

Title: Assessing the potential for prevention or earlier detection of on-site monitoring findings from randomised controlled trials: further analyses of findings from the prospective TEMPER triggered monitoring study

Running Head: Alternatives to on-site monitoring

Version + Date: Version 1.0, 25th January 2020

Word count: Abstract: 425
Text (excluding abstract, references, appendix, tables and figures): 3517

Tables/Figures: 3 Tables, 2 Figures, 3 Supplementary Tables

Supplement: 2 files

Authors:	Name	Affiliation
	William J Cragg ORCID: 0000-0002-1274-8521	1: MRC Clinical Trials Unit at UCL, London, UK; 2: Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
	Caroline Hurley ORCID: 0000-0002-3793-3408	Health Research Board-Trials Methodology Research Network (HRB-TMRN), National University of Ireland, Galway, Ireland
	Victoria Yorke-Edwards ORCID: 0000-0002-6465-7330	MRC Clinical Trials Unit at UCL, London, UK
	Sally P Stenning ORCID: 0000-0001-7009-5488	MRC Clinical Trials Unit at UCL, London, UK

Correspondence: William Cragg
w.cragg@leeds.ac.uk

Tel: 0113 343 8398

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Clinical Trials Research Unit
Leeds Institute of Clinical Trials Research
University of Leeds
Leeds
LS2 9JT,
United Kingdom

Research support: Medical Research Council (MC_UU_12023/24)
The TEMPER study was funded by a grant from Cancer Research UK (C1495/A13305)

Summary of figures and tables

Table 1: summary of Major and Critical findings at TEMPER monitoring visits.

Table 2: summary of stakeholder groups invited to participate in the consultation exercise, completion rates, clinical trials experience and number of processes reviewed.

Table 3: on-site monitoring findings judged not centrally detectable or preventable, following consensus exercise on feasibility of **suggested processes**.

Figure 1: a suggested relationship between preventative measures, monitoring and risk.

Figure 2: flow diagram describing the current study.

Summary of supplementary information

1. Online Tables S1

- **Online Table S1a:** all processes agreed after the consensus exercise to be feasible or feasible if adjusted in specific ways, and the number and proportion of findings each addresses, overall and by subtype.
- **Online Table S1b:** all processes agreed after the consensus exercise to be not feasible or of uncertain feasibility, and the number and proportion of findings each addresses, overall and by subtype.
- **Online Table S1c:** processes that could address on-site monitoring findings, but indirectly or without being completely fool-proof, with feasibility rating and,

where applicable, suggested adjustments to improve feasibility or reasons why infeasible.

2. Consultation exercise text

1

1 **Abstract**

2

3 **Background/Aims**

4 Clinical trials should be designed and managed to minimise important errors with potential
5 to compromise patient safety or data integrity, employ monitoring practices that detect and
6 correct important errors quickly, and take robust action to prevent repetition. Regulators
7 highlight the use of risk-based monitoring, making greater use of centralised monitoring and
8 reducing reliance on centre visits. The TEMPER study was a prospective evaluation of
9 triggered monitoring (a risk-based monitoring method), whereby centres are prioritised for
10 visits based on central monitoring results. Conducted in three UK-based randomised cancer
11 treatment trials of investigational medicine products with time-to-event outcomes, it found
12 high levels of serious findings at triggered centre visits but also at visits to matched control
13 centres that, based on central monitoring, were not of concern. Here, we report a detailed
14 review of the serious findings from TEMPER centre visits. We sought to identify feasible,
15 centralised processes which might detect or prevent these findings without a centre visit.

16 **Methods**

17 We considered a representative example of each finding type through a three-staged
18 consensus exercise. 1: Three trialists independently reviewed the issues and graded their
19 potential for detection by central monitoring or prevention by some described process, each
20 on a five-point scale. 2. Results of round 1 were shared anonymously and each trialist re-
21 viewed the list choosing whether or not to change their grading. 3. The three trialists
22 discussed and agreed grades for any issues without consensus, and proposed a feasibility
23 score for each proposed process. In a consultation exercise, the proposed processes were
24 reviewed and rated for feasibility by an invited external trialist group. The primary outcome

1 was the proportion of all major and critical findings theoretically detectable or preventable
2 through a feasible, centralised process.

