Citation for published version (APA):
A systematic review of fit note use for workers in the UK

Sarah Dorrington¹,², Emmert Roberts¹,², Arnstein Mykletun³, Stephani Hatch¹, Ira Madan⁴, Matthew Hotopf¹,²

¹King’s College London, Department of Psychological Medicine, Institute of Psychiatry Psychology and Neuroscience.
²South London and Maudsley NHS Foundation Trust, BR3 3BX.
³Norwegian Institute of Public Health, Oslo
+ University of Tromsø, Department of Community Medicine & Centre for Work and Mental Health, Health Region North, Norway
+ Centre for Research and Education in Forensic Psychiatry, Haukeland University Hospital, Bergen, Norway
⁴Guy’s and St Thomas’ NHS Foundation Trust.

The authors jointly conceived the study question, and designed the review. SD and ER conducted the review and SD drafted the manuscript, which all authors edited. The study did not require ethical approval.

Funding: Sarah Dorrington is funded by the Donald Dean Fellowship in Occupational Mental Health from the Royal College of Psychiatrists.
This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, King’s College London and the Royal College of Psychiatrists. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, or the Royal College of Psychiatrists.

Contributorship: SD, SH, IM and MH planned the review. SD and ER assessed the titles and abstracts identified by the search and reviewed the full text of the remaining articles for inclusion. Any discrepancy was resolved by discussion, and where agreement could not be reached MH was consulted. SD drafted the text, which was commented on by ER, SH, IM, AM and MH.

Competing interests: None
Correspondence to Sarah Dorrington: sarah.dorrington@kcl.ac.uk

**Transparency declaration** The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.
Abstract

Objectives: The fit note, introduced in England, Wales and Scotland in 2010, was designed to change radically the sickness certification process from advising individuals on their inability to work to advising them on what they could do if work could be adapted. Our review aimed to evaluate: (1) is the ‘maybe fit’ for work option being selected for patients? (2) are work solutions being recommended? (3) has the fit note increased return to work? and (4) has the fit note reduced the length of sickness absence? We considered the way in which outcomes vary according to patient demographics including type of health problem.

Methods: Studies were identified by a systematic search. We included all studies of any design conducted in the UK with working-aged adults, aged 16 or over, from 1 April 2010 – 1st Nov 2017. Risk of bias was assessed using a modified Newcastle Ottawa Scale.

Results: Thirteen papers representing 7 studies met inclusion criteria. In the largest study, ‘maybe fit’ for work was recommended in 6.5% of fit notes delivered by GPs (n=361,801) between April 2016 and March 2017. ‘Maybe fit’ recommendations were made in 8.5%-10% of fit notes received by primary care patients in employment, and in 10%-32% of patients seen by GPs trained in the diploma in occupational medicine. ‘Maybe fit’ was recommended more for women, those with higher socioeconomic status, and for physical, as opposed to psychiatric disorders. The majority of fit notes with the ‘maybe fit’ option selected included work solutions. There was inconclusive evidence to suggest that the introduction of the fit note has reduced sickness absence amongst patients in employment.

Conclusions: Fit notes represent a major shift in public policy. Our review suggests that they have been incompletely researched and not implemented as intended.
What this paper adds

What is already known about this subject?

- The fit note, introduced in England, Wales and Scotland in 2010, was designed to radically change the sickness certification process from advising on individuals’ inability to work to emphasize what they may be able to do if work solutions were made available. This is the first systematic review of the literature evaluating the implementation and impact of the fit note.

What are the new findings?

- Our review has found little quantitative research into the impact of this major policy change. Available research suggests that fit notes have been incompletely researched and not implemented as intended; the results are inconclusive.
- The largest study found that 6.5% of fit notes had ‘maybe fit’ recommended. Across all studies, the majority of fit notes with ‘maybe fit’ recommended included work solutions.

How might this impact on policy or clinical practice in the foreseeable future?

- On the basis of available evidence, it is unclear whether fit notes would, if properly implemented, give a desired change for patients. Evidence suggests that fit note implementation could be improved by legislation which encourages employers to adapt to the needs of patients. Low cost staggered implementation could be used to evaluate the impact of future policy change on patients.
Introduction

The impact of long-term sickness absence from employment is usually expressed in financial terms, with an estimated £14bn lost to the UK economy every year(1). However it also has devastating consequences for affected individuals and families, leading to social exclusion, widened health inequalities and financial insecurity(2). Patterns of sickness absence in the UK have changed considerably over the past few decades(3). There has been a decline in sickness certification for musculoskeletal conditions; mental illness has become the leading cause of long-term sickness absence in the UK(4). The relationship between sickness absence and health status is non-linear – many people with significant disability participate in the workplace(5). The lack of a clear correlation between disease severity and sickness absence suggests that sickness absence is not an inevitable consequence of mental or physical illness(6, 7).

