Cochrane Library

Cochrane Database of Systematic Reviews

Monitoring strategies for clinical intervention studies (Protocol)

Klatte K, Pauli-Magnus C, Love S, Sydes M, Benkert P, Bruni N, Ewald H, Arnaiz Jimenez P, Bonde MM, Briel M

Klatte K, Pauli-Magnus C, Love S, Sydes M, Benkert P, Bruni N, Ewald H, Arnaiz Jimenez P, Bonde MM, Briel M.
Monitoring strategies for clinical intervention studies.
Cochrane Database of Systematic Reviews 2019, Issue 11. Art. No.: MR000051.
DOI: 10.1002/14651858.MR000051.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Methodology Protocol]

Monitoring strategies for clinical intervention studies

Katharina Klatte[1], Christiane Pauli-Magnus[1], Sharon Love[2], Matthew Sydes[2], Pascal Benkert[1], Nicole Bruni[1], Hannah Ewald[3], Patricia Arnaiz Jimenez[1], Marie Mi Bonde[1], Matthias Briel[1]

[1]Department of Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland. [2]Medical Research Council (MRC) Clinical Trials Unit, University College London, London, UK. [3]University Medical Library, University of Basel, Basel, Switzerland

Contact address: Christiane Pauli-Magnus, Department of Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland. christiane.pauli-magnus@anti-clutterusb.ch.

Editorial group: Cochrane Methodology Review Group
Publication status and date: New, published in Issue 11, 2019.

Citation: Klatte K, Pauli-Magnus C, Love S, Sydes M, Benkert P, Bruni N, Ewald H, Arnaiz Jimenez P, Bonde MM, Briel M. Monitoring strategies for clinical intervention studies. *Cochrane Database of Systematic Reviews* 2019, Issue 11. Art. No.: MR000051. DOI: 10.1002/14651858.MR000051.

ABSTRACT

This is a protocol for a Cochrane Review (Methodology). The objectives are as follows:

The main objective of this systematic review of prospective empirical studies is to evaluate the benefits and disadvantages of trial-specific, risk-based monitoring strategies compared with a traditional intensive on-site monitoring strategy or other monitoring strategies for randomized and non-randomized prospective intervention studies.

BACKGROUND

Description of the problem or issue

Trial monitoring is requested by the Good Clinical Practice (GCP) Guideline of the International Conference on Harmonisation of Technical Requirements For Registration of Pharmaceuticals for Human Use (ICH) to ensure the safety and rights of study participants, confidentiality of personal information, and quality of data (ICH 1996). Source data verification (SDV) during monitoring visits was estimated to use up to 25% of the sponsor's entire clinical trial budget, even though the association between data quality/participant safety and the extent of monitoring and SDV has not been clearly demonstrated (Funning 2009). Consistent application of intensive on-site monitoring creates financial and logistical barriers to the design and conduct of clinical trials, with no evidence of participant benefit or increase in the quality of clinical research (Baigent 2008; Duley 2008; Hearn 2007; Tudur Smith 2012; Tudur Smith 2014).

Recent developments at international bodies and regulatory agencies such as the European Medicines Agency (EMA), the Organisation for Economic Co-operation and Development (OECD), the European Commission (EC) and the Food and Drug Administration (FDA), as well as the 2016 addendum to ICH E6 GCP, have supported the need for risk-proportionate approaches to clinical trial monitoring and overall trial management (EC 2014; EMA 2013; FDA 2013; ICH 2016; OECD 2013). This has encouraged study sponsors to implement risk assessments in their monitoring plans and to utilize alternative monitoring approaches. There are several publications reporting on the experience of using a risk-based monitoring approach, often including central monitoring, in specific clinical trials (Edwards 2014; Heels-Ansdell 2010; Valdés-Márquez 2011). The conduct of "lower risk" trials — which optimise the use of already authorized medicinal products, validated devices, and implemented interventions — may particularly benefit from a risk-based approach to clinical trial monitoring in terms of timely completion and cost efficiency. Such "lower risk" trials are typically investigator-initiated clinical trials conducted in the academic setting, which either compare therapeutic options within their marketing authorization (comparative effectiveness trials), or explore new indications for already marketed products (OECD 2013). Different risk assessment strategies for clinical trials have been developed, with the objective of defining risk-proportionate monitoring plans (Hurley 2016). Although there is no standardized approach for examining the baseline risk in a clinical protocol, different risk assessment approaches evaluate risks associated with the safety profile of the investigational medicinal product (IMP), the phase of the clinical trial, and the data collection process, and include a combination of central and on-site monitoring components. Central monitoring is based on the evaluation of electronically available study data in order to identify data inconsistencies and study sites with poor data quality or problems in trial conduct (Venet 2012), whereas on-site monitoring comprises site inspection, investigator/staff contact, SDV, observation of study procedures, and the review of regulatory elements of a trial. The OECD classifies risk assessment strategies into stratified approaches and trial-specific approaches, and proposes a harmonized two-pronged strategy based on internationally validated tools for risk assessment and risk mitigation (OECD 2013). The effectiveness of these new risk-based approaches in terms of quality assurance, patient rights and safety and reduction of cost, needs to be empirically assessed. We recently examined the risk-based monitoring approach followed at our own institution (the Clinical Trial Unit and Department of Clinical Research, University Hospital Basel, Switzerland) using mixed methods (von Niederhausern 2017). In addition, three large prospective studies evaluating risk-based monitoring - ADAMON (Brosteanu 2017), OPTIMON (Journot 2015), and TEMPER (Stenning 2018) - have been completed and the first results have been presented, and two further studies evaluating monitoring strategies are being conducted at present: START (Hullsiek 2015) and MONITORING (Fougerou-Leurent 2018).

