PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Preparing the prescription: A review of the aim and measurement of social referral programmes
AUTHORS	Rempel, Emily; Wilson, Emma; Durrant, Hannah; Barnett, Julie

VERSION 1 - REVIEW

REVIEWER	Helen Chatterjee University College London Collaborators and I currently have a mansucript in review which provides a systematised review of social prescribing. This is based on a piece of grey lietrature which we published in 2015: Thomson LJ, Camic PM, Chatterjee HJ. Social presc ribing: A review of community referral schemes. London: University College London, 2015. The MS under review here draws heavily on our work and although our review cited at the end, this review only adds moderately to what is already published.
REVIEW RETURNED	23-Feb-2017

GENERAL COMMENTS	This review provides a useful overview of SP and appraises aims of social referral initiatives and explores the measures used to evaluate if the aims have been met. Overall the paper is well written and accessible to a wide audience. In relation to the first aim, Table 1 is useful in providing a summary of aims of SP schemes and the exploration of these aims adds value. The authors usefully identify the challenges in widespread uptake of SP and barriers to impact due to the poor evidence base and rightly identify that this is problematic due to the lack of consideration for the causal pathways from treatment programmes to health outcomes. The investigation of the methods/measures used to evaluate SP schemes is less robust and rather descriptive. The overview of findings is helpful (Table 2) but the discussion regarding why there have been such a variety of methods used to evaluate SP schemes does not tackle the core issue of the multifarious approaches to SP, neither in terms of types of schemes (arts, exercise, books, learning, eco, etc) nor different referral mechanisms, definitions or interpretations of SP currently in use. This section does also not consider the challenges of assessing non-clinical interventions which are inherent in community referral; a more critical assessment in this section would improve the manuscript and afford an opportunity to make recommendations regarding future evaluation approaches.

REVIEWER	Christian Blickem Public Health Institute Liverpool John Moores University UK
REVIEW RETURNED	02-Mar-2017

GENERAL COMMENTS	Social referral programmes have a lot of potential to improve outcomes for people with long-term health problems and address problems associated with health inequalities and social isolation. This is a very well written review about social referral programmes and I think it makes an important contribution to this field as it points out the inconsistencies in the way social referrals schemes are evaluated. I congratulate the authors for conducting an excellent review on this topic.

REVIEWER	Kelly Blockley University of Exeter, UK
REVIEW RETURNED	16-Mar-2017

GENERAL COMMENTS	Thank you for submitting this paper. Considering the current interest in social prescribing, this comes as a timely piece of work in a relevant subject area. I particularly liked the introduction to the subject area using a quote from Walt Disney and the building of the argument for the need for this piece of work is relatively strong. However, I think that the reporting of the methods and results need further clarity and alternative conclusions should be considered. In particular, more attention needs to be paid to the necessity of social referral programmes having vastly different aims and measurement tools. As currently written, social referral programmes seem to be lumped together as one single type of intervention, when in reality they are used for a wide variety of different population groups and conditions and referral can be to numerous different types of activities, which makes use of the same measures unsuitable. Without acknowledging this and presenting the review findings in light of the disparate nature of social referral, the findings have little relevance to the field. However, if the articles found through your searches can be grouped together, for instance to describe the interventions and measures used to address each individual core aim, which you have done in part, then this could provide a useful point of reference for those interested in, for example, improving mental well-being. In my opinion, the following points should be addressed. INTRODUCTION 1) “these types of programmes improve social well-being and, ultimately, physical and mental health.” – How? Much of the social prescribing literature draws out the mechanisms by which the programmes are envisaged to improve physical, mental and social well-being so it would be useful to reference some of these here to build the argument for why this is considered a potentially effective new way of working in the UK. You could also draw on the relevance of the topic area for the NHS e.g. anticipated reduced service use. Also give more detail on the criticisms of the current evidence and

	present a balanced argument. At present, the introduction covers the relevant aspects but lacks depth and explanation, which the above points would add. 2) What are you using as a definition of social referral? You have said in the introduction that the "programmes link health care to opportunities and events provided by third sector organisations," but further detail would be helpful, for example from which aspect of health care (e.g. primary care, secondary/tertiary mental health services, community-provided or private health services, etc.)? A definition from one of the more impactful social prescribing papers could be used (e.g. University of York CRD, 2015). METHODS 3) Why were the reference and citation lists of included grey reports not hand-searched? These may have included other relevant grey sources that you had not identified through Google searching. Please provide justification for your search methods. 4) Inclusion and exclusion criteria need to be stated explicitly to allow the steps to be followed by external researchers, e.g. define or give examples for "health context" and "social context." Who will be making the referral? Are studies part of routine care or academic research programmes, e.g. if a GP passes details onto a researcher who then contacts the patient as part of a RCT, is this included as a referral? 5) "After identification of relevant articles and reports…" – please provide more detail on how these were deemed 'relevant' and the screening processes used. 6) "initial coding framework…" – was this refined after the 'initial' development? If so, how and why? The final framework should be made available to the reader, e.g. as an appendix/supplementary material. 7) "particularly taking note where aims of the empirical research did not fit in to the established framework." – 'taking note' is not informative – what did you actually do with these studies? Did it result in any revisions to the framework? 8) What was the percentage agreement between the two coders? RESULTS 9) What is a "title sort for relevance?" Figure 1 states that title and abstract screening was conducted – did you screen by both title and abstract or just title to find these articles? It is not clear what was done in either this part of the results or the methods. 10) Please give examples of the types of non-empirical articles that were excluded so that the reader can see why they were not relevant. 11) Figure 1 does not follow the usual structure for a PRISMA diagram. The number of records identified through both database and grey are usually put in separate boxes at the top, then down to TiAb, FT, includes. Your Figure seems to be a chronological representation of what you did, with grey articles identified last rather than included as a main search strategy.