The report by UK government advisor, Dame Carol Black suggested that a contributor to the problem of long-term sickness absence was the process of sick certification by general practitioners (GPs)(8). In the UK, sickness absence beyond seven days requires certification, most often delivered by a general practitioner (GP). Black’s review recommended the introduction of a “fit note” (figure 1), a potentially powerful intervention, which was implemented in 2010 to replace the sick note in the GP setting, followed secondary care settings. This was based on clear evidence that fitness for work is not a binary decision(9). A growing body of research from Nordic countries suggests that the introduction of reduced working hours, known as partial sickness absence, can improve long term occupational and health outcomes for patients(10). Partial sickness absence and similar interventions such as the fit note could be implemented in other countries, particularly in settings with welfare systems in place(10).

The fit note expands the ‘fit to work’ or ‘not fit to work’ options, by including a third ‘maybe fit’ option. In addition to this third option doctors and patients are invited to indicate circumstances under which the patient could work(11), and to identify potential ‘work solutions’, similar to partial sickness absence. These ‘work solutions’ include a graded return to work, altered hours, amended duties, and workplace adaptations(12, 13). The fit note is designed to enable the GP to be an advisor on work, rather than an adjudicator(14), the GP can provide guidance for the patient and the employer. The fit note also has a role in providing evidence of long term sickness absence in applications for employment and support allowance. Importantly, GP ‘maybe fit’ advice and work solutions on the fit note do not prevent individuals from obtaining benefits.
Initially the fit note completely replaced the sick note only in primary care, followed by its introduction in secondary care settings. To support this major shift in UK policy, the fit note was introduced with a specially designed national educational program for GPs, distinct from the broader diploma in occupational medicine (DipOccMed) also available to GPs (15). The introduction of the fit note coincided with the development of a government-funded ‘Fit For Work’ pilot scheme which delivers general health and work advice to employees, employers and GPs via a website and telephone line; the scheme is currently undergoing major changes due to a lack of uptake(16, 17). There have been numerous publications guiding GPs, employees and employers on how best to use the fit note(18, 19). Previous studies have explored the acceptability of fit notes for GPs, employers and patients(20-22). A recent UK government white paper published November 2017, Improving Lives: The Future of Work, Health and Disability(17) has drawn attention to the fit note, and raised questions about its future. This is the first systematic review of the literature evaluating the implementation and impact of the fit note.

We aimed to address whether the fit note is meeting its stated aims: (1) is the ‘maybe fit’ for work option being selected for patients? (2) are work solutions being recommended? (3) has the fit note increased return to work? and (4) has the fit note reduced the length of sickness absence? We consider the way in which outcomes vary according to patient demographics including type of health problem.

**Methods**

Search strategy and inclusion criteria

The review is reported according to the Preferred Reporting Items for Systematic Reviews (PRISMA) group(23). The search strategy can be found in supplementary files 1 and 2. We searched Embase, Cochrane CENTRAL, Medline, HMIC, Social Policy and Practice, Pubmed and PsychInfo from 1 April 2010 –1st November 2017. The search terms were ‘fitnote(s)’ or ‘fit note(s)’ or ‘fit-note(s)’. Studies were included if they were reported in English and the study population was working age adults (16-65) in England, Scotland or Wales. We included any study design, that reported one of the following outcomes (1) the use of ‘maybe fit’, (2) work solutions on fit notes and (3) changes in return to work and (4) length of sickness absence after the introduction of the fit note. We excluded qualitative studies and case series.
Two authors (SD and ER) initially assessed the titles and abstracts identified by the search and reviewed the full text of the remaining articles for inclusion. Any discrepancy was resolved by discussion, and where agreement could not be reached a third author (MH) was consulted. All relevant references were checked for additional citations.

Data extraction
Two authors (SD and ER) extracted the following data from each study using a data extraction proforma (see supplementary file 3): the first author’s last name, publication year, country or region where the study was conducted, study period, sample size, number of participants, funders, whether exclusions were made, effect sizes and 95% confidence intervals for our outcome variables: ‘maybe fit’, work solutions, recommended length of sickness absence and return to work.

Quality Assessment
Two authors (SD and ER) used a 10-point quality assessment tool adapted from the Newcastle-Ottawa Scale (24) for non-randomised studies to assess risk of bias. This scale is used to assess the methodological quality of observational studies and has acceptable validity and reliability. Methodological considerations in each cross-sectional study were scored as follows: 0-5 points for adequate selection of study subjects, 0-2 points for adequate comparability of study subjects, 0-3 points for adequate outcome assessment. For longitudinal studies there were two additional points, one for length of follow-up and one for adequacy of follow-up. Overall study quality for cross sectional studies was scored as follows: 0-3 = low quality; 4-7 = medium quality; 8-10 = high quality. Longitudinal studies were scored 0-4 = low quality 5-8 = medium quality 9-12 = high quality.

