Appendix

A1 Eligibility criteria.

Table A1 Eligibility criteria.

|  |  |
| --- | --- |
| **Inclusion criteria** | **Exclusion criteria** |
| Cohort or cross-sectional observational studies  Studies set in primary care or the general population  Studies published since 1990  Studies where prevalence data for CWP can be extracted or calculated  Adult study population (mean age greater than 18 years)  Studies using ACR-1990 or ACR-2010 criteria to define CWP. | Conference proceedings, editorials and letters.  Studies quoting incidence rather than prevalence figures  Sample in a specific subset of the general population (e.g. women only, students, employed population only, hospital outpatient clinic patients) |

A2 Search strategy

**Table A2.1 Keywords included in the search strategy for all four databases; terms searched for in the title and abstract of papers.**

|  |  |
| --- | --- |
| **Pain term** | chronic widespread pain **OR** fibromyalgia **OR** chronic pain syndrome **OR** diffuse pain **OR** fibrositis **OR** fibromyositis **OR** myofascial pain syndrome |
|  | **AND** |
| **Study type term** | epidemiology **OR** cohort stud\* **OR** cohort analys\* **OR** cross sectional stud\* **OR** cross sectional analys\* **OR** observational analys\* **OR** prevalence **OR** disease frequency |

**Table A2.1 Database specific subject heading terms.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Medline - MeSH Headings** | **CINAHL - Subject Headings** | **AMED - Subject Headings** | **Embase - Emtree Subject Headings** |
| **Pain term** | fibromyalgia  myofascial pain syndromes | fibromyalgia  myofascial pain syndromes | fibromyalgia pain | fibromyalgia/ epidemiology myofascial pain/ epidemiology |
| **Study type term** | prevalence  cross-sectional studies epidemiology  epidemiologic methods epidemiologic research design  epidemiologic studies epidemiologic measurements  cohort studies | prevalence cross-sectional studies epidemiology epidemiological research prospective studies | epidemiology | epidemiology prevalence cross sectional study |

A3 Database specific search strategies

**Table A3.1 Medline – limit to human, 1990–current.**

|  |  |  |
| --- | --- | --- |
| **Pain term** | Keywords searched for in abstract and title | (TI “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”)  **OR**  (AB “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”) |
|  |  | **OR** |
|  | MeSH headings | (MH "Fibromyalgia") OR (MH "Myofascial Pain Syndromes") |
|  |  | **AND** |
| **Study type term** | Keywords searched for in abstract and title | “epidemiology” OR “cohort stud\*” OR “cohort analys\*” OR “cross sectional stud\*” OR “cross sectional analys\*” OR “observational analys\*” OR “prevalence” OR “disease frequency” |
|  |  | **OR** |
|  | MeSH headings | (MH "Prevalence") OR (MH "Cross-Sectional Studies") OR (MH "Epidemiologic Measurements") OR (MH "Epidemiologic Methods") OR (MH "Epidemiologic Research Design") OR (MH "Epidemiology") OR (MH "Cohort Studies") |

**Table A3.2 AMED – 1990–current.**

|  |  |  |
| --- | --- | --- |
| **Pain term** | Keywords searched for in abstract and title | (TI “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”)  **OR**  (AB “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”) |
|  |  | **OR** |
|  | Subject headings | (DE "FIBROMYALGIA") OR (DE "PAIN") |
|  |  | **AND** |
| **Study type term** | Keywords searched for in abstract and title | “epidemiology” OR “cohort stud\*” OR “cohort analys\*” OR “cross sectional stud\*” OR “cross sectional analys\*” OR “observational analys\*” OR “prevalence” OR “disease frequency” |
|  |  | **OR** |
|  | Subject headings | (DE "EPIDEMIOLOGY") |

**Table A3.3 EMBASE – limit to human, 1990–current.**

|  |  |  |
| --- | --- | --- |
| **Pain term** | Keywords searched for in abstract and title | (chronic widespread pain OR fibromyalgia OR chronic pain syndrome OR diffuse pain OR fibrositis OR fibromyositis OR myofascial pain syndrome).ab,ti. |
|  |  | **OR** |
|  | Emtree subject headings | (fibromyalgia/epidemiology OR myofascial pain/epidemiology).sh. |
|  |  | **AND** |
| **Study type term** | Keywords searched for in abstract and title | (epidemiology OR cohort stud\* OR cohort analys\* OR cross sectional stud\* OR cross sectional analys\* OR observational analys\* OR prevalence OR disease frequency).ab,ti. |
|  |  | **OR** |
|  | Emtree subject headings | (epidemiology OR prevalence OR cross sectional study).sh. |

