BMJ 2014;349:g5219 doi: 10.1136/bmj.g5219

RESEARCH

Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection

Céline Habre *research fellow*[1], Martin R Tramèr *professor in anaesthesia*[2 3], Daniel M Pöpping *anaesthetist*[4], Nadia Elia *public health epidemiologist*[2 5]

[1]Department of Radiology, Geneva University Hospitals, 4 rue Gabrielle-Perret-Gentil, CH-1211 Geneva 14, Switzerland; [2]Division of Anaesthesiology, Geneva University Hospitals, Geneva, Switzerland; [3]Faculty of Medicine, University of Geneva, Geneva, Switzerland; [4]Department of Anaesthesiology and Intensive Care, University Hospital Münster, Münster, Germany; [5]Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Abstract

Objective To examine whether, according to the conclusions of a 2000 systematic review with meta-analysis on interventions to prevent pain from propofol injection that provided a research agenda to guide further research on the topic, subsequently published trials were more often optimally blinded, reported on children, and used the most efficacious intervention as comparator; and to check whether the number of new trials published each year had decreased and whether the designs of trials that cited the review differed from those that did not.

Study design Systematic review comparing old trials (published before, and included in, the review) with new trials (published afterwards).

Data sources Medline, Cochrane, Embase, and bibliographies to January 2013.

Eligibility criteria for study selection Randomised studies testing any intervention to prevent pain from propofol injection in humans.

Results 136 new trials (19 778 patients) were retrieved. Compared with the 56 old trials (6264 patients), the proportion of optimally blinded trials had increased from 10.7% to 38.2% (difference 27.5%, 95% confidence interval 16.0% to 39.0%, P<0.001), and the proportion of trials that used the most efficacious intervention as comparator had increased from 12.5% to 27.9% (difference 15.4%, 4.0% to 26.9%, P=0.022). The proportion of paediatric trials had increased from 5.4% to 12.5%, although this was not significant (difference 7.1%, −1.0% to 15.2%, P=0.141). The number of new trials published each year was significantly higher (median number/year 12 (range 7-20) *v* 2.5 (0-9), P<0.001) with no obvious decreasing trend. 72.8% (n=99) of the new trials cited the review, with their designs similar to trials not citing the review. Only 36.0% (n=49) of the new trials were considered clinically relevant since they used the most efficacious intervention as comparator or included a paediatric population.

Conclusions The impact of the systematic review on the design of subsequent research was low. There was an improvement in the reporting of optimal blinding procedures and a tendency towards an increase in the proportion of paediatric trials. The most efficacious intervention was more often chosen as comparator but remained marginally used, and the number of trials published per year had not decreased. The use of systematic reviews should be encouraged to inform rational, and thus ethical, trial design and improve the relevance of new research.

Introduction

Systematic reviews often identify gaps in knowledge and methodological flaws in the existing literature.[1] They should also guide researchers in assessing the need for further investigations,[2] although this is often overlooked. For example, a systematic review published in 2005 and including 64 trials found that in 1992 it could already have been shown, after the 12th trial had been published, that aprotinin reduced the risk of bleeding in patients undergoing cardiac surgery; thus a timely performed systematic analysis of the published literature could have prevented 52 further trials from being performed.[3] It remains unclear though to what extent published systematic reviews influence the design of subsequent trials.

Designing a trial comprises, among other things, the choice of population and outcomes of interest, methods of data collection,

Correspondence to: C Habre celine.habre@hcuge.ch

Extra material supplied by the author (see http://www.bmj.com/content/349/bmj.g5219?tab=related#datasupp)

Effective reference interventions according to Picard review
Summary of analysed randomised controlled trials
References of included trials

or a comparator against which a new, potentially useful experimental intervention ought to be tested. Administering the anaesthetic propofol intravenously may be distressing for patients as it is often associated with pain at the injection site.[4] In 2000, a systematic review by Picard and Tramèr (a coauthor of the present analysis), including data from 6264 patients from 56 randomised placebo controlled trials, tested the analgesic efficacy of interventions to prevent the pain from propofol injection.[5] The systematic review, which was published in one of the top five anaesthesiology journals, indexed in all major medical databases, provided six main messages. Firstly, evidence showed that the most efficacious analgesic intervention was to administer a small intravenous dose of lidocaine (lignocaine) with venous occlusion before the propofol injection; with lignocaine 40 mg, the best documented regimen, the number needed to treat to prevent any pain compared with placebo was 1.8 (95% confidence interval 1.5 to 2.2). Secondly, although alternative interventions were also efficacious (for example, an intravenous bolus of lignocaine without venous occlusion, lignocaine mixed with propofol, or a variety of opioids or non-opioid analgesics administered concomitantly), none of these showed a similar degree of efficacy compared with lignocaine with venous occlusion (see supplementary file 1). Thirdly, for several experimental interventions, such as intravenous ondansetron, droperidol, or ketamine, or the dilution of propofol with homologous blood, no meaningful conclusions could be drawn owing to a lack of valid data. Fourthly, more data on children were needed to allow definite conclusions in this population. Fifthly, blinding procedures needed to be improved as only 11% of the trials were optimally blinded. And finally the authors of the Picard review concluded their report by questioning the necessity of performing further trials to identify yet another analgesic intervention to prevent pain from propofol injection.