Results
Our initial search returned 137 articles. Fifty-seven papers were original research about fit notes. After applying our inclusion and exclusion criteria we were left with 10 papers. Three additional papers were identified in the grey literature when references were checked for additional citations, making 13 papers in all (see figure 2)(25-37). Several papers explore different study questions within the same population. The 13 papers identified describe 7 studies. The second largest study (study 6 in table 1) was analysed in 5 different papers, which defined the population in different ways. For example one paper analysed the entire primary care population(32) and another restricted the population to employed patients with mental health problems(34).
We report findings in relation to the type of study population presented: 11 out of the 13 papers explored the use of the fit note in primary care, one explored its use by employers (30) and one by employees (25) (table 1). Two of the primary care based papers analysed the use of the fit note by GPs with a training in the Diploma in Occupational Medicine. Seven papers analysed a study funded by the UK government Department of Work and Pensions, 3 papers were funded by the Institute of Occupational Safety and Health, one was funded by the UK Health and Safety Executive, one was funded by the National Institute of Health Research and the most recent was funded by National Health Service (NHS) Digital.

Three papers compared data from ‘before and after’ the fit note’s introduction (27, 31, 33); the remainder were conducted after the introduction of the fit note (25, 26, 28-30, 32, 34-37). Four papers were limited to populations of patients in employment. Study sizes ranged from 94 fit notes in the smallest study, to 5 million fit notes in the largest (28, 37). See table 1 for more information on study design.

**Outcome 1: Recommendation of ‘maybe fit’ for work**

The prevalence of ‘maybe fit’ use varied from 3.5% to 32% of fit notes depending on the population studied (table 2). Several studies analysed demographic variation in ‘maybe fit’ use; they found variation at both individual and practice level (25, 27, 31).

**All primary care patients**

Two studies analysed fit note use in all primary care patients. In a cross-sectional analysis of GP patients from 68 practices, ‘maybe fit’ was recommended in 6.4% (n=5080) of all fit notes (32). The same patients could receive multiple fit notes – when this is taken into account, 12% of all patients receiving fit notes had a ‘maybe fit’ recommendation at least once (n=2990) (31, 32). These findings were replicated by NHS digital: 6.5% of patients receive a ‘maybe fit’ recommendation in over 5 million fit notes.

**Primary care patients in employment**

Six papers analysed ‘maybe fit’ use amongst employed patients. After excluding one study which intentionally oversampled fit notes (28), we found the prevalence of ‘maybe fit’ ranged from 7-10% of all fit notes. Gabbay et al 2015 restricted primary care study 6 to include only GP patients who were known to be in employment because employment was recorded by
the GP (32.5% of fit notes n=25,061)(33). In this population 8.5% (2,151) of fit notes included a ‘maybe fit’ recommendation.

Shiels et al 2016 defined the employed population by excluding patients recorded by their GP as ‘not in work’ (7% n=2,315 patients with 3,468 discrete episodes). Patient episodes were analysed, rather than individual fit notes. In 91% (n = 28,588) of first episodes the GP recommended that the patient should abstain from all work. In the remaining 9% (n = 2,865) the episode had concluded with the GP advising that the patient ‘may be fit’ to work provided adjustments were made to normal working conditions.

Three other papers in our sample analysing fit notes given to patients in employment found the use of ‘maybe fit’ to be 9-10% of fit notes; two of these papers analysed the same study data (25, 29, 30). The sixth report restricted the population to fit notes for employed patients with mental health problems and found a lower percentage, 7% (n=562), of fit notes recommending ‘maybe fit’(34).

**GPs with a diploma in occupational medicine**

Two studies analysed fit notes provided by GPs with a diploma in occupational medicine (DipOccMed)(26, 27). The DipOccMed is undertaken by medical practitioners who are working part-time in occupational medicine or who have an interest in occupational medicine. It is most frequently taken by general practitioners who often add to their practice portfolio by providing occupational health services to local companies. Around 4% of GPs complete the DipOccMed(27). The first study included both GPs with DipOccMed delivering the new Fit For Work service, and untrained GPs; fit notes were measured(26). The second study included only GPs training or trained in the DipOccMed, and measured cases(27). In the first study, overall 10% (n=73) of all fit notes had ‘maybe fit’ recommended(26), compared to 6.5% of patients seen by all GPs(37). In the population referred to the Fit For Work (FFW) service, a much higher percentage, 32%, of fit notes had a ‘maybe fit’ recommendation(26). In the second study in which ‘maybe fit’ was measured by cases, 25% of cases (n=209) were given ‘maybe fit’ notes, compared to 12% of cases seen by all GPs (27, 31).