Description of the methods being investigated

Traditional trial monitoring consists of intensive on-site monitoring strategies comprising frequent on-site visits and up to 100% of SDV. Risk-based monitoring is a new strategy that recognizes that not all clinical trials require the same approach to quality control and assurance (Stenning 2018) and allows for a stratification based on risk indicators assessed during the trial or before it starts. Risk-based strategies differ in their risk assessment approaches as well as in their implementation and extent of on-site and central monitoring components; they are also referred to as risk-adapted or risk-proportionate monitoring strategies. In this review, different monitoring methods will be investigated in terms of their effectiveness in ensuring patient rights and safety, and the validity of trial data. These key elements of clinical trial conduct are assessed by monitoring critical or major violation of GCP objectives, according to the classification of GCP findings described in EMA 2017.

Monitoring strategies

1. The risk-adapted strategy proposed by Brosteanu and colleagues is based on the assessment of the risk associated with an individual trial protocol, classified using a three-level risk graduation that considers the potential risk of a study intervention compared to the standard medical care for the indication in question (Brosteanu 2009). The implementation of risk assessments thus focuses on critical data and procedure describing the risk associated with therapeutic intervention. The implementation of risk assessments is further based on questionnaires that assess the existing knowledge about the investigated therapies as well as indicators for patient-related risks, indicators of robustness, and indicators for site-related risks. Trial-specific risk analysis then proposes a monitoring plan of three different monitoring classes, containing on-site elements as well as central and statistical monitoring methods to a different extent. On-site monitoring should focus on trial aspects that cannot be influenced by other quality management measures or would require substantial resources to do so. According to this approach, monitoring adapted to the risk identified should be used in conjunction with other measures for quality management, including continuous supervision of recruitment and study conduct.
2. The triggered on-site monitoring strategy suggested by the Medicines and Healthcare products Regulatory Agency, Medical Research Council, and UK Department of Health includes an initial risk assessment on the basis of the intervention and design of the trial and a resultant monitoring plan for different trial sites that is continuously updated through centralized monitoring. Over the course of the pre-selected trials, sites are prioritised for on-site visits based on predefined central monitoring triggers (Meredith 2011; Stenning 2018).

3. The consensus risk-assessment scale (RAS) and risk-adapted monitoring plan (RAMP) developed by Journot and colleagues in 2010 consists of a four-level initial risk assessment, leading to monitoring plans of four levels of intensity (Journot 2011). The most intensive monitoring plan for studies with a high risk to participants corresponds to traditional trial monitoring, whereas less intensive plans (for studies with no or a low risk to participants and capabilities for intensive remote monitoring) correspond to no on-site visits (Journot 2011). The intermediate risk level on-site visit with verification of 100% of key data is carried out for 10% of patients, according to a sampling plan determined before the start of the study. The optimized monitoring strategy concentrates on the main scientific and regulatory aspects, compliance with requirements for patient consent and serious adverse events, and the frequency of serious errors concerning the validity of the study's main results, and the trial's eligibility criteria (Chene 2008).
4. A strategy that is mainly based on central monitoring, combined with a local quality control provided by qualified personnel on site, is being evaluated by Hullsiek and colleagues in the START Monitoring sub-study (Hullsiek 2015). In this study, targeted central monitoring uses descriptive statistics on the consistency and quality of the data and data completeness. Semi-annual performance reports are generated for each site, focusing on the key variables/endpoints regarding patients' safety (serious adverse events, eligibility violations) and data quality.
5. The monitoring strategy developed for the MONITORING study is characterized by a targeted SDV in which only regulatory and scientific key data were verified. This strategy is compared to full SDV and assessed based on final data quality and costs.
6. The value of SDV has also been assessed in a clinical trial of cancer, where the generation of the data base for the outcome data is accompanied only by central monitoring techniques and compared to a 100% SDV of the trial outcome data (Tudur Smith 2012).
7. A new strategy of remote SDV is being assessed in a pilot study where documents are accessed via electronic health records, clinical data repositories, web-based-access technologies, or authentication and auditing tools (Mealer 2013).

We will also include further methods of risk-based monitoring for prospective intervention studies if these are identified through our systematic literature search.

How these methods might work

The hope is that risk-based monitoring methods with reduced on-site monitoring will increase the cost-effectiveness of trials while being non-inferior for major or critical violation of essential GCP objectives, according to EMA 2017. The risk assessment preceding the risk-based monitoring plan should consider the likelihood of errors occurring in key aspects of study performance, and the anticipated effect of such errors on the protection of human participants and the reliability of the trial results (Landray 2012). Trials within a certain risk category are initially assigned to a defined monitoring strategy which remains adjustable throughout the conduct of the trial and should always match the needs of the study and specific trial sites. This flexibility is an advantage, considering the heterogeneity of study designs and participating trial sites. In addition, central monitoring would allow for continuous verification of data quality based on pre-specified triggers and thresholds, and would enable early intervention in the trial conduct in cases of procedural or data-recording errors. Besides the detection of missing or invalid data, trial entry procedures and protocol adherence, as well as other performance indicators, can be monitored through a continuous analysis of electronically captured data (Baigent 2008). In addition, comparison with external sources may be undertaken to validate information contained in the data set; and the identification of poorly performing sites would ensure a more targeted application of on-site monitoring resources. Utilization of methods that take advantage of the increasing use of electronic systems (e.g. electronic Case Report Forms (CRFs)) would allow data to be checked by automated means and would apply entry rules supporting up-to-date, high-quality data; these methods would also ensure patient rights and safety while simultaneously improving trial management and optimizing trial conduct. Adaptations in the monitoring approach towards a reduction of on-site monitoring visits, provided that patient rights and safety are ensured, could allow the application of resources to the most crucial study components (Journot 2011).