**Table A3.4 CINAHL – limit to human, 1990–current.**

|  |  |  |
| --- | --- | --- |
| **Pain term** | Keywords searched for in abstract and title | (TI “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”)  **OR**  (AB “chronic widespread pain” OR “fibromyalgia” OR “chronic pain syndrome” OR “diffuse pain” OR “fibrositis” OR “fibromyositis” OR “myofascial pain syndrome”) |
|  |  | **OR** |
|  | Subject headings | (MH "Fibromyalgia") OR (MH "Myofascial Pain Syndromes") |
|  |  | **AND** |
| **Study type term** | Keywords searched for in abstract and title | “epidemiology” OR “cohort stud\*” OR “cohort analys\*” OR “cross sectional stud\*” OR “cross sectional analys\*” OR “observational analys\*” OR “prevalence” OR “disease frequency” |
|  |  | **OR** |
|  | Subject headings | (MH "Prevalence") OR (MH "Cross Sectional Studies") OR (MH "Epidemiology") OR (MH "Epidemiological Research") OR (MH "Prospective Studies") |

A4 Risk of bias assessment

**Table A4.1 Guidelines for assessing risk of bias – based on study participation and outcome measurement domains of the QUIPS tool [16]. After reflection on all items risk of bias is rated in each domain as high, moderate or low.**

|  |  |
| --- | --- |
| **Potential bias** | **Items to be considered for assessment of potential opportunity for bias** |
| **Study participation** |  |
| Does the study sample represent the population of interest on key characteristics sufficient to limit potential bias to the results? | **Target population:** The source population or population of interest is adequately described for key characteristics.  **Sampling frame:** The sampling frame and recruitment are adequately described, possibly including methods to identify the sample (number and type used, e.g. referral patterns in health care), period of recruitment, and place of recruitment (setting and geographic location).  The sampling frame and procedures used to sample subjects should not lead to selection of participants that are systematiclly different from eligible non-participants.  **Inclusion criteria:** Inclusion and exclusion criteria are adequately described (e.g. including explicit diagnostic criteria or “zero time” description). Inclusion/exclusion criteria should not select participants that are systematically different from eligible non-participants.  **Baseline study population:** The baseline study sample (i.e. individuals entering the study) is adequately described for key characteristics.  **Adequant study participation:** There is adequate participation in the study by eligible individuals. Studies should report factors associated with non-response, quantify and interpret these associations to determine if it is a selective sample. For example, a low participation raises suspicion that there may be a barrier to participating that may influence outcomes. |
| **Outcome measurement** |  |
| Is the outcome of interest adequately measured in study participants sufficeint to limit potential bias. | **Definition of outcome:** A clear definition of the outcome of interest is provided, including duration of follow-up and level and extent of the outcome construct.  **Valid and reliable measure of outcome:** The outcome measure and method used are adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test). Measures that are uncommon or have been modified should provide evidence of reliability and validity. Whenever possible, validated instruments should be used.  **Method and setting of outcome measurement:** The method and setting of measurement are the same for all study participants. The measurement approach, timing, and setting of assessment should be standardised across subjects, or conducted in a way that limits systematically different measurement. If there are differences, this should be reported and the implications should be considered.  **Estimation of population parameters:** Estimates of population parameters should be calculated using data observed in the whole sample, not extrapolated from rates observed in a sub-sample (For example, are all participants examined?). |

