PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cost-effectiveness analysis of a Community Paramedicine Program for low-income seniors living in subsidized housing: The Community Paramedicine at Clinic Program (CP@clinic)
AUTHORS	Agarwal, Gina; Pirrie, Melissa; Angeles, Ricardo; Marzanek, Francine; Thabane, Lehana; O'Reilly, Daria

VERSION 1 – REVIEW

REVIEWER	Inna Feldman Uppsala University
REVIEW RETURNED	23-Feb-2020

GENERAL COMMENTS	This is a very interesting work aims to examine cost-effectiveness of a Community Paramedicine Program for seniors. The topic is very important since all health care providers for senior people do many efforts to optimize health care services in order to improve health outcomes and decrease societal costs. However, some important methodological issues should be addressed. 1. My main concern is the treatment of uncertainty while presenting cost-utility results (ICER): a. It looks like the authors did not use bootstrap analyses putting together costs and outcomes (QALYs). b. The results should be presented on cost-effectiveness plane or using CEA – curve, to show the probability of cost-effectiveness depending on wiliness to pay. c. While including costs for EMS calls in sensitivity analysis, the results should be also presented as a bootstrapping, with a probability to be cost-effective/cost saving. d. My guess is that the authors do not have access to the individual data for EMS callas and that is why they use the estimates. If they have the data on individual level, it is necessary to present the distribution. Other remarks 1) In the abstract: I am not agree that the ICER is outcome measurer. The outcome measure for the intervention are QALYs. ICER is a result of cost-utility analysis It is possible to divide this section into “outcomes” and “methods”

	2) Introduction, line 93-95: “County, Ontario, performed an economic evaluation of a home visit program model (Aging at Home) and was able to demonstrate an incremental cost per quality-adjusted life year(QALY)” – what was the ICER? Difficult to follow. 3) Method “Main Trial Results” are used in CUA later and both QALYs and EMS fit better to the Results – section. While presenting, it is necessary to indicate missing data (%) Probabilistic sensitivity analysis should be included. 4) Results CEA – curve alt CE plane have to be presented 5) Discussion The limited cost-perspective has to be acknowledge.

REVIEWER	Hildegard Seidl Helmholtz Zentrum München Germany
REVIEW RETURNED	27-Feb-2020

GENERAL COMMENTS	Cost-effectiveness analysis of a Community Paramedicine Program for low-income seniors living in subsidized housing: The Community Paramedicine at Clinic Program (CP@clinic) Summary This paper evaluates a Community Paramedicine Program for low-income seniors living in subsidized housing from the cost-effectiveness perspective. The authors economically evaluate a program that is intended to reduce EMS calls from low-income seniors living in social housing. Within the intervention, risk-assessment and health promotion sessions were delivered. The cost-effectiveness analysis compares QALY, ICER, and net program costs. The authors find that the intervention results in better health-related quality of life and reduced EMS calls. Overall Assessment The paper tackles an important research topic: the cost-effectiveness of an intervention program that provides support for vulnerable target group. This topic is important in light of demographic change and an aging population. The effectiveness as well as the economic evaluation of paramedicine programs has been limited in the literature. The paper claims to fill this gap with the help of data from the CP@clinic randomized trial. I appreciate this intention and I think there is a lack of RCTs in this field as well as economic evaluation of paramedicine programs. But several methodological aspects have to be clarified before a decision on acceptance for publication can be made. I recommend a further revision of the manuscript after addressing major comments. Major Comments 1. The main objective of this paper was to conduct a cost-utility analysis. Unfortunately, it is unclear how QALYs are computed. The authors do not describe how the health states of EQ-5D-3L are converted into utility scores and which scoring algorithm was used.

