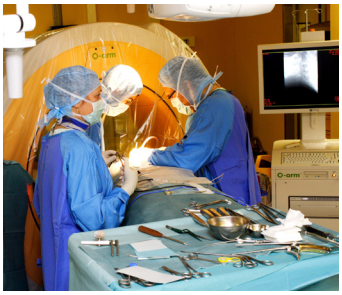
research

Diagnostic accuracy of self-tests p 77



Impact of cervical neurotomy on language function p 80

Physician assistants and quality of care p 82

The pitfalls of diagnostic self-tests

ORIGINAL RESEARCH Cross sectional review of regulation and evidence of performance

Direct-to-consumer self-tests sold in the UK in 2023

Hillier B, Deeks JJ, Alderman J, et al

Cite this as: *BMJ* 2025;390:e085547
Find this at doi: 10.1136/bmj-2025-085547

Study question What is the evidence supporting performance claims of self-tests sold on the UK high street, and how usable and safe are they?

Methods A cross sectional review was conducted of a comprehensive sample of self-tests sold in supermarkets, pharmacies, and health shops in the UK. The instructions for use leaflets, packaging, sampling, testing equipment, and evidence documents (clinical and lay person study reports) were reviewed. Details of accuracy claims on boxes and in instructions for use documents and clinical and lay person study reports were summarised. Quality of study reports was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Studies 2) tool. Ergonomics, usability, and safety concerns about the equipment and instructions were evaluated.

Study answer and limitations 30 self-tests for 19 different conditions were obtained. Accuracy claims were made in instructions for use for 24/30 tests: accuracy for 19, sensitivity for 17, and specificity for 16. Performance claims of ≥98% were made on accuracy for 53% (10/19) of tests, 41% (7/17) on sensitivity, and 63% (10/16) on specificity. Where reference standards were reported in instructions for use, 29% (5/17) evaluated the accuracy of self-tests against similar rapid tests. Manufacturers' study reports were only available for 12/30 tests; all were rated as having some unclear risk of bias because of poor reporting, and most were rated as having high applicability concerns because of inappropriate study designs. More than half of the 30 self-tests had at least one high risk concern about usability or safety. The review assessment process was necessarily subjective, but used consensus methods involving multidisciplinary experts.

What this study adds Not all manufacturers are willing to share evidence on the performance of their self-tests. Some self-tests have been approved based only on laboratory studies with unrepresentative and poor descriptions of study populations, unsuitable choices of comparator tests, and lack of blinding. Issues over sampling and test equipment, which might lead to errors, were identified.

Funding, competing interests, and data sharing Funded by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. No competing interests declared. No additional data available.

ORIGINAL RESEARCH Cross sectional review of intended use, instructions, and decision making

Direct-to-consumer self-tests sold in the UK in 2023

Davenport C, Richter A, Hillier B, et al

Cite this as: *BMJ* 2025;390:e085546
Find this at doi: 10.1136/bmj-2025-085546

Study question Is the information provided for self-tests sold in shops on UK high streets suitable for consumers to make informed decisions about use, interpretation, and subsequent health actions?

Methods A cross sectional review was conducted of 30 self-tests for 19 different conditions sold in supermarkets, pharmacies, and health and wellbeing shops within a 10 mile radius of the University of Birmingham's campus at Edgbaston in 2023. Claims on the packaging and instructions for use leaflets were assessed for quality of information about intended use of the test, the biomarker and clinical condition to be detected, interpretation of results, and recommendations for post-test actions. The coherence of intended use and post-test recommendations with evidence based guidance was also assessed.

Study answer and limitations 30 self-tests assessing 20 biomarkers for 19 different conditions were included. Information to guide purchase was present on a few boxes: who should use the test and when (8/30, 27%), action after the test result (7/30, 23%), and numerical test performance (10/30, 33%). From the information provided either on the packaging or in the instructions for use leaflets, 21 (70%) tests were judged to be used for diagnosis and 15 (50%) to be used for screening, although 3/21 (14%) did not provide any information about symptoms and 10/15 (67%) did not provide any information about risk factors to guide use. Use of the tests as advertised was judged contrary to evidence based guidance for 11 of the 19 (58%) conditions studied. Although data extraction was checked by a second, independent person and a multidisciplinary team, it could have resulted in overestimation of the accessibility of information available to lay people to support decision making after use of self-tests.