We examined whether the Picard review had had any impact on subsequent research on pain from propofol injection. Specifically, we checked whether the number of new trials on the subject published per year had decreased over time, and whether the Picard review had influenced the design of subsequently published trials. For example, we expected that most new trials would focus on children, that the proportion of optimally blinded trials would increase, and that the most efficacious analgesic intervention identified in the Picard review would be chosen as a comparator against which new experimental interventions were tested. We also checked whether subsequently published trials cited the Picard review, and whether the designs of trials that cited the review differed from those that did not.

Methods

This systematic review was written according to the PRISMA statement for reporting systematic reviews and meta-analyses.[6] The study protocol was not registered but is available from the authors.

Eligibility criteria

We included all trials that had been analysed in the Picard review[5] and added a new search to identify all trials that had since been published. As in the Picard review, we searched for full published reports of randomised trials testing the analgesic efficacy of any intervention compared with placebo or no treatment to prevent pain from propofol administered intravenously. Since the necessity (and ethical acceptability) of a placebo arm may be questioned as an efficacious intervention had been identified, we additionally searched for trials that did not include a placebo arm. We included trials in adults, children, or volunteers undergoing general anaesthesia or sedation. We considered drug interventions (for example, pretreatment with a drug, or an alternative emulsion of propofol) and non-drug interventions (for example, cooling of propofol). To be included, trials had to report on the incidence of pain on injection of propofol as the primary outcome. We did not consider letters, conference abstracts, or studies in animals.

Information sources and searches

We performed searches in Medline (via Pubmed), Embase, and the Cochrane Library. We additionally identified trials from bibliographies of retrieved trials and checked references of a further relevant systematic review by Jalota and colleagues that was published 11 years after the Picard review.[7] We limited the search period from January 2002 (to ensure that trialists had the scope to read the Picard review, published in 2000) to January 2013. We did not search for unpublished trials. Trials were identified using the same search strategy and key words as in the Picard review—namely, "propofol", "pain", "injection", and "random", sought in the titles and abstracts, with a limit to humans but no limit to language.

Study selection and risk of bias assessment

One author (CH) assessed the eligibility of retrieved articles by screening the titles and abstracts. Queries were resolved through discussion with two other authors (NE, MRT). As in the Picard review, we scored new trials for quality of data reporting using the five point Oxford scale, which considers the three items randomisation, blinding, and flow of patients.[8] Blinding was rated optimal (2 points) when drugs were matched. Since we analysed exclusively randomised trials, the minimum score of an included trial was 1.

Data collection process

One author (CH) entered all data into an excel spreadsheet, which was developed for the purpose of this analysis. One of three other authors (NE, DMP, or MRT) independently checked the data. We contacted the authors of the original reports when we needed to clarify the nature of the data or were unable to access a report.

Data items

We extracted the characteristics of the trials, including year of publication, journal impact factor (Journals Citation Reports 2011; we analysed journals without an impact factor separately), open access status of the journal (yes/no), study population (adults, children, volunteers), number of analysed participants, and sources of funding (none, academic, industry, not declared).

According to the Picard review, the most efficacious intervention was an intravenous bolus injection of lignocaine with venous occlusion (manually or with a tourniquet) about 20 seconds before administering propofol into the same vein.[5] For the purpose of our analysis we classified this method as the primary reference treatment. Alternative interventions that had some proved efficacy, although less so compared with the lignocaine occlusion technique (for instance, intravenous injection of lignocaine without occlusion), were classified by us as secondary reference treatments. We classified interventions without proved efficacy according to the Picard review, and new, potentially useful interventions that had not yet been retrieved in that systematic review, as experimental interventions. We regarded no treatment controls as placebos.

Additionally, we classified the new trials according to their clinical relevance. We considered that for rational decision making clinicians needed to know not only how well a new experimental intervention performed compared with placebo but also how it performed compared with the primary reference treatment. Therefore we regarded study designs to be clinically relevant that compared an experimental intervention with the primary reference treatment, with or without an additional arm. We regarded all other comparisons—for example, experimental intervention versus placebo only, or experimental intervention versus secondary reference treatment, to be not clinically relevant because these trials were unlikely to contribute importantly to existing knowledge. If a trial showed that an experimental intervention offered advantages over placebo or over a secondary reference treatment, it would still leave the question unanswered as to whether that experimental intervention should be preferred or not to the currently most efficacious intervention (that is, the primary reference treatment). Also, we regarded trials that compared the primary reference treatment with a secondary reference treatment only or the primary or a secondary reference treatment with placebo as redundant and therefore not clinically relevant, as these comparisons had already been analysed in the Picard review. Furthermore, we considered any trial performed in children as clinically relevant since there was a lack of evidence for the best intervention to prevent pain from propofol injection in this subgroup. Finally, when the original trials cited the Picard review, we checked the context in which the review was cited and whether it was explicitly used to design the trial.

Data synthesis

We compared the characteristics and designs of trials published before and included in the Picard review (old trials) with those published after the Picard review (new trials). The main outcomes were the proportion of trials performed in children, the proportion that were optimally blinded (Oxford quality score 2), and the proportion that included the primary reference treatment. Results are reported as numbers and percentages. Continuous variables are reported as medians and ranges. We applied the χ^2 test to determine statistically significant differences between categorical variables, and we reported the difference in proportions between groups, with 95% confidence intervals for each category. For comparisons of non-Gaussian distributions we used the Mann-Whitney test.