	Furthermore, I think, no QALYs were estimated (area under the curve) but rather differences in index scores. In case of this, it is not a cost-utility analysis. 2. The authors stated that significant reductions in EMS calls were the main trail results (difference of -0.9 calls) after adjustment for the study design and baseline calls. The published main article (Reducing 9-1-1 emergency medical service calls by implementing a community paramedicine program for vulnerable older adults in public housing……) only found in the sensitivity analysis significantly reduced EMS calls but not in the intention to treat analysis. This should be mentioned. 3. Baseline characteristics of residents should be provided and a detailed flow chart to show the amount of buildings, apartment units, and residents for intervention and control group as well as drop-outs. 4. There are two different numbers of residents: 356 residents in the intervention group (table 2) and in line 171 n=595. How many residents were in the intervention group and what is the reason for this difference? This is not described (line 170: When restricted to those who attended the program n=595?) and should be presented in the flow chart. 5. Please give more information about the intervention and inclusion/exclusion criteria because readers do not want to search for former publication (also see comment 6). 6. The paramedic service perspective is not appropriate because the economic evaluation should inform decision maker and therefore include associated costs. In the study protocol, the authors stated, “Using a prespecified algorithm, participants will be directed to appropriate services. Those identified as high risk will be immediately referred to appropriate healthcare resources, such as their primary healthcare provider” and in the main publication “Where urgent medical assistance was indicated, paramedics facilitated immediate connection with the family physician, urgent care, or ED.” Therefore, the intervention is certainly associated with subsequent resource use e.g. outpatient ore inpatient treatment. Furthermore, in the study protocol cost measurement are described as follow: “Rates of EMS calls, ED visits, primary care visits and other healthcare utilisation will be collected from the administrative database of the local paramedic services, hospital ED databases, Institute for Clinical Evaluative Sciences (ICES) databases, the CHAP-EMS database and consenting participants’ primary care charts. Administrative data will be collected preintervention and postintervention, and retrospectively for the 12 months of intervention.” For this reason, please conduct the cost analysis including associated costs. 7. The intervention program were offered in the building common areas. Please give information about subsidized housing building and its equipment features – are common areas standard? 8. Please give more information about building matching. You write that buildings were matched by socio-demographics. I think you mean by social-demographic characteristics from residents – but which social-demographics variables were taken? 9. What is modified stuff? Please give a short explanation. 10. Was the intervention equal in all sites? You say that each site had different staffing arrangements. Are there differences in qualification? 11. Table 1 shows monthly EMS calls per 100 units at baseline and after one year. The means appear low in contrast to table 1 in the main paper. 12. The analysis was conducted at various levels: EMS reduction and program costs at building level and health related quality of life

	at resident level. Then a synthesis was made. However, the assumptions were different. At building level all residents were included but at resident level only residents 55 years and older were included. Change in health related quality of life in residents 55 years and older was assumed to can be generalized. Furthermore, only one quarter of residents (358 of 1461) participated in intervention (participation bias). Is there any information about non-participants? In order to reflect real world, program costs should be allocated to participants only. From which kind of residents (elderly residents?) EMS calls were made? You can conduct an analysis at individual level only. You need program costs per participant, EMS calls per participant, other health care resource use per participant, and health related quality of life. Then an appropriate analyses including uncertainty (cost-effectiveness plane according to bootstrapping) is possible. 13. Please give information about drop outs (deceased?) and its handling in the imputation process.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

However, some important methodological issues should be addressed.

1. My main concern is the treatment of uncertainty while presenting cost-utility results (ICER):

a. It looks like the authors did not use bootstrap analyses putting together costs and outcomes (QALYs).

b. The results should be presented on cost-effectiveness plane or using CEA – curve, to show the probability of cost-effectiveness depending on wiliness to pay.

c. While including costs for EMS calls in sensitivity analysis, the results should be also presented as a bootstrapping, with a probability to be cost-effective/cost saving.

d. My guess is that the authors do not have access to the individual data for EMS callas and that is why they use the estimates. If they have the data on individual level, it is necessary to present the distribution.

Thank you for your helpful comments. You are correct in that we do not have individual data on EMS calls, only at a building-level. However, we have completed a bootstrapping analysis to account for the uncertainty for the main QALYs, and now present a cost effectiveness acceptability curve. We have bootstrapped our complete case adjusted data, as bootstrapping incomplete case data would have required including results from multiple imputation which could result in statistically invalid results.*

We have amended the manuscript in the Methods, Results, Table 2 and Discussion sections to reflect and also added Figures 1 and 2.