In order to evaluate whether these new risk-based monitoring approaches are non-inferior to the traditional extensive on-site monitoring, an assessment of differences in critical and major findings during monitoring activities is essential. Monitoring findings will be determined with respect to patient safety, patient rights and reliability of the data, and classified as critical and major according to the classification of GCP findings described in the *Procedures for reporting of GCP inspections requested by the Committee for Medicinal Products for Human Use* (EMA 2017). Critical findings are conditions, practices or processes that adversely affect the rights, safety or well-being of the participants and/or the quality and integrity of data. Major findings are conditions, practices or processes that might adversely affect the rights, safety or well-being of the participants and/or the quality and integrity of data.

Why it is important to do this review

There is insufficient information to guide the choice of monitoring methods consistent with the ICH GCP to be used in a specific trial set up, and there is a lack of evidence on the effectiveness of various new monitoring approaches. This has resulted in high heterogeneity of monitoring practices used by research institutions, especially in the academic setting. A guideline describing which kind of monitoring strategy is most effective for clinical trials in terms of patient rights and safety, and data quality, is urgently needed for the academic clinical trial setting. Evaluating the benefits and disadvantages of different risk-based monitoring strategies, incorporating components of central and/or targeted and triggered monitoring versus intensive on-site monitoring, might lead to a consensus on how effective these new approaches are. In addition, evaluating the evidence of effectiveness could provide information on the extent to which on-site monitoring content (such as SDV or frequency of site visits) can be adapted or supported by central monitoring interventions. In this regard, we will explore whether monitoring that incorporates central and statistical components can be extended to support the overall management of study quality in terms of participant recruitment and follow-up.

The first three large prospective studies (ADAMON, OPTIMON, and TEMPER) evaluating different approaches to risk-based monitoring for the academic setting have recently been completed (Brosteanu 2017; Journot 2015; Stenning 2018). Further studies are being conducted (START and MONITORING (Fougerou-Leurent 2018; Hullsiek 2015)). The risk-based monitoring interventions being evaluated in-

corporate on-site and central monitoring components, which may vary in terms of extent and procedural structure. In line with the recommendation from the Clinical Trials Tranformation Initiatative (Grignolo 2011), it is crucial to systematically analyze and compare the existing evidence so that best practices may be established. This review will facilitate the sharing of current knowledge on effective monitoring strategies, helping trialists, support units, and monitors to choose the best strategy for their trials. Evaluation of the impact of a change of monitoring approaches on data quality and study cost is relevant for the effective adjustment of current monitoring strategies. In addition, evaluating the effectiveness of these new monitoring approaches in comparison with intensive on-site monitoring might reveal possible methods to replace or support on-site monitoring strategies by taking advantage of the increasing use of electronic systems and resulting opportunities to implement statistical analysis tools.

OBJECTIVES

The main objective of this systematic review of prospective empirical studies is to evaluate the benefits and disadvantages of trial-specific, risk-based monitoring strategies compared with a traditional intensive on-site monitoring strategy or other monitoring strategies for randomized and non-randomized prospective intervention studies.

METHODS

Criteria for considering studies for this review

Types of studies

We will include randomized or non-randomized prospective, empirical evaluation studies of different monitoring strategies in one or more prospective intervention studies. These types of embedded studies have recently been called "studies within a trial" (SWAT) (Anon 2012).

We will follow the Cochrane Effective Practice and Organisation of Care (EPOC) Group definitions of these study designs (EPOC 2016)

Types of data

We will extract information about monitoring processes as well as evaluations of the comparison and advantages/disadvantages of different monitoring approaches. We will include data from published and unpublished studies, and grey literature, that compare different monitoring strategies (e.g. standard monitoring versus a risk-based approach).

Study characteristics of interest are:

1. monitoring interventions;
2. risk assessment characteristics;
3. finding rates of serious/critical audits;
4. impact on patient recruitment and follow-up; and
5. costs.

Types of methods

We will include studies that compare:

1. a risk-based monitoring strategy versus an intensive on-site monitoring strategy for prospective intervention studies; or
2. any other monitoring strategies for prospective intervention studies.

Types of outcome measures

Specific outcome measures might differ between studies included in this review. They are therefore not part of the eligibility criteria.

Primary outcomes

The primary outcomes are critical and major monitoring findings of prospective intervention studies.

Critical and major findings will be defined according to the classification of GCP findings described in EMA 2017, as follows.

1. Critical findings: conditions, practices or processes that adversely affect the rights, safety or well-being of the study participants or the quality and integrity of data. Observations classified as critical may include a pattern of deviations classified as major, bad quality of the data and/or absence of source documents. Manipulation and intentional misrepresentation of data re included in this group.
2. Major findings: conditions, practices or processes that might adversely affect the rights, safety or well-being of the study participants and/or the quality and integrity of data. Major observations are serious deficiencies and are direct violations of GCP principles. Observations classified as major may include a pattern of deviations or numerous minor observations (or both).