COMMENTARY Tests should be clinically useful and part of an evidence based pathway

Rapid advances in diagnostic technology, coupled with persuasive advertising, have resulted in a surge of direct-to-consumer self-tests, often sold under the banner of "wellness." Marketed as tools for empowerment and early detection, these self-tests promise convenience and autonomy and are promoted as tools for individuals to proactively manage their health. Self-testing can also offer an anticipated alternative route for health management, given the challenges in accessing primary care. But behind the glossy marketing lie multiple difficulties with real world use, and considerable potential for harm.

NICK ROBINSON

Two analyses in this issue illuminate the current state of self-tests available on the UK market. In one, Davenport and colleagues examined the information provided to consumers in information leaflets,[1] and in the other, Hillier and colleagues interrogated manufacturers' claims about test performance.[2] The findings are concerning as consumers are not given the information they need to enable informed decision making. Claims about test accuracy are often unsubstantiated, and the evidence base is worryingly thin. When tests are available, both direct to consumer and through the NHS, the pathways differ substantially. For instance, NHS guidance makes it clear that screening for prostate specific antigen should be done only after discussing the complexity of the meaning of the test result—and not simply ordered.[3]

Tests alone do not improve health

Poor quality tests can cause real harm to patients. As well as the impact on individuals, healthcare systems are likely to be affected by the downstream consequences of interpreting and acting on dubious test results. False positive results may lead to anxiety, unnecessary investigations, overdiagnosis, and overtreatment. False negative results can offer erroneous reassurance, which could potentially lead to delays in seeking appropriate medical

Jessica Watson
Jessica.Watson@bristol.ac.uk
Margaret McCartney
See bmj.com for author details

What this study adds The UK self-test market fails to support informed use. Most of the tests lack essential information about who should use them, how to interpret results, and what actions to take next. The effectiveness of regulatory oversight is a serious concern. Coherent guidance and improved regulation to protect both individuals and healthcare systems are needed.

Funding, competing interests, and data sharing
Funded by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. No competing interests declared. No additional data available.

Summary of direct-to-consumer self-tests in study

Test details	All obtained self-tests (n=30)
Sample type	Capillary blood (n=11), urine (n=6), nasal swab (n=4), semen (n=3), vaginal swab (n=3), faeces (n=2), throat swab (n=1)
Type of condition tested*	Respiratory (n=5), menopause (n=4), nutritional deficiencies (n=4), male fertility (n=3), vaginal infections (n=2), bowel health (n=2), prostate health (n=1), stomach ulcer (n=1), gluten sensitivity (n=1), urinary tract infection (n=1), chlamydia (n=1), HIV (n=1), kidney health (n=1), thyroid health (n=1), cholesterol level (n=1), blood glucose level (n=1)
Price†	Less than £5 (n=5), £5 to £9.99 (n=7), £10 to £14.99 (n=13), £15 to £19.99 (n=2), more than £20 (n=3)
Accuracy claims of ≥98% from instructions for use	Accuracy (10 of 19 claims, 53%), sensitivity (7 of 17 claims, 41%), specificity (10 of 16 claims, 63%)
Contact from test manufacturer or distributor‡	No response (n=12), provided by distributor (n=10), refused by manufacturer (n=6), provided by manufacturer (n=2)

£1.00 (€1.16; $1.36).
Data represent number of tests unless otherwise stated.
*Categorised based on claim in test product name, test box, or instructions for use.
†Original purchase cost in 2023.
‡In response to requesting access to evidence documentation (clinical and lay person study reports).

care, with risks of delayed diagnosis and treatment. For example, using a "bowel health" self-test to detect "early stages of colon cancer" is not only misleading, it can also be dangerous. Even accurate results may not be useful to those wanting to self-monitor, owing to physiological variation.[4]

Where do these tests stand?