3 **Results**

4 In TEMPER, 312 major or critical findings were identified at 94 visits. These findings
5 comprised 120 distinct issues, for which we proposed 56 different centralised processes.
6 Following independent review of the feasibility of the proposed processes by 87
7 consultation respondents across different stakeholder groups, we conclude that 306/312
8 (98%) findings could theoretically be prevented or identified centrally.

9 **Conclusions**

10 This work presents a best-case scenario, where a large majority of monitoring findings were
11 deemed theoretically preventable or detectable by central processes. Caveats include the
12 cost of applying all necessary methods, and the resource implications of enhanced central
13 monitoring for both centre and trials unit staff. Of processes not currently deemed feasible,
14 use of electronic health records has the largest potential benefit.

15 Our results will inform future monitoring plans and emphasise the importance of continued
16 critical review of monitoring processes and outcomes to ensure they remain appropriate.

17 **Key words:** monitoring, central monitoring, on-site monitoring, risk-based monitoring,
18 efficient trial conduct

19

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Introduction

A well-run clinical trial is designed and managed to minimise damaging errors in conduct.¹

Monitoring is done to detect important errors in a reasonable timescale and to enable action to prevent repetition. Cumulatively, this helps ensure the safety of trial participants and the integrity of trial results. **Figure 1** shows a suggested relationship between risk, prevention and monitoring.

Historically, and partly in response to regulatory guidance,² trial monitoring has relied on frequent sponsor visits to trial centres.³ These visits have clear benefits in terms of building rapport between sponsor representatives and trial centres, delivering training and achieving trial promotion, as well as the opportunity for in-person review of facilities and source documents.⁴ However, on-site monitoring has some significant limitations. The cost of travel and staff time required for regular centre visits is considerable,⁵⁻⁷ and may not be justified given the acknowledged limited benefit of source data verification, a common driver of intensive centre visit strategies.⁸⁻¹⁰ Depending on how frequent visits are, on-site monitoring may detect issues less quickly than central monitoring, i.e. monitoring conducted without centre visits, using data collected from trial centres. Finally, direct access to individual participant source data, while a strength of on-site monitoring, is less useful than central monitoring when looking for trial-wide issues in multicentre trials.

Acknowledging this, current regulatory guidance now encourages risk-based monitoring¹¹⁻¹³ with greater emphasis on, and even suggested methods for, central monitoring.¹⁴ There remains, however, a lack of evidence to support different monitoring practices.^{3,15,16}

1 The TEMPER study (TargetEd Monitoring: Prospective Evaluation and Refinement)¹⁷
2 assessed whether centrally monitored threshold-based rules – “triggers” – could be used as
3 a means to distinguish clinical trial centres with high or low rates of concerning on-site
4 monitoring findings. In a prospective matched-pair design, centres participating in one of
5 three phase III randomised cancer treatment trials of investigational medicinal products
6 (IMPs) with time-to-event outcomes were prioritised for visits based on the trigger findings.
7 Each triggered centre was matched with one from the same trial that was not triggered, i.e.
8 not of current concern, and both centres were visited. Centre visits were conducted
9 according to each trial’s monitoring plan, but generally included similar activities across all
10 the trials: review of some or all informed consent forms, source data verification and
11 medical notes review for a sample of participants, facility review (including pharmacy), and
12 review of quality and completeness of essential documents. Findings were categorised in
13 terms of seriousness, according to a standardised system, as Critical, Major or Other, using
14 similar definitions to those of UK regulators.¹⁸ We considered the collection of Critical and
15 Major findings to be those of interest, i.e. ‘errors that matter’.¹

16 The full methods and results are reported elsewhere.¹⁷ TEMPER found that the majority of
17 centres in both triggered visits and matched visits (those without concern) had at least one
18 Major or Critical finding, questioning the efficacy of triggered monitoring as employed in
19 these trials. Here, we report an exploratory review of all the Major and Critical findings
20 reported in TEMPER. We sought to propose centralised monitoring processes or trial
21 process changes which might detect or prevent these findings prior to centre visit. Through
22 this, we aimed to inform and improve future trial conduct by developing an evidence-based
23 (or at least experience-based) central monitoring and quality assurance plan.

