PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Invasiveness of pharmacokinetic studies in children: a systematic review
AUTHORS	Altamimi, Mohammed; Choonara, Imti; Sammons, Helen

VERSION 1 - REVIEW

REVIEWER	Soumitra Shankar Datta Tata Medical Centre, Kolkata
	India
REVIEW RETURNED	04-Dec-2015

GENERAL COMMENTS	This is an interesting review paper and done using a robust methodology. However it may not be of interest in its present form to clinicians across specialities. One idea is to just rewrite the paper in light of various ethical issues involved in doing PK studies (e.g. autonomy of children to decide about thier own participation in trials vs parental consent etc) so that it becomes a bit more lively for general medical readers. Also comments on the reasons for changing trends in PK studies would be useful for some of the readers. I do feel the paper is definitely worth publishing. The authors have been precise about their results and I do not have any comments about the methods.

REVIEWER	Steven Lane
	Liverpool University, UK
REVIEW RETURNED	16-Feb-2016

GENERAL COMMENTS	Why was the decade 1981-90 chosen?
	Why not do 40 year study with 4 ten year decades? This would
	allowed some investigation of the trends over time (1975-2014)
	Summary statistics median IQR suggests data may not be normally
	distributed, so why not use Mann-Whitney U test rather than t-test
	Authors seem to be ignoring the fact (from Table 1) that there has
	been a large increase in the number of <18yr olds taking part in PK
	studies. This is probably more important than number of samples or
	amount of blood being taken. It would add to the paper if authors
	included this in their analysis.

REVIEWER	Stefan Lange
	Cologne
	Germany
	Deputy director

REVIEW RETURNED	28-Feb-2016

GENERAL COMMENTS	I have only a few minor comments which should not compromise the scientific and statistical integrity of the manuscript.
	(i) The flow of information (flow chart) is not presented according to the PRISMA statement; it is not specified how many full texts were
	reviewed, and what were the reasons to exclude reviewed full texts
	from the publication (study) pool.
	(ii) The authors state that they have studies excluded where data for blood samples were not presented. The exclusion of such studies
	may have introduced bias as long as one cannot assume that the
	missing data are not associated with the research question. Hence,
	the authors should present the number of studies which were
	excluded for this reason per decade (i.e. 1981-1990, 2005-2014,
	and in between).
	(III) The authors do not well justify why they don't present the study
	outcomes for the years between 1991 and 2004. It should be
	clarified that this was not a data driven decision.
	(iv) The authors used t-tests for independent samples for the
	inferential statistical analysis of continuous variables. This is
	appropriate. However, the authors present medians and interquartile
	ranges only for the descriptive statistical analysis. Given the
	parametric nature of the statistics used for inferential analysis, I
	would prefer to see also means and standard deviations.

REVIEWER	Sue J. Lee Mahidol-Oxford Tropical Medicine Research Unit (MORU) Faculty of Tropical Medicine, Mahidol University 3rd Floor, 60th Anniversary Chalermprakiat Building 420/6 Ratchawithi Road Ratchathewi District Bangkok 10400 THAILAND
REVIEW RETURNED	07-Mar-2016

GENERAL COMMENTS	This manuscript asks whether pediatric (0 to 18 years old) PK studies have become less invasive.
	The primary outcome measures were the number of samples collected per child and the volume of blood collected, both per sample and total per child.
	N=501 studies were identified in a literature review and n=85 were included for the 1981 to 1990 period and n=205 for the 2005 to 2014 period.
	Major comments: 1. Overall, there should be better labeling of units throughout the manuscript to provide clarity for the reader, e.g., p. 8, line 44, "The median was 9 samples PER CHILD" and in Table 2, "Number of blood samples PER CHILD" (presumably this is what is meant rather than number of samples per study?).
	2. In general, when a Student's t-test is used to compare numbers, means and standard deviation (SD) are reported. This is because the t-test assumes a normal distribution of the data and, more importantly, it assumes approximately equal variances. It is unclear why medians are reported and I think it would be more appropriate

