Effectiveness of shared medical appointments delivered in primary care for improving health outcomes in patients with long-term conditions: a systematic review of randomised controlled trials

Mei Yee Tang¹⁺, Fiona Graham¹⁺⁺, Amy O'Donnell¹, Fiona R Beyer¹, Catherine Richmond², Falko F Sniehotta¹,₃⁺, Eileen Kaner¹⁺⁺

¹ NIHR Policy Research Unit in Behavioural Science, Newcastle University, Newcastle upon Tyne, UK
² Population Health Science Institute, Newcastle University, Newcastle upon Tyne, UK
³ Department of Public Health, Preventive and Social Medicine, Mannheim Medical Faculty Heidelberg University, Germany

*Corresponding author: Fiona.graham@newcastle.ac.uk

¹⁺Mei Yee Tang and Fiona Graham contributed equally to this paper as first co-authors. Falko Sniehotta and Eileen Kaner contributed equally to this paper as last co-authors.

Funding
This study is funded by the National Institute for Health Research (NIHR) [Policy Research Unit in Behavioural Science (project reference PR-PRU-1217-20501)]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Acknowledgements
The authors would like to thank the following authors for providing additional data used in this review: Dr Renee E Cole, Dr Connor Drake, Dr William Yancy, and Professor Tracey Taveria. We would also like to thank Professor Dawn Craig for her comments on an earlier draft of this manuscript.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Background: Shared medical appointments (SMAs) have the potential to address interlinked challenges of limited capacity in primary healthcare and rising prevalence of patients with multiple long-term conditions (LTCs). This review aimed to examine the effectiveness of SMAs compared to one-to-one appointments in primary care at improving health outcomes and reducing demand on healthcare services.

Methods: We searched for randomised controlled trials (RCTs) of SMAs involving patients with LTCs in primary care across six databases from 2013 and added eligible papers identified from previous relevant reviews. Data were extracted and outcomes narratively synthesised, meta-analysis was undertaken where possible.

Results: Twenty-three unique trials were included. SMA models varied in terms of components, mode of delivery and target population. Most trials recruited patients with a single LTC, mostly commonly diabetes (n=13), although eight trials selected patients with multiple LTCs. There was substantial heterogeneity in outcome measures which we categorised into health outcomes (biomedical indicators, psychological and well-being measures), behavioural outcomes, and resource use. Meta-analysis showed that participants in SMA groups had lower diastolic blood pressure than those in usual care (d=-0.123, 95%CI = -0.22, -0.03, k=8). No statistically significant differences were found across other outcomes. Where individual studies showed significant differences (patient self-efficacy), these trended in favour of SMAs. Compared with usual care, SMAs had no significant effect on healthcare service use.

Conclusions: SMAs were at least as effective as usual care in terms of health outcomes and did not lead to increased healthcare service use in the short-term. They show some potential in improving self-efficacy which may boost self-care. To strengthen the evidence base, future studies should target standardised behavioural and health outcomes and clearly report SMA components so key behavioural ingredients can be identified. Similarly, transparent approaches to measuring costs would improve comparability between studies.

PROSPERO: CRD42020173084

Keywords: shared medical appointments, long-term conditions, systematic review, meta-analysis
INTRODUCTION

Shared Medical Appointments (SMAs), also known as group consultations, are a model of care with the potential to address the interlinked challenges of limited capacity in primary care and rising prevalence of patients with multiple long-term conditions (LTCs)[1,2]. SMAs are longer appointments (typically 60-120 minutes) whereby patients with the same LTCs meet with their physician together[3]. SMAs are typically co-led and/or facilitated by healthcare professionals, such as nurses, pharmacists, psychologists, and physiotherapists. The group typically consists of between 6-10 patients and may include family members and caregivers[4]. There are various models of SMA but generally they retain some features of a standard one-to-one appointment such as physical examinations and personalised review of medical charts[2]. In addition, SMAs provide participants an opportunity to ask questions of clinicians and other patient and receive formal education and counselling during the group session. SMAs have been used to deliver care for a range of health conditions including diabetes, hypertension, and chronic pain; though there is potential for wider application, including multimorbidity[5].

A recent synthesis of qualitative literature found that most patients and primary care practitioners regarded SMAs positively[6]. Key benefits included improved patient self-confidence and motivation for self-management; whilst practitioners felt that SMAs had the potential to provide a more efficient and effective way of delivering care[6]. Previous reviews of effectiveness were inconclusive but evidence, largely from the United States (US) and Australia, reported a promising effect of group consultations for some biomedical measures. For example, improvements in glycated haemoglobin A1C (HbA1C) and systolic blood pressure (SBP) were greater in patients attending SMAs compared to usual care for diabetes[4,7]. However, previous reviews include studies that evaluate the use of SMAs as follow-up appointments after secondary care treatment[4]. It is unclear whether these types of group appointments are effective in supporting improved ongoing management of LTCs in primary care.

Responding to these evidence gaps, we aimed to examine the effects of SMAs delivered in primary care on health outcomes and healthcare service use in patients with LTCs. We sought to answer two overarching research questions:

1. Are SMAs effective in improving health outcomes for patients with one or more long-term conditions?
2. Do SMAs reduce healthcare service use by patients with one or more long-term conditions?

METHOD
This systematic review follows Cochrane Handbook Guidance[8] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[9].

Protocol and registration
This study was registered on PROSPERO (CRD42020173084). Regarding protocol changes, we proposed coding of Behaviour Change Techniques Taxonomy V1 (BCTTv1)[10] used in SMAs and associated with changes in outcomes. However, most included studies did not report the required detail and so instead we narratively described this information.

Inclusion/exclusion criteria
Studies were included if they met the criteria outlined in Table 1.