1 **Methods**

2 **Source data: Monitoring Findings from TEMPER study**

3

4 We used findings from all 94 on-site monitoring visits conducted for TEMPER. There were
5 312 individual Major or Critical findings (298 Major, 14 Critical); these are summarised in
6 **Table 1**. Some findings had been detected several times within and across centres. In total
7 there were 120 distinct issues; a representative example of each was reviewed as described
8 below. **Figure 2** summarises all the stages of the current study.

9

10

11 **Initial review and development of suggested processes (consensus exercise)**

12

13 **Step 1:** independently of each other, three authors (WC, CH, SS) reviewed all 120 issues to
14 consider whether, hypothetically: (1) each issue could be identified through central
15 monitoring; (2) the issue could be prevented, or the consequences prevented, through
16 some change in trial processes; and (3) if the answer to either (1) or (2) was yes, what the
17 central monitoring method or trial process change would be. We considered “prevention”
18 to include both complete prevention of the issue (e.g. process to prevent patient being
19 approached for trial entry if ineligible) and prevention of its consequences (e.g. process to
20 identify erroneously randomised ineligible patient at the time of randomisation and
21 therefore prevent them starting study intervention if not appropriate or safe to do so). Both

1 prevention and central detectability were rated against a five-point scale: 5 = definitely, 4 =
2 possibly, 3 = not sure, 2 = probably not, 1 = definitely not.

3 **Step 2:** the independent review results were combined into one file. Each reviewer saw
4 their initial responses alongside those of the other, anonymised reviewers and either
5 confirmed or updated their response for each issue.

6 **Step 3:** reviewers met in-person to agree final results for each issue, using majority votes in
7 cases of any disagreement. The reviewers discussed the potential prevention and central
8 monitoring processes, brainstormed further ideas for these, finalised the process list and
9 agreed a consensus feasibility score for each, using the same five-point scale mentioned
10 above.

11

12

13 **Stakeholder consultation to assess feasibility of suggested processes**

14

15 We sought independent views on the feasibility of our list of proposed central monitoring
16 and prevention processes through a consultation exercise involving staff at the University
17 College London Institute of Clinical Trials and Methodology, and collaborators working with
18 the Institute's trials (mainly those working on the three trials involved in TEMPER). We
19 aimed to capture the views of experienced stakeholders both from trial centres (clinicians,
20 research nurses, pharmacists, radiologists) and clinical trials units (trial managers, data
21 managers, trialists and statisticians).

1 We used Opinio survey software¹⁹ to develop the consultation exercise. Invitees were not
2 sent any reminders after the initial invite. We asked respondents to review each of our
3 central monitoring and prevention processes (without disclosing our prior feasibility score)
4 and provide a feasibility score from options “Feasible and easy to achieve in current
5 practice” (score 5), “Feasible but expensive or challenging” (4), “Not sure” (3), “Possible but
6 cost or practical issues make it unworkable” (2), “Not possible at present” (1). We also asked
7 respondents whether any of a pre-defined list of challenges might apply to implementing a
8 given process: cost, time for trials unit staff, time for centre staff, logistical issues, legal or
9 regulatory issues, other (specify). There was an opportunity to add free text comments after
10 each of these questions, for each process; these informed our interpretation of process-
11 level and individual results, but are not summarised for this report. We also collected
12 information on respondents’ professional role and their experience with clinical trials and
13 with trial monitoring.