In a further subgroup analysis we grouped trials according to whether or not they cited the Picard review and we compared the characteristics of their designs. We compared these subgroups using similar statistical tests. We also compared the characteristics of clinically relevant new trials with those that were not clinically relevant on the outcomes: published in an open access journal (yes/no), impact factor, citing the Picard review (yes/no), and sources of funding.

Results

Study selection

We identified 360 new reports (fig 1⇓). Through screening of titles and abstracts, we excluded 189 reports. The remaining 171 were studied in detail and a further 35 were subsequently excluded. We eventually included 136 new randomised trials that had been published at least two years after the publication of the Picard review (see supplementary files 2 and 3). Of these 136 new trials, 94 were placebo controlled.

Characteristics of new trials

The 136 new trials included data from 19 778 patients (table 1⇓). The trials originated from 30 countries and were published in 51 different journals, of which 29 (56.9%) had an impact factor and 14 (27.5%) were open access (see supplementary file 2). Eighty trials (58.8%) tested the efficacy of a variety of drugs administered before, or concomitantly with, propofol, 40 (29.4%) tested the analgesic efficacy of different emulsions of propofol, and 16 (11.8%) tested non-drug interventions.

Synthesis of differences between old and new trials

General characteristics

Compared with the old trials, the new trials were larger (median number of analysed participants 125.5 *v* 100, P<0.001), published in journals with lower impact factors (median 2.23 *v* 2.96, P<0.001), more often published in journals without an impact factor (30.1% *v* 16.1%, difference 14.1%, 95% confidence interval 1.7% to 26.4%, P=0.043), and scored higher for quality of data reporting (median 3 *v* 2, P<0.001, table 1). Eighty nine (65.4%) new trials scored 2 for randomisation compared with 6 (10.7%) old trials (difference 54.7%, 95% confidence interval 43.3% to 66.1%, P<0.001). The flow of patients was optimally described in 29 (21.3%) new trials compared with five (8.9%) old trials (difference 12.4%, 2.2% to 22.6%, P=0.041).

Main outcomes

Compared with the old trials, the number of new trials published per year increased (median 12 *v* 2.5; P<0.001, fig 2⇓, table 2⇓). The number of new trials performed in children had also increased, from three (5.4%) to 17 (12.5%), although the difference in proportions did not reach significance (P=0.141). Blinding procedures had improved (P<0.001) and a greater proportion of new trials were optimally blinded (10.7% *v* 38.2%; difference 27.5%, 16.0% to 39.0%). New trials used the primary reference treatment more often as a comparator (27.9% *v* 12.5%; difference 15.4%, 4.0% to 26.9%, P=0.022) and used a secondary reference treatment less often (35.3% *v* 62.5%; difference −27.2%, −42.2% to −12.2%, P<0.001).

Ninety nine of the 136 new trials (72.8%) cited the Picard review (table 3⇓). Compared with the 37 trials that did not, trials citing the review were published more often per year (median 9 *v* 3, P<0.001), were not more often performed in children (14.1% *v* 8.1%, P=0.344), the distribution of blinding scores were not statistically significantly different although optimal blinding was more common (43.4% *v* 24.3%, difference 19.1%, 2.2% to 36.0%), and reporting quality tended to be higher (median Oxford score 3 *v* 2, P=0.011). Reporting of randomisation procedures scored 2 in 70 (70.7%) of the trials citing the review compared with 19 (51.4%) of the trials not citing the review (difference 19.4%, 0.9% to 37.8%, P=0.035), and the flow of patients was optimally described in 24 (24.2%) of the trials citing the review compared with five (13.5%) of the trials not citing the review (P=0.174). The proportion of trials including the primary reference treatment did not differ (29.2% *v* 24.3%, P=0.565, table 4⇓). However, trials citing the review included a secondary reference treatment more often (40.4% *v* 21.6%; difference 18.8%, 2.4% to 35.2%, P=0.041) and a design without a primary or a secondary reference treatment less often (30.3% *v* 54.1%; difference −23.8%, −42.2% to −5.3%, P=0.011).

Additional findings

Clinical relevance

Forty nine (36.0%) new trials were regarded by us as clinically relevant. Of the 87 (64.0%) trials considered not clinically relevant, 47 (54.0%) compared experimental interventions with or without a placebo, 32 (36.8%) compared an experimental intervention with a secondary reference treatment with or without a placebo, four (4.6%) compared the primary reference treatment with a secondary reference treatment with or without a placebo, three (3.5%) compared secondary reference treatments with or without a placebo, and one (1.1%) compared the primary reference treatment with placebo.

There were no significant differences between clinically relevant and non-relevant trials for the proportion that cited the Picard review (79.6% *v* 69.0%, P=0.181), the proportion published in journals without an impact factor (28.6% *v* 31.0%, P=0.764), the median impact factor of those published in journals with an impact factor (2.23 *v* 2.19, P=0.418), or the proportion published in open access journals (22.4% *v* 20.7%, P=0.810, table 5⇓).

Of the 49 clinically relevant trials, two (4.1%) were funded by industry; of the 87 clinically non-relevant trials, seven (8.0%) were funded by industry (difference −4.0%, −11.9% to 4.0%, table 5). Of the nine new trials funded by industry, two were optimally blinded, two were not double blinded at all, five (55.5%) compared an experimental intervention with a secondary reference treatment, three (33.3%) compared an experimental intervention with placebo, and one (11.1%) compared two experimental interventions. None used the primary reference treatment (see supplementary file 2).