Table 1: Inclusion and exclusion criteria

<table>
<thead>
<tr>
<th>PICOS Criterion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td>Adult patients (≥18 years of age) with one or more long-term conditions who have attended at least one shared medical appointment (SMA) in a primary care setting were eligible for inclusion. Patients in primary care settings who were seen by a secondary care specialist during the SMA were excluded.</td>
</tr>
<tr>
<td><strong>Intervention (Model of care)</strong></td>
<td>All countries were eligible for inclusion. SMAs/group consultations/group visits conducted in primary care setting, delivered by a primary care healthcare professional (e.g. nurse, doctor, pharmacist), were eligible for inclusion. The present review considered SMAs to be clinical encounters in which groups of patients with the same/similar long-term condition meet with a healthcare professional for routine care. The SMA must have included one-to-one time for every patient in attendance. Therefore, peer support groups were excluded.</td>
</tr>
<tr>
<td><strong>Comparison/control</strong></td>
<td>No restrictions – usual care, active control (e.g., another SMA model).</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Behavioural outcomes – e.g., healthcare utilisation, physical activity, medication adherence. Disease-specific measures – e.g., Haemoglobin A1C (HbA1C). Biophysical health indicators – e.g., cholesterol, weight, BMI. Cost/resource use.</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Only randomised controlled trials that were published in peer-reviewed journals were eligible for inclusion.</td>
</tr>
</tbody>
</table>
Search strategy
A comprehensive search strategy was developed, based on the approach described in Booth et al.[4], to search for trials published after their search, namely the period 2013-2020. Key changes included the removal of the terms “group outpatient”, “GMV” or “GMA”, “group processes” and “Group care” (see S1). The search strategy was first used to search MEDLINE (via OVID) and then translated for the following databases: EMBASE (via OVID), Science citation index (via Web of Knowledge) Social Science Citation Index (via Web of Knowledge), Cumulative Index to Nursing and Allied Health Literature (via EBSCOhost), Cochrane Central Register of Controlled Trials (Wiley), DARE, NHS EED, and HTA (Centre for Reviews and Dissemination). Any relevant pre-2013 trials identified by forward and backward citation searches of the included trials, including those in relevant systematic reviews[4,5,11,12], were also included in the review.

Screening
Screening and data extraction was facilitated using the systematic review management tool, Covidence[13]. Two reviewers independently screened all titles and abstracts against the inclusion criteria, and a third reviewer adjudicated any disagreements. This process was also applied to the screening of full-text papers.

Data extraction
Information relating to the study design, population, intervention, form of delivery[14] (e.g. experience/training of the providers and facilitators), outcomes, and results were extracted from all relevant papers using a data extraction form (see S2). All information was double-extracted by two researchers, with disagreements resolved through discussion or third-party moderation.

Where data were reported for several time-points, the data-point closest to the end of the SMA intervention was used to calculate the effect size as this would be when the largest effects attributable to the SMA is expected. If available, intention-to-treat data were used to calculate effect sizes.

Quality assessment
Two researchers independently assessed the quality of all included studies using the Cochrane Risk of Bias Tool[15]. Percentages of judgements (high, low, or unclear) for each domain was calculated across the studies.

**Data analysis/synthesis**

We mapped all reported outcomes into the following categories agreed by the wider research team to best reflect SMA effectiveness and efficiency: biomedical health indicators, psychological and well-being measures, behavioural outcomes, and cost and resource.

Meta-analyses were performed in StataIC 15. Given the heterogeneity between studies, a random-effects model was used[16]. Meta-analyses were conducted where there were at least two studies reporting a specific outcome[17]. Outcome effect sizes were calculated as Cohen’s D (standardised mean difference). Heterogeneity was assessed using Higgins I-Square ($I^2$), whereby 50-90% was considered as representing substantial heterogeneity[18]. Authors were contacted for additional information if data needed to calculate effect sizes were not sufficiently reported in the published paper(s). Where this information could not be obtained from authors, $p$-values and confidence intervals were used to calculate effect sizes[8]. Only studies in which the comparator was usual care were pooled into the meta-analysis.

Using meta-regression, sensitivity analyses were conducted to explore whether results differed according to sources of bias identified from the risk of bias assessment and the omission of outliers.

Studies that were too heterogeneous to perform meta-analysis (where the comparator was not usual care) were synthesised narratively. For each outcome measure, evidence of an effect was determined by the $p$-values reported in the papers. To assess the certainty of the evidence, number of study participants, confidence intervals and the consistency of effects across studies, the risk of bias of the studies, how directly the included studies address the planned question (directness) were taken into consideration.