Secondary outcomes

1. Impact of data monitoring strategy on patient recruitment and follow-up.
2. Economic data (costs) will be assessed when available, to evaluate the effect of different monitoring strategies on resource use.
3. Method characteristics within the group of risk-proportionate monitoring will be collected to provide an overview of existing methods with a special focus on central monitoring aspects.
4. Qualitative research data or process evaluations of the monitoring interventions will be collected, if available.
5. Individual components of the primary outcome:
   a. major eligibility violations;
   b. major informed-consent violations;
   c. findings that raise doubt about the accuracy or credibility of key trial data and deviations of intervention from the trial protocol (with impact on patient safety or data validity);
   d. errors in endpoint assessment; and
   e. errors in serious adverse event reporting.

Search methods for identification of studies

Electronic searches

We will conduct a comprehensive search using a search strategy that we developed together with an experienced librarian scientist (HE). MEDLINE, Embase, and CENTRAL will be systematically searched for relevant published literature, using the search strategy shown below. We will also search the online SWAT repository (go.qub.ac.uk/SWAT-SWAR) for SWAT examining different methods for trial monitoring. We will not apply any restrictions in the search or the selection process regarding language or date of publication.

We will use the following terms to identify prospective studies that compare different strategies for trial monitoring:

1. triggered monitoring;
2. targeted monitoring;
3. risk-adapted monitoring;
4. risk adapted monitoring;
5. risk-based monitoring;
6. risk based monitoring;
7. centralized monitoring;
8. centralised monitoring;
9. statistical monitoring;
10. on site monitoring;
11. on-site monitoring;
12. monitoring strategy;
13. monitoring method;
14. monitoring technique;
15. trial monitoring; and
16. central monitoring.

The search for host trials is intended to identify prospective intervention studies, based on their title or abstract. In addition, we will apply a high-sensitivity filter for randomized trials in humans to identify host randomized trials (Higgins 2019) to produce the following search strategy for MEDLINE.

("on site monitoring"[tiab] OR "on-site monitoring"[tiab] OR "monitoring strategy"[tiab] OR "monitoring method"[tiab] OR "monitoring technique"[tiab] OR "triggered monitoring"[tiab] OR "targeted monitoring"[tiab] OR "risk-adapted monitoring"[tiab] OR "risk adapted monitoring"[tiab] OR "risk-based monitoring"[tiab] OR "risk based monitoring"[tiab] OR "risk proportionate"[tiab] OR "centralized monitoring"[tiab] OR "centralised monitoring"[tiab] OR "statistical monitoring"[tiab] OR "central monitoring"[tiab]) AND ("prospective" [tiab] OR "prospectively" [tiab] OR randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans[mh])

Searching other resources

We will search reference lists of included studies and similar systematic reviews to find additional relevant resources (Horsley 2011). In addition, we will search the grey literature (i.e. conference proceedings of the Society for Clinical Trials and the International Trials Methodology Conference), and trial registries (ClinicalTrials.gov, the WHO International Clinical Trials Registry Platform, the European Union Drug Regulating Authorities Clinical Trials Database, and the ISRCTN) for ongoing or unpublished prospective studies. Finally, we plan to collaborate closely with researchers of already identified eligible studies (e.g. OPTIMON, ADAMON, START, MONITORING) to identify additional studies (and unpublished data, if needed).

Data collection and analysis

Data collection and analysis methods will be based on the recommendations described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019) and *Methodological Expectations for the Conduct of Cochrane Intervention Reviews* (Higgins 2016).

Selection of studies

After elimination of duplicate records, two review authors will independently screen titles and abstracts for eligibility. Potentially relevant studies will be retrieved as full-text reports; two review authors will independently assess these for eligibility, applying prespecified criteria (see: Criteria for considering studies for this review). Any disagreements between review authors will be resolved by discussion until consensus is reached, or by involving a third review author (MB). We will document the study selection process in an appropriate flow diagram, as described in the PRISMA statement (Moher 2009).

Data extraction and management

For each eligible study (and where unpublished data are provided by the original researchers) two review authors will independently extract information on a number of key characteristics, using electronic data collection forms. Data will be extracted in Eppi-Reviewer 4 (Thomas 2010). Any disagreements will be resolved by discussion until consensus is reached, or by involving a third review author. We will contact authors of studies directly when target information is unreported or unclear, in order to clarify or complete extracted data. Collected data will be summarized qualitatively (and quantitatively, if possible) in the 'Results' section of the review. If meta-analysis of the primary or secondary outcomes is not applicable due to considerable methodological heterogeneity between studies, the results will be reported qualitatively only.

Study characteristics to be extracted include the following.

1. General information about the study: title, authors, year of publication, language, country, funding sources.
2. Methods: study design, allocation method, study duration, stratification of sites (stratified on risk level, country, projected enrolment, etc.).
3. Population of trials (characteristics of host trials will be extracted):
   a. design (randomized clinical trial (RCT) or other prospective intervention trial);
   b. setting (primary care, tertiary care, community, etc.);
   c. national or multinational;
   d. study population;
   e. total number of sites randomized/analyzed;
   f. inclusion/exclusion criteria;
   g. IMP risk category;
   h. support from clinical trial unit or clinical research organization for host trial or evidence for experienced research team; and
   i. trial phase.