Context of citation of Picard review

Since we were unable to find significant differences between the designs of trials citing and not citing the Picard review, we made further investigations regarding the context in which the Picard review was cited in the original trials. Among the 99 trials that cited the Picard review, 39 (39.4%) did so to illustrate the large variety of analgesic interventions that had been tested in this setting (five used the primary reference treatment) and 21 (21.2%) to document the underlying risk of pain (four used the primary reference treatment). Fifteen trials explicitly reported having used the review as a basis for the choice of the primary reference treatment as a comparator (of which one criticised the fact that the studies included in the review were too disparate and that they had methodological gaps, and therefore they chose to repeat the comparison). Seven trials explicitly reported not having chosen the primary reference treatment for different reasons: the tourniquet was difficult or threatening in children (n=4 trials), the primary reference treatment was not often used although it was the best option available (n=1), the primary reference treatment was judged to be awkward (n=1), and the Picard review was regarded as not-conclusive (n=1). Eleven further trials acknowledged the identification of the primary reference treatment by the review; 10 ignored it without any explanation and one chose the primary reference treatment, but on the basis of the authors' own previous pilot study. The authors of four trials chose the primary reference treatment without explanation but eventually compared their results with those of the Picard review. Finally, the authors of two trials stated as a limitation of their study that they did not use the primary reference treatment that was identified by the Picard review.

All analyses were performed on the subgroup of 94 trials including a placebo or no treatment group. The results were similar.

Discussion

This study, which aimed to examine the impact of a systematic review with meta-analysis on the design and relevance of subsequent research, illustrates four major problems. Firstly, although the systematic review had identified a simple, effective, and low cost intervention and strongly suggested that additional trials on this specific issue were no longer necessary, the publication of trials has not decreased. Secondly, although the systematic review provided a clear research agenda, its influence on the design of further trials has remained poor. Thirdly, the proportion of subsequently published trials that could have had an impact on clinical practice has remained low. Finally, citing the systematic review had no clear influence on the design or relevance of subsequently published research.

Comparison with other studies

There are numerous examples where systematic reviews, if performed in a timely manner, could have provided evidence of the effectiveness of an intervention and thus prevented redundant research.[9 10] There is also evidence that knowledge from systematic reviews is underused to inform future research.[3 11] Our study confirms these findings. These raise ethical concerns not only because patients are unnecessarily randomised in worthless trials but also because resources are wasted.[2]

It remains unclear why the number of published trials on the prevention of pain from propofol injection has not decreased; in fact the number has actually increased. Individual motivations may explain this finding. Doctors are challenged to engage in research for their career progression (publish or perish policy), but clinically relevant large studies with long follow-up periods are difficult to achieve.[12] Studies focusing on pain from propofol injection are easy and quick to perform and are straightforward to publish. As long as academic promotion and funding systems are based on simplistic counts of published papers, favouring prolific authors regardless of the relevance and validity of their work, this is unlikely to change. Ethics committees are supposed to ensure scientific soundness and relevance of clinical investigations, preventing enrolment of participants in non-ethical trials. This should include the rational choice of a comparator intervention[13] and be supported by systematic reviews to check how new protocols fit in with the current state of medical knowledge. In real life, however, ethics committees seem to stand alone, with limited resources and sometimes not enough scientific credit or knowledge to identify, and stop, the performance of irrelevant research. It also remains unclear why editors accepted these new trials for publication. The new trials were published in journals with lower impact factors, and a higher proportion was published in journals without any impact factor. This suggests that editors of higher quality journals were unwilling to publish articles on an already solved problem. It has been suggested that open access journals may apply less stringent criteria for publication.[14] We cannot confirm this hypothesis based on our sample of trials; a similar proportion of relevant and non-relevant trials were published in open access journals.

The Picard review provided a clear research agenda; however, its influence on the design of further research has remained marginal. More data on children were deemed necessary, and although 17 new paediatric trials have been published, the proportional increase in paediatric trials did not reach statistical significance and may have occurred by chance. Interestingly, although a total of 20 trials performed in children is now available and may be sufficient to draw conclusions on this

population, these trials were excluded from the updated systematic review by Jalota and colleagues,[7] and to our knowledge no systematic review has yet summarised the available evidence on the prevention of pain from propofol injection in children. The research agenda in the Picard review also suggested that more optimally blinded trials were needed, which was achieved in subsequent trials. Moreover, blinding was more often optimal in trials citing than not citing the Picard review. However, since new studies scored higher in all aspects of quality of data reporting, the specific impact of the Picard review may be questioned, since this improvement may actually reflect an increase in the implementation of the CONSORT statement.[15] Interestingly, the source of funding was still not declared in about 65% of the trials, and only three have been registered, although these items are included in the CONSORT checklist.

The identification of the currently most effective intervention to prevent pain from injection of propofol was the main message of the Picard review. It is plausible that in some of the subsequently published trials the review had an influence on the choice of the comparator against which a new, potentially innovative, experimental intervention was tested. It has been argued that the choice of a comparator intervention should be supported by a systematic review of the relevant literature.[16] However, the authors of less than one third of subsequently published trials chose the primary reference treatment as a comparator.