**Table A4.2 Risk of selection and outcome measurement bias and justification for rating using the QUIPS tool.**

| **Study** | **Risk of selection bias** | **Risk of outcome measurement bias** |
| --- | --- | --- |
| Ablin et al. 2012 [1] | **High:** Very small response rate (30%) and respondents recruited by telephone which systematically excludes those without a home telephone. In addition no information is provided regarding the time of day of the call (daytime only calls systematically exclude daytime workers) or any effort made to reach those not answering a call first time. There is no information provided regarding the target population or non-responders to compare with the study sample to establish if the study sample is different to non-responders or the target population. Estimates of prevalence using extrapolation from data observed in rheumatology outpatients in not a robust estimate of general population prevalence as rheumatology outpatients are likely to be systematically different to the general population. | **High:** Used positive predictive value (PPV) ascertained from a different population to that under investigation to calculate prevalence. Therefore unlikely to be a reliable measure of prevalence. Also no statement made regarding validation of examination used in the rheumatology outpatients department to determine FM status and therefore calculate the PPV used to calculate community prevalence. |
| Aggarwal et al. 2006 [2] | **Low:** High participation rate, clear comparison of responders vs. non-responders. Only question is whether residents of Handforth, Manchester are representative of the UK general population. | **Low:** Measurement of outcome is valid, reliable and similar for all subjects. Might be some underestimation of ACR-1990 CWP status due to exclusion of head pain from case criteria, however, likely to be small. |
| Assumpção et al. 2009 [3] | **High:** Non-random sample selection for screening phase; recruitment of sample by telephone when 27% of target population do not have a telephone and 30% of those called did not answer the phone; recruitment to examination phase based on participant agreement – highly likely to lead to a systematically different group of patients. | **Moderate:** Clear operationalisation of outcome measure but unclear validation of examination/diagnosis and not all participants examined. Insufficient information about screening questionnaire to judge validity. |
| Bergman et al. 2001 [6,7] | **Low:** Random sampling from an appropriate sampling frame. However, unclear if two regions selected are representative of the whole of the general population of Sweden). Good response rate and clear description of non-responders sufficient to provide evidence that participation unrelated to outcome measure. | **Low:** Measurement of outcome is similar for all subjects. While the questionnaire is not validated is uses a standard approach to assessing self-reported CWP status which has been used by other studies (i.e. question regarding pain duration and body manikin to locate pain). |
| Branco et al. 2010 [4,8] | **Moderate:** Unclear sampling frame, insufficient information presented regarding non-responders, recruitment conducted by telephone (excludes those without a landline and those not in at the time of call, therefore responders likely to be systematically different to non-responders). However, sample stated to be representative of the general population, therefore moderate rather than high risk of selection bias. | **High:** Uses positive predictive value (PPV) ascertained from a different population to that under investigation to calculate prevalence, this is unlikely to lead to a reliable prevalence estimate. Also no statement made regarding validation of examination used in the rheumatology outpatient department to determine case status and therefore calculate the PPV used to calculate prevalence. |
| Buskila et al. 2000 [9] | **Low:** Good participation rate, non-responders described, responders compared with 1993 Israel census data, gender distribution of sample found to deviate from census data, however analysis takes this into account (figures adjusted for age and gender). | **Low:** Outcome measurement conducted in a similar way for all participants using an accepted approach to measurement and using clearly documented case criteria. |
| Carnes et al. 2007 [10] | **Moderate:** A reasonable, but not ideal response rate of 60%. The study population is demonstrated to be older (mean age 52 years) and more likely to be female (56% female). No discussion regarding non-responders is provided. Must therefore conclude at least a moderate risk of selection bias. | **Low:** Measure of outcome similar for all subjects, criteria are clearly stated and while the questionnaire is not validated it uses a standard approach to assessing pain location using a blank body manikin. However, there is no detail provided regarding how chronicity of pain was elicited, it is assumed that patients were asked a simple question about pain duration given the use of the chronic pain grade scale. |
| Choudhury et al. 2013 (long) [11] | **High:** While response rate was high (94%), patients were recruited from a GP waiting room thereby automatically selecting a group of the population more likely to be unwell. Quota sampling was used to help ensure a representative sample, but who to approach was still at the discretion of the interviewer and was therefore likely to be biased.  The paper argues that since the aim of the study was not to calculate robust prevalence estimates, but to make comparisons between prevalence in the white and Bangladeshi populations thereby making some of my arguments irrelevant however, I would counter that to make meaningful comparisons between prevalence estimates for groups from different ethnic backgrounds sampling of these groups needs to be representative. | **Moderate:** While the questionnaire uses a standard method of ascertaining CWP status (body manikin) and the setting is the same for all participants, there is no statement regarding the interview validity and it is unclear how much a possible lack of confidentiality in the interview setting (GPs waiting rooms) might have influenced participants responses. |