	12) Your footnote for Figure 1 explains why the number of articles citation-chased is higher than the number included; however, this is not reported in either the methods or results sections, i.e. how the inclusion criteria were refined. Please add detail about this in the body of the article. 13) In Figure 1, the box for excluded full-text articles has omitted the one article removed for being a book chapter (as specified in the body of the text). 14) N=33 – how many different studies/programmes does this correspond to, i.e. are all 33 for different studies or do some report on the same study? 15) What types of studies do the included articles report on, e.g. RCTs, before-after controlled studies, evaluations, etc.? At present you have only specified the methodology of the study in Table 1 (qualitative, mixed methods, quantitative), rather than the actual study type/design. 16) You have included only patient populations, but can you be more specific about the populations in your included studies, e.g. adults, children, condition (e.g. cancer patients, overweight, etc.)? How many participants are included in the studies? This will give an indication of the size of the programmes and whether the measures they used were suitable to detect changes in the desired outcomes. 17) "Physical well-being, social well-being and optimised service use were less frequently cited with 15, 17 and 20 studies, respectively." – Optimised service use had 18 citations, not 20. Please correct. Also, the phrase 'less frequently' makes it sound as though these were not important or common aims, whereas they still scored highly. It might read better as "…and optimised service use were also frequently cited, with 15, 17 and 20 studies, respectively." 18) Table 2: "**not applicable due to inconclusive reporting on the choice of qualitative measures" – this part of the table is reporting the aims addressed in the studies that used particular measures/methods. Therefore, the aims for these qualitative studies are likely to have been reported in the articles – it is not for judgement by the authors to discount these aims based on the study author's choice of measures used. DISCUSSION 19) The review has looked at both the aims and measures used in social referral programmes. The findings for both of these objectives have been reported in the results separately and it would be nice to bring them both together in the discussion. For example, which measures are commonly used for the studies aiming to improve mental well-being? This could also be combined n the results section and reported in a table. Otherwise it seems like two separate pieces of work that do not talk to each other. 20) "An alternative explanation is that researchers and evaluators alike do not have a definitive understanding of what social referral services can or will do." – I do not agree that it is a lack of understanding. More a lack of a consistently applied definition of social prescribing – several reports have discussed this issue which should be cited here.

	21) “Currently, researchers appear to be using a wide variety of measures in the hope of finding a statistically valid, quantitative proof of effect.” This seems a very harsh statement and links with my main issue of this work, described below. The phrase “quantitative proof of effect” should not be used as even quantitative measures cannot prove an effect, only imply probability of a likely effect. Additionally, many social referral evaluation studies use qualitative methods to measure the programmes’ progress towards their aims and these do not appear to be considered in this review, other than their representation in tables. GENERAL 22) This is a minor comment - numbers <10 should be written out in full, e.g. one, two, three, etc. Numbers 10 and above should be numerical, e.g. 10, 11, 12, etc. 23) The fundamental issue in this piece of work is the lack of attention paid to the necessity of the different aims of social referrals, which consequently makes measurement tools necessarily varied. Social referral programmes are used for a wide variety of patient groups and conditions, and the referral can be to a wide variety of different programmes, so it would not make sense for the measures used to be the same across studies. For example, you cannot compare the aims or measures used in a programme referring overweight patients to an exercise group, with the aims or measures of a programme referring elderly patients experiencing social isolation to a community arts on prescription group. It is therefore necessary for different measures to be used in light of the different aims. I do believe that this could be a useful piece of work and inform the design of future social referral programmes, but using ‘social referral’ as a whole will not work in practice because of the reasons explained above. I would be more inclined to accept this paper if the included studies had been kept in their ‘core aim’ categories and the measures presented in respect to these so that readers wishing to know what measures are used in programmes to e.g. improve social well-being can easily identify them. Although this is partly illustrated in Table 2, it could be clearer and made as the focus of the work. It would also be improved if the type of programme is reported alongside this information, e.g. exercise referral, arts on prescription, etc. to demonstrate the interventions used to address the aims. Additionally, no judgement on the quality of the studies is presented or the time points at which the measures are applied (e.g. baseline plus follow-ups or at just one point in time). It feels like a lot of useful information coming out of the included studies could have been presented but has been omitted.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Helen Chatterjee

Institution and Country: University College London

Competing Interests: Collaborators and I currently have a mansucript in review which provides a systematised review of social prescribing. This is based on a piece of grey lietrature which we published in 2015: Thomson LJ, Camic PM, Chatterjee HJ. Social prescribing: A review of community referral schemes. London: University College London, 2015. The MS under review here draws heavily

on our work and although our review cited at the end, this review only adds moderately to what is already published.

This review provides a useful overview of SP and appraises aims of social referral initiatives and explores the measures used to evaluate if the aims have been met. Overall the paper is well written and accessible to a wide audience.