4. Intervention (components related to the applied monitoring strategy, including theoretical basis):
   a. number of sites randomized/allocated to groups (specifying number of sites or clusters);
   b. duration of intervention period;
   c. risk assessment characteristics (follow-up questions)/triggers or thresholds that induce on-site monitoring (follow-up questions);
   d. frequency of monitoring visits;
   e. extent of on-site monitoring;
   f. frequency of central monitoring reports;
   g. number of monitoring visits per patient;
   h. cumulative monitoring time on-site;
   i. average number of monitoring visits per site;
   j. delivery (procedures used for central monitoring: structure/components of on-site monitoring/triggers/thresholds);
   k. who performed the monitoring (study team, trial staff; qualifications of monitors);
   l. degree of source data verification (median number of patients undergoing SDV); and
   m. co-interventions (site/study-specific co-interventions)
5. Outcomes: primary and secondary outcomes, components of combined primary outcome, outcome measures and scales, time points of measurement, statistical analysis of outcome data.
6. Data to assess the risk of bias of included studies, e.g. random sequence generation, allocation concealment, blinding of outcome assessors, performance bias, selective reporting, or other sources of bias.

Assessment of risk of bias in included studies

Two review authors will independently assess the risk of bias in each included study using the criteria described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019) and by the Cochrane EPOC Review Group (EPOC 2017). The domains provided by these criteria will be evaluated for all included studies and ratings of low, high or unclear risk of bias will be assigned. Further assessment of methodological quality of the studies will be performed and evaluated. We will assess randomized and non-randomized studies, as follows.

Selection bias

Generation of the allocation sequence

1. If sequence generation is truly random (e.g. computer generated): low risk.
2. If sequence generation is not specified and we are unable to obtain relevant information from study authors: unclear risk.
3. If there is a quasi-random sequence generation (e.g. alternation): high risk.
4. Non-randomized trials: high risk.

Concealment of the allocation sequence (steps taken prior to the assignment of intervention to ensure that knowledge of the allocation is not possible)

1. If opaque, sequentially numbered envelopes were used or central randomization was performed by a third party: low risk.
2. If the allocation concealment is not specified and we are unable to ascertain whether the allocation concealment was protected before and until assignment: unclear risk.
3. Non-randomized trials and studies using inadequate allocation concealment: high risk.

For non-randomized studies we will further assess:

1. attempt to balance groups by design (control for selection bias); and
2. control for risk of confounding.

Performance bias

It will not be possible to blind participating sites and monitors to the intervention to which they have been assigned because of the procedural differences of monitoring strategies.

Detection bias (blinding of the outcome assessor)

1. If the assessors performing audits have knowledge of the intervention and thus outcomes were not assessed blindly: high risk.
2. If we cannot ascertain whether assessors were blinded and study authors do not provide information to clarify: unclear risk.
3. If outcomes were assessed blindly: low risk.

Attrition bias

We do not expect to have missing data for our primary outcome (the primary outcome will be the rates of serious/critical audit findings at the end of host randomised trials or other prospective intervention studies; and as missing patients will not be audited, missing data in the proportion of critical findings is not expected). However, for the statistical power of the individual study outcomes, missing data for participants and site accrual is an issue and will be discussed in the review.

Selective reporting bias

We will investigate whether all outcomes mentioned in available study protocols, registries, or methodology sections of study publications are reported in results sections.

1. If all outcomes in the methodology or outcomes specified in the study protocol are not reported in the results, or if outcomes reported in the results are not listed in the methodology or in the protocol: high risk.
2. If outcomes are only partly reported in the results, or if an obvious outcome is not mentioned in the study: high risk.
3. If all outcomes are listed in the protocol/methodology section and reported in the results: low risk.

Other potential sources of bias

1. If there is one or more important risk of bias (e.g. flawed study design): high risk.
2. If there is incomplete information regarding a problem which may lead to bias: unclear risk.
3. If there is no evidence of other sources of bias: low risk.

Measures of the effect of the methods

We will conduct a comparative analysis of the impact of different risk-based monitoring strategies on data quality and patient rights and safety measures, for example by the proportion of critical find-

ings. It is not clear if statistical pooling of results of studies will be appropriate or feasible given the heterogeneity of studies we expect to include.

Dichotomous data will be analysed using a risk ratio or odds ratio with a 95% confidence interval (CI). Continous data will be analyzed using mean differences with a 95% CI if the measurement scale is the same. If the scale is different, standardized mean differences, with 95% CIs, will be used.

Unit of analysis issues

Included studies may differ in outcomes chosen to assess the effects of the respective monitoring strategy. Critical/serious audit findings may be reported on a patient level, per finding event, or per site. Furthermore, components of the primary endpoints may vary between studies. We will specify the study outcomes as defined in the study protocols or reports, and only pool outcomes that are based on similar definitions. In addition, we might compare individual components of the primary outcome if these are consistently defined across studies (e.g. eligibility violations).

Any cluster-randomized trials that we include in the review will be identified as such. We will report the baseline comparability of clusters and consider statistical adjustment, if this might help to reduce an imbalance. We will estimate the intra-cluster correlation coefficient (ICC), as described by Higgins 2019, using information from the study (if it is available) or from an external estimate from a similar study. If we do this, we will conduct sensitivity analyses to explain variation in ICC values.

Dealing with missing data

Authors of included studies will be contacted in an attempt to obtain unpublished data or additional information of value for this review (Young 2011). Where a study has been registered and a relevant outcome is specified in the study protocol but no results were reported, we will contact the authors and sponsors to request study reports. We will create a table to summarize the results for each outcome. We will narratively explore the potential impact of missing data in the 'Discussion' section of the review.