to either report means (SD) here OR to use a non-parametric test for comparison (such as the Mann Whitney U test) if medians are to be reported.
3. Although the tables report the total number of studies for each age group, it would be more useful to split this into 2 columns so that we know how many studies were included for each decade for each age group. This (with the means and SD) would allow the reader to reproduce/check the p-values reported. Furthermore, for table 3 and 5, if we knew the number of studies for each decade, we would know if "NA" meant that there were no studies for this age group in the first decade or rather that there WERE studies but that the median volume (total and per sample) was simply not reported.
4. There is no explanation as to why these 2 decades are chosen (with a 14 year gap in between). In particular, why was the decade from 1995-2004 NOT chosen?? Indeed, it seems that the search criteria included this time frame and that the differences in the number of patients and number of studies (Table 1) might have been smaller if decades that were closer to each other were chosen.
Minor comments: 1. p. 7, line 24. The groups presented are separate for preterm and term neonates (and without an age definition), yet in all tables, neonates are presented as one group. In addition, how was the "mixed" age group defined?
2. Please provide some consistency with the number of decimal places reported.
3. p. 10, line 36 and p. 11, lines 3-7. For the median, please report the IQR (or, if changing to means, report SDs). For the comparisons across age groups, please adjust for time (i.e., the two decades). This applies to comparison of population PK vs. non-population PK studies, as well. (p. 11, lines 10- 15).
4. p. 11, line 15. Please report this p-value to 3 decimals. While it is not incorrect to say that 0.04 is a "statistical difference", it is important to keep in mind that 0.05 is an arbitrarily chosen cut-off and that both 0.04 and 0.06 are therefore "borderline" significant. In fact, if the authors felt that this was indeed an important statistical finding then they should also consider adjusting for this factor in the analysis for the results presented in Tables 2, 3 and 5.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

One idea is to just rewrite the paper in light of various ethical issues involved in doing PK studies (e.g. autonomy of children to decide about thier own participation in trials vs parental consent etc) so that it becomes a bit more lively for general medical readers.

Also comments on the reasons for changing trends in PK studies would be useful for some of the readers.

The purpose of this review is to describe the invasiveness based on frequency and volume of blood collected. We however discussed the changing trend of PK studies.

Reviewer: 2

-Why was the decade 1981-90 chosen?

-Why not do 40 year study with 4 ten year decades? This would allowed some investigation of the trends over time (1975-2014)

We have done what the reviewers suggested and have studied the 4 decades

-Summary statistics median IQR suggests data may not be normally distributed, so why not use Mann-Whitney U test rather than t-test

We have now used the Kruskal-Wallis test

Authors seem to be ignoring the fact (from Table 1) that there has been a large increase in the number of <18yr olds taking part in PK studies. This is probably more important than number of samples or amount of blood being taken. It would add to the paper if authors included this in their analysis.

We have included this in our analysis Reviewer: 3

(i) The flow of information (flow chart) is not presented according to the PRISMA statement; it is not specified how many full texts were reviewed, and what were the reasons to exclude reviewed full texts from the publication (study) pool.

We have created new chart

(ii) The authors state that they have studies excluded where data for blood samples were not presented. The exclusion of such studies may have introduced bias as long as one cannot assume that the missing data are not associated with the research question. Hence, the authors should present the number of studies which were excluded for this reason per decade (i.e. 1981-1990, 2005-2014, and in between).

We have clarified this in the results

(iii) The authors do not well justify why they don't present the study outcomes for the years between 1991 and 2004. It should be clarified that this was not a data driven decision.

We have analysed all the studies for 4 decades

(iv) The authors used t-tests for independent samples for the inferential statistical analysis of continuous variables. This is appropriate. However, the authors present medians and interquartile ranges only for the descriptive statistical analysis. Given the parametric nature of the statistics used for inferential analysis, I would prefer to see also means and standard deviations. We have now used the Kruskal-Wallis test

Reviewer: 4

Major comments:

1. Overall, there should be better labeling of units throughout the manuscript to provide clarity for the reader, e.g., p. 8, line 44, "The median was 9 samples PER CHILD..." and in Table 2, "Number of blood samples PER CHILD" (presumably this is what is meant rather than number of samples per study?).

We have made the suggested changes

2. In general, when a Student's t-test is used to compare numbers, means and standard deviation (SD) are reported. This is because the t-test assumes a normal distribution of the data and, more importantly, it assumes approximately equal variances. It is unclear why medians are reported and I think it would be more appropriate to either report means (SD) here OR to use a non-parametric test for comparison (such as the Mann Whitney U test) if medians are to be reported. We have now used the Kruskal-Wallis test

3. Although the tables report the total number of studies for each age group, it would be more useful to split this into 2 columns so that we know how many studies were included for each decade for each

age group. This (with the means and SD) would allow the reader to reproduce/check the p-values reported. Furthermore, for table 3 and 5, if we knew the number of studies for each decade, we would know if "NA" meant that there were no studies for this age group in the first decade or rather that there WERE studies but that the median volume (total and per sample) was simply not reported. Changes have been made

4. There is no explanation as to why these 2 decades are chosen (with a 14 year gap in between). In particular, why was the decade from 1995-2004 NOT chosen?? Indeed, it seems that the search criteria included this time frame and that the differences in the number of patients and number of studies (Table 1) might have been smaller if decades that were closer to each other were chosen. We have now studied 4 decades

Minor comments:

1. p. 7, line 24. The groups presented are separate for preterm and term neonates (and without an age definition), yet in all tables, neonates are presented as one group. In addition, how was the "mixed" age group defined?