In relation to the first aim, Table 1 is useful in providing a summary of aims of SP schemes and the exploration of these aims adds value. The authors usefully identify the challenges in widespread uptake of SP and barriers to impact due to the poor evidence base and rightly identify that this is problematic due to the lack of consideration for the causal pathways from treatment programmes to health outcomes.

The investigation of the methods/measures used to evaluate SP schemes is less robust and rather descriptive. The overview of findings is helpful (Table 2) but the discussion regarding why there have been such a variety of methods used to evaluate SP schemes does not tackle the core issue of the multifarious approaches to SP, neither in terms of types of schemes (arts, exercise, books, learning, eco, etc) nor different referral mechanisms, definitions or interpretations of SP currently in use. This section does also not consider the challenges of assessing non-clinical interventions which are inherent in community referral; a more critical assessment in this section would improve the manuscript and afford an opportunity to make recommendations regarding future evaluation approaches.

- **We would like to first thank Reviewer 1 for their helpful comments and critiques. As both reviewers 1 and 3 highlighted our need to discuss the varied nature of social referral programmes we have added reflection on this in both our results and discussion. We have also edited the discussion in general to more fairly reflect the challenges in evaluating non-medical interventions.**

Reviewer: 2

Reviewer Name: Christian Blickem

Institution and Country: Public Health Institute, Liverpool John Moores University, UK

Competing Interests: None declared

Social referral programmes have a lot of potential to improve outcomes for people with long-term health problems and address problems associated with health inequalities and social isolation. This is a very well written review about social referral programmes and I think it makes an important contribution to this field as it points out the inconsistencies in the way social referrals schemes are evaluated.
I congratulate the authors for conducting an excellent review on this topic.

- **We would like to thank Reviewer 2 for their encouraging comments.**

Reviewer: 3

Reviewer Name: Kelly Blockley

Institution and Country: University of Exeter, UK

Competing Interests: None declared

Thank you for submitting this paper. Considering the current interest in social prescribing, this comes as a timely piece of work in a relevant subject area. I particularly liked the introduction to the subject area using a quote from Walt Disney and the building of the argument for the need for this piece of work is relatively strong. However, I think that the reporting of the methods and results need further clarity and alternative conclusions should be considered. In particular, more attention needs to be

paid to the necessity of social referral programmes having vastly different aims and measurement tools. As currently written, social referral programmes seem to be lumped together as one single type of intervention, when in reality they are used for a wide variety of different population groups and conditions and referral can be to numerous different types of activities, which makes use of the same measures unsuitable. Without acknowledging this and presenting the review findings in light of the disparate nature of social referral, the findings have little relevance to the field. However, if the articles found through your searches can be grouped together, for instance to describe the interventions and measures used to address each individual core aim, which you have done in part, then this could provide a useful point of reference for those interested in, for example, improving mental well-being.

- **Thank you to Reviewer 3 for their constructive criticism, we feel that by addressing their concerns the paper reads more clearly. We have added in a small section in the introduction on varied aims of social referral and expanded this in a new discussion section on pages 12 through 14.**

In my opinion, the following points should be addressed.

INTRODUCTION

1) "these types of programmes improve social well-being and, ultimately, physical and mental health." – How? Much of the social prescribing literature draws out the mechanisms by which the programmes are envisaged to improve physical, mental and social well-being so it would be useful to reference some of these here to build the argument for why this is considered a potentially effective new way of working in the UK. You could also draw on the relevance of the topic area for the NHS e.g. anticipated reduced service use. Also give more detail on the criticisms of the current evidence and present a balanced argument. At present, the introduction cover3s the relevant aspects but lacks depth and explanation, which the above points would add.

- **We have addressed your first point about mechanisms by added commentary in the introduction on page 4.**
- **We have addressed your second point about NHS relevancy by adding in the context of NHS prioritisation of joined up health and care services on Page 3.**

2) What are you using as a definition of social referral? You have said in the introduction that the "programmes link health care to opportunities and events provided by third sector organisations," but further detail would be helpful, for example from which aspect of health care (e.g. primary care, secondary/tertiary mental health services, community-provided or private health services, etc.)? A definition from one of the more impactful social prescribing papers could be used (e.g. University of York CRD, 2015).

- **We have discussed a clearer definition of social prescribing, citing the York review as recommended, on page 4.**

METHODS

3) Why were the reference and citation lists of included grey reports not hand-searched? These may have included other relevant grey sources that you had not identified through Google searching. Please provide justification for your search methods.

- **We did not initially search the references of the grey literature, however we have since done so and added in an additional eight references.**

4) Inclusion and exclusion criteria need to be stated explicitly to allow the steps to be followed by external researchers, e.g. define or give examples for "health context" and "social context." Who will be making the referral? Are studies part of routine care or academic research programmes, e.g. if a GP passes details onto a researcher who then contacts the patient as part of a RCT, is this included as a referral?

- **Thank you this comment, upon review we have updated our examples of health and social contexts on page 6. In addition we have created an additional Appendix which highlights more details about individual studies in order to produce a richer discussion section.**

5) "After identification of relevant articles and reports…" – please provide more detail on how these were deemed 'relevant' and the screening processes used.