**Brand J, van Buuren S, le Cessie S, van den Hout W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Statistics in Medicine. 2019;38:210–220. https://doi.org/10.1002/sim.7956*

Other remarks from Reviewer 1

1) In the abstract:

I am not agree that the ICER is outcome measurer. The outcome measure for the intervention are QALYs. ICER is a result of cost-utility analysis It is possible to divide this section into "outcomes" and "methods"

We have altered the abstract to reflect this, and moved the description of the QALY as the main outcome to be at the start of the outcome section and moved the ICER statement to the end. We have also reflected this in the main paper and have moved information about the QALY outcome from the methods section to the results section.

2) Introduction, line 93-95:

"County, Ontario, performed an economic evaluation of a home visit program model (Aging at Home) and was able to demonstrate an incremental cost per quality-adjusted life year(QALY)" – what was the ICER? Difficult to follow.

The purpose of this comment is not to discuss the values of the ICERs but to describe the fact that this method is infrequently used in community paramedicine. We do describe the ICER value for this paper in the discussion:

The ICER of a home visit program in Renfrew County, Ontario has been described to be between $67,000 and $76,000 [8] compared to the CP@clinic ICER of $2,200

3) Method

"Main Trial Results" are used in CUA later and both QALYs and EMS fit better to the Results – section. While presenting, it is necessary to indicate missing data (%)

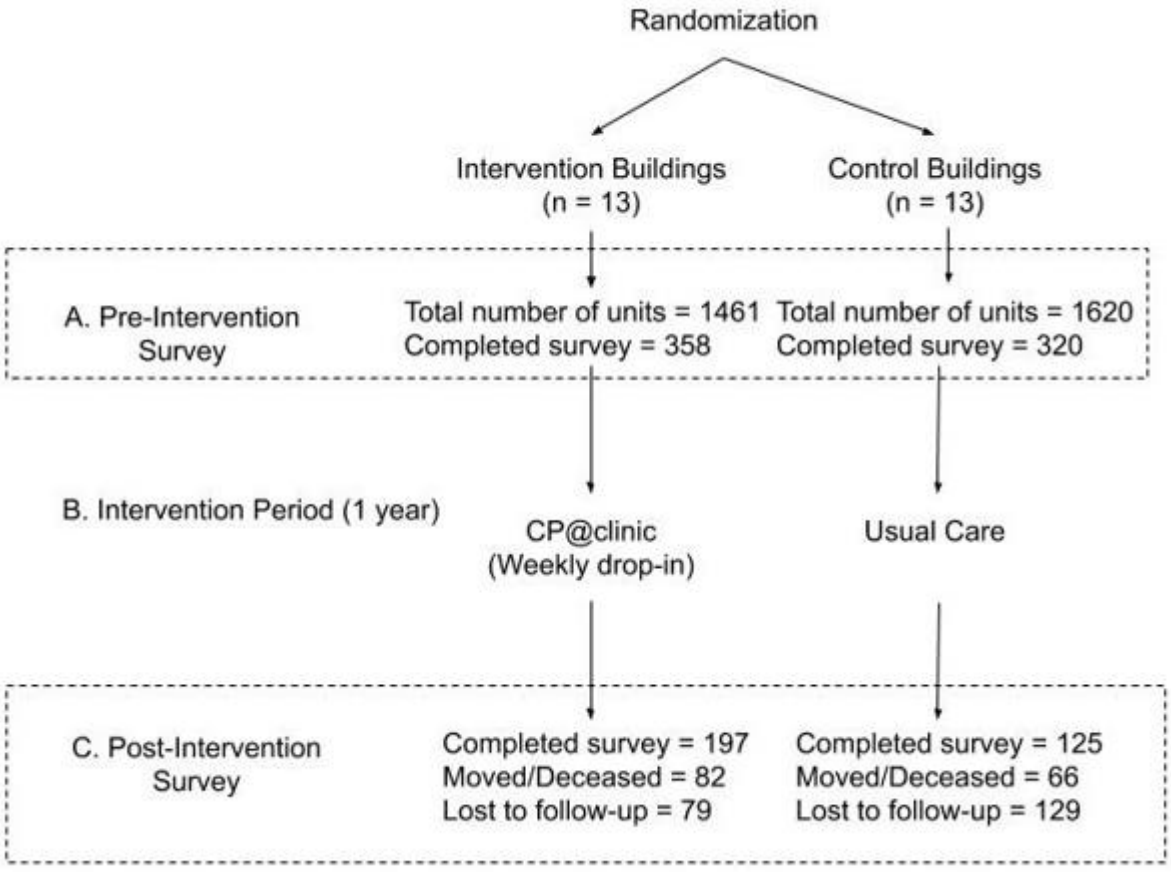
Probabilistic sensitivity analysis should be included.