| Choudhury et al. 2013 (short) [11] | **High:** Very low response rate (27%) and the exclusion of patients with pain due established pathophysiological diagnoses, since CWP can coexist with its differential diagnoses. There is insufficient documentation regarding non-responders and no comparison made between study sample and target population, therefore given the very low response rate we must consider this study to be at high risk of participation bias. | **Low:** Measure of outcome similar for all subjects (although some were assessed with postal questionnaire, late responders were assessed using a telephone questionnaire), criteria are clearly stated and while the questionnaire is not validated it uses a standard approach to assessing pain location using a blank body manikin. |
| Croft et al. 1993 [13] | **Low:** Good (corrected) response rate, comparison been responders and non-responders shows some difference (responders were more likely to be female and in current employment). Short questionnaire completed by a sample of non-responders showed that they were less likely to have chronic pain compared to those who did (suggests that study may slightly overestimate prevalence) but examination of rates of consultation for pain complaints shows little difference between responders and non-responders so effect of participation bias unlikely to be large. | **Low:** Clear operationalisation of CWP criteria, use of a study specific questionnaire that, while not stated to have been validated or tested for reproducibility, uses a standard approach to assessing CWP. Measurement of outcome is similar for all subjects. |
| Hardt et al. 2008 [14] | **Moderate:** The cohort from which the large sample is drawn is stated to be representative of the US general population with an oversampling of ethnic minorities. This is accounted for in analysis by standardising the figures to the US population. This would seem a reasonable approach to take, however, no evidence is provided regarding the representativeness of the study sample, neither are response rates or recruitment strategies documented, it is assumed that these are provided in the publications referenced in the paper, however, since sample is stated to oversample ethnic minorities it cannot be taken to be representative of the whole population. There is also a lack of documentation regarding what constitutes a 'personal' interview; if this is conducted by phone it may systematically exclude individuals without a phone. | **Moderate:** Due to insufficient documentation of interview process and validity/reliability of questionnaire unable to give this study a low risk of outcome measurement bias and must conclude a moderate risk of bias. |
| Häuser et al. 2013 [15] | **Low:** While response rate is low (56%), reasons for non-participation are presented and while the sample is stated to differ from the general population on age and gender distribution and education it is not significantly different from the general population. | **Low:** Validated measure of outcome used, applied in a consistent method to all participants. |
| Hunt et al. 1999 [19,26] | **Low:** Sample selected randomly, reasonable response rate, description of responders provided and comparison of figures derived from sample against figure standardised to the UK general population show little difference, there is no evidence that any differences in the study sample will influence the outcome. | **Low:** Measurement of outcome is similar for all subjects, it uses a standard approach and the coder of manikins is blinded to patient's psychosocial status. |
| Jacobsson et al.1996 [20] | **Moderate:** No documented details are given regarding sample selection, recruitment methods, response rate, non-responders, or baseline study population (with the exception of gender and diabetic status). No comparison is made to the target population. Must therefore conclude at least a moderate risk of selection bias, since insufficient information to make the judgment. | **Moderate:** The operationalisation of the outcome measure is reasonable, but no statement is made regarding any validity/reliability testing of the questionnaire/interview process, must therefore conclude moderate risk of outcome measurement bias. |
| Kim et al. 2006 [22] | **High:** Non-random voluntary recruitment is highly likely to lead to participation bias. This is supported by the observation that 66% of the sample were female, this is unlikely to be representative of the underlying population. Response rate is not applicable since participants were recruited by voluntary response to a mass health screening. | **Moderate:** Clear case criteria have been used, but there is insufficient documentation regarding the reliability of the screening questionnaire and setting of outcome measurement to infer low risk of bias. There is also inadequate documentation regarding testing for reliability of interviewers and examinations, so must conclude at least moderate risk of bias. |
| Klemp et al. 2002 [23] | **High:** Non-random recruitment by tribal elders of tribal constituent of the study, while necessary due to cultural beliefs, will nonetheless lead to bias. Effort has been made to counter this by adjusting figures for age (would result in a moderate risk of bias), however, in addition, the study had very poor response rates and there was no analysis of non-responders and no comparison made with target population, must therefore conclude a high risk of participation bias. | **Moderate:** While there is clear documentation of the examination procedure and the examination process is shown to be reliable and reproducible by pre-study standardisation, there is no documentation regarding the validity/reliability of the interview procedure to ascertain CWP status, must therefore conclude a moderate risk of outcome measurement bias. |
| Lindell et al. 2000 [24] | **High:** While response rate is reasonable, and sample selection would seem to be fair, there seems to be a problem with response-bias: non-participants are shown to be different to participants with respect to the outcome measure (non-responders shown to be less likely to complain of chronic pain). No comparison is made to target population. Study only examines a stratified sample of positive responders. | **High: L**ong delay between questionnaire and examination, may have resulted in resolution of symptoms for some subjects leading to underestimation of prevalence. In addition only examining a stratified sample of screen positives and extrapolating to the rest of the population is not robust. |