Self-tests should not be dismissed outright though. History offers cautionary tales: when home pregnancy tests were first introduced, some doctors argued that women could not be trusted to use them.[5] Clearly that is not the case. The use of HIV self-tests has been extensively and carefully evaluated, with decades of research, including large randomised controlled trials.[6 7] UK guidelines now recommend self-testing for HIV in at risk groups in areas of high seroprevalence, to increase uptake and frequency of testing and help overcome barriers to testing.[8] These tests are, however, "binary," with a yes or no answer, integrated within healthcare systems, and with clear actions to be taken based on the results. Many of the tests evaluated by Davenport and colleagues and Hillier and colleagues yielded ambiguous results, meaning that the interpretation, and the actions resulting from them, were more complex.[1 2]

Giving people access to information and tools to understand their health is not in itself inherently problematic. But tools must be fit for purpose. Currently, consumers do not routinely have access to independent information before purchase, are not diverted from purchasing inappropriate tests when symptoms are present, and do not usually receive the result in a system designed to support understanding of the results. In the meantime, NHS general practitioners could be called upon to assist with the interpretation of a test that may not have been warranted, or, alternatively, could have been justified within the NHS.

To be useful, tests must be more than accurate—they should also have clinical utility. To be of genuine value, tests are one part of a system where the meaning of a result can be contextualised with the rationale for performing it, leading to evidence based actions and improved outcomes. Tests alone do not improve health.

Regulatory implications

To make informed decisions about self-testing, consumers must be provided with clear, balanced information about what a test can and cannot do. This information should be independently created, tested with the help of people, and mandatorily supplied. Meantime, the sale of tests lacking clinical utility should not be allowed. The NHS should not be expected to provide a "free" follow-up service for companies offering inappropriate, oversold, and low value tests. However, it could do more to explain to people what tests are available when symptoms are not manifest, and what the evidence is for other commonly marketed self-tests. This information could be placed prominently on NHS websites and placed on social media platforms where much of the advertising takes place, offering people a trusted source of informed advice.

Regulation should ensure that tests meet appropriate standards for accuracy, not just based on controlled laboratory settings but on real world use, building on existing guidance for best practice in test evaluation.[9] Commercial developers should work in partnership with clinicians and patients to ensure that innovations address genuine clinical need and do not prey on or create health anxieties to generate profit. Research to measure the impact of self-testing on patient outcomes and workload effects on healthcare systems would be welcomed. Empowering individuals to take an active role in their health is an important goal, but if self-tests are to be sold directly to the public, they must be supported by high quality evidence, robust regulation, trustworthy public information, and clear pathways for interpretation and follow up.

Cite this as: *BMJ* 2025;390:r1476

Find the full version with references at http://dx.doi.org/10.1136/bmj.r1476

Cervical neurotomy for post-stroke aphasia

ORIGINAL RESEARCH Multicentre, randomised controlled trial

Right C7 neurotomy at the intervertebral foramen plus intensive speech and language therapy versus intensive speech and language therapy alone for chronic post-stroke aphasia

Feng J, Hu R, Lyu M, et al

Cite this as: *BMJ* 2025;389:e083605
Find this at doi: 10.1136/bmj-2024-083605

Study question Can right neurotomy of the seventh cervical nerve (C7) at the intervertebral foramen plus intensive speech and language therapy (SLT) improve language function compared with intensive SLT alone in patients with chronic aphasia after stroke?

Methods In this assessor blinded randomised controlled trial in four centres in Shanghai, China, 50 patients aged 40-65 years with aphasia for more than one year after a single left hemispheric stroke were recruited and randomised 1:1 to receive either C7 neurotomy plus intensive SLT or intensive SLT alone (control group), stratified by treatment centre. The primary outcome was change in score on the 60 item Boston naming test (scores 0-60, with higher scores indicating better naming function) from baseline to one week after neurotomy plus intensive SLT for three weeks or intensive SLT for three weeks after one week's deferral. Secondary outcomes included change in severity of aphasia using the aphasia quotient, calculated using the western aphasia battery (scores 0-100, with higher scores indicating better function), and patient reported outcomes on quality of life and depression after stroke.