Assessment of heterogeneity

We have specified that we will include non-randomized trials in this review, which may lead to increased statistical heterogeneity. A subgroup analysis might be appropriate, looking at monitoring strategies using very similar approaches and consistent outcomes. If we identify methodological heterogeneity, we will not pool results in a meta-analysis. Instead we will qualitatively synthesize results, grouping studies with similar designs and interventions together, and describing existing methodological heterogeneity (e.g. use of different methods to assess outcomes). If study characteristics, methodology and outcomes are sufficiently similar across studies, we will quantitatively pool results in a meta-analysis and assess heterogeneity by visually inspecting forest plots of included studies (location of point estimates and the degree to which confidence intervals overlap), and considering the results of the Chi^2 test for heterogeneity and the I^2 statistic. We will follow the guidance outlined in Higgins 2019 to quantify statistical heterogeneity using the I^2 statistic:

1. 0% to 40% might not be important;
2. 30% to 60% may represent moderate heterogeneity*;
3. 50% to 90% may represent substantial heterogeneity*;
4. 75% to 100%: considerable heterogeneity*.

*The importance of the observed value of I^2 depends on the magnitude and direction of effects, and the strength of evidence for heterogeneity (e.g. P value from the Chi^2 test, or a confidence interval for I^2). If our I^2 value indicates that heterogeneity is a possibility and either the Tau^2 is greater than zero, or the P value for the Chi^2 test is low (less than 0.10), heterogeneity may be due to a factor other than chance.

Possible sources of heterogeneity from the characteristics of host trials include:

1. design (randomized or other prospective intervention trial);
2. setting (primary care, tertiary care, community, etc.);
3. IMP risk category;
4. trial phase;
5. national or multinational;
6. support from a clinical trial unit or clinical research organization for host trial or evidence for an experienced research team; and
7. study population.

Possible sources of heterogeneity from the characteristics of methodology studies include:

1. study design;
2. components of outcome;
3. method of outcome assessment;
4. level of outcome (patient/site); and
5. classification of monitoring findings.

Assessment of reporting biases

To decrease the risk of publication bias affecting the findings of the review, we will apply various search approaches using different resources. These include grey literature searching and checking reference lists. If 10 or more studies are included in a meta-analysis, we will create a funnel plot to investigate whether bias may exist unless all studies are of a similar size. If we notice asymmetry, we cannot conclude that reporting biases exist, but we will consider the sample sizes and presence (and possible influence) of outliers. We will discuss potential explanations, such as publication bias or poor methodological quality of included studies, and perform sensitivity analyses.

Data synthesis

Data will be synthesized using tables to compare different monitoring strategies. We will also report results by different study designs. This will be accompanied by a descriptive summary in the 'Results' section of the review. We will use Review Manager 5 software (Review Manager 2014) to conduct our statistical analysis and undertake meta-analysis, if it is deemed appropriate. If meta-analysis of the primary or secondary outcomes is not applicable because of considerable methodological heterogeneity between studies, the results will be reported qualitatively.

Two review authors will assess the quality of the evidence. Based on the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019) and GRADE (Guyatt 2013a; Guyatt 2013b), we will create 'Summary of findings' tables for the

main comparisons of the review. We will present the following primary and secondary outcomes for each comparison: critical/serious audit findings at the end of host trials; economic data; impact of monitoring strategy on data management in terms of recruitment and follow-up; and other main outcomes of interest, including specification of subgroups for critical/serious findings, as outlined in Types of outcome measures (e.g. major eligibility violations). We will describe the study settings and number of sites addressing each outcome. For each assumed risk cited in the table(s), we will provide a source and rationale, and we will implement the GRADE system to assess the quality of the evidence using GRADEpro GDT software or the GRADEpro GDT app (GRADEpro GDT 2015). If meta-analysis is not appropriate or the units of analysis cannot be compared, we will present results in a narrative 'Summary of findings' table. If we do this, the imprecision of the evidence will be an issue of concern due to the lack of a quantitative effect measure.

Subgroup analysis and investigation of heterogeneity

If visual inspection of the forest plots, Chi^2 test, I^2 statistic, and Tau^2 indicate that statistical heterogeneity could be present, we will carry out a subgroup analysis. A subgroup analysis will be deemed appropriate if the included studies satisfy criteria assessing the credibility of subgroup analyses (Oxman 1992; Sun 2010).

The following are our a priori subgroups: type if intervention characteristics (e.g. predominantly central monitoring versus risk-based monitoring), and type of study (randomized versus non-randomized).

Sensitivity analysis

We plan to conduct sensitivity analyses restricted to:

1. peer-reviewed and published studies only (i.e. excluding unpublished studies); and
2. studies at low risk of bias only (i.e. excluding non-randomized studies and randomized studies without allocation concealment; Assessment of risk of bias in included studies).

ACKNOWLEDGEMENTS

The authors would like to thank the monitoring team of the Department of Clinical Research (Clinical Trial Unit Basel), including Klaus Ehrlich, Petra Forst, Emilie Müller, and Astrid Roesler, for sharing their experience and contributing to discussions on monitoring procedures.

REFERENCES

Additional references

Anon 2012

Anon. Education section—Studies Within A Trial (SWAT). *Journal of Evidence-based Medicine* 2012;**5**:44-5.

Baigent 2008

Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials* 2008;**5**:49-55.

Brosteanu 2009

Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. *Clinical Trials* 2009;**6**:585-96.