**RESULTS**

<Figure 1 – PRISMA diagram of study selection process here>

**Characteristics of the included studies**
Twenty-three unique trials (reported in thirty-four papers) were identified, for PRISMA details see Figure 1 and Table 2. See S3 for list of included papers.
<table>
<thead>
<tr>
<th>Study</th>
<th>Condition(s)</th>
<th>Country</th>
<th>Setting</th>
<th>Sample size (N)</th>
<th>Model</th>
<th>Comparator</th>
<th>Characteristics of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott (2004)[19]</td>
<td>Chronic conditions</td>
<td>USA</td>
<td>Large health maintenance organisation</td>
<td>295</td>
<td>CHCC model- 'group visits'</td>
<td>Usual care</td>
<td>SMA: Age*, 74.2 (7.6) UC: Age*, 74.1 (7.4) Female %, 58.8</td>
</tr>
<tr>
<td>Coleman (2001)[20]</td>
<td>^Chronic conditions</td>
<td>USA</td>
<td>Community Health Centres serving low-income, racially and ethnically diverse populations</td>
<td>159</td>
<td>Integrated medical group visit</td>
<td>Usual care</td>
<td>Age* 50.5 (12.3) Female %, 86.5 White%, 81.6</td>
</tr>
<tr>
<td>Gardiner (2017)[21]</td>
<td>Chronic pain and depression</td>
<td>USA</td>
<td>Community Health Centres serving low-income, racially and ethnically diverse populations</td>
<td>159</td>
<td>Integrated medical group visit</td>
<td>Usual care</td>
<td>Age* 51.4 (8.5) Age range: 32-65 years Female %, 89.3 Ethnicity %: non-Hispanic Black 77.4, non-Hispanic White 17.9, Asian Pacific 1.2, American Indian, 1.2</td>
</tr>
<tr>
<td>Gardiner (2019)[22]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>Community-based health center serving low-income population who are working</td>
<td>80</td>
<td>Group visits</td>
<td>Usual care</td>
<td>Age*, 54.0 (10.44) (range 22–83 years); Female %, 78.3 Ethnicity %: African American 77.5, Caucasian 21.7</td>
</tr>
<tr>
<td>Berry (2016)[23]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>University Primary Care Center serving uninsured or inadequately insured patients</td>
<td>120</td>
<td>CHCC model- 'group visits'</td>
<td>Usual care</td>
<td>Age, 56.1 (range, 26.5-80.7), Female %, 72 African American %, 82.8</td>
</tr>
<tr>
<td>Clancy 2003a[24]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>University Primary Care Center serving mainly minority, inadequately insured patients</td>
<td>186</td>
<td>CHCC model- 'group visits'</td>
<td>Usual care</td>
<td>Age*, 58.3 (9.6) Female %, 46 Ethnicity, %: Caucasian 64, African American 17, Hispanic 19</td>
</tr>
<tr>
<td>Clancy 2003b[25]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>University Primary Care Center serving mainly minority, inadequately insured patients</td>
<td>186</td>
<td>CHCC model- 'group visits'</td>
<td>Usual care</td>
<td>Age*, 58.3 (9.6) Female %, 46 Caucasian 64, African American 17, Hispanic 19</td>
</tr>
<tr>
<td>Clancy 2007a[26]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>University Primary Care Center serving mainly minority, inadequately insured patients</td>
<td>186</td>
<td>CHCC model- 'group visits'</td>
<td>Usual care</td>
<td>Age*, 58.3 (9.6) Female %, 46 Caucasian 64, African American 17, Hispanic 19</td>
</tr>
<tr>
<td>Cole (2013)[29]</td>
<td>Diabetes (Prediabetes)</td>
<td>USA</td>
<td>Military health system</td>
<td>65</td>
<td>SMA</td>
<td>Individual counselling</td>
<td>Age*, 58.3 (9.6) Female %, 46 Caucasian 64, African American 17, Hispanic 19</td>
</tr>
<tr>
<td>Drake (2018)[30]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>Family medicine centre</td>
<td>33</td>
<td>SMA + personalised health planning</td>
<td>SMA</td>
<td>SMA: Age* 59.4 (10.5) UC: Age* 55.1 (14.5) Female %, 68.4 Ethnicity, %: White non-Hispanic 21.1, Black non-Hispanic 73.7, Hispanic 5.3</td>
</tr>
<tr>
<td>Study (Year)</td>
<td>Condition</td>
<td>Country</td>
<td>Setting</td>
<td>Sample Size</td>
<td>Type of Care</td>
<td>Age</td>
<td>Gender</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td>--------------</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td>Ee (2020)[31]</td>
<td>Diabetes (Type 2)</td>
<td>Australia</td>
<td>University primary care center</td>
<td>18</td>
<td>SMA</td>
<td>Usual care</td>
<td>SMA: Age* 55.9 (11.6), Female %, 33 &lt;br&gt; Ethnicity %: European 33.3, Oceanic 27.3, Asian 36.8</td>
</tr>
<tr>
<td>Liu (2012)[32]</td>
<td>Diabetes (Type 2)</td>
<td>The People’s Republic of China</td>
<td>General practices</td>
<td>208</td>
<td>CHCC model-‘group visits’</td>
<td>Usual care</td>
<td>SMA: Age, 61.99 (9.8) &lt;br&gt; UC: Age, 62.49 (9.97) &lt;br&gt; Female %, 62.0 &lt;br&gt; Ethnicity %: Han 96.2, Zhuang 2.9, Other 0.96</td>
</tr>
<tr>
<td>Naik (2011)[33]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>Veterans Affairs Medical Centres</td>
<td>87</td>
<td>Group clinic</td>
<td><strong>Enhanced usual care</strong></td>
<td>SMA Age, 63.82 (7.9) &lt;br&gt; UC: Age, 63.45 (7.8) &lt;br&gt; African American %, 31.0</td>
</tr>
<tr>