14 We asked each stakeholder group only about processes we felt were relevant to that group
15 (e.g. we asked pharmacists only about issues of drug management). In cases where many
16 processes (e.g. >25) applied to a single group, the list was split into two and a separate
17 survey link was sent to each half of the whole distribution list.

18 We carried out data cleaning prior to analysis. 16 respondents had provided free-text
19 comments about the feasibility of particular processes, but no feasibility score; we discussed
20 these and, where we felt the comment was clear, imputed a feasibility score for that
21 respondent. ‘Not sure’ answers were in one of two groups: either respondents understood
22 the proposed process but were not sure of its feasibility, or had said ‘not sure’ to express
23 uncertainty about what the proposed process would entail. Where comments were agreed

1 to unambiguously suggest uncertainty about the nature of the process, we removed the
2 'not sure' response from the final dataset.

3 No ethical approval was required for the consultation exercise (confirmed using the Health
4 Research Authority's decision tool²⁰).

5

6

7 **Outcomes and analysis**

8

9 For each process we calculated the median feasibility score from the consultation per
10 stakeholder group, and across all respondents. For comparison with our prior feasibility
11 grade, we grouped these into broader categories: broadly feasible (median score ≥ 4),
12 broadly not feasible (median score ≤ 2) and not sure (median score 3). Where consensus and
13 consultation agreed (including across individual stakeholder groups), the broad category
14 was taken as the final answer on the process' feasibility. Where there was any disagreement
15 in overall or inter-stakeholder medians, we reviewed the consultation data in detail and
16 made a final decision about what broad feasibility category to ascribe. Non-integer,
17 borderline medians, for example a score of 3.5 for one stakeholder group among others ≥ 4 ,
18 were taken as disagreement worth further, detailed review. Our general approach was that
19 we would defer to the views of the stakeholders unless we felt they had not understood our
20 description of the proposed process. When respondents suggested a process was feasible
21 only in certain conditions, we categorised as 'feasible with caveats'.

1 Returning to the original list of 120 distinct issues, we applied the final feasibility scores to
2 any processes that might address each one. From this, we were able to determine our main
3 overall outcomes, namely the proportion of individual Major and Critical findings (n=312)
4 that could be identified through a feasible central monitoring process, or prevented (or their
5 consequences prevented) through a feasible change to trial processes.

6 Aside from the survey software, all data collection and analysis took place in Microsoft
7 Excel, and WC managed the data. Independent checks were carried out by SS on a 10%
8 sample of findings and proposed processes respectively, to check feasibility scores pre- and
9 post-consultation.

1 Results

2

3 In our consensus exercise, we agreed that 114 (95%) of the 120 distinct issues were
4 potentially detectable through central monitoring. A different, mostly overlapping list of 114
5 (95%) were theoretically preventable through simple changes in trial processes. We
6 proposed 56 processes (or process changes), 43 of which could directly address on-site
7 findings from TEMPER and 13 of which we thought could potentially have an impact, but
8 less directly or without being completely fool-proof, e.g. additional training for trial centre
9 staff.

10 All 56 processes were included in the consultation exercise, which was run between 20
11 December 2017 and 26 January 2018. **Table 2** summarises the stakeholder recipients and
12 responders to each process. 87 people completed the exercise, an overall response rate of
13 19% of those to whom we sent the survey (n=450). We excluded 10 additional responses
14 that provided monitoring experience information only, without answers to the remaining
15 questions. Mean number of years' experience in clinical trials was 11 (range 2-31). Most
16 respondents had not personally conducted on-site monitoring (54%) or central monitoring
17 (59%), and had worked at a centre that had undergone central or on-site monitoring (55%).
18 For 26/56 of our proposed processes (46%), the median consultation feasibility was in the
19 same broad category (i.e. broadly feasible, broadly not feasible, not sure), across all
20 stakeholder groups, as our prior consensus score. Of the remaining 30/56 processes, 14
21 were in the same broad category overall, but with differences in some stakeholder groups.
22 After discussing and agreeing final feasibility assessments for these remaining 30, the total

1 56 were divided into: 43 processes feasible or feasible with a caveat (77%), 5 not feasible
2 (9%) and 8 uncertain (14%).