- **We have highlighted in more detail our inclusion criteria on pages 5 and 6.**

6) "initial coding framework…" – was this refined after the 'initial' development? If so, how and why? The final framework should be made available to the reader, e.g. as an appendix/supplementary material.

- **For clarity, we have added in additional details on coding on page 6. All initial and final coding criteria are included now in Appendix 1 and Appendix 2.**

7) "particularly taking note where aims of the empirical research did not fit in to the established framework." – 'taking note' is not informative – what did you actually do with these studies? Did it result in any revisions to the framework?

- **We have reviewed this phrase and decided to remove it because the second coding did not result in any revisions to the framework. The coding was the same for both initial and final coding.**

8) What was the percentage agreement between the two coders?

- **Due to the qualitative nature of the review, e.g. no study quality assessment, we did not calculate a percentage agreement. We have added in a comment on this on page 6 to ensure transparency.**

RESULTS

9) What is a "title sort for relevance?" Figure 1 states that title and abstract screening was conducted – did you screen by both title and abstract or just title to find these articles? It is not clear what was done in either this part of the results or the methods.

- **We have clarified our search criteria on page 6 and 7.**

10) Please give examples of the types of non-empirical articles that were excluded so that the reader can see why they were not relevant.

- **We have added in an additional example of non-empirical articles on page 6.**

11) Figure 1 does not follow the usual structure for a PRISMA diagram. The number of records identified through both database and grey are usually put in separate boxes at the top, then down to TiAb, FT, includes. Your Figure seems to be a chronological representation of what you did, with grey articles identified last rather than included as a main search strategy.

- **Initially, we as you have suggested presented the PRISMA diagram with a few changes as we had seen other articles in BMJ journals do the same. We have since edited the Prisma diagram to ensure it follows the traditional format.**

12) Your footnote for Figure 1 explains why the number of articles citation-chased is higher than the number included; however, this is not reported in either the methods or results sections, i.e. how the inclusion criteria were refined. Please add detail about this in the body of the article.

- **We have now added in a sentence on this on page 6.**

13) In Figure 1, the box for excluded full-text articles has omitted the one article removed for being a book chapter (as specified in the body of the text).

- **Thank you for this comment, upon review we realised that the text box had hidden that record. We have edited this in the diagram.**

14) N=33 – how many different studies/programmes does this correspond to, i.e. are all 33 for different studies or do some report on the same study?

- **Based on our new supplementary appendices, we have addressed this in the results section, page 7. As well we have added details on programme name in Appendix 1.**

15) What types of studies do the included articles report on, e.g. RCTs, before-after controlled studies, evaluations, etc.? At present you have only specified the methodology of the study in Table 1 (qualitative, mixed methods, quantitative), rather than the actual study type/design.

- **We have added this in to Appendix 1.**

16) You have included only patient populations, but can you be more specific about the populations in your included studies, e.g. adults, children, condition (e.g. cancer patients, overweight, etc.)? How many participants are included in the studies? This will give an indication of the size of the

programmes and whether the measures they used were suitable to detect changes in the desired outcomes.

- **We have also added this in to Appendix 1, as well as a commentary in the results on pages 7 and 8.**

17) "Physical well-being, social well-being and optimised service use were less frequently cited with 15, 17 and 20 studies, respectively." – Optimised service use had 18 citations, not 20. Please correct. Also, the phrase 'less frequently' makes it sound as though these were not important or common aims, whereas they still scored highly. It might read better as "…and optimised service use were also frequently cited, with 15, 17 and 20 studies, respectively."

- **Thank you for that note, we have corrected it in the text. As well we have changed 'less' to 'also' as recommended.**

18) Table 2: "**not applicable due to inconclusive reporting on the choice of qualitative measures" – this part of the table is reporting the aims addressed in the studies that used particular measures/methods. Therefore, the aims for these qualitative studies are likely to have been reported in the articles – it is not for judgement by the authors to discount these aims based on the study author's choice of measures used.

- **Thank you for this comment, upon reflection we realise that our choice of wording was poor. The reason the aim was not stated was because semi-structured interviews that explore patient experience necessitate not asking participants about whether specific aims were met e.g. did your mental health improve? Did your physical health improve? These interviews by their nature did not address a specific study aim. We have reworded this in Table 2.**

DISCUSSION

19) The review has looked at both the aims and measures used in social referral programmes. The findings for both of these objectives have been reported in the results separately and it would be nice to bring them both together in the discussion. For example, which measures are commonly used for the studies aiming to improve mental well-being? This could also be combined n the results section and reported in a table. Otherwise it seems like two separate pieces of work that do not talk to each other.

- **We have rewritten our discussion section to discuss comparisons between aims and measures. This included going in more depth to explore Table 2. We have not included a commentary on what measures are most often used in which aims because the vast majority of studies contained multiple aims. It would therefore be our value judgment on which measures were intended to address each aim. We have kept Table 2 quite broad, e.g. listing the potential aims that each measure addressed, in order to avoid incorrectly assigning an aim to a specific measure.**

20) "An alternative explanation is that researchers and evaluators alike do not have a definitive understanding of what social referral services can or will do." – I do not agree that it is a lack of understanding. More a lack of a consistently applied definition of social prescribing – several reports have discussed this issue which should be cited here.