<td>Schillinger (2008)[34]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>University affiliated-Safety net settings-community health network</td>
<td>339</td>
<td>Group medical visit</td>
<td>Automated telephone self-management</td>
<td>Age* 56.1 (12) &lt;br&gt; Female %, 59 &lt;br&gt; Ethnicity, %: Asian 23.3, African American 20.6, White/Latino 46.9, White non-Latino 7.7, Other/unknown 1.5</td>
</tr>
<tr>
<td>Vaughan (2017)[37]</td>
<td>Diabetes (Prediabetes/Type 2)</td>
<td>USA</td>
<td>Community health clinic serving low-income Hispanic adults</td>
<td>62</td>
<td>Group visit</td>
<td>After 6 months, the control group received the intervention</td>
<td>SMA: Age* 51.3 (NR) &lt;br&gt; UC: Age* 48 (NR) &lt;br&gt; Female %, 86 Hispanic 98%</td>
</tr>
<tr>
<td>Vaughan (2020)[38]</td>
<td>Diabetes (Type 2)</td>
<td>USA</td>
<td>Community health clinic serving low-income Hispanic adults</td>
<td>89</td>
<td>Group visits</td>
<td>Usual care</td>
<td>SMA: Age* 55.99 (7.12) &lt;br&gt; UC: Age* 53.86 (9.07) &lt;br&gt; Female %, 71.9 &lt;br&gt; Self-identified as Latino(a), and Spanish-speaking</td>
</tr>
<tr>
<td>Wagner (2001)[39]</td>
<td>Diabetes (Type 1 &amp; 2)</td>
<td>USA</td>
<td>Primary care practices in health maintenance organisation</td>
<td>708</td>
<td>Chronic care clinic</td>
<td>Usual care</td>
<td>SMA: Age* 61.2 (NR) &lt;br&gt; UC: Age*, 60.4 (NR) &lt;br&gt; Female% 47.1 Caucasians %, 68.9</td>
</tr>
<tr>
<td>Cohen (2011)[40]</td>
<td>Diabetes (Type 2) and cardiovascular risk</td>
<td>USA</td>
<td>Veterans Association Medical Center</td>
<td>99</td>
<td>SMA programme (2 phases)</td>
<td>Usual care</td>
<td>SMA: Age* 69.8 (10.7) &lt;br&gt; UC: Age*, 67.2 (9.4) &lt;br&gt; Female %, 2</td>
</tr>
<tr>
<td>Taveira (2014)[41]</td>
<td>Diabetes and cardiovascular risk</td>
<td>USA</td>
<td>Veterans Association Medical Center</td>
<td>200</td>
<td>Group medical visit</td>
<td>CRRC Individual clinic</td>
<td>Usual care</td>
</tr>
<tr>
<td>Study</td>
<td>Disease</td>
<td>Location</td>
<td>Setting</td>
<td>Sample Size</td>
<td>Intervention</td>
<td>Comparator</td>
<td>Characteristics</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------</td>
<td>-------------</td>
<td>--------------------------------------------------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Wu (2018)[42]</td>
<td>Diabetes (Type 2) and cardiovascular risk</td>
<td>USA Veterans Health Administration Hospital</td>
<td>primary care services</td>
<td>250</td>
<td>Group medical visit</td>
<td>Usual care</td>
<td>SMA: Age* 65.8 (8.7) UC: Age* 65 (9.8) Female% 3.6 Ethnicity %: Caucasian 76.6, African American 10.8, Asian pacific islanders 9.45, Hispanic 1.4, Unknown 1.5</td>
</tr>
<tr>
<td>Taveira (2011)[43]</td>
<td>Diabetes (Type 2) and depression</td>
<td>USA Veterans Affairs Medical Centres</td>
<td></td>
<td>88</td>
<td>SMA/ Group medical appointment</td>
<td>Usual care</td>
<td>SMA: Age* 60.2 (9.3) UC: Age* 61.4 (9.9), Female %, 2.3 White %, 98.9</td>
</tr>
<tr>
<td>Yancy (2020)[44]</td>
<td>Diabetes (Type 2) and overweight</td>
<td>USA Veterans Association Medical Center</td>
<td>Group medical visit</td>
<td>263</td>
<td>Group medical visit with IWM</td>
<td>Age* 60.7 (8.2) Female % 10.6 Ethnicity %: White 42.6, African American 54.4, Other 3</td>
<td></td>
</tr>
<tr>
<td>Crowley (2017)[45]</td>
<td>Diabetes (Type 2) and overweight</td>
<td>USA Veterans Association Medical Center</td>
<td>Group medical visit</td>
<td>239</td>
<td>Group clinic</td>
<td>Usual care</td>
<td>Age, 62.0 Male %, 96.9 African American %, 57.1</td>
</tr>
<tr>
<td>Edelman (2010)[46]</td>
<td>Diabetes (Type 2) and hypertension</td>
<td>USA Veterans Association Medical Center</td>
<td>Group clinic</td>
<td>239</td>
<td>Group clinic</td>
<td>Usual care</td>
<td>Age, 66.5 (9.8) Female %, 58.6</td>
</tr>
<tr>
<td>Crowley (2013)[47]</td>
<td>Diabetes (Type 2) and hypertension</td>
<td>USA Veterans Association Medical Center</td>
<td>Group clinic with IWM</td>
<td>239</td>
<td>Group clinic</td>
<td>Usual care</td>
<td></td>
</tr>
<tr>
<td>Crowley (2014)[48]</td>
<td>Diabetes (Type 2) and hypertension</td>
<td>USA Veterans Association Medical Center</td>
<td>Group clinic with IWM</td>
<td>239</td>
<td>Group clinic</td>
<td>Usual care</td>
<td></td>
</tr>
<tr>
<td>Eisenberg (2019)[49]</td>
<td>Diabetes (Type 2) and hypertension</td>
<td>USA Veterans Association Medical Center</td>
<td>Group clinic with IWM</td>
<td>239</td>
<td>Group clinic</td>
<td>Usual care</td>
<td></td>
</tr>
<tr>
<td>Gao (2015)[50]</td>
<td>Hypertension</td>
<td>The People’s Republic of China Community Health Care Centre</td>
<td>CHCC model+ ‘group visits’</td>
<td>1346</td>
<td>CHCC model + ‘group visits’</td>
<td>Usual care</td>
<td>Age, 66.5 (9.8) Female %, 58.6</td>
</tr>
<tr>
<td>Simon (2015)[51]</td>
<td>Hypertension</td>
<td>Germany Physician practices</td>
<td>CHCC model ‘group medical visit’</td>
<td>48</td>
<td>CHCC model ‘group medical visit’</td>
<td>Usual care</td>
<td>NR</td>
</tr>
</tbody>
</table>