Overall, the number of clinically relevant trials (those that are likely to have an impact on clinical practice) remained low. Although the Picard review identified a primary reference treatment, researchers may have wanted to test yet another experimental intervention that may have been even more efficacious, or simpler to use. One would expect this experimental intervention to be compared with the current primary reference treatment. In our study, about 1 in 3 subsequently published trials only used the primary reference treatment as a comparator. Authors that aim for academic recognition may embark on research that does not necessarily improve existing knowledge.[17] Also, they may want to reach statistically significant results to ensure publication, since journals are more prone to publish such results,[18] and therefore refrain from comparing an experimental intervention with the currently most efficacious treatment. Interestingly, nine trials were industry sponsored, and none of those used the primary reference treatment as a comparator. This suggests that in these trials a comparator was chosen to favour a priori the experimental intervention, at the cost of threatening the principle of equipoise.[19] Citing the Picard review did not clearly reflect its influence on study design. Although almost 73% of the new trials cited the Picard review, their characteristics and clinical relevance were largely similar to those that did not. Only about 32% of the trials mentioning the review explicitly described having used it to inform study design; 15 cited it to justify the choice of the primary reference treatment, whereas seven cited it to explain why they had chosen not to use the primary reference treatment. Ignoring the recommendations of the Picard review may be justified as long as it is explicitly explained.

Strengths and weaknesses of this study

This study has several limitations. Firstly, we did not analyse some trials that were published in Chinese, Japanese, or Persian, because of problems with translation, and we did not include studies published between January 2000 and December 2001. It is unlikely that including those trials would have changed the results of our analysis. Secondly, we cannot exclude that some authors remained unaware of the Picard review even though we selected trials that had been published from 2002. A two year delay between the publication of the review and the inclusion of trials into our analyses may be short. Seven of 13 trials (54%) published in 2002 cited the review and it is possible that some of these trials had started enrolment of patients before the publication of the review. This would explain why their design was not influenced by the Picard review. Thirdly, the relevance of the Picard review may be questioned since it may be considered out of date. The median survival time of a systematic review has been estimated to be about six years.[20] However, the biological basis underlying pain from propofol injection has not changed, and the Picard review has remained the only systematic review on this topic for almost 12 years. The relevance of the primary reference treatment identified in the review may also be questioned since it was based on about 400 patients only. However, to find an even smaller treatment effect (for example, a decrease from a conservative baseline risk of 60% pain with placebo to 20% pain with an experimental intervention, number needed to treat 2.5), only about 35 patients would be necessary in each group of a single trial to obtain 90% power for detecting such a level of analgesic efficacy (two sided test, α level 0.05). In 2011, a new systematic review on the same subject, published by different authors,[7] included data from 25 260 patients from 177 trials and came to the same conclusion—namely, that administering lidocaine (lignocaine) along with venous occlusion was the most efficacious intervention for the prevention of pain from propofol injection. This second review also reported that injection of propofol through a large vein provided a similar degree of pain relief, an intervention that had remained ill described at the time of the Picard review. Fourthly, we did not contact the authors of the new trials to ask them whether they knew about the Picard review, why they had chosen to cite it or not, and on what grounds they had designed their study. It may be that some authors had actually used the Picard review to inform their study design but did not overtly refer to it. It is also possible that the new trials, which were designed according to the Picard review, subsequently influenced further trials too. Indeed, some trials used a clinically relevant study design but did not explicitly justify their choice. Fifthly, our definition of what constitutes a clinically relevant study design may be challenged. For example, seven trials had chosen, based on the Picard review, not to use the primary reference treatment for different reasons. Of these, two were classified by us as being not clinically relevant. There may be an argument to classify those as clinically relevant as the authors used the Picard review to defend the choice of their comparator. And finally, the generalisability of our findings remains unknown. This analysis is based on a specific subject of perioperative medicine. It is possible that in research areas where studies of longer follow-up times and more complex infrastructure are required, fewer irrelevant trials are produced.

Unanswered questions and future research

Our study raises questions about the dissemination of results identified through a systematic review. It seems that dissemination of the main results of the Picard review and their implementation into practice and methodological guidelines has not been sufficient. The extraordinary degree of analgesic efficacy of intravenous lignocaine with venous occlusion, its ease of use, and low cost, may have led the authors of the Picard review to believe that the primary reference treatment would be widely accepted both in clinical practice and as a comparator for subsequent research. They were wrong. It was surprising that so many new trials had been published on a topic for which

an efficacious intervention was known. Even more disturbing was that most new trials ended up with clinically non-relevant designs, which meant that these trials were unable, regardless of the results, to generate important new knowledge and thus to considerably change clinical practice. The profusion of clinically non-relevant studies reflects the considerable imbalance between the strong pressure to publish and the weakness of barriers that prevent useless research. Methodological implementation strategies are needed to define an appropriate and ethical research agenda based on the best currently available evidence. Dissemination and implementation of scientific knowledge is a challenge and has been widely discussed.[21 22] There is a difference, however, between implementation of clinical knowledge and implementation of methodological knowledge. While clinicians remain free to treat patients as they want, research protocols have to go through a range of barriers that could avoid waste of resources if they were adequately used. Barriers include funders (for funded research), ethics committees, and clinical trial registries. Unfortunately, clinical trial registries do not yet exert any quality control over registered protocols; this is supposed to be done by funding agencies and ethics committees. However, funders are more easily impressed by a long list of publications than by relevant content, and ethics committees are more concerned about patient protection than down-regulation of non-relevant research, although scientific soundness is a prerequisite of ethical soundness (for example, Helsinki Declaration article 21).[23] We suggest that both funders and ethics committees start asking authors explicitly whether a systematic review on their research topic exists, and if yes, whether that systematic review proposes a research agenda. If this is the case, authors should explain how their proposed research project fits into that research agenda. This would require a wider acceptance of the strengths of systematic reviews and meta-analyses in defining research agendas but would avoid unnecessary, redundant, or invalid, and thus unethical, research. Finally, journal editors may inform their readers that they will stop considering reports of trials that ignore such knowledge for publication.