**Additional studies**

Coole et al 2015 used a sampling method which intentionally over sampled ‘maybe fit’ notes(28). GPs were asked to record the first ten ‘new’ fit notes issued to employed patients including a minimum of five ‘maybe fit’ notes. As expected from this sampling strategy, ‘maybe fit’ use was overrepresented in their sample, 27% (25) of all fit notes.
Outcome 2: Work solutions: structured and free text advice on the fit note

Primary care population
In the primary care population, eighty-three percent (n=4216) of ‘maybe fit’ notes had structured advice, and 84% (n=3542) of ‘maybe fit’ notes with structured advice had free text advice to guide patients and employers(32). Amongst fit notes with structured advice, 35% indicated a phased return, 20% altered hours and 9% workplace adaptations (31).

Primary care patients in employment
Amongst the 8.5% of employed patients recommended a fit note, a large proportion had work solutions recommended: 44% amended duties, 34% phased return, 20% altered hours and 9% workplace adaptations. In the study which oversampled fit notes with ‘maybe fit’ recommended, of the 98 fit notes received, free text comments were divided into 3 groups: 23 fit notes had ‘work-related advice’, 11 had comments on the ‘functional effects of a patient’s condition’ and 8 had ‘additional information on a health condition’(27).

GPs with a diploma in occupational medicine
Amongst patients seen by occupational health trained GPs, 16% of cases had fit notes with both ‘may be fit’ and structured advice(27). In a small study of GPs with DipOccMed, structured or unstructured advice was given on 72% of ‘maybe fit’ notes in primary care and 98% of ‘maybe fit’ notes in the FFW (Fit For Work) service (26). GPs with a DipOccMed were more likely to recommend ‘maybe fit’ and to give advice on work solutions when seeing patients referred to the FFW service (26, 27).

Outcome 3: Return to work outcomes

Return to work following fit note use was examined in two cross-sectional studies of employers and employees, both contain patients in employment. In these studies 42%-82% of people returned to work after a fit note(25, 30). The results were not compared to return to work before the introduction of the fit note, only to ‘not fit’ notes: Coole et al 2015 found that more than 80% (44) employees returned to work after the expiry date of a ‘may be fit’ note compared with 43% (167) of those issued with a ‘not fit’ note(30)

Outcome 4: Change in duration of sickness absence
The most recent study found that between 2016 and 2017, a third of fit notes (33.6%) were for 5 weeks or longer (37). Two “before and after” papers analysed the same cohort of GPs and patients from 2001-2002 and 2011-2013 to explore changes in length of sickness absence, before and after the introduction of the fit note (31, 33). Shiels et al. 2013 found no change in long-term sickness absence across all seven GP practices. Gabbay et al. 2015 studied the same population, restricted to the 31% of patients whose occupation was recorded by the GP, with the aim of identifying a patient population in employment. These selected patients were pooled in a multi-level logistic regression analysis, which found a reduction in the risk of a long term (over 12 weeks) certified sickness episode [OR 0.65 (0.58-0.72)](33) for patients who had occupational information recorded by their GP.

The use of ‘maybe fit’ advice was associated with shorter time off work recommended by GPs(31). When the population recorded as ‘not in work’, were excluded from the primary care population, patients whose initial episode ended with a ‘may be fit’ note had a reduced rate of follow up fit note with the same diagnosis as their first fit note (IRR = 0.72, 95% CI = 0.63–0.81)(34).

Employed patients who had received ‘may be fit’ advice on a fit note at the end of their first episode were significantly less likely to have another episode of sickness absence within the study period(35). However, in the report of patients in employment with a mental disorder on their fit note, ‘maybe fit’ episodes were longer than episodes where no return to work advice was offered (median weeks 5.5 weeks compared with 4 weeks, P < 0.001)(34). Longer duration of these ‘maybe fit’ episodes was statistically significant for both stress (5.4 versus 3.0 weeks, P < 0.001) and bereavement (4.2 versus 2.1 weeks, P < 0.001)(34).

Demographic variation

Most likely to present for a fit note
In the general primary care population fit notes were most frequently given to women (56%) and patients with mental and behavioural disorders (31%) (37). In the employed population the pattern was similar, patients most likely to present for a fit note were women, patients with mental health problems, people working in public sector and semi-routine or routine employment, and people with a disability discrimination act (DDA) disability (25).