Multiple papers for the same trial were found therefore we bolded the one used as the index paper throughout the rest of the paper. *Chronic conditions (including asthma, COPD, heart failure, diabetes, arthritis, deafness, blindness), *Age in years mean (standard deviation), **enhanced usual care- patients required to attend 2 diabetes group education sessions. CRRC- Cardiovascular Risk Reduction Clinic, IWM- intensive weight management, SM- Self-management, SMA- Shared medical appointment, UC- usual care.
Sixteen trials (70%) were for a single LTC, of these: 13 were for diabetes[23,25,26,29–34,37–39,53], two for hypertension[50,51]; and one was for osteoporosis[52]. Eight trials considered multiple LTCs: three were diabetes and hypertension/cardiovascular risk[40–42,46]; one was diabetes and depression[43], one was for overweight patients with diabetes[44]; one was chronic pain and depression[22]; and one included multiple LTCs including: arthritis, hypertension, cancer, deafness and diabetes[19]. Overall, 20/23 (87%) of trials focused on patients with diabetes.

Nineteen trials (83%) were conducted in the US[19,22,23,25,26,29,30,33,34,37–44,46,53], two in China[32,50] and one each in Australia[31], Germany[51] and the United Kingdom[52]. Twelve trials were measured the effectiveness, impact or efficacy of SMAs compared to usual care[19,22,23,26,29,32,33,38,39,42,46,53], nine trials examined feasibility parameters[24,30,31,34,37,38,43,50], and two trials were non-inferiority/superiority trials[44,45,52].

In eight trials (35%) participants were veterans or military personnel[29,33,40–44,46]. Participants were from low-income communities in four trials[22,23,37,38], and uninsured communities in three US trials[25,26,34]. Three trials were tailored for non-English speaking participants, where written materials were available in Spanish[34,38,53]. The majority of participants were over 50 years old, the mean age of participants ranged between 50.5[21,22] to 74 years old[52]. Two trials were specifically for older patients over 55 or 60 years respectively[19,51], and two trials excluded patients over 75[44] and 80 years[32]. Twelve trials (52%) had a majority of female participants[19,23,24,32,37,38,50,52]. Ten trials had majority of male participants[29–31,33,40–44,46]. Two trials did not report the gender balance of participants[51,53]. Five studies had a majority White population[22,29,39,42,43], six trials had a majority Black population[23,25,26,30,44,46], three trials had a majority Hispanic population[37,38,53], two trials had a majority Asian population[31,32], and one trial had a majority White-Latino population[34]. Five studies did not report the ethnicity of participants[19,40,50–52].

Most trials (n=18, 78%) had a two-arm design that compared SMAs with usual care, typically a one-to-one (1:1) appointment with primary care physician[19,22,23,25,26,31,32,37–43,46,50,52,53]. In three other two-arm trials, the comparator was a 1:1 appointment plus two diabetes group education sessions[33], an SMA without an integrated weight management programme[44] or an SMA with a personal health planning component (PHP)[30]. Two trials had a delayed six-month waitlist control design[37,38]. Two trials had a three-arm design; one examined the effectiveness of a cardiovascular risk reduction clinic
compared with group medical visits and usual care[41], a second compared an automated telephone self-management service with SMAs and usual care[34].

Trial outcome measures varied across trials though most included behavioural measures, biomedical health indicators and psychological and well-being measures. Full details of all outcome measures reported by the studies are presented in S4.

Risk of bias
Risk of bias item across the studies was generally low across the items, except for ‘Blinding of participants and personnel’ (83% of trials) (See S5).

Sensitivity analyses
There were no differences for any of the outcomes according to the risk of bias assessment criteria relating to random sequence generation, allocation concealment, blinding, incomplete outcome data, and selected reporting (see S5).

Effectiveness of SMAs

Biomedical measures

_Glycated haemoglobin A1C (HbA1c) (%)_

Of the 16 trials measuring HbA1c (%), nine trials[23,26,31,34,38,41–43] which compared SMA to usual care were included in a meta-analysis. No statistically significant difference between SMAs and usual care was found for HbA1c (%) at follow-up \( (d=-0.098, 95\% CI = -0.34, 0.14, k=9) \) \( (p=0.420) \) (see Figure 2a). Substantial heterogeneity was observed \( (I^2 = 70.9\%) \). Of five other SMA trials reporting HbA1c (%) but not in the meta-analysis[25,39,40,46], only one reported significant group differences, whereby the SMA group had significantly higher odds of attaining HbA1c goals (< 7%) compared to usual care[23]. However, this was a high risk of bias study, scoring ‘unclear’ across the six domains. The pooled effect remained similar when Ee et al. (2020)[31] was removed as an outlier in the sensitivity analysis due to wide 95% CIs, \( (d=-0.120, 95\% CI = -0.37, 0.13, k=8) \)

_Diastolic blood pressure_

Of 12 studies which reported diastolic blood pressure (DBP)[23,26,29–32,34,37,38,41,46,50], eight were included in meta-analysis[23,29,31,34,37,38,41,50]. A very small statistically significant pooled effect was found at follow-up \( (d=-0.123, 95\% CI = -
0.22, -0.03, k=8) (p=0.008), whereby participants in the SMA group had lower DBP than those in usual care (see Figure 2b). The pooled effect remained similar (d=-0.194, 95%CI = -0.35, -0.04, k=7) (p=0.016) when Gao et al. (2015)[50], the study with the largest sample size (1204), was omitted in the sensitivity analysis.

Of the three studies not included in the meta-analysis[26,32,46], one trial of SMAs for diabetes and hypertension reported that mean DBP was lower in the SMA group (78.3 mmHg) than in the usual care group (82.1 mmHg) at 12 months[46].