Conclusions and recommendations

Much effort has been put into the conviction of academia and policymakers that systematic reviews are vital instruments to improve efficacy and safety of healthcare.[24] These efforts should be extended to the research area. Authors should justify their trial design in the context of the current state of knowledge.[11 25] Our findings highlight the role that systematic reviews should play in guiding trialists in their choice of the most appropriate study design, avoiding ill designed and clinically none relevant trials and thus a waste of resources.

We thank P Picard (Centre d'Étude et de Traitement de la Douleur, Clermont-Ferrand, France) and M R Tramèr (Geneva University Hospitals, Geneva, Switzerland) who shared with us original data of their systematic review; T Perneger (Geneva University Hospitals) for his thoughtful advice; R Abbasivash (Department of Anaesthesiology, Faculty of Medicine, Urmia University of Medical Sciences, Urmia, Iran) who responded to our enquiry and provided a copy of his paper; M Öztürk (Department of Surgery, Geneva University Hospitals) who translated two Turkish papers into English. We assume full responsibility for the analyses and interpretation of these data.

Presentation: Summary data of this analysis have been presented as an oral presentation at the 7th International Congress on Peer Review and Biomedical Publication, Chicago, September 8-10, 2013 (abstract No PRC-13-0043).

Contributors: CH, MRT, and NE conceived the study, contributed to study design, conducted the analyses, and prepared the manuscript. CH performed systematic literature searches. CH and DP extracted data from original trials. DMP contributed to the preparation of the manuscript. All authors read and approved the manuscript. CH and NE had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. NE is guarantor.

Funding: This study did not receive specific funding.

Competing interests: All authors completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: MRT is a coauthor of the Picard review; no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The lead author (NE) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

1 Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ* 1997;315:1371-4.
2 Young C, Horton R. Putting clinical trials into context. *Lancet* 2005;366:107-8.
3 Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials* 2005;2:218-29; discussion 229-32.
4 Eger EI 2nd. Characteristics of anesthetic agents used for induction and maintenance of general anesthesia. *Am J Health Syst Pharm* 2004;61(suppl 4):3-10.
5 Picard P, Tramèr MR. Prevention of pain on injection with propofol: a quantitative systematic review. *Anesth Analg* 2000;90:963-9.
6 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1-34.
7 Jalota L, Kalira V, George E, Shi YY, Hornuss C, Radke O, et al. Prevention of pain on injection of propofol: systematic review and meta-analysis. *BMJ* 2011;342:d1110.
8 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
9 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54.
10 Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *Int J Epidemiol* 2005;34:874-87.
11 Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *J R Soc Med* 2007;100:187-90.
12 Yuan HF, Xu WD, Hu HY. Young Chinese doctors and the pressure of publication. *Lancet* 2013;381:e4.
13 Garattini S, Bertele V, Li Bassi L. How can research ethics committees protect patients better? *BMJ* 2003;326:1199-201.
14 Bohannon J. Who's afraid of peer review? *Science* 2013;342:60-5.
15 Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
16 Mann H, Djulbegovic B. Choosing a control intervention for a randomised clinical trial. *BMC Med Res Methodol* 2003;3:7.
17 Krishnan V. Etiquette in scientific publishing. *Am J Orthod Dentofacial Orthop* 2013;144:577-82.
18 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;1:MR000006.
19 Ashcroft R. Equipoise, knowledge and ethics in clinical research and practice. *Bioethics* 1999;13:314-26.
20 Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007;147:224-33.
21 Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *BMJ* ;347:f6753.
22 Contopoulos-Ioannidis DG, Alexiou GA, Gouvias TC, Ioannidis JP. Medicine. Life cycle of translational research for medical interventions. *Science* 2008;321:1298-9.
23 Gandevia B, Tovell A. Declaration of Helsinki. *Med J Aust* 1964;2:320-1.
24 Clarke M. Doing new research? Don't forget the old. *PLoS Med* 2004;1:e35.
25 Chalmers I. Academia's failure to support systematic reviews. *Lancet* 2005;365:469.

Accepted: 3 August 2014

Cite this as: *BMJ* 2014;349:g5219

What is already known on this topic

- It is unethical to embark on new research without first analysing what can be learnt from existing literature
- Systematic reviews guide researchers in assessing the need for further investigations, to avoid unnecessary and redundant research
- Systematic reviews often provide a research agenda to guide future research

What this study adds

- A systematic review that had identified an efficacious analgesic intervention to prevent pain on injection of propofol and provided a clear agenda for future research had only little impact on the design of subsequently published trials on the same subject
- Implementation strategies are needed to ensure dissemination of results and methodological issues identified through systematic reviews
- New trials should be designed according to the findings and recommendations of the research agenda of previously published systematic reviews

Tables

Table 1| Characteristics of trials published before and after the Picard review. Values are numbers (percentages) unless stated otherwise

Characteristics	Old trials*	New trials†	Difference in proportions (95% CI)	P value
Trials	56	136		
Participants	6264	19 778		
Median (range) year of publication	1995 (1982-99)	2007 (2002-12)		
Median No (range) of participants per trial	100 (28-368)	125.5 (16-500)		<0.001
Published in journal without impact factor	9 (16.1)	41 (30.1)	14.1 (1.7 to 26.4)	0.043
Median (range) impact factor‡	2.96 (1.21-5.36)	2.23 (0.03-5.36)		<0.001
Oxford quality score§:				<0.001
1 or 2	43 (76.8)	45 (33.1)	–43.7 (–57.3 to –30.1)	
3	12 (21.4)	43 (31.6)	10.2 (–3.1 to 23.5)	
4 or 5	1 (1.8)	48 (35.3)	33.5 (24.8 to 42.3)	

*Published before (and therefore included in) the Picard review.