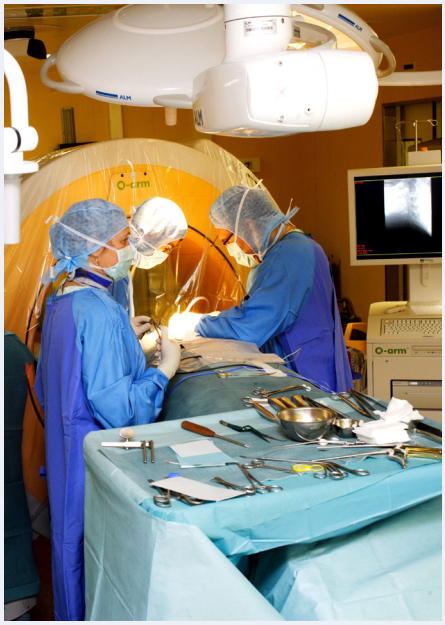
Study answer and limitations Mean increase in score on the Boston naming test was 11.16 points in the neurotomy plus intensive SLT group and 2.72 points in the control group at one month (difference 8.51 points, 95% confidence interval 5.31 to 11.71, P<0.001). The between group difference in score remained stable at six months (difference 8.26 points, 4.16 to 12.35, P<0.001). In addition, the aphasia quotient improved significantly in the neurotomy plus intensive SLT group versus control group (difference at one month 7.06 points, 4.41 to 9.72, P<0.001), as did patient reported activities of daily living and post-stroke depression. No treatment related severe adverse events were reported. The trial was conducted in a single geographical region, so results may not be generalisable elsewhere.

What this study adds C7 neurotomy plus three weeks of intensive SLT was associated with greater improvement in language function than three weeks of intensive SLT alone over a period of six months. No severe adverse events or long term troublesome symptoms or functional loss were reported

COMMENTARY Functional improvements seen after golden period of stroke recovery

The recovery journey after stroke often plateaus after the initial "golden period" of about 3-6 months, with further substantial functional improvements typically limited.[1 2] Aphasia, a language impairment resulting from damage to the brain's language centres, leaves many patients with few treatment options beyond this window, and they are often stigmatised as non-responders to standard interventions.[3 4] The randomised controlled trial by Feng and colleagues tackled this challenge.[5] The study investigated the adjunctive role of right neurotomy of the seventh cervical nerve (C7) combined with intensive speech and language therapy (SLT), building on previous research by the same research group.[6 7] C7 neurotomy involves a surgical intervention targeting the C7 nerve root at the right intervertebral foramen. Although C7 neurotomy is typically used to treat spasticity of the left arm, Feng and colleagues propose it as a novel treatment for chronic aphasia after stroke.

Supattana Chatromyen
dfondsupattana@gmail.com
See bmj.com for author details



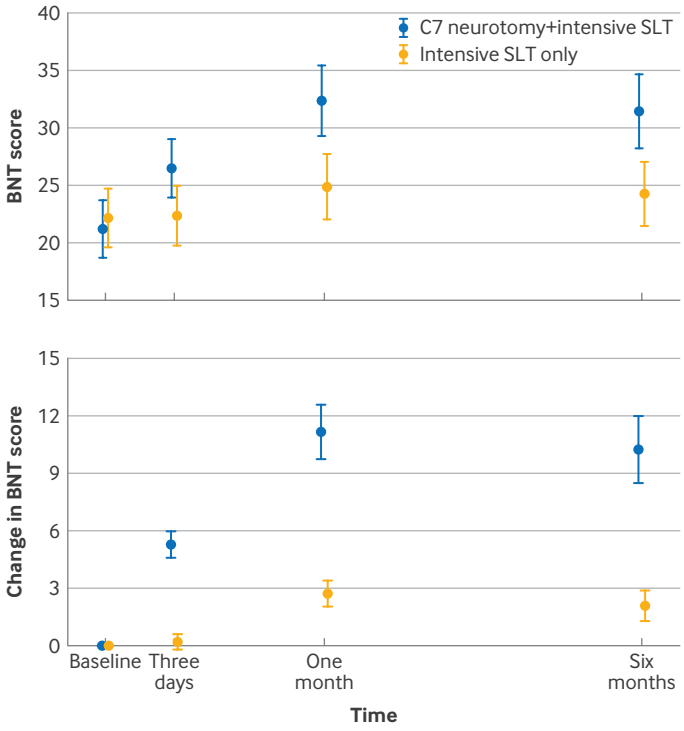
BURGER/PHANIE/SPL

The trial, conducted in China, involved 50 people with chronic aphasia (>1 year) after a single left hemispheric stroke, who were randomised to receive either C7 neurotomy plus three weeks of intensive SLT or three weeks of intensive SLT alone (control group). The primary outcome was change in the 60 item Boston naming test score from baseline to three days, one month, and six months after intervention. Secondary outcomes included broader language function (western aphasia battery-aphasia quotient), patient reported quality of life (Barthel index), and measures of post-stroke depression. The neurotomy plus intensive SLT group showed statistically significant improvements across all measured outcomes compared with the control group.