**Systolic blood pressure**

Of 13 trials reporting SBP[23,26,29,31,32,34,37,38,40–42,46,50], nine could be meta-analysed[23,29,31,34,37,38,41,42,50]. No statistically significant difference between SMAs and usual care was found for SBP at follow-up (d=-0.018, 95%CI = -0.11, 0.08, k=9) (p=0.709), with low level of heterogeneity observed (I² = 3.8%) (see Figure 2c). Of the four trials not in the meta-analysis[26,32,40,46], two moderately robust studies reported statistically significant group differences in SBP at follow-up[32,46], whereby the SMA group showed greater decreases in SBP compared to usual care.

No statistically significant effect of SMAs compared to usual care was found for other biomedical health outcomes including: total cholesterol, high density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, weight, and BMI (see S6).

**Trials with non-usual care comparators**

Three trials had enhanced SMAs as their comparator[30,33,44]. Drake et al., (2018) reported that participants in the PHP SMA group had lower DBP (M=86 mmHg) at 8 months follow-up compared to participants in the standard SMA group (M=79.8 mmHg)[30]. Naik et al. (2011) also reported significant group differences at follow-up, whereby participants (US veterans) in the SMA arm demonstrated significantly greater improvements in HbA1c (%) than those in the control group receiving usual care and group education[33]. Yancy et al. (2020) reported patient weight loss in the SMAs with intensive weight management was comparable to weight loss amongst patients attending SMAs but statistical significance was unclear[44]. It is possible that other studies could not detect statistically significant differences between arms due to small sample sizes.

<Insert Figure 2 a,b,c, here>

Psychological and well-being measures
Quality of life
Six trials reported quality of life (QoL) outcomes[19,22,31,34,40,42] of which two trials reported significant group differences[19,34]. One trial of SMAs for chronically ill patients with multiple LTCs found that participants in the SMA group (M=7.2, SD=1.8) reported significantly better QoL than the usual care group (M=6.3, SD=2.0) (p=.002) at 24 months[19]. Schillinger et al., (2008) measured QoL using the short form (SF)-12 instrument which comprised of mental health and physical health subscales. Improvements in SF-12 mental health was observed for SMA group compared to SMA (effect size=0.31, p=0.03) and usual care (effect size=0.18, p=0.2)[34]. However, this was considered as a high risk of bias study, with high/unclear judgements across four out of six domains.

Patient satisfaction
Four trials measured patient satisfaction[19,30,39,52]. Scott et al. (2004) reported significant differences at follow-up with SMA patients reporting higher satisfaction with practitioner discussions compared to controls[19]. The other three studies found no differences.

Patient self-efficacy
Self-efficacy was measured in nine trials[19, 22, 23,32,34,40,43,46,50], of which five studies were included in meta-analysis[19,22,23,34,50]. No statistically significant effect was found (d=0.167, 95%CI = -0.08, 0.41, k=5) (p=0.182) (See S7). High levels of heterogeneity were observed (I²=78.9%). Of the four other studies not included in meta-analysis[32,40,43,46], two reported that SMA patients had significant improvements in self-efficacy to manage diabetes compared to usual care[32,46].

No statistically significant effect of SMAs compared to usual care was found for other depression (see S7).

Trials with non-usual care comparator
Drake et al. (2018) reported significant improvements in self-efficacy, as measured using the Diabetes Empowerment Scale, for the PHP SMA group compared to the standard SMA group[30]. Naik et al. (2011) did not find any differences in diabetes self-efficacy scores between the SMA group and the traditional diabetes group education group[33].

Behavioural outcomes

Hospital admissions
Seven trials reported hospital admissions within six to 24 months\cite{19,23,25,39,41,43,46} and three were included in a meta-analysis\cite{19,23,41}. There was no difference between SMAs and usual care in terms of hospital admissions at follow-up ($d=0.016$, 95\%CI = -0.38, 0.35, $k=3$; ($p=0.931$) (see Figure 3a). Substantial heterogeneity was observed ($I^2 = 71.1\%$).

None of the other four trials\cite{25,39,43,46} reported significant group differences for hospital admissions at follow-up.

**Emergency department use**

Of eight relevant trials\cite{19,22,23,25,39,41,43,46}, four were included in a meta-analysis\cite{19,23,39,41}. No difference between SMAs and usual care was found for admissions to emergency departments at follow-up ($d=-0.083$, 95\%CI = -0.30, 0.13, $k=4$, $p=0.453$ (Figure 3b). Considerable heterogeneity was observed ($I^2 = 61.7\%$).

Of four trials not in the meta-analysis \cite{22,25,43,46}, only Edelman et al. (2010) reported significant group differences in emergency department use favouring SMAs with 0.4 (95\%CI = 0.20, 0.70) fewer emergency care visits than the usual care group over the 12-month study period \cite{46}.

**Primary care visits**

Four trials reported the number of primary care visits participants made during the study period\cite{19,39,43,46}. Three were pooled in a meta-analysis\cite{19,39,43} showing no statistically significant difference ($d=0.034$, 95\%CI = -0.09, 0.16, $k=3$, $p=0.575$) (see Figure 3c).

Edelman et al. (2010), which could not be included into the meta-analysis, reported that SMA participants had significantly fewer primary care visits than controls (5.3 vs. 6.2 per patient-year) at 12 months\cite{46}.