3 **Online Table S1a** lists the processes deemed feasible, and the number and proportion of
4 on-site monitoring findings each addresses; **Online Table S1b** lists the processes deemed
5 not currently feasible or of uncertain feasibility, with summarised reasoning for each
6 judgement. **Online Table S1c** summarises the feasibility of the indirect or not fool-proof
7 processes.

8 Based on the consultation exercise, 304/312 (97%) TEMPER visit findings could be detected
9 through feasible central monitoring methods, and 260/312 (83%) were theoretically
10 preventable, or their consequences preventable, through feasible changes to trial processes.
11 306/312 (98%) were either centrally detectable, preventable, or both. 256/312 (82%)
12 findings were addressed by more than one suggested process or process change, although
13 this varied across the types of finding (informed consent findings: 222/222, 100%; source
14 data review findings: 21/71, 30%).

15 **Table 3** lists abridged summaries of the six remaining findings (three each from triggered
16 visits and non-triggered, matched control centre visits) that could not be centrally detected,
17 or prevented, by a feasible process. All were classified Major; no Critical findings remained.
18 All were from Case Report Form checks and source data verification, with no findings
19 remaining from pharmacy, essential document or informed consent form checks. As with
20 the TEMPER study findings as a whole,¹⁷ there is no indication that these findings would
21 have had any significant impact on the results or interpretation of the trials involved.

22

1 Discussion

2 We found that a large majority of important on-site monitoring findings – including all
3 categorised as “Critical” – could theoretically be detected through feasible central
4 monitoring processes or prevented altogether through feasible changes to trial processes.
5 These results corroborate those of the previous, similar work in a different setting.²¹ A large
6 number of findings, especially in informed consent monitoring, could be addressed by more
7 than one process. Exclusion of the 306 “preventable” or “centrally detectable” findings
8 reduces the number of TEMPER monitoring visits with ≥ 1 Major or Critical finding from
9 81/94, as in the primary TEMPER study results, to just six (three at triggered visits, three at
10 untriggered).

11 It is important to note that this represents a best case scenario. Some of the processes will
12 already be in place for some RCTs, including to some degree in the TEMPER trials. We
13 cannot, therefore, easily say how many potential on-site monitoring findings were already
14 successfully avoided through use of these processes in these trials. We did not aim to prove
15 that the TEMPER on-site monitoring findings would definitely have been found earlier
16 through our suggested processes, only to use the findings to develop a more comprehensive
17 plan for future trials. Nonetheless, this may imply that, as currently used, some processes
18 may be insufficient to prevent or centrally identify all issues, or that they cannot be – or
19 have not so far been - implemented consistently across centres and time. Furthermore, an
20 unavoidable limitation of many central monitoring processes is their reliance on good
21 quality and timely reporting from centres.

22 An increased focus on central rather than on-site monitoring may have implications for
23 resourcing. From free-text responses to our consultation exercise (data not shown), it is

1 clear some centre research staff already feel that some central monitoring processes make
2 large demands of their time, without adequate resourcing or recognition (as noted by
3 others²²). Central monitoring also displaces resource within the trials unit from dedicated
4 on-site monitors towards the database programmers and statisticians responsible for
5 developing reports and reviewing data, and trial management staff responsible for following
6 up highlighted issues.²³ Resolving these resourcing questions, perhaps partly through
7 different financial arrangements with centres in trials that rely more on central monitoring,
8 could be a necessary precursor to a more widespread adoption of central monitoring
9 methods.

10 Our Major and Critical findings were dominated by errors relating to informed consent,
11 because of (1) the importance of consent forms in clinical trial legislation, and therefore also
12 in our monitoring plans and findings categorisation scheme, (2) the relative frequency of
13 errors (not apparently atypical^{15,24,25}), (3) the monitoring approach: only one of the three
14 TEMPER trials centrally monitored consent forms, and then not as a pre-randomisation
15 check, and (