We have moved the QALYs information in the methods section as suggested, also as suggested by reviewer #2.

We have included probabilistic sensitivity analysis (bootstrapping) as already mentioned above.

We have included information about missing data in our results in the form of a flow diagram (Figure 1). Individuals who did not complete the survey post intervention had either moved, were deceased or were lost to follow up.

Figure 1. CP@clinic study design and data collection flow diagram



4) Results

CEA – curve alt CE plane have to be presented

We have now presented a Cost Effectiveness Acceptability curve as recommended (Figure 2).

5) Discussion

The limited cost-perspective has to be acknowledged.

Thank you, we have now added a line at the end of the discussion to acknowledge this:

We have also only considered the perspective of the paramedic service since in Ontario they determine how to allocate staff and resource funding to extra programs. The perspective of society or other payers could be considered in future work.

Reviewer: 2

Major Comments

1. The main objective of this paper was to conduct a cost-utility analysis. Unfortunately, it is unclear how QALYs are computed. The authors do not describe how the health states of EQ-5D-3L are converted into utility scores and which scoring algorithm was used. Furthermore, I think, no QALYs were estimated (area under the curve) but rather differences in index scores. In case of this, it is not a cost-utility analysis.

We have added more description into the methods about how we used a previously validated Canadian data set (Feng et al, 2016) to calculate our QALYs:

The raw EQ5D-3L survey responses were treated as five-digit vectors (e.g. 13415) and transformed into index scores using previously validated Canadian EQ5D-3L value sets.

We also have demonstrated the analysis we did below to confirm that it was in fact a cost utility analysis:*

*We were not able to perform an area under the curve analysis since our baseline utility score values for intervention and control groups were significantly different from each other. Literature (Manca et al, 2005**) states that in this circumstance the area under the curve method is inaccurate. Manca et al. suggests the alternative methods for calculating QALY in these circumstances are to look at the differences between utility scores over time or use regression techniques. We had opted to use a combination of these methods, calculating the difference in utility score over time, and adjusting for the baseline differences using regression.*

We respectfully argue that our analysis is therefore a CUA as presented.

**Sarah J. Whitehead, Shehzad Ali, Health outcomes in economic evaluation: the QALY and utilities, British Medical Bulletin, Volume 96, Issue 1, December 2010, Pages 5–21, https://doi.org/10.1093/bmb/ldq033*

***Manca, A., Hawkins, N. and Sculpher, M.J. (2005), Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ., 14: 487-496. doi:10.1002/hec.944*

2. The authors stated that significant reductions in EMS calls were the main trail results (difference of -0.9 calls) after adjustment for the study design and baseline calls. The published main article (Reducing 9-1-1 emergency medical service calls by implementing a community paramedicine program for vulnerable older adults in public housing……) only found in the sensitivity analysis significantly reduced EMS calls but not in the intention to treat analysis. This should be mentioned.

Thank you for reading the initial RCT paper, we have included the fact you mention in our paper in the introduction:

The Community Paramedicine at clinic program (CP@clinic) has been evaluated in the format of a rigorous randomized controlled trial (RCT), in which it was found to have positive effects on the reduction of EMS calls from implementation sites, with a reduction of -0.88 calls/month/100 apartment units in Hamilton, and a reduction of -0.9 calls/month/100 apartment units in the sensitivity analysis for the whole RCT.[2,3]

3. Baseline characteristics of residents should be provided and a detailed flow chart to show the amount of buildings, apartment units, and residents for intervention and control group as well as drop-outs.

We have provided a flow chart (Figure 1).

We have fully described the baseline characteristics of residents in the full RCT paper which is referenced. Due to limitations on the number of tables, we had omitted this table - however, we can supply the table below to be included as an appendix or additional file if you deem it necessary.