Brosteanu 2017

Brosteanu O, Schwarz G, Houben P, Paulus U, Strenge-Hesse A, Zettelmeyer U, et al. Risk-adapted monitoring is not inferior to extensive on-site monitoring: results of the ADAMON cluster-randomised study. *Clinical Trials* 2017;**14**:584-96.

Chene 2008

Chene G. Evaluation of the efficacy and cost of two monitoring strategies for public clinical research. OPTIMON study: OPTImisation of MONitoring. ssl2.isped.u-bordeaux2.fr/OPTIMON/DOCS/OPTIMON%20-%20Protocol%20v12.0%20EN%202008-04-21.pdf. Available from: https://ssl2.isped.u-bordeaux2.fr/OPTIMON/Documents.aspx: Unité de Soutien Méthodologique à la Recherche Clinique et Epidémiologique du, (accessed 2 October 2019).

Duley 2008

Duley L, Antman K, Arena J, Avezum A, Blumenthal M, Bosch J, et al. Specific barriers to the conduct of randomised trials. *Clinical Trials* 2008;**5**:40-8.

EC 2014

European Commission. Risk proportionate approaches in clinical trials. Recommendations of the expert group on clinical trials for the implementation of Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use. ec.europa.eu/health/files/clinicaltrials/2016_06_pc_guidelines/gl_4_consult.pdf (accessed 10 November 2018).

Edwards 2014

Edwards P, Shakur H, Barnetson L, Prieto D, Evans S, Roberts I. Central and statistical data monitoring in the clinical randomisation of an antifibrinolytic in significant haemorrhage (CRASH-2) trial. *Clinical Trials* 2014;**11**:336-43.

EMA 2013

European Medicines Agency. Reflection paper on risk based quality management in clinical trials. ema.europa.eu/docs/en_GB/document_library/Scientific_guidelines/2013/11/WC500155491.pdf (accessed 10 November 2018).

EMA 2017

European Medicines Agency. Procedures for reporting of GCP inspections requested by the Committee for Medicinal Products for Human Use. ema.europa.eu/en/documents/regulatory-procedural-guideline/ins-gcp-4-procedure-reporting-good-clinical-practice-inspections-requested-chmp_en.pdf 2017 (accessed 16 September 2019).

EPOC 2016

Effective Practice Organisation of Care. What study designs should be included in an EPOC review and what should they be called? EPOC resources for review authors. epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/EPOC%20Study%20Designs%20About.pdf 2016 (accessed 16 September 2019).

EPOC 2017

Effective Practice Organisation of Care. Suggested risk of bias criteria for EPOC reviews. EPOC Resources for review authors. epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/suggested_risk_of_bias_criteria_for_epoc_reviews.pdf 2017 (accessed 16 September 2019).

FDA 2013

US Department of Health and Human Services Food and Drug Administration. Guidance for industry oversight of clinical investigations - a risk-based approach to monitoring. www.fda.gov/downloads/Drugs/Guidances/UCM269919.pdf (accessed 10 November 2018).

Fougerou-Leurent 2018

Fougerou-Leurent C, Laviolle Bruno, Bellissant E. Cost-Effectiveness of Full Versus Targeted Monitoring of Randomized Controlled Trials. Toulouse, France: Fundamental and Clinical Pharmacology, French Society of Pharmacology and Therapeutics, and INSERM Clinical Research Centers (CIC), 2018.

Funning 2009

Funning S, Grahnén A, Eriksson K, Kettis-Linblad A. Quality assurance within the scope of Good Clinical Practice (GCP) - what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members. *Qualty Assurance Journal* 2009;**12**(1):3-7. [DOI: 10.1002/qaj.433]

GRADEpro GDT 2015 [Computer program]

GRADE Working Group. GRADEpro GDT: GRADEpro Guideline Development Tool. McMaster University (developed by Evidence Prime Inc). Available from gradepro.org, 2015.

Grignolo 2011

Grignolo A. The Clinical Trials Transformation Initiative (CTTI). *Annali dell'Istituto Superiore di Sanita* 2011;**47**:14-8. [DOI: 10.4415/ANN_11_01_04; PUBMED: 21430332]

Guyatt 2013a

Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings

tables-binary outcomes. *Journal of Clinical Epidemiology* 2013;**66**:158-72.

Guyatt 2013b

Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *Journal of Clinical Epidemiology* 2013;**66**:173-83.

Hearn 2007

Hearn J, Sullivan R. The impact of the 'Clinical Trials' directive on the cost and conduct of non-commercial cancer trials in the UK. *European Journal of Cancer* 2007;**43**:8-13.

Heels-Ansdell 2010

Heels-Ansdell D, Walter S, Zytaruk N, Guyatt G, Crowther M, Warkentin T, et al. Central statistical monitoring of an international thromboprophylaxis trial. *American Journal of Respiratory and Critical Care Medicine* 2010;**181**:A6041.

Higgins 2016

Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews. London: Cochrane, 2016.

Higgins 2019

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 (updated July 2019). Available from training.cochrane.org/handbook. Cochrane, 2019.