| Macfarlane et al. 2005 [25] | **Moderate:** Using GP registered population to reflect general population of the UK is valid due to high proportion of individuals registered with a GP in the UK, however, GP practices were chosen for high density of South Asian population, therefore unlikely to represent the general population of the UK. Response rates are relatively poor. Comparison of age/gender distribution in the two study groups demonstrates that samples not similar, therefore also unlikely to be representative of the underlying population. However, prevalence figures standardised for age and gender. | **Low:** Measurement of outcome is similar for all subjects, while study makes no statement regarding validity/reliability of questionnaire; it uses a standard approach to assessing CWP. |
| Mundal et al. 2014 [28] | **Moderate:** by attempting to recruit all residents in an area sample selection methods were unlikely to lead to bias, however restricting the sample to only those who participated in both waves of a large study reduced the overall participation to 54%. The baseline characteristics of the sample revealed that 57% of the sample were female and 70% were older than 40 suggesting that the sample is unlikely to be representative of the underlying population. | **Low:** The study uses a stand method of outcome measurement and is similar for all subjects. |
| Papageorgiou et al. 2002 [32] | **Low:** Reasonable response rate and sample selected randomly from GP register (those registered with a GP in the UK are felt to be representative of the general population in the UK), however, need to be careful with generalising this prevalence to the population as a whole, since only used one general practice in Handford. | **Low:** The study uses a stand method of outcome measurement and is similar for all subjects. |
| Raspe & Baumgartner 1993 [33] | **High:** The sample is selected randomly, there is a reasonable response rate of 81% to the initial postal questionnaire and a reasonable response of 76% to examination, however, there is no evidence provided that the sample is representative of the target population and no information is provided regarding non-responders to both the questionnaire and examination phase. Must therefore conclude at least a moderate risk of sample selection bias, however, the assumptions that have been made in the estimation of population parameters mean this study must be rated at high risk of bias. The assumptions made in the calculation of prevalence are very likely to result in inaccurate estimates. It is not appropriate to assume that non-responders are the same to responders, cannot therefore assume that there is an equal frequency of widespread pain in non-responders, cannot therefore extrapolate anything regarding prevalence among those who have not responded. | **High:** Definition of outcome not adequately documented. Unclear if questionnaire is capable of picking up all subjects with CWP, therefore only examining those that are defined as having multi location chronic pain according to the questionnaire may miss relevant patients. Examination is well documented, with the exception of the location of the control points. |
| Scudds et al 2006 [35] | **Moderate:** Sample selected randomly from telephone directory, with reasonable response rate of 60%. Sample stated to be representative of Hong Kong census in terms of age. No other comparisons offered. No information provided regarding non-responders. Low response rate to examination phase also a problem. Sampling frame excludes those without a telephone and those who have chosen not to have their numbers in the telephone directory, therefore likely to lead to a systematically different set of individuals in the sample. | **Low:** The study uses clearly defined case definition criteria and uses a back translated version of a validated instrument to screen participants for CWP. Interviews are standardised and examination protocol clearly documented. |
| Storozhenko et al. 2004 [37] | **Moderate:** Good response rate of 76% and the age and gender distribution of the sample is provided however, no statement is made regarding whether Yekaterinburg is representative of the general population of Russian and no comparison made with target population and no description of non-responders provided. | **Low:** Method and setting of outcome measurement similar for all respondents. Questionnaire based on validated questionnaires. Russian and English speakers translated and back translated from a validated questionnaire. |
| White et al. 1999 [39] | **Moderate:** Sample shown to be different to census data, while figures are standardised against census data, this does not remove possible systematic bias as a result of telephone interviewing (excludes those without a home phone). | **Low:** Study uses validated questionnaire, interviews are quality controlled and examination is clearly documented and consensus between examiners checked. |
| White et al. 2003 [40] | **Low:** Amish arm of study low risk of selection bias due to 74% participation by those eligible to participate. Amish community in Aylmer likely to be representative of other Canadian Amish.  **Moderate:** Non-Amish rural arm of study moderate risk of selection bias due to telephone recruitment of participants excluding those without a telephone and those not in at the time of a call, no information about non-responders, and figure for FM likely to be biased due to low participation rate for examination arm of study (42%) | **Moderate:** There is insufficient documentation regarding the reliability of the examination phase of the study, must therefore conclude at least a moderate risk of outcome measurement bias. |
| Wolfe et al. 1995 [42] | **Moderate:** While the study samples the general population of Wichita city, with a good response rate of 86% for questionnaire and a reasonable response of 61% to the examination phase, it is not clear how representative Wichita city is of the general population in the US, must therefore conclude a moderate risk of selection bias. | **Low:** Questionnaire used is not validated but uses a standard approach, examination is standardised and clearly documented. |