Supplementary Table 1: Individual-level characteristics for intervention and control buildings at baseline

Descriptive Variables	Intervention building n=358 n (%)	Control building n=320 n (%)
Age years: mean (SD)	73.90 (9.05)	70.44 (7.94)
Female	286 (79.9)	229 (71.6)
Lives alone	322 (90.70)	287 (89.97)
Education		
Some High School or lower	160 (45.1)	146 (45.8)
High School Diploma	83 (23.4)	75 (23.5)
Some College/University or Higher	56 (15.8)	50 (15.7)
College or University	56 (15.8)	48 (15.0)
Poor Health Literacy[a]	80 (84.2)	84 (81.6)
With Chronic Diseases		
Heart Problems	111 (31.1)	80 (25.0)
Hypertension	192 (53.6)	177 (55.3)
High Cholesterol	135 (37.7)	119 (37.2)
Stroke	43 (12.0)	39 (12.2)
Diabetes	96 (26.8)	90 (28.1)
Risk Factors		
Low Physical Activity	148 (41.9)	166 (51.9)
Low Fruits and Vegetable intake	123 (34.6)	106 (33.2)
High Alcohol Intake	5 (1.4)	11 (3.4)
Smoker	87 (24.5)	122 (38.4)
	247 (69.6)	221 (69.0)

High BMI		
CANRISK[b]	104 (39.8)	98 (42.6)
Moderate	151 (57.9)	123 (53.5)
High		
Health Status and Quality-of-Life		
Reported Poor to Fair health	135 (38.0)	139 (43.5)
With mobility problems	218 (61.4)	192 (60.0)
With self-care problems	83 (23.4)	59 (18.4)
With problems doing usual activities	166 (46.8)	133 (41.6)
With pain/discomfort	249 (70.1)	239 (74.9)
With anxiety/depression	176 (48.5)	154 (48.1)
Has a Family Doctor	327 (91.3)	298 (93.1)

Notes: [a]For the health literacy assessment n= 89; for intervention 143 for control in Hamilton site only; [b]Only for participant not previously diagnosed with Diabetes

4. There are two different numbers of residents: 356 residents in the intervention group (table 2) and in line 171 n=595. How many residents were in the intervention group and what is the reason for this difference? This is not described (line 170: When restricted to those who attended the program n=595?) and should be presented in the flow chart.

This information is presented in Figure 1 as a flow chart.

The sentence that has the different n=595 in results, has been deleted, since it was in fact from our RCT paper, and was a sensitivity analysis of only those who attended the program. This paper is not concerned with only the attendees of the program, as we are concerned with a building level effect and cost (regardless of attendance). The inclusion of the sensitivity analysis data from the attendees is therefore confusing and irrelevant here. The program was open to all building residents, and we have hypothesized that the program would have an effect on any resident, not limited to just attendance ta the program, because of the social connections that might be developed with non-attendees.

We feel that with the flow diagram and the clearer results section, this is now in fact easier to understand and less ambiguous for the reader.

5. Please give more information about the intervention and inclusion/exclusion criteria because readers do not want to search for former publication (also see comment 6).

We have completely overhauled the whole Intervention section to include more information (see paper). We have added the following information to our Design section in the Methods:

Inclusion criteria were that each building required more than 60% of residents aged 55 years and older,more than 50 residential units, a unique postal code, and had at least one building of similar size and demographic to form a matched pair. There were no exclusion criteria.

6.a) The paramedic service perspective is not appropriate because the economic evaluation should inform decision maker and therefore include associated costs.

In this situation, we beg to differ from your opinion and confirm that indeed, the paramedic services are actually the decision makers here. They are the determining factor in what programs they decide to implement and what programs they decide to not implement.

6b) In the study protocol, the authors stated, "Using a prespecified algorithm, participants will be directed to appropriate services. Those identified as high risk will be immediately referred to appropriate healthcare resources, such as their primary healthcare provider" and in the main publication "Where urgent medical assistance was indicated, paramedics facilitated immediate connection with the family physician, urgent care, or ED." Therefore, the intervention is certainly associated with subsequent resource use e.g. outpatient ore inpatient treatment. Furthermore, in the study protocol cost measurement are described as follow: "Rates of EMS calls, ED visits, primary care visits and other healthcare utilisation will be collected from the administrative database of the local paramedic services, hospital ED databases, Institute for Clinical Evaluative Sciences (ICES) databases, the CHAP-EMS database and consenting participants' primary care charts. Administrative data will be collected preintervention and postintervention, and retrospectively for the 12 months of intervention." For this reason, please conduct the cost analysis including associated costs.