Most likely to receive a ‘maybe fit’ note
‘Maybe fit’ use varied from practices, with a range of 1% to 15% of fit notes given across 68 practices(31). Patient variables associated with receiving a ‘maybe fit’ note were female sex
[OR 1.09 (95%CI 1.01-1.21)], lower deprivation (defined by index of multiple deprivation (38)) [more deprived OR 0.86 (95%CI 0.73-0.94)], and having a physical as opposed to a mental disorder (OR 1.69 (95%CI 1.58-1.81) (32). Only 3.5% of fit notes given to patients for depression had a ‘maybe fit’ recommendation, compared to 6.5% of all fit notes(36, 37). In a study of GPs with a DipOccMed patients with mental health problems were also less likely to receive ‘maybe fit’ than patients with physical health problems(27).

**Least likely to receive advice on work adjustments**

A study of GPs training in the DipOccMed found that patients with mental health problems were least likely to receive a structured workplace adjustment: 6% of cases compared with 21% musculoskeletal and 23% of other cases(27). ‘Maybe fit’ notes issued to patients with a mild-to-moderate mental health disorder had a lower than average proportion (37%) of additional written advice from the GP(31). However, a study which analysed only employed patients with a fit note for common mental disorders (CMD) found that although only 7% of employed patients with CMD had a ‘maybe fit’ recommendation, of those patients 57% had phased return recommended, 30% altered hours, 23% amended work duties, 8% work place adaptation and 57% had some written advice. This suggests that once employed patients with a mental health problem are recommended ‘maybe fit’, work solutions are recommended(34).

**Least likely to return to work or longer sickness absence**

Gabbay et al 2017 found that over 35% (8127/3361) fit notes issued to patients for depression were over four weeks in duration. Predictors of having over four weeks sickness absence recommended was associated with having previous fit notes for depression, not having “maybe fit” advice on the fit note, being older, living in a deprived neighbourhood and having higher practice deprivation status. Fit notes for depression issued by female GPs to female patients were less likely to be long-term. Other GP factors were not significant predictors of a long-term fit note (36).

The rate of GPs issuing workplace adjustment advice in the occupational health trained group increased consistently (IRR = 1.12) in all diagnoses except mental ill-health, which indicated a decrease over the 3 years of 12% per year(27).

**Quality Assessment**
Using the Newcastle Ottawa Scale to assess risk of bias in each data source (see supplementary file 4) we found six papers from the same study were high quality (table 3). Of the remaining studies five were medium quality and two were high quality. Several study designs were mixed-methods, their quality score reflects the quantitative section of these papers and is not representative of their qualitative methods(26, 28, 30). The main limitations in medium quality studies were low sample size, non-respondents and the representativeness of sample. Two studies defined the primary care population according to employment status and benefit receipt based on GP records, which are known for poor recording of employment related variables(34, 35). We could not test for publication bias due to the limited number of studies in each population: primary care patients, employed patients and patients seen by GPs trained in the DipOccMed.

**Discussion**

The 'maybe fit' option is recommended in only a small minority of patients who receive fit notes in UK primary care(25, 27, 31-35). Wide variation in practice in the use of ‘maybe fit’ is seen between GP practices and according to GP training and setting(25-35). The highest use of ‘maybe fit’ and advice on work solutions is by GPs trained in the DipOccMed.

Patients who are more deprived and have a mental disorder are most likely to present for a fit note and least likely to receive a ‘maybe fit’ recommendation(27). Amongst patients who receive a ‘maybe fit’ recommendation, the majority are given further advice on adaptations to the workplace (25, 27, 31-34). The low proportion of individuals receiving ‘maybe fit’ suggests that there are some who would be eligible for this advice who do not receive it. Why might this be? Qualitative research into the acceptability of the fit note, describes obstacles to ‘maybe fit’ note implementation, including insufficient communication and understanding between the triad of employee, GP and employer(22). The second white paper from the Health at Work Policy Unit argues that shared knowledge and decision making could improve occupational support for patients, particularly patients with fluctuating chronic conditions(15, 39, 40).

Stigma continues to be a problem for patients, who report concerns about disclosure of health status on a fit note, which will be seen by their employer (9, 41-43). Some GPs seem apprehensive about the impact discussing work related issues may have on their relationship with patients, and are sceptical about whether the fit note will have any positive impact (44-47). GPs report concerns about the limits to their expertise in occupational health
and their understanding of patients’ employment (48-50). Therefore there are significant barriers in implementation. However, these concerns from GPs are amenable to training: GPs who receive some form of work and health training have more positive attitudes to patients’ returning to work and to the fit note(51). The National Education Programme implemented in 2009 to support the introduction of the fit note, addressed both knowledge and skills for the management of the health and work consultation. The 3 hour interactive workshops increased GP confidence and led to GPs giving higher priority to consultations involving discussions around sickness absence(15, 52). The evidence reviewed here indicates that more indepth training in the DipOccMed impacts on the use of ‘maybe fit’ certification(27). Research suggests that a new models of case-specific colleague guidance may also help GPs manage the most challenging sickness absence cases(53). There is evidence that training to support use of the fit note by other health care professionals is also effective(54, 55). Acceptability of the fit note amongst employers in the UK is high, with many reporting that the fit note had been useful in planning the patient’s return to work and modifying their jobs (56-58).