- **We have edited this section and added in a commentary on the varied kinds of social referral. As well we have cited additional literature that discusses the reviews social prescribing on page 14.**

21) "Currently, researchers appear to be using a wide variety of measures in the hope of finding a statistically valid, quantitative proof of effect." This seems a very harsh statement and links with my main issue of this work, described below. The phrase "quantitative proof of effect" should not be used as even quantitative measures cannot prove an effect, only imply probability of a likely effect. Additionally, many social referral evaluation studies use qualitative methods to measure the programmes' progress towards their aims and these do not appear to be considered in this review, other than their representation in tables.

- **We have re-written this commentary to avoid strong language and have shifted away from discussing quantitative measures. Instead we discuss the proliferation of measures in general.**

GENERAL

22) This is a minor comment - numbers <10 should be written out in full, e.g. one, two, three, etc. Numbers 10 and above should be numerical, e.g. 10, 11, 12, etc.

- **We have edited the paper to align with this standard formatting.**

23) The fundamental issue in this piece of work is the lack of attention paid to the necessity of the different aims of social referrals, which consequently makes measurement tools necessarily varied. Social referral programmes are used for a wide variety of patient groups and conditions, and the referral can be to a wide variety of different programmes, so it would not make sense for the measures used to be the same across studies. For example, you cannot compare the aims or measures used in a programme referring overweight patients to an exercise group, with the aims or measures of a programme referring elderly patients experiencing social isolation to a community arts on prescription group. It is therefore necessary for different measures to be used in light of the different aims. I do believe that this could be a useful piece of work and inform the design of future social referral programmes, but using 'social referral' as a whole will not work in practice because of the reasons explained above. I would be more inclined to accept this paper if the included studies had been kept in their 'core aim' categories and the measures presented in respect to these so that readers wishing to know what measures are used in programmes to e.g. improve social well-being can easily identify them. Although this is partly illustrated in Table 2, it could be clearer and made as the focus of the work. It would also be improved if the type of programme is reported alongside this information, e.g. exercise referral, arts on prescription, etc. to demonstrate the interventions used to address the aims. Additionally, no judgement on the quality of the studies is presented or the time points at which the measures are applied (e.g. baseline plus follow-ups or at just one point in time). It feels like a lot of useful information coming out of the included studies could have been presented but has been omitted.

- **Thank you for this useful commentary, we have thus added in a discussion of the varied structure of different kinds of social prescribing in the introduction and discussion. We have not added in the suggested core aim table due to the discussion above, e.g. that the majority of studies had multiple aims and it would likely lead to errors if we made a judgment about which study instrument addressed which aims.**
- **We have also added in the type of programme and study design, among other columns, in the new Appendix 1.**
- **We have not added in a judgement on quality of studies in interest of brevity and due to the qualitative nature of the review as discussed above. Although we agree that study quality is a key determinant in systematically reviewing the impact of interventions, we feel that our focus on aims and measures, and not results, does not necessitate study quality exploration.**
- **To specify that our review follows systematic review protocols but does not focus on study outcome we have removed 'systematic' from the title of our article.**

Additional Comments from the Editorial Team:

The manuscript currently reads more like a report than a scientific paper – better use of side headings in the Discussion would help and make it read less like a campaign.

- **We have edited the headings in the discussion to be less suggestive, as well we hope the other edits outlined above allow for a more scientific voice.**

The reporting of your methods needs improving. Were databases searched from inception? Was the search restricted to articles written in English? What data was extracted? You need to be clearer about your inclusion criteria. Importantly how did you assess study quality? What was your planned analysis? Supplementary Table 1 would be easier to read with some horizontal lines. Your Discussion in particular needs some more thought.

- **We have edited the methods section to outline exact data extracted and inclusion criteria. We have also added a commentary on why we did not assess study quality. We also added horizontal lines to our Supplementary appendices. We hope that the edits as outlined above address concerns around the discussion.**

VERSION 2 - REVIEW

REVIEWER	Kelly Blockley University of Exeter. UK
REVIEW RETURNED	02-Jun-2017

GENERAL COMMENTS	INTRODUCTION This is much improved from the first submission and the more detailed explanation of social prescribing, its relevance to the UK context and previous research provides a clear introduction to the area for the reader and sets the scene for your research well. Addressing the following two minor points may improve this further: 1) Page 4: "It is theorised that these types of programmes…." – Please reference this, i.e. theorised by who? 2) You state in your response to reviewers that you have addressed the comment about mechanisms. However, this is still not clear i.e. why or how do group and community activities improve social wellbeing and physical and mental health? For example, depression may be linked to social isolation so enabling people to attend a community group-based activity provides them with the opportunity to meet other people and build social networks, potentially reducing their feelings of isolation and subsequently improving their mental health. So the actual mechanisms by which these programmes are expected to work would be helpful. LITERATURE SEARCH METHODOLOGY 3) "We identified PubMed suggested terms associated with SP" – Before its use SP should be written in brackets after your first use of 'social prescribing'. However, this is the only time you have used it in the paper so I would suggest writing it in full for consistency. 4) Why did you search on 'social prescribing services'? Surely these results would have been picked up in your search for 'social prescribing' as they contain exactly the same phrase. 5) You state in the introduction that you are using the term 'social referral', yet you have not used this as one of your search terms – why? 6) The search terms used are very specific and limited, e.g., no 'social referral' and I do not understand why 'community referring physicians' was a term because 'physician' is more of a US term, which means that UK studies looking at similar things may be missed as they would likely use doctor or health professional instead. You state that you "identified PubMed suggested terms" but how were the final terms decided on, e.g. did you consult an expert advisory group containing professionals working in the social prescribing field (as is common practice for systematic reviews), or other peers or previous publications? Is this the full search strategy as used in every database or e.g. did you use functions such as prescri$ and prescri* to pick up all articles using prescription, prescribing, prescribe, prescribed, etc.? If not, why not? I am not convinced that this search strategy is systematic or picked up all relevant articles (particularly as 603 is rather small) so you might consider re-labelling the methodology as a literature review, rather than a systematic review. 7) Were any date restrictions put on the searches? 8) Page 7 "developed the initial coding framework" – if the framework was not changed after the second coding then remove