We applaud the reviewers for having sought to read the study protocol, however though the protocol details are correct, specific information regarding the rates of ED visits, primary care visits and other healthcare utilization data, is not available currently, but is underway from ICES. There is a considerable time delay in the procurement of such data, and this paper is still warranted until the data might be available in the future. Please note, we have been transparent about this and have mentioned it as a limitation in the paper. Furthermore, paramedic services, who are the payers in this situation, are not responsible for the costs of individuals' healthcare utilization costs, therefore it is most appropriate to present the cost analysis as we have.

7. The intervention program were offered in the building common areas. Please give information about subsidized housing building and its equipment features – are common areas standard?

We do not have enough space to cover descriptions of all facts related to the intervention in this paper, due to word limitations. However, common areas of buildings are spaces into which all residents have access, such as a sitting room, or games room. We have described this in detail in the RCT protocol however.

8. Please give more information about building matching. You write that buildings were matched by socio-demographics. I think you mean by social-demographic characteristics from residents – but which social-demographics variables were taken?

We did not use residents' individual socio-demographics to match buildings. We have inserted this line in the methods:

Housing organizations provided building level information which was used in the matching process: proportion of 'older aged' residents, number of units in the building, number of 911 calls per month per 100 units (baseline), and presence of building-level wellness programming.

9. What is modified staff? Please give a short explanation.

We added a short explanation of staff on modified duties under 'staffing costs' in the Data Collection section.

10. Was the intervention equal in all sites? You say that each site had different staffing arrangements. Are there differences in qualification?

Yes, the interventions were delivered in a uniform manner across all intervention sites regardless of the staffing model. We conducted fidelity assessments to assess uniformity of intervention. We added the word "Standardized " to the description of the intervention to emphasize the uniformity of the sessions.

All sites had paramedics staff the sessions, but in some cases the paramedics were staffing it as part of their modified duties (e.g. if they were pregnant or had a mental/physical injury). Supplementary Table 2 has been prepared to show the differences across sites and the associated staffing costs.

11. Table 1 shows monthly EMS calls per 100 units at baseline and after one year. The means appear low in contrast to table 1 in the main paper.

In the RCT tables the adjusted numbers are presented which correspond to some of the information in Table 1 here; but we are also presenting unadjusted information which is why the numbers may appear different between the papers.

12. The analysis was conducted at various levels: EMS reduction and program costs at building level and health related quality of life at resident level. Then a synthesis was made.

However, the assumptions were different. At building level all residents were included but at resident level only residents 55 years and older were included. Change in health related quality of life in residents 55 years and older was assumed to can be generalized.

Buildings were all seniors' buildings therefore only had residents who were aged 55 years and older, therefore no one was excluded in the building-level analyses.

Furthermore, only one quarter of residents (358 of 1461) participated in intervention (participation bias). Is there any information about non-participants? In order to reflect real world, program costs should be allocated to participants only. From which kind of residents (elderly residents?) EMS calls were made?

Even though only some of the building residents attended, the rates of EMS calls we obtained were for the whole building (the whole seniors building of people aged 55 years and older). We have assumed that the program would have effects on seniors who did not attend the program, due to social interactions that might have occurred as a result, for example.

You can conduct an analysis at individual level only. You need program costs per participant, EMS calls per participant, other health care resource use per participant, and health related quality of life. Then an appropriate analyses including uncertainty (cost-effectiveness plane according to bootstrapping) is possible.

We do not have individual EMS calls per participant, nor the health care utilization costs per participant and therefore cannot perform this analysis.

However, we have now performed a Bootstrapping PSA in a section under the results and modified Table 5 (see paper) to include these results:

Probabilistic Sensitivity Analysis using Bootstrapping

After the bootstrapping analysis was performed, the CP@clinic RCT found a QALY gain of 0.03 per intervention building resident (see Table 2). The mean ICER with Fieller's 95% CI was $4645 ($2489, $10,127). The Cost Effectiveness Acceptability curve is presented in Figure 2 with a willingness-to-pay threshold of $50,000 demonstrating that 100% acceptability was achieved well below willingness-to-pay of $15,000.