No statistically significant effect of SMAs compared to usual care were found for other behavioural outcomes including medication adherence and physical activity (see S8)

<Insert Figure 3a,b,c here>

**Cost and resource use**

Few studies reported the costs involved in the delivery of the SMA and those that did were unclear about cost parameters (i.e whether scheduling and preparation time was included or not) or how the costs were attributed. One trial of diabetes SMAs reported that overall costs
per patient were higher in SMAs than those in usual care group for the study period of six months[25]. However, another trial found no significant difference between SMA and usual care in terms of total costs incurred during the 24 months study period, but showed a positive effect of the SMA at 13 months post-study where cost decreased by 6% for the SMA but increased by 13% for usual care p<0.01[39]. The SMA trial for osteoporosis reported that the costs incurred during the study period were lower for the SMA group compared to control groups[52]. A trial of chronic condition SMAs (reported that total costs incurred by the SMA group were lower than the usual care group[19].

**SMA components and mode of delivery**

There was much variation in the SMAs models reported by studies (see S9 for detailed description). Key features of SMA models were: the opportunity for each patient to have individual 1:1 time with a clinician (23 trials), facilitated group discussion or group question and answer session (15 trials)[19,22,25,26,29,30,32–34,37–39,40,43,46,50], ‘group education’ (14 trials)[19,23,25,31–33,37,38,41,42,43,44,46,50], and the opportunity to socialise (11 trials)[19,22,25,26,29,32,34,37,38,46,50].

SMAs were delivered face-to-face in all trials, although three SMA models included digital technologies, namely website and telephone support[22] and phone calls or text message support and/or reminders[37,38]. In three trials[33,41,52] SMAs were delivered by a single healthcare professional though mostly they were delivered by multidisciplinary teams. Professionals most commonly involved were family physicians[19,23,24,30–34,39,44,46,50,51,53], nurse practitioners[23,26–29,32,44,45] or nurses[19,20,39,40,42,46–50,53]. It was not always possible to tell what role each member of staff had in the delivery of the SMA. Provider characteristics other than profession or role were rarely reported, though two trials involving a majority Hispanic/Latino participants reported that the community health worker and or physician were bilingual[34,38]. Four trials reported that the same interventionists attended all visits for a particular group[40,44,46,51]. One trial of diabetes SMAs reported that group assignments were maintained for all SMAs to facilitate peer interactions and relationships within groups[33]. The consistency in group composition in terms of patient and interventionists attending each session was not reported by most studies.

**DISCUSSION**

Our systematic review identified 23 unique RCTs comparing SMAs for one or more LTCs to usual care or an enhanced SMA. We found that SMAs significantly improved diastolic blood pressure for diabetes patients and there were non-significant trends favouring SMAs for
QoL, patient satisfaction and patient self-efficacy. Improved self-efficacy may help boost self-care. In line with the findings of previous reviews[12], no harm was observed for the use of SMAs across these outcomes and there was no effect on healthcare service use compared to usual care. This indicates that whilst SMAs may not be superior to usual care in terms of most health outcomes or reducing demand on services, they do not appear to increase demand at least in the short-term. Evidence reporting costs is too heterogeneous to draw firm conclusions.

Comparison with previous literature

Like previous reviews of SMAs for LTCs[4,12,54], more than half of the included RCTs included patients with diabetes, and as such the most commonly reported outcome measure was HbA1C. However, unlike previous reviews[7,55], we did not observe any significant improvements in HbA1c. This may be because previous reviews included trials in secondary care. Our meta-analysis showed that SMA participants had lower DBP compared to patients who received usual care. This effect remained even when a large study[50] with 1204 participants was omitted in the sensitivity analyses. Whilst significant differences were not found for the other biomedical outcomes between the SMA and usual care groups, individual trials reporting a significant difference suggest trends favouring SMA for outcomes including SBP and cholesterol.

Previous systematic reviews have been inconclusive with regards to the impact of SMAs on healthcare utilisation. Edelman et al.’s (2012) review[7] of SMAs for patients with chronic medical conditions in older adults found a lower pattern of healthcare utilisation, whilst Booth et al. (2015)[4] reported a mixed pattern of changes. Our meta-analyses show that SMAs do not differ from usual care in terms of healthcare utilisation. There is no evidence in the present data to suggest that patients compensate for a lack of privacy by returning to primary care again or that they risk hospitalisation because issues are not adequately addressed during the SMA session(s). However, it should be noted that the key source of bias across the included studies was the lack of blinding of participants and personnel. Therefore, possible selection bias may result in recruitment of SMA participants with less concern about sharing their personal/medical information.

In comparison to biomedical outcomes and psychological outcomes, behavioural outcomes were less frequently reported by studies. This echoes the findings of Edelman et al. (2015)[46] which found there to be limited data on key patient-centred outcomes such as patient satisfaction. Behavioural outcomes such as medication adherence are important across many LTCs and are key to understanding how patients are self-managing their
conditions. In line with Kelly et al.’s (2017)[12] recommendation, future studies should report outcome effectiveness measures that are common or comparable across different LTCs such as physical activity, self-efficacy, medication adherence, and quality of life. It would be advantageous to agree a Core Outcome Set (COS), consisting of a standardised group of outcomes, to be reported by all SMA trials. This can help with future evaluations of SMAs through reducing heterogeneity and facilitating meta-analysis and ensuring that outcome measures are relevant to key stakeholders[56].

**Strengths and limitations**

Although previous reviews have explored the effectiveness of SMAs in improving health outcomes, this review provides a focus on primary care which is key to managing LTCs. We found ten additional trials with 1160 participants since the comprehensive work by Booth et al. (2015)[4] indicating a rapidly growing field. We used robust methods whereby our search strategy was developed with input from Information Specialists through an iterative process and key stages of the review (including screening, data extraction, and quality appraisal) were undertaken independently by two reviewers. We included studies regardless of type of LTC so that we were able to summarise all the available evidence on effectiveness of SMAs for LTCs in primary care in one analysis. However, evidence of an effect was determined by $p<.0.5$ in the papers. This assumes that studies were adequately powered, which may not be the case, particularly for some of the secondary outcomes of the included studies.