	the word 'initial'. 9) You have detailed your search and coding processes, but not the screening of the articles. I.e. were the titles and abstracts of all 645 articles screened by both ER and EW, or just the final includes coded by both? RESULTS 10) Add in to the first line of the results the number of articles retrieved through grey searches and that the total number of articles retrieved was 645 – at present this number is only presented in Figure 1 and not in the text, so on a quick read it looks like the total number was only 603. 11) Page 11 "there were around 33…" – be exact. If there were 33 different measures then remove the word 'around'. DISCUSSION 12) Page 13 "…how they measured and evaluated that 'linking up'" – Looking at the measures used in the studies, it seems likely that they were used to measure patient outcomes, rather than the linking up i.e. how to get from the health professional to the activity. This sentence therefore contradicts what you have actually reviewed. 13) Page 14 "diversity of measures evident in social referral initiatives, often associated with a series of vaguely similar aims, suggests that what programmes are aiming to do is often unclear." – This goes back to my original critique – many individual programmes are likely to be clear about their aims but due to the heterogeneity in terms of populations, activities, design etc., they are not all aiming to do the same thing. This is not the same as saying that what they are aiming to do is unclear. 14) Page 14 "it is less clear how GP visits would relate to physical wellbeing" – Was this not stated in the studies that used administrative counts, e.g. there is less need to go to the GP if physical wellbeing is improved? If not then there should be some literature to support this assumption. (After a very quick Google Scholar search I found a couple of papers that would support it). 15) Page 14 "…researchers and evaluators do not have a definitive understanding of what exactly the exact aim of their social referral service is" – based on your findings, it seems likely that people delivering and evaluating interventions know what their aims are but are less clear on the best ways of measuring progress towards them. 16) Page 15 "…not possible while aims and measures are so inconsistent" – Again it is not possible for programmes to have the same aims and use the same measures due to the different populations, reasons for referral, activities, etc. PRISMA CHECKLIST 17) If you are going to put "Reported on page #" then put the actual page number, not a range. APPENDIX 1 18) In the study design column, for some articles you have included measurement time points and in others you have not. It might be worth adding a separate column for this because the frequency of data collection time points gives some idea about the suitability of the study or evaluation design and the measures that they have used to determine the impact. APPENDIX 2 19) Remove the columns for peer reviewed and study type as these

	are in the previous table, so no need for repetition. Also, the table is quite difficult to read with such narrow columns so this may give you more space. 20) I do not know what the column for 'individual & system aims?' is for. Is this in addition to the individual and system level aims from the core aim column?

VERSION 2 – AUTHOR RESPONSE

Reviewer Name: Kelly Blockley

Institution and Country: University of Exeter, UK

Competing Interests: None declared

Thank you for submitting this paper. Considering the current interest in social prescribing, this comes as a timely piece of work in a relevant subject area. I particularly liked the introduction to the subject area using a quote from Walt Disney and the building of the argument for the need for this piece of work is relatively strong. However, I think that the reporting of the methods and results need further clarity and alternative conclusions should be considered. In particular, more attention needs to be paid to the necessity of social referral programmes having vastly different aims and measurement tools. As currently written, social referral programmes seem to be lumped together as one single type of intervention, when in reality they are used for a wide variety of different population groups and conditions and referral can be to numerous different types of activities, which makes use of the same measures unsuitable. Without acknowledging this and presenting the review findings in light of the disparate nature of social referral, the findings have little relevance to the field.

However, if the articles found through your searches can be grouped together, for instance to describe the interventions and measures used to address each individual core aim, which you have done in part, then this could provide a useful point of reference for those interested in, for example, improving mental well-being.

- Thank you to Reviewer 3 for their constructive criticism, we feel that by addressing their concerns the paper reads more clearly. We have added in a small section in the introduction on varied aims of social referral and expanded this in a new discussion section on pages 12 through 14.

In my opinion, the following points should be addressed.

INTRODUCTION

1) “these types of programmes improve social well-being and, ultimately, physical and mental health.” – How? Much of the social prescribing literature draws out the mechanisms by which the programmes are envisaged to improve physical, mental and social well-being so it would be useful to reference some of these here to build the argument for why this is considered a potentially effective new way of working in the UK. You could also draw on the relevance of the topic area for the NHS e.g. anticipated reduced service use. Also give more detail on the criticisms of the current evidence and present a balanced argument. At present, the introduction cover3s the relevant aspects but lacks depth and explanation, which the above points would add.