Horsley 2011

Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews* 2011, Issue 8. [DOI: 10.1002/14651858.MR000026.pub2]

Hullsiek 2015

Hullsiek KH, Kagan JM, Engen N, Grarup J, Hudson F, Denning ET, et al. Investigating the efficacy of clinical trial monitoring strategies: design and implementation of the cluster randomized START monitoring substudy. *Therapeutic Innovation and Regulatory Science* 2015;**49**(2):225-33. [DOI: 10.1177/2168479014555912; PUBMED: 25973346]

Hurley 2016

Hurley C, Shiely F, Power J, Clarke M, Eustace JA, Flanagan E, et al. Risk based monitoring (RBM) tools for clinical trials: A systematic review. *Contemporary Clinical Trials* 2016;**51**:15-27.

ICH 1996

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6 (R1). www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf (accessed 10 November 2018).

ICH 2016

International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. Intergrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6R(2). database.ich.org/sites/default/files/E6_R2_Addendum.pdf (accessed 2 October 2019).

Journot 2011

Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Metro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies--the Pre-Optimon study. *Contemporary Clinical Trials* 2011;**32**:16-24.

Journot 2015

Journot V. OPTIMON - first results of the French trial on optimisation of monitoring. ssl2.isped.u-bordeaux2.fr/OPTIMON/docs/Communications/2015-Montpellier/OPTIMON%20-%20EpiClin%20Montpellier%202015-05-20%20EN.pdf (accessed 2 October 2019).

Landray 2012

Landray MJ, Grandinetti C, Kramer JM, Morrison BW, Ball L, Sherman RE. Clinical trials: rethinking how we ensure quality. *Drug Information Journal* 2012;**46**:657-60. [DOI: 10.1177/0092861512464372]

Mealer 2013

Mealer M, Kittelson J, Thompson BT, Wheeler AP, Magee JC, Sokol RJ, et al. Remote source document verification in two national clinical trials networks: a pilot study. *PLoS One* 2013;**8**:e81890.

Meredith 2011

Meredith S, Ward M, Booth G, Fisher A, Gamble C, House H, Landray M. Risk-adapted approaches to the management of clinical trials: guidance from the Department of Health (DH) / Medical Research Council (MRC)/Medicines and Healthcare Products Regulatory Agency (MHRA) Clinical Trials Working Group. *Trials* 2011;**12**:A39.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**:1006-12.

OECD 2013

Organisation for Economic Co-operation and Development. OECD recommendation on the governance of clinical trials. oecd.org/sti/inno/oecdrecommendationonthegovernanceofclinicaltrials.htm. Available from: https://www.pharmamedtechbi.com/~/media/Supporting%20Documents/The%20Gray%20Sheet/39/9/oecdrecommendationgovernanceofclinicaltrials.pdf, (accessed 10 November 2018).

Oxman 1992

Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Annals of Internal Medicine* 1992;**116**:78-84.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Stenning 2018

Stenning SP, Cragg WJ, Joffe N, Diaz-Montana C, Choudhury R, Sydes MR, et al. Triggered or routine site monitoring visits for randomised controlled trials: results of TEMPER, a prospective, matched-pair study. *Clinical Trials* 2018;**15**:600-9.

Sun 2010

Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;**340**:c117.

Thomas 2010 [Computer program]

Thomas J, Brunton J, Graziosi S. EPPI-Reviewer: software for research synthesis. EPPI-Centre Software. Version 4.0. London: Social Science Research Unit, Institute of Education, University of London, 2010.

Tudur Smith 2012

Tudur Smith C, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, et al. The value of source data verification in a cancer clinical trial. *PLoS One* 2012;**7**:e51623.

Tudur Smith 2014

Tudur Smith C, Williamson P, Jones A, Smyth A, Hewer SL, Gamble C. Risk-proportionate clinical trial monitoring: an example approach from a non-commercial trials unit. *Trials* 2014;**15**:127.

Valdés-Márquez 2011

Valdés-Márquez E, Hopewell CJ, Landray M, Armitage J. A key risk indicator approach to central statistical monitoring in multicentre clinical trials: method development in the context of an ongoing large-scale randomized trial. *Trials* 2011;**12**(Suppl 1):A135.

Venet 2012

Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, et al. A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials* 2012;**9**:705-13.

von Niederhausern 2017

von Niederhausern B, Orleth A, Schadelin S, Rawi N, Velkopolszky M, Becherer C, et al. Generating evidence on a risk-based monitoring approach in the academic setting - lessons learned. *BMC Medical Research Methodology* 2017;**17**:26.

Young 2011

Young T, Hopewell S. Methods for obtaining unpublished data. *Cochrane Database of Systematic Reviews* 2011, Issue 11. [DOI: 10.1002/14651858.MR000027.pub2]

CONTRIBUTIONS OF AUTHORS

KK, CPM, and MB conceived the study and wrote the first draft of the protocol. SL, MS, PB, NB, HE, PAJ, and MMH reviewed the protocol and suggested changes for improvement. KK and HE developed the search strategy and will conduct the search. KK, CPM, and MB will select the studies and assess risk of bias and certainty of evidence. KK and MMH will extract relevant data from included studies.

DECLARATIONS OF INTEREST

Katharina Klatte declares to have no conflicts of interest. Christiane Pauli-Magnus declares to have no conflicts of interest. Sharon Love declares to have no conflicts of interest. Matthew Sydes is a co-investigator in a potentially eligible study (TEMPER), but will have no role in study selection, risk of bias or certainty of evidence assessment. He has no other conflicts to declare. Pascal Benkert declares to have no conflicts of interest. Nicole Bruni declares to have no conflicts of interest. Patricia Arnaiz Jimenez declares to have no conflicts of interest. Marie Mi Bonde declares to have no conflicts of interest. Matthias Briel declares to have no conflicts of interest.