The motivation behind the introduction of the fit note was to reduce unnecessary sickness absence and improve return to work rates, thereby avoiding the individual and societal costs of long term sickness absence. The evidence we have reviewed about the impact of the fit note on these outcomes is sketchy. Whist there is some evidence from before-and-after studies that the introduction of the fit note reduced sickness absence amongst patients in employment (27, 32, 33), return to work outcomes have not been evaluated against the sick note(25, 30). These data are from observational studies, and it is impossible to know whether the fit note or other unrelated secular changes are responsible for the observed differences. The results are therefore inconclusive – they support the possibility of a favourable impact, but do not allow more definite inferences to be drawn.

There is evidence from other countries supporting a model similar to the fit note. Partial sickness absence was introduced in Norway in 2004 and in Finland in 2007. Like the fit note, it introduces a third option of reduced working hours instead of the binary ‘fit for work’ versus ‘not fit for work’. In contrast to the fit note, partial sickness absence legislation places the burden formally on the employer to adapt to the patient; the employer is expected to pay a salary for the hours the employee is at work and is reimbursed for time that the employee is not working(10). In Finland the use of partial sick leave is voluntary for the individual, in both Finland and Norway the employer is only able to decline if the work arrangements needed at the work place are not feasible(59).
Partial sickness absence faces similar challenges to the fit note, yet it has been introduced successfully in Nordic countries and is an effective way to improve return to work (10, 59, 60). Demographic variation in use of partial sickness absence is similar to ‘maybe fit’ use: women, older patients, patients with a slightly higher education, and higher earnings prior to the absence spell are more likely to receive partial sickness absence. In Finland partial sick leave reduced the decline in work participation in patients, especially among those aged 45-65 and in patients with mental disorders (59, 60).

Partial sickness absence was implemented with a change in law demanding adjustments at the workplace to facilitate part-time work. By contrast, the UK fit note is a weaker instrument because there is no legislation to encourage employers directly to put recommended work solutions in place for the employee (59).

**Strengths and Limitations**

We used a robust search strategy and adhered to PRISMA guidelines, however it is not possible to rule out publication biases, particularly from analyses reported in the grey literature. We avoided summarizing the results in a meta-analysis because the numbers of studies in each category was small, and in some cases papers drew from the same study population. Whilst the formal assessment of quality did not unearth significant problems, the studies were generally limited by their reliance on routine and incomplete data – such as patients’ employment status – and the fact that they were universally observational in nature. Ideally, there would have been cluster randomised trials or at least phased introductions using stepped wedge designs to allow the impact of the introduction of fit notes to be understood.

**Conclusion**

The fit note represents a major shift in policy in an area of urgent public interest. It is therefore remarkable how little published research there is on fit notes. The studies we review indicate that it is incompletely implemented, with only a small minority of GPs using the ‘maybe fit’ recommendation and significant variation in practice. Whilst it may have had an impact on long term sickness absence amongst patient in employment, the research to date does not do justice to the scale of the problem the fit note was designed to address. It isn’t clear whether the current system could be more effective or whether more regulatory
support is needed. Evidence from Nordic countries suggest that fit note implementation could be improved by legislation to encourage employers to adapt to the needs of patients. Low cost staggered implementation could be used to evaluate the impact of future policy change on patients.

13. Health at work – an independent review of sickness absence Dame Carol Black and David Frost CBE. 2011.
45. Hann M, Sibbald B, National Primary Care R, Development C. General practitioners' attitudes towards patients' health and work. 2011.
49. Welsh VK. Exploration of GP's views and the use of the fit note: a qualitative study in primary care.
### Table 1: Search results: 13 papers evaluating the fit note