- We have addressed your first point about mechanisms by added commentary in the introduction on page 4.

- We have addressed your second point about NHS relevancy by adding in the context of NHS prioritisation of joined up health and care services on Page 3.

2) What are you using as a definition of social referral? You have said in the introduction that the “programmes link health care to opportunities and events provided by third sector organisations,” but further detail would be helpful, for example from which aspect of health care (e.g. primary care, secondary/tertiary mental health services, community-provided or private health services, etc.)? A definition from one of the more impactful social prescribing papers could be used (e.g. University of York CRD, 2015).

- We have discussed a clearer definition of social prescribing, citing the York review as recommended, on page 4.

METHODS

3) Why were the reference and citation lists of included grey reports not hand-searched? These may have included other relevant grey sources that you had not identified through Google searching. Please provide justification for your search methods.

- We did not initially search the references of the grey literature, however we have since done so and added in an additional eight references.

4) Inclusion and exclusion criteria need to be stated explicitly to allow the steps to be followed by external researchers, e.g. define or give examples for “health context” and “social context.” Who will be making the referral? Are studies part of routine care or academic research programmes, e.g. if a GP passes details onto a researcher who then contacts the patient as part of a RCT, is this included as a referral?

- Thank you this comment, upon review we have updated our examples of health and social contexts on page 6. In addition we have created an additional Appendix which highlights more details about individual studies in order to produce a richer discussion section.

5) “After identification of relevant articles and reports…” – please provide more detail on how these were deemed ‘relevant’ and the screening processes used.

- We have highlighted in more detail our inclusion criteria on pages 5 and 6.

6) “initial coding framework…” – was this refined after the ‘initial’ development? If so, how and why? The final framework should be made available to the reader, e.g. as an appendix/supplementary material.

- For clarity, we have added in additional details on coding on page 6. All initial and final coding criteria are included now in Appendix 1 and Appendix 2.

7) “particularly taking note where aims of the empirical research did not fit in to the established framework.” – ‘taking note’ is not informative – what did you actually do with these studies? Did it result in any revisions to the framework?

- We have reviewed this phrase and decided to remove it because the second coding did not result in any revisions to the framework. The coding was the same for both initial and final coding.

8) What was the percentage agreement between the two coders?

- Due to the qualitative nature of the review, e.g. no study quality assessment, we did not calculate a percentage agreement. We have added in a comment on this on page 6 to ensure transparency.

RESULTS

9) What is a “title sort for relevance?” Figure 1 states that title and abstract screening was conducted – did you screen by both title and abstract or just title to find these articles? It is not clear what was done in either this part of the results or the methods.

- We have clarified our search criteria on page 6 and 7.

10) Please give examples of the types of non-empirical articles that were excluded so that the reader can see why they were not relevant.

- We have added in an additional example of non-empirical articles on page 6.

11) Figure 1 does not follow the usual structure for a PRISMA diagram. The number of records identified through both database and grey are usually put in separate boxes at the top, then down to TiAb, FT, includes. Your Figure seems to be a chronological representation of what you did, with grey articles identified last rather than included as a main search strategy.

- Initially, we as you have suggested presented the PRISMA diagram with a few changes as we had seen other articles in BMJ journals do the same. We have since edited the Prisma diagram to ensure it follows the traditional format.

12) Your footnote for Figure 1 explains why the number of articles citation-chased is higher than the number included; however, this is not reported in either the methods or results sections, i.e. how the inclusion criteria were refined. Please add detail about this in the body of the article.

- We have now added in a sentence on this on page 6.

13) In Figure 1, the box for excluded full-text articles has omitted the one article removed for being a book chapter (as specified in the body of the text).

- Thank you for this comment, upon review we realised that the text box had hidden that record. We have edited this in the diagram.

14) N=33 – how many different studies/programmes does this correspond to, i.e. are all 33 for different studies or do some report on the same study?

- Based on our new supplementary appendices, we have addressed this in the results section, page 7. As well we have added details on programme name in Appendix 1.

15) What types of studies do the included articles report on, e.g. RCTs, before-after controlled studies, evaluations, etc.? At present you have only specified the methodology of the study in Table 1 (qualitative, mixed methods, quantitative), rather than the actual study type/design.

- We have added this in to Appendix 1.

16) You have included only patient populations, but can you be more specific about the populations in your included studies, e.g. adults, children, condition (e.g. cancer patients, overweight, etc.)? How many participants are included in the studies? This will give an indication of the size of the programmes and whether the measures they used were suitable to detect changes in the desired outcomes.

- We have also added this in to Appendix 1, as well as a commentary in the results on pages 7 and 8.

17) “Physical well-being, social well-being and optimised service use were less frequently cited with 15, 17 and 20 studies, respectively.” – Optimised service use had 18 citations, not 20. Please correct. Also, the phrase ‘less frequently’ makes it sound as though these were not important or common aims, whereas they still scored highly. It might read better as “…and optimised service use were also frequently cited, with 15, 17 and 20 studies, respectively.”

- Thank you for that note, we have corrected it in the text. As well we have changed ‘less’ to ‘also’ as recommended.

18) Table 2: “**not applicable due to inconclusive reporting on the choice of qualitative measures” – this part of the table is reporting the aims addressed in the studies that used particular measures/methods. Therefore, the aims for these qualitative studies are likely to have been reported

in the articles – it is not for judgement by the authors to discount these aims based on the study author's choice of measures used.