†Published after the Picard review, from 2002 onwards.

‡Calculated excluding trials that were published without an impact factor.

§Randomisation: none (score 0), mentioned (1), described and adequate (that is, computer generated list of random numbers, or sealed envelopes) (2); blinding: none or observer blinding only (score 0), double blinding described but not optimal (1), optimal (double dummy or convincing blinding procedures) (2); drop-outs: not described or incomplete (score 0), clear follow-up of each patient (1).

Table 2| Comparisons of outcomes in trials published before and after the Picard review. Values are numbers (percentages) unless stated otherwise

Outcomes	Old trials*	New trials†	Difference in proportions (95% CI)	P value
Publication rate:				
Median No (range) of trials published per year	2.5 (0-9)	12 (7-20)		<0.001
Population:				
Trials in children	3 (5.4)	17 (12.5)	7.1 (–1.0 to 15.2)	0.141
Blinding (Oxford quality score):				<0.001
No attempt, or single blinding only (score 0)	19 (33.9)	37 (27.2)	6.7 (21.2 to 7.8)	
Described but not optimal (score 1)	31 (55.4)	47 (34.6)	–20.8 (–36.1 to –5.5)	
Optimal (score 2)	6 (10.7)	52 (38.2)	27.5 (16.0 to 39.0)	
Trial design including primary reference treatment:				
PT *v* ST *v* experimental (*v* placebo)	3 (5.4)	5 (3.7)		
PT *v* experimental (*v* placebo)	0 (0.0)	28 (20.6)		
PT *v* ST (*v* placebo)	4 (7.1)	4 (2.9)		
PT *v* placebo	0 (0.0)	1 (0.7)		
Total	7 (12.5)	38 (27.9)	15.4 (4.0 to 26.9)	0.022
Trial design without PT but including ST:				
ST *v* experimental (*v* placebo)	19 (33.9)	45 (33.1)		
ST *v* ST (*v* placebo)	16 (28.6)	3 (2.2)		
Total	35 (62.5)	48 (35.3)	–27.2 (–42.2 to –12.2)	<0.001
Trial design without PT or ST:				
Experimental *v* experimental (*v* placebo)	14 (25.0)	50 (36.8)	11.8 (–2.2 to 25.7)	0.116

PT=primary reference treatment (intravenous lidocaine (lignocaine) with venous occlusion); ST=secondary reference treatment (that is, alternative intervention of proved efficacy, for example, lignocaine added to propofol).

*Published before (and therefore included in) the Picard review.

†Published after the Picard review, from 2002 onwards.

Table 3| Comparison of characteristics of trials according to their reference to the Picard review. Values are numbers (percentages) unless stated otherwise

Characteristics	Reference to review	No reference to review	Difference in proportions (95% CI)	P value
Trials	99 (72.8)	37 (27.2)		
Participants	14 590	5188		
Median (range) year of publication	2007 (2002-12)	2007 (2002-12)		0.992
Median No (range) of participants per trial	127 (22-500)	120 (16-335)		0.839
Published in journal without impact factor	26 (26.3)	15 (40.5)	−14.3 (−32.3 to 3.8)	0.106
Median (range) impact factor*	2.23 (0.03-5.36)	2.11 (0.52-3.29)		0.385
Oxford quality score†:				0.011
1 or 2	26 (26.3)	19 (51.4)	−25.1 (−43.4 to −6.8)	
3	32 (32.3)	11 (29.7)	2.6 (−14.8 to 20.0)	
4 or 5	41 (41.4)	7 (18.9)	22.5 (6.6 to 38.4)	

*Calculated excluding trials that were published without an impact factor.

†Randomisation: none (score 0), mentioned (1), described and adequate (that is, computer generated list of random numbers, or sealed envelopes) (2); blinding: none or observer blinding only (score 0), double blinding described but not optimal (1), optimal (double dummy or convincing blinding procedures) (2); drop-outs: not described or incomplete (score 0), clear follow-up of each patient (1).