**Limitations of evidence base and implications for future research**

Our narrative synthesis is limited by what is reported thus important contextual factors may be missing from authors' descriptions of the SMAs. Similar components may be described differently by different authors or, conversely, similar descriptions are used to describe different components. Standardised taxonomies for describing mode of delivery and intervention content would help to identify important behavioural components and key implementation processes that contribute to intervention effectiveness. However, for this to be possible, it is also important that interventionists clearly specify which target behaviours (e.g. to increase physical activity) the SMAs aim to change and also measure fidelity to ensure that the session(s) are delivered as intended to achieve optimum effects[57]. Further, theoretical underpinnings were lacking in the included SMA interventions, making it difficult to identify ‘mechanisms of action’ through which interventions bring about change[58]. Future SMAs interventions should be theory-based and be explicit in reporting its theoretical underpinnings.
Where multiple healthcare professionals are involved in the SMA, their key role and purpose in the SMA were rarely clearly defined. There was also limited reporting on the composition of the SMA groups across some of the included studies (i.e. how patients were selected for recruitment and size). Therefore, it is unclear which groups of patients might benefit from attending the same SMAs together and what implications SMAs may have for intervention-generated health inequalities. One third of the included studies were conducted on US veterans and one third of studies have involved participants from low-income/uninsured population groups. Generalisability of these groups to other healthcare settings in other countries is unclear. Further investigations into SMAs for patients with one or more LTCs is required, including a wider variety of LTCs (such as asthma and chronic obstructive pulmonary disease) and with more diverse population groups. For example, including low-income and disadvantaged groups in other countries, including the UK.

Conclusions
This review is the first to examine the effects of SMAs delivered in primary care on health outcomes and healthcare service use in patients with one or more LTCs. Our review suggests that SMAs are unlikely to result in less favourable outcomes to patients with LTCs compared to usual care. To identify key intervention components that contribute to effectiveness, future studies will benefit from using standardised taxonomies to report intervention content. The use of an evaluation framework, with a core outcome set, is recommended to improve evidence in this field.
REFERENCES


17. Valentine JC, Pigott TD, Rothstein HR. How many studies do you need? A primer on


52 Baqir W, Gray WK, Blair A, et al. Osteoporosis group consultations are as effective as


**Figure legends**

Figure 1: PRISMA diagram of study selection process

Figure 2: Forest plot for a) HbA1C(%), b) diastolic blood pressure, c) systolic blood pressure

Figure 3: Forest plot for a) hospital admissions, b) emergency department use, c) primary care visits
a) Study | Effect (95% CI) | Weight
--- | --- | ---
Berry et al. (2016) | -0.66 (-1.11, -0.21) | 10.53
Clancy et al. (2007a/2007b/2008) | 0.41 (0.12, 0.70) | 13.48
Cole et al. (2013) | -0.25 (-0.74, 0.24) | 9.87
Ee et al. (2020) | 0.34 (-0.81, 1.30) | 4.52
Schillinger et al. (2008/2009) - GMV vs Usual Care | 0.00 (-0.28, 0.28) | 13.71
Taveira et al. (2011) | -0.58 (-1.01, -0.16) | 10.95
Taveira et al. (2014) - GMV vs Usual Care | 0.15 (-0.23, 0.52) | 11.86
Vaughan et al. (2020) | -0.34 (-0.76, 0.08) | 11.09
Wu et al. (2018) | 0.07 (-0.19, 0.33) | 13.98
Overall, DL ($\chi^2 = 70.9\%$, $p = 0.001$) | -0.10 (-0.34, 0.14) | 100.00

NOTE: Weights are from random-effects model

b) Study | Effect (95% CI) | Weight
--- | --- | ---
Berry et al. (2016) | -0.37 (-0.81, 0.07) | 4.31
Cole et al. (2013) | -0.17 (-0.66, 0.31) | 3.54
Ee et al. (2020) | 0.39 (-0.57, 1.36) | 0.91
Gao et al. (2015) | -0.09 (-0.20, 0.03) | 65.88
Schillinger et al. (2008/2009) - GMV vs Usual Care | -0.19 (-0.46, 0.08) | 11.55
Taveira et al. (2014) - GMV vs Usual Care | 0.02 (-0.35, 0.38) | 6.31
Vaughan et al. (2017) | -0.38 (-0.94, 0.18) | 2.69
Vaughan et al. (2020) | -0.33 (-0.75, 0.09) | 4.81
Overall, DL ($\chi^2 = 0.0\%$, $p = 0.624$) | -0.12 (-0.22, -0.03) | 100.00

NOTE: Weights are from random-effects model

c) Study | Effect (95% CI) | Weight
--- | --- | ---
Berry et al. (2016) | -0.13 (-0.57, 0.31) | 4.48
Cole et al. (2013) | 0.13 (-0.36, 0.62) | 3.64
Ee et al. (2020) | 0.41 (-0.55, 1.38) | 0.94
Gao et al. (2015) | -0.03 (-0.15, 0.08) | 53.11
Schillinger et al. (2008/2009) - GMV vs Usual Care | -0.12 (-0.39, 0.15) | 11.50
Taveira et al. (2014) - GMV vs Usual Care | 0.17 (-0.20, 0.53) | 6.38
Vaughan et al. (2017) | 0.15 (-0.41, 0.70) | 2.82
Vaughan et al. (2020) | -0.43 (-0.85, -0.01) | 4.87
Wu et al. (2018) | 0.14 (-0.12, 0.40) | 12.25
Overall, DL ($\chi^2 = 3.8\%$, $p = 0.403$) | -0.02 (-0.11, 0.08) | 100.00

NOTE: Weights are from random-effects model