• Thank you for this comment, upon reflection we realise that our choice of wording was poor. The reason the aim was not stated was because semi-structured interviews that explore patient experience necessitate not asking participants about whether specific aims were met e.g. did your mental health improve? Did your physical health improve? These interviews by their nature did not address a specific study aim. We have reworded this in Table 2.

DISCUSSION

19) The review has looked at both the aims and measures used in social referral programmes. The findings for both of these objectives have been reported in the results separately and it would be nice to bring them both together in the discussion. For example, which measures are commonly used for the studies aiming to improve mental well-being? This could also be combined n the results section and reported in a table. Otherwise it seems like two separate pieces of work that do not talk to each other.

• We have rewritten our discussion section to discuss comparisons between aims and measures. This included going in more depth to explore Table 2. We have not included a commentary on what measures are most often used in which aims because the vast majority of studies contained multiple aims. It would therefore be our value judgment on which measures were intended to address each aim. We have kept Table 2 quite broad, e.g. listing the potential aims that each measure addressed, in order to avoid incorrectly assigning an aim to a specific measure.

20) "An alternative explanation is that researchers and evaluators alike do not have a definitive understanding of what social referral services can or will do." – I do not agree that it is a lack of understanding. More a lack of a consistently applied definition of social prescribing – several reports have discussed this issue which should be cited here.

• We have edited this section and added in a commentary on the varied kinds of social referral. As well we have cited additional literature that discusses the reviews social prescribing on page 14.

21) "Currently, researchers appear to be using a wide variety of measures in the hope of finding a statistically valid, quantitative proof of effect." This seems a very harsh statement and links with my main issue of this work, described below. The phrase "quantitative proof of effect" should not be used as even quantitative measures cannot prove an effect, only imply probability of a likely effect. Additionally, many social referral evaluation studies use qualitative methods to measure the programmes' progress towards their aims and these do not appear to be considered in this review, other than their representation in tables.

• We have re-written this commentary to avoid strong language and have shifted away from discussing quantitative measures. Instead we discuss the proliferation of measures in general.

GENERAL

22) This is a minor comment - numbers <10 should be written out in full, e.g. one, two, three, etc. Numbers 10 and above should be numerical, e.g. 10, 11, 12, etc.

- We have edited the paper to align with this standard formatting.

23) The fundamental issue in this piece of work is the lack of attention paid to the necessity of the different aims of social referrals, which consequently makes measurement tools necessarily varied. Social referral programmes are used for a wide variety of patient groups and conditions, and the referral can be to a wide variety of different programmes, so it would not make sense for the measures used to be the same across studies. For example, you cannot compare the aims or measures used in a programme referring overweight patients to an exercise group, with the aims or measures of a programme referring elderly patients experiencing social isolation to a community arts on prescription group. It is therefore necessary for different measures to be used in light of the different aims. I do believe that this could be a useful piece of work and inform the design of future social referral programmes, but using 'social referral' as a whole will not work in practice because of the reasons explained above. I would be more inclined to accept this paper if the included studies had been kept in their 'core aim' categories and the measures presented in respect to these so that readers wishing to know what measures are used in programmes to e.g. improve social well-being can easily identify them. Although this is partly illustrated in Table 2, it could be clearer and made as the focus of the work. It would also be improved if the type of programme is reported alongside this information, e.g. exercise referral, arts on prescription, etc. to demonstrate the interventions used to address the aims. Additionally, no judgement on the quality of the studies is presented or the time points at which the measures are applied (e.g. baseline plus follow-ups or at just one point in time). It feels like a lot of useful information coming out of the included studies could have been presented but has been omitted.

- Thank you for this useful commentary, we have thus added in a discussion of the varied structure of different kinds of social prescribing in the introduction and discussion. We have not added in the suggested core aim table due to the discussion above, e.g. that the majority of studies had multiple aims and it would likely lead to errors if we made a judgment about which study instrument addressed which aims.

- We have also added in the type of programme and study design, among other columns, in the new Appendix 1.

- We have not added in a judgement on quality of studies in interest of brevity and due to the qualitative nature of the review as discussed above. Although we agree that study quality is a key determinant in systematically reviewing the impact of interventions, we feel that our focus on aims and measures, and not results, does not necessitate study quality exploration.

- To specify that our review follows systematic review protocols but does not focus on study outcome we have removed 'systematic' from the title of our article.

Additional Comments from the Editorial Team:

The manuscript currently reads more like a report than a scientific paper – better use of side headings in the Discussion would help and make it read less like a campaign.

- We have edited the headings in the discussion to be less suggestive, as well we hope the other edits outlined above allow for a more scientific voice.

The reporting of your methods needs improving. Were databases searched from inception? Was the search restricted to articles written in English? What data was extracted? You need to be clearer about your inclusion criteria. Importantly how did you assess study quality? What was your planned analysis? Supplementary Table 1 would be easier to read with some horizontal lines. Your Discussion in particular needs some more thought.

- We have edited the methods section to outline exact data extracted and inclusion criteria. We have also added a commentary on why we did not assess study quality. We also added horizontal lines to our Supplementary appendices. We hope that the edits as outlined above address concerns around the discussion.