What do the results mean?

For patients in the chronic phase of stroke recovery with spasticity in the right upper arm and coexisting aphasia, this study seems to offer a glimmer of hope. It suggests a potential for further recovery exceeding the golden period.

Some caution is, however, warranted. Firstly, despite C7 neurotomy's association with improvements in motor function and reduced spasticity in patients with spastic arm paralysis due to cerebral injury,[8 9] it is a surgical procedure that should not be considered as a first line treatment for acute or subacute stroke, even for treating spasticity. This is primarily



Mean change in BNT score (standard error) from baseline in C7 neurotomy plus intensive SLT group and intensive SLT only group. Data are from baseline to three days after surgery in the neurotomy plus intensive SLT group or three days after start of one week's deferral of intensive SLT in the control group, one month from surgery or start of deferral, and baseline to six months after start of intensive SLT. BNT=Boston naming test; C7 neurotomy=right neurotomy of the seventh cervical nerve at the intervertebral foramen; SLT=speech and language therapy

Fundings, competing interests, and data sharing This study was funded by the National key R&D programme of China, National Natural Science Foundation of China, Shanghai Municipal Clinical Medical Centre Project, and Shanghai Municipal Health Commission. No competing interests declared. All data requests should be submitted to the corresponding author for consideration.

Trial registration Chinese Clinical Trial Register ChiCTR2200057180.

because during these earlier stages, the potential for spontaneous motor recovery is considerable, and a range of less invasive, multimodal treatments is available to promote rehabilitation.[10]
These include physiotherapies, occupational therapies, drugs, and botulinum toxin injections. Surgical interventions such as NC7 neurotomy are typically reserved for chronic, established spasticity when these conservative measures have been exhausted or are deemed insufficient. This is especially true when the potential for spontaneous motor recovery in the affected arm exists, as C7 neurotomy is mostly used in people with spasticity for more than one year, and the procedure may induce muscle weakness, sensory loss, and pain in the arm.[10-12]

Secondly, the highly specific inclusion criteria of this trial—chronic phase, left hemispheric stroke, aphasia with co-existing spasticity of the right arm—mean that the results apply to a select subgroup of people after stroke. Generalisation beyond this profile requires further research. Lastly, further studies with larger sample sizes, diverse settings, different investigator groups, and long term data on efficacy and safety in real world populations are needed. Replicating these findings in different languages and contexts would be a useful next step. Further studies, including animal and neuroimaging studies, could also explore the possible mechanisms of how the peripheral neurotomy influences cortical reorganisation and language networks. Feng and colleagues' findings also raise the question of whether C7 neurotomy could help with other stroke related impairments beyond aphasia and spasticity.

C7 neurotomy could become a potential adjunctive option for carefully selected individuals in the future

Where do we go from here?

In conclusion, Feng and colleagues' trial is an interesting step forward with room to explore further, suggesting that C7 neurotomy, when combined with intensive SLT, may offer benefits for certain people with chronic aphasia after stroke in whom improvement from treatment was previously thought to be limited. The concept of the golden period often leads to healthcare policies that restrict funding or access to intensive rehabilitation services, including intensive SLT, for people with chronic stroke. This often results in persistent disabilities such as severe communication impairments (eg, chronic aphasia), functional limitations, reduced quality of life, and increased caregiver burden, representing important long term sequelae after stroke.[13-15] If further evidence supports these findings, there could be a possible reason to rethink funding or reimbursement for people with chronic stroke who meet the appropriate criteria.

Although intensive SLT remains the cornerstone of aphasia treatment,[16] C7 neurotomy could become a potential adjunctive option for carefully selected individuals in the future. This research should spark further scientific research and a critical re-evaluation of rehabilitation paradigms and policies for chronic stroke care, fostering a more optimistic and proactive approach to long term recovery.

Cite this as: *BMJ* 2025;389:r1232

Find the full version with references at http://dx.doi.org/10.1136/bmj.r1232

Impact of physician assistants on quality of care

Cooper N, Agius S, Freeman K, et al

Cite this as: *BMJ* 2025;390:e086358
Find this at doi: 10.1136/bmj-2025-086358

Study question What is the impact of physician assistants on quality of care compared with physicians?