<table>
<thead>
<tr>
<th>Study</th>
<th>Paper</th>
<th>Year of data collection</th>
<th>Recruitment and sampling</th>
<th>Data collection</th>
<th>Funder</th>
<th>Study design and Method</th>
<th>Demographic variation</th>
<th>FNs</th>
<th>Study population</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chenery 2013(25)</td>
<td>Jan-June 2012</td>
<td>Adults selected from the Labour Force Survey (LFS). The Fit Note Survey was attached to the last of five LFS interviews.</td>
<td>Survey – computer assisted telephone interviews</td>
<td>Department for Work and Pensions</td>
<td>Cross sectional. Logistic regression</td>
<td>Age, gender, SES, health condition, size of organization, sector worked in, self report health, disability status, health condition</td>
<td>1398 employees</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>Coole 2013</td>
<td>June 2011-Dec 2011</td>
<td>Selected from a service evaluation.</td>
<td>Fit note comments</td>
<td>National Institute for Health Research</td>
<td>Cross sectional. Content analysis</td>
<td></td>
<td>1212</td>
<td>1 practice (712 FNs) 1 FFW (500 FNs)</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>Hussey 2015(27)</td>
<td>4 years before (April 2006-March 2010) and 3 years (July 2010-June 2013)</td>
<td>Selected all available data from the THOR-GP surveillance scheme</td>
<td>Information reported by GPs with training in occupational health to diploma level.*</td>
<td>UK Health and Safety Executive (THOR contract number: HSE JN4243).</td>
<td>Longitudinal. Multi-level random effects poisson regression model fitted to the monthly data count</td>
<td>Type of health condition</td>
<td>5517 patients (835 patients since introduction of the fit note)</td>
<td>250 GPs</td>
<td>Y Y</td>
</tr>
<tr>
<td>4</td>
<td>Coole 2015(29)</td>
<td>November 2013 to May 2014</td>
<td>A total of 272 GP practices were invited to participate in the study</td>
<td>Data from fit note copies and questionnaires.</td>
<td>Institute of Occupational Safety and Health</td>
<td>Cross sectional Quantitative data were analysed descriptively. Free-text comments using thematic content analysis</td>
<td>Type of health condition</td>
<td>94</td>
<td>11 GPs</td>
<td>Y Y</td>
</tr>
<tr>
<td>4+5</td>
<td>Coole 2015(28)</td>
<td>Study 4+5 combined</td>
<td>Study 4 + 5 combined</td>
<td>Study 4+5 combined</td>
<td></td>
<td></td>
<td>94 + 498</td>
<td>13 employer organisations</td>
<td>Y Y Y</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Coole 2015 (30)</td>
<td>May 2013 to March 2014</td>
<td>A combination of opportunistic, and random sampling</td>
<td>Postal questionnaires</td>
<td></td>
<td></td>
<td>498</td>
<td></td>
<td>Y Y Y</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Shiels 2013(31)</td>
<td>2001-2002 + Oct 2011- Jan 2013</td>
<td>Practices were invited to take part from 5 areas of UK</td>
<td>Carbonised fit note pads</td>
<td>Department for Work and Pensions</td>
<td>Longitudinal Logistic regression: Multilevel mixed-effects model</td>
<td>Patient (age, gender, type of health condition, social deprivation), GP (sex age partner/locum status, full time/part time)</td>
<td>58700</td>
<td>25000</td>
<td>49 practices</td>
</tr>
<tr>
<td></td>
<td>Shiels 2014(32)</td>
<td>late 2011-Jan 2013</td>
<td>The second project involved additional data from an evaluation of 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gabbey 2018(34)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Time Period</td>
<td>Practices / Characteristics</td>
<td>GP Practice (size, location, GP deprivation)</td>
<td>Episodes</td>
<td>Patients</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shiels 2016 (35)</td>
<td>late 2011-Jan 2013</td>
<td>practices sited in three FFWS pilot sites</td>
<td></td>
<td>38,934</td>
<td>31,453</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabbay 2017 (36)***</td>
<td>late 2011-Jan 2013</td>
<td></td>
<td></td>
<td>8127</td>
<td>3361</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHS digital 2017 (37)</td>
<td>Dec 2014-March 2017</td>
<td>207 CCGs across the UK, 70 had more than 95% coverage, 25 had less than 5%</td>
<td>NHS digital fit note data</td>
<td>5,803.98</td>
<td>6*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bold = grey literature  
FFW = Fit For Work schemes  
THOR = The Health and Occupational Research Network  
FN= fit note  
NHS = National Health Service  
*62.1% of all fit notes prescribed in England during this period to people aged 18-65  
**patients with CMD  
***patients with depression only  
****Outcome 1: Maybe fit for work outcome  
Outcome 2: Work solutions  
Outcome 3: Return to work outcomes  
Outcome 4: Length of sickness absence
### Table 2: Outcomes 1 to 4 by population sample