The age groups have been classified based on the definition in our method. The mixed age group is also being defined

2. Please provide some consistency with the number of decimal places reported. We have made required changes

3. p. 10, line 36 and p. 11, lines 3-7. For the median, please report the IQR (or, if changing to means, report SDs). For the comparisons across age groups, please adjust for time (i.e., the two decades). This applies to comparison of population PK vs. non-population PK studies, as well. (p. 11, lines 10-15).

We have made suggested changes

4. p. 11, line 15. Please report this p-value to 3 decimals. While it is not incorrect to say that 0.04 is a "statistical difference", it is important to keep in mind that 0.05 is an arbitrarily chosen cut-off and that both 0.04 and 0.06 are therefore "borderline" significant. In fact, if the authors felt that this was indeed an important statistical finding then they should also consider adjusting for this factor in the analysis for the results presented in Tables 2, 3 and 5.

The suggested changes have been made

VERSION 2 – REVIEW

REVIEWER	Soumitra Shankar Datta
	Kings College Hospital, UK
REVIEW RETURNED	16-May-2016

GENERAL COMMENTS	The reviewer completed the checklist but made no further
	comments.

REVIEWER	Steven Lane Department of Biostatistics University of Liverpool UK
REVIEW RETURNED	25-Apr-2016

GENERAL COMMENTS	The authors appears to have addressed both my previous
	comments and those of the other reviewers

REVIEWER	Stefan Lange
	Institute for Quality and Efficiency in Health Care (IQWiG)
	Cologne
	Germany
	Deputy director
REVIEW RETURNED	06-May-2016

GENERAL COMMENTS The (inferential) statistical analysis should be better described. While it is mentioned that Kruskal-Wallis tests and subsequently Dunn's post-hoc tests were used, it remains unclear whether and how the multiple tests problem was handled in light of the 6 possible pairwise comparisons between the 4 decades. In table 2 the results of the statistical tests are rather unclear presented. I assume that the last column represents the results of the Kruskal-Wallis tests. Exact P- values should be given here with 3 decimal places. Designations as 'NS' should be avoided (this is also relevant for table 3 and 4). In the single cells of table 2 some cells are marked with a '*' which indicates 'statistically significant' according to the legend. I assume that results of the Dunn's post-hoc tests are addressed here. However, as Dunn's post-hoc tests are addressed here.	GENERAL COMMENTS The (inferential) statistical analysis should be better described. While it is mentioned that Kruskal-Wallis tests and subsequently Dunn's post-hoc tests were used, it remains unclear whether and how the multiple tests problem was handled in light of the 6 possible pairwise comparisons between the 4 decades. In table 2 the results of the statistical tests are rather unclear presented. I assume that the last column represents the results of the Kruskal-Wallis tests. Exact P- values should be given here with 3 decimal places. Designations as 'NS' should be avoided (this is also relevant for table 3 and 4). In the single cells of table 2 some cells are marked with a '*' which indicates 'statistically significant' according to the legend. I assume that results of the Dunn's post-hoc tests are addressed here. However, as Dunn's post-hoc test compares two groups, it remains unclear which pairs were compared. Finally, in the last row ('Total')
unclear which pairs were compared. Finally, in the last row ('Total')	

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

The (inferential) statistical analysis should be better described. While it is mentioned that Kruskal-Wallis tests and subsequently Dunn's post-hoc tests were used, it remains unclear whether and how the multiple tests problem was handled in light of the 6 possible pairwise comparisons between the 4 decades. In table 2 the results of the statistical tests are rather unclear presented. I assume that the last column represents the results of the Kruskal-Wallis tests. Exact P-values should be given here with 3 decimal places. Designations as 'NS' should be avoided (this is also relevant for table 3 and 4). In the single cells of table 2 some cells are marked with a '*' which indicates 'statistically significant' according to the legend. I assume that results of the Dunn's post-hoc tests are addressed here. However, as Dunn's post-hoc test compares two groups, it remains unclear which pairs were compared. Finally, in the last row ('Total') all decades are marked with a '*', which is totally confusing.

We have amended Tables 2, 3 and 4 accordingly to make the statistical analysis clear.

REVIEWER	Stefan Lange
	Cologne
	Germany
	Deputy director
REVIEW RETURNED	14-Jun-2016

VERSION 3 - REVIEW

comments.