Methods This rapid systematic review included empirical studies published between January 2005 and January 2025 that quantitatively compared care delivered by physician assistants with care delivered by physicians in economically developed countries. The main outcomes were measures of outcomes of care, as defined by the Institute of Medicine's definition of quality: safety, effectiveness, patient centredness, timeliness, efficiency, and equity. Two reviewers independently extracted data on study design, samples, methods, and findings, and each study was assessed using a risk of bias tool. Owing to the heterogeneity of included studies, a narrative synthesis of the main findings was conducted. An assessment of confidence in the body of evidence for each outcome was based on the number and quality of relevant studies and the consistency of results between similar studies.

Study answer and limitations Of 3636 studies screened, 167 were eligible and 40 met the inclusion criteria. The greatest number of studies with the most consistent results were those which found that physician assistants practised safely and effectively when working under direct supervision and in post-diagnostic care. Evidence on cost effectiveness was limited. Patients in the UK were more likely to see a physician assistant if they lived in a socioeconomically deprived area. Included studies consisted mainly of retrospective observational studies of weak quality; 32/40 studies were from the US and no data from a post-covid-19 context were found.

What this study adds In the context of an ongoing UK policy review, the evidence is limited and does not support the safety or effectiveness of indirect supervision of physician assistants in undifferentiated (not yet diagnosed) settings.

Funding, competing interests, and data sharing No specific funding. No competing interests declared. No additional data available.

Study registration PROSPERO CRD42024614992.

Summary of findings according to study category

Category	No of studies (countries)	Summary of findings
Primary care	6 (5 US; 1 UK)	In the US, physicians were significantly more likely than PAs to prescribe newly approved drugs for chronic diseases. No difference was found in diabetes outcomes between patients managed primarily by a PA or a physician. In the UK, patients seen by PAs were younger, were more likely to present with minor problems, and had fewer chronic diseases, repeat prescriptions, and visits to the practice in the previous three months compared with patients seen by GPs. Patients were more likely to have smoking cessation and education/counselling services documented if seen by a PA
Secondary care	5 (3 US; 2 UK)	In all studies (conducted in the emergency department), physicians and PAs saw significantly different patients. Patient flow measures (eg, wait times) and process measures (eg, diagnostic test use) had mixed results
Residents *v* PAs in hospitals	14 (12 US; 2 Netherlands)	In the US, no significant differences existed between PAs and residents acting as first assistant in elective low risk surgeries. In medicine, PA-attending models had similar outcomes to resident-attending models, apart from one study of specialist haematology care in which the PA-attending model had shorter length of stay, lower readmission rates, and fewer consult requests. In the Netherlands, no difference was found in prescribing quality or patients' outcomes; patients' experiences of care (communication, continuity, cooperation, and medical care) were rated higher in the PA-attending model
Diagnosis/ performance	8 (7 US; 1 UK)	In the US, in ambulatory care, physicians and PAs saw significantly different patients; no differences were found in the quality of prescribing practices for 10/13 quality standards. PAs were more likely than physicians to overprescribe opioids. PAs prescribed clotrimazole-betamethasone (as a surrogate of poor dermatological care) at a rate of 16.9% in primary care versus 4.9% for primary care physicians; direct supervision significantly lowered the prescribing rates of PAs. PAs performed better than GI fellows/gastroenterologists in some colonoscopy metrics. PAs needed more skin biopsies to diagnose skin cancer and were less likely to diagnose melanoma in situ than were dermatologists. PAs were significantly more likely to have diagnosis related and treatment related malpractice allegations as a proportion of overall claims than were physicians. In the UK, GPs saw more complex patients but outperformed PAs in all aspects of the consultation, especially in diagnosis and management
Patient satisfaction	3 (2 US; 1 Netherlands)	Satisfaction was consistently high, with no significant differences between physicians and PAs
Cost effectiveness	4 (3 US; 1 Netherlands)	US studies analysed costs not cost effectiveness. In the Netherlands, no difference was found in hospital costs and patient quality of life scores between the PA-attending and resident-attending models

GI=gastrointestinal; GP=general practitioner; PA=physician assistant.