Table 4| **Comparison of outcomes in trials according to their reference to the Picard review. Values are numbers (percentages) unless stated otherwise**

Outcomes	Reference to review	No reference to review	Difference in proportions (95% CI)	P value
Publication rate:				
Median No (range) of trials published per year	9 (6-15)	3 (0-6)		<0.001
Population:				
Trials in children	14 (14.1)	3 (8.11)	6.0 (−5.1 to 17.2)	0.344
Blinding (Oxford quality score):				0.118
No attempt, or single blinding only (score 0)	24 (24.2)	13 (35.1)	−10.9 (−28.4 to 6.6)	
Described but not optimal (score 1)	32 (32.3)	15 (40.5)	−8.2 (−26.5 to 10.1)	
Optimal (score 2)	43 (43.4)	9 (24.3)	19.1 (2.2 to 36.0)	
Trial designs including PT:				
PT *v* ST *v* experimental (*v* placebo)	3 (3.0)	2 (5.4)		
PT *v* experimental (*v* placebo)	23 (23.2)	5 (13.5)		
PT *v* ST (*v* placebo)	3 (3.0)	1 (2.7)		
PT *v* placebo	0 (0.0)	1 (2.7)		
Total	29 (29.2)	9 (24.3)	5.0 (−11.5 to 21.5)	0.565
Trial designs without PT but including ST:				
ST *v* experimental (*v* placebo)	37 (37.4)	8 (21.6)		
ST *v* ST (*v* placebo)	3 (3.0)	0 (0.0)		
Total	40 (40.4)	8 (21.6)	18.8 (2.4 to 35.2)	0.041
Trial designs without PT or ST:				
Experimental *v* experimental (*v* placebo)	30 (30.3)	20 (54.1)	−23.8 (−42.2 to −5.3)	0.011

PT=primary reference treatment (intravenous lidocaine (lignocaine) with venous occlusion). ST=secondary reference treatment (that is, alternative intervention of proved efficacy, for example, lignocaine added to propofol).

Table 5| **Relevant versus non-relevant trial designs among 136 new trials. Values are numbers (percentages) unless stated otherwise**

Characteristics	Relevant designs	Non-relevant designs	Difference in proportions (95% CI)	P value
Trials	49 (36.0)	87 (64.0)		
Median (range) year	2007 (2002-12)	2007 (2002-12)		0.667
Reference to Picard review	39 (79.6)	60 (69.0)	10.6 (–4.3 to 25.5)	0.181
Published in journal without impact factor	14 (28.6)	27 (31.0)	–2.4 (–18.4 to 13.5)	0.764
Median (range) impact factor*	2.23 (0.03-5.36)	2.19 (0.32-4.24)		0.418
Published in open access journal	11 (22.4)	18 (20.7)	1.8 (–12.7 to 16.2)	0.810
Funding source:				0.143
Not declared	32 (65.3)	55 (63.2)	2.1 (–14.7 to 18.8)	
None	10 (20.4)	8 (9.2)	11.2 (–1.6 to 24.0)	
Academic	5 (10.2)	17 (19.5)	–9.3 (–21.2 to 2.5)	
Industry	2 (4.1)	7 (8.0)	–4.0 (–11.9 to 4.0)	

*Calculated excluding trials that were published without an impact factor.

Figures

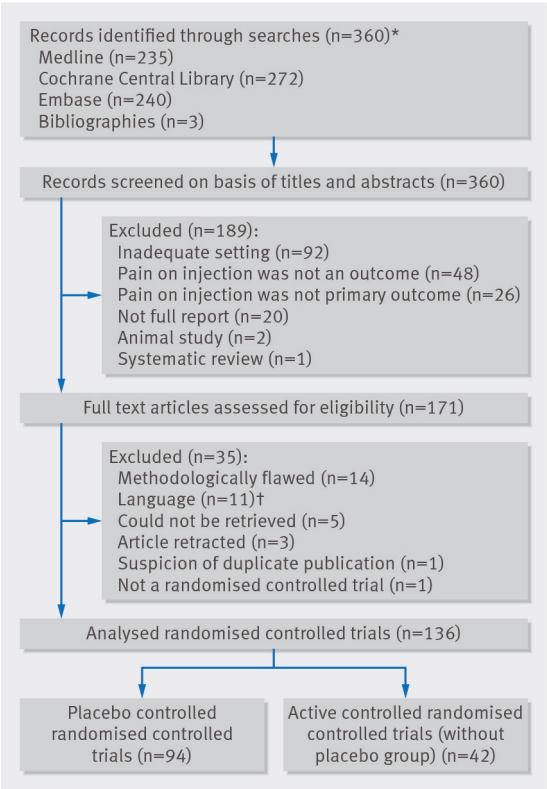

Fig 1 Flow chart of study selection. *Number does not add up as some titles were from more than one database. †Chinese (n=3 trials), Japanese (n=7), and Persian (n=1)

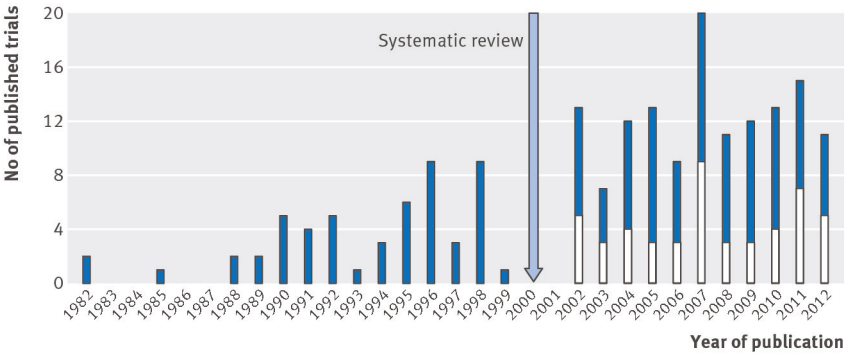

Fig 2 Number of published randomised controlled trials studying efficacy of interventions for prevention of pain from propofol injection. Trials published before 2000 are those included in the Picard review.[5] For the present analysis, searches for trials published after the Picard review included references from 2002 onwards. White bars represent clinically relevant trials