13. Please give information about drop outs (deceased?) and its handling in the imputation process.

We have presented this information in the new Figure 1. All data that we were unable to collect was accounted for in a multiple imputation analysis (iterative Markov Chain Monte Carlo method). Age, education, presence of chronic diseases (hypertension, heart disease, diabetes, high cholesterol, previous stroke), gender, living arrangement (living alone, marital status), baseline EQ5D measures (by individual domains), and baseline utility were used in the imputation process.

We have added this information to the analysis section of the paper.

VERSION 2 – REVIEW

REVIEWER	Inna Feldman Uppsala University, Sweden
REVIEW RETURNED	01-May-2020

GENERAL COMMENTS	Some minor comments to the abstract: Page 3, line 24: An incremental cost effectiveness ratio (ICER) based on incremental costs and health outcomes between the groups was calculated. Page 3, line 19:Quality-adjusted life years (QALYs) gained, measured with EQ-5D-3L. QALY gained were analysed between the groups at post-intervention, controlling for pre-intervention values and building pairings

REVIEWER	Hildegard Seidl Helmholtz Zentrum Munich Germany
REVIEW RETURNED	28-Apr-2020

GENERAL COMMENTS	Major Comments 1. You have added more information about QALY calculation and you wrote that you are not able to perform an area under the curve (AUC) analysis by citing Manca et al.. Manca described in detail how to calculate the AUC based on the area defined by the change from baseline utility using the linear interpolation model. Therefore, I do not understand the comment of the authors. You wrote: “For each individual, the difference in the pre-intervention and post-intervention index scores was calculated and multiplied by 1 year to get the QALY gained over the 1 year intervention.” You do calculate the AUC. But you do not describe how to handle discrete changes in quality of life between assessments. The linear change is the most commonly used approach (see Manca). If you multiply the difference by 1 year you make the assumption that the index score immediately “jumps” at t0 from pre-intervention to post-intervention index score. This is not an appropriate approach. 2. I worte in my comments: The authors stated that significant reductions in EMS calls were the main trail results (difference of -0.9 calls) after adjustment for the study design and baseline calls. The published main article (Reducing 9-1-1 emergency medical service calls by implementing a community paramedicine program for vulnerable older adults in public housing……) only found in the sensitivity analysis significantly reduced EMS calls but not in the intention to treat analysis. This should be mentioned. Please write clearly that only in the sensitivity analysis significantly reduced EMS calls were found. 3. In the supplemental table 1 (baseline characteristics of residents) the utility index scores have to be included!

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

Major Comments
1. You have added more information about QALY calculation and you wrote that you are not able to perform an area under the curve (AUC) analysis by citing Manca et al.. Manca described in detail how to calculate the AUC based on the area defined by the change from baseline utility using the linear interpolation model. Therefore, I do not understand the comment of the authors.
You wrote: “For each individual, the difference in the pre-intervention and post-intervention index scores was calculated and multiplied by 1 year to get the QALY gained over the 1 year intervention.”
You do calculate the AUC. But you do not describe how to handle discrete changes in quality of life between assessments. The linear change is the most commonly used approach (see Manca). If you multiply the difference by 1 year you make the assumption that the index score immediately “jumps” at t0 from pre-intervention to post-intervention index score. This is not an appropriate approach.

Thank you for your comment. We will try to explain better.
We did in fact use one of the alternative methods suggested by Manca, as opposed to his traditional method because we used multiple imputation methods when comparing the utility values. We used a regression based adjustment method for the post intervention measurement as described by Manca*: pg 490:
“The more appropriate method of dealing with imbalance in mean baseline utilities is the use of multiple regression. This approach allows the estimation of differential QALYs, as well as the prediction of adjusted QALYs, while controlling for baseline utility values.”

We only compared the endpoints adjusting for baseline
We do not have complete data for T2 therefore used a multiple imputation method based on T1, and therefore it is not recommended to calculate an AUC.

*Manca, A., Hawkins, N. and Sculpher, M.J. (2005), Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ., 14: 487-496. doi:10.1002/hec.944

Thank you for your comment. We have now modified our analysis to reflect those of Manca Method #3 exactly rather than utilising a mixed approach.
We have modified the analysis section, to remove the incorrect mixed approach and to include a more accurate description of the methods as follows:
QALYs gained were analysed between the groups at post-intervention, controlling for pre-intervention values and building pairing.

