendicitis and renal colic does not generalize to patients presenting to the emergency department with undifferentiated upper abdominal pain, where the negative predictive value is only 64%.[4] That means as many as 1 of 3 normal abdominal computed tomographic scans in this population may represent a false-negative test result, with the most commonly missed pathologic changes being inflammatory conditions of the biliary tract and upper gastrointestinal tract systems.

In addition, reflexive test ordering may marginalize the clinical examination. Preliminary data suggest that that the history typically accounts for 75% or more of the diagnostic yield when evaluating common symptoms; the physical examination, 10% to 15%; and testing, generally less than 10%.[5] Ironically, the US reimbursement system financially incentivizes these components in the reverse order. Diagnostic testing and procedures receive the highest remuneration and, even within the clinical encounter, evaluation and management coding favors from a billing standpoint the physical examination of more bodily parts (even if irrelevant to the presenting complaint) over a detailed and more diagnostically informative interview.

Despite these limitations of diagnostic testing in patients with a low probability of disease, a conventional justification is reassurance of the patient. However, the meta-analysis by Rolfe and Burton[3] suggests that even this benefit may be overestimated. The authors included only trials in which patients with a low probability of disease were randomized to receive initial diagnostic testing vs a nontesting approach. The patient sample is appropriate because most would agree that diagnostic testing is warranted to rule in or confirm a suspected disease, determine its extent or sever-