<table>
<thead>
<tr>
<th>Population sample</th>
<th>Paper</th>
<th>Outcome 1: ‘May be fit for’ work option selected</th>
<th>Outcome 2: Work solution: structured and free text advice</th>
<th>Outcome 3 + 4: Return to work outcomes and length of sickness absence</th>
<th>Newcastle Ottawa Scale (adapted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All GP patients</td>
<td>Shiels 2013(31)</td>
<td>6% (3670/58700)</td>
<td>12% (2990/25000)</td>
<td>82% structured advice 7% (273) no structured or free text advice 64% free text advice 10.2% (376) free text advice only</td>
<td>Reduction in certifications length amongst those &gt;4 weeks and &gt;12 weeks in 3/7 individual practices but not overall***</td>
</tr>
<tr>
<td></td>
<td>Shiels 2014(32)</td>
<td>6.4% (5080/79375)</td>
<td></td>
<td>83% of MBF notes had some structured advice but no free text advice</td>
<td>High 10/10</td>
</tr>
<tr>
<td></td>
<td>Gabbay 2017****</td>
<td>3.5% (286/8127) of fit notes for depression</td>
<td></td>
<td>16.8% of may be fit notes and 34.5% of all fit notes were for over 4 weeks</td>
<td>High 10/10</td>
</tr>
<tr>
<td></td>
<td>NHS digital 2017 (37)</td>
<td>6.5%(361801/5603 986)</td>
<td></td>
<td>More than 80% recommended an adaptation in the workplace, working hours or duties.</td>
<td>High 10/12</td>
</tr>
<tr>
<td>Occ health trained GPs seeing all GP patients</td>
<td>Coole 2013</td>
<td>FFW 32% (160/500) GP 10% (73/712)</td>
<td>FFW: 98% (157) advice of any type GPs: 72% (53) advice of any type</td>
<td></td>
<td>Medium 4/10</td>
</tr>
<tr>
<td></td>
<td>Hussey 2015(27)</td>
<td>25% (209/835)**</td>
<td>16% (131/835)** of cases had structured advice</td>
<td></td>
<td>Medium 8/12</td>
</tr>
<tr>
<td>Employed patients only</td>
<td>Coole 2015(28)***</td>
<td>27% (25/94)</td>
<td>Free text comment types: 23 work-related advice 11 functional effects of a patient’s condition additional 8 information on a health condition</td>
<td>84% (21/25) of GP MBF notes had structured advice</td>
<td>Medium 4/10</td>
</tr>
<tr>
<td></td>
<td>Coole 2015(30)***</td>
<td>10% (51/498)</td>
<td></td>
<td>42% (211/498) RTW</td>
<td>Medium 5/10</td>
</tr>
<tr>
<td></td>
<td>Chenery2013(25)</td>
<td>9% overall (4% of first fit notes 19% of secondary fit notes)</td>
<td></td>
<td>&gt;78% MBF had structured advice</td>
<td>High 8/10</td>
</tr>
<tr>
<td></td>
<td>Gabbay 2015(33)</td>
<td>8.5% (2151)</td>
<td>12% (1602) included MBF and return to work advice.</td>
<td>44% (955) amended duties, 34% phased return, 20% altered hours and 9% workplace adaptations.</td>
<td>High 11/12</td>
</tr>
</tbody>
</table>
17% (210/1205) of certified sickness due to a fracture or other injury
16% (295/1858) of back or other musk
9.9% of MMMD category had return to work advice from the GP (393/3950)

<table>
<thead>
<tr>
<th>Study</th>
<th>Details</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabbay 2016(34)****</td>
<td>7% (562/8074) of employed patients with CMD</td>
<td>57% (318) phased return, 30% (169) altered hours, 23% (129) amended work duties, 8% (43) workplace adaptations. 57% (321) some written advice.</td>
</tr>
<tr>
<td>Shiels 2016(35)</td>
<td>9% (2865) of first fit note episodes ended with a fit note</td>
<td>High 9/10</td>
</tr>
</tbody>
</table>

*patient  ** first year of study, not reported after that  
***Coole et al 2015 combines these 2 studies  
**** Fit notes for CMD only  
*****Fit notes for depression only  
MBF = maybe fit  
CMD= common mental disorder  
FFW= fit for work  
RTW = return to work  
MMMD = Mild-moderate mental disorder  
NHS = National Health Service
### Table 3: Study Quality assessed by the Newcastle-Ottawa Scale

<table>
<thead>
<tr>
<th>Study</th>
<th>Paper</th>
<th>Representativeness of sample</th>
<th>Sample size</th>
<th>Non respondents</th>
<th>Ascertainment of exposure</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Statistical test</th>
<th>Length of follow up</th>
<th>Adequacy of follow up</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-2</td>
<td>0-1</td>
<td>0-1</td>
<td>0-1</td>
<td>0-2</td>
<td>0-1</td>
<td>0-1</td>
<td>0-1</td>
<td>0-1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Chenery 2013(25)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Coole 2013</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Hussey 2015(27)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Coole 2015 GPs*(28)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Coole 2015 RTW* (30)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>Shiels 2013 (31)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Chiels 2014 (32)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Gabbay 2015 (33)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Gabbay 2016 (34)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Shiels 2016 (35)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Gabbay 2017(36)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>NHS digital 2017(37)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

*Study 5 by Coole 2015(29) contains studies 4 and 6  
NHS= National Health Service