For each individual, the difference in the pre-intervention and post-intervention index scores was calculated and multiplied by 1 year to get the QALY gained over the 1 year intervention. These values were then adjusted for baseline differences using regression.

Due to this we have also updated Table 2, and a portion of the results section:
Over the course At the end of the 1-year intervention, there was an unadjusted 0.06 QALY change (from baseline) per person (95% CI, 0.01 to 0.12) in favour of the intervention buildings. Wwhen adjusting for baseline differences in the EQ-5D index score between the intervention and control buildings and for building pairing using regression, there was a significant adjusted mean 0.04 QALY change per person (95% CI, 0.01 to 0.07 0.00 to 0.08).
2. I wrote in my comments:

The authors stated that significant reductions in EMS calls were the main trail results (difference of -0.9 calls) after adjustment for the study design and baseline calls. The published main article (Reducing 9-1-1 emergency medical service calls by implementing a community paramedicine program for vulnerable older adults in public housing……) only found in the sensitivity analysis significantly reduced EMS calls but not in the intention to treat analysis. This should be mentioned. Please write clearly that only in the sensitivity analysis significantly reduced EMS calls were found.

Intro: Although we had clearly stated that it was in the sensitivity analysis, we have moved wording around now to make this statement at the beginning of the sentence in question:
The Community Paramedicine at clinic program (CP@clinic) has been evaluated in the format of a rigorous randomized controlled trial (RCT), in which the sensitivity analysis it was found CP@clinic to have positive effects on the reduction of EMS calls from implementation sites, with a reduction of -0.88 calls/month/100 apartment units in Hamilton, and a reduction of -0.90 calls/month/100 apartment units in the sensitivity analysis for the whole RCT.

Main Trial Results: we have added the wording requested:
As published previously, the CP@clinic RCT demonstrated significantly reduced EMS calls after 1 year of implementation when adjusted for the study design (i.e. building pairing) and baseline calls in the sensitivity analysis.

3. In the supplemental table 1 (baseline characteristics of residents) the utility index scores have to be included!

We have included this information in the supplemental table now.

Reviewer: 1
Some minor comments to the abstract:

Page 3, line 24: An incremental cost effectiveness ratio (ICER) based on incremental costs and health outcomes between the groups was calculated.
Page 3, line 19:Quality-adjusted life years (QALYs) gained, measured with EQ-5D-3L. QALY gained were analysed between the groups at post-intervention, controlling for pre-intervention values and building pairings

We have made the requested changes in the abstract at those 2 locations.

VERSION 3 – REVIEW

REVIEWER	Hildegard Seidl Helmholtz Zentrum München Germany
REVIEW RETURNED	24-Jul-2020

GENERAL COMMENTS	Major Comments You did not calculate QALYs. There is a difference between QALY calculation and using only index scores from EQ-5D! QALYs are always based on performing an area under the curve (AUC) regardless of baseline adjustment or multiple imputation. See Manca page 493. You only used the index score of post-intervention and controlled for pre-intervention values. Either you calculate QALYs (e.g. using linear change approach: (baseline utility index scor + utility index score after 1 year)/2 and multiply by 1 year) and adjust

	these values for baseline utility scores using regression or you take post-intervention values and adjust these values. The former is an effectiveness analysis using QALYs for outcome measurement – the latter is an effectiveness analysis using utility index scores after one year. You either calculate QALYs or withdraw the wording. If you want to calculate QALYs, describe how to handle discrete changes in quality of life between assessments. The linear change is the most commonly used approach (see Manca).

VERSION 3 – AUTHOR RESPONSE

Thank you for the opportunity to revise the QALY calculation as per the reviewer suggestion.
We have made the requested changes to our QALY calculation method and have revised the manuscript and other documents as required.

VERSION 4 – REVIEW

REVIEWER	Hildegard Seidl Helmholtz Zentrum München IGM Institute of Health Economics and Health Care Management Germany
REVIEW RETURNED	11-Sep-2020

GENERAL COMMENTS	The requested changes to the QALY calculation are made and I recommend to accept the manuscript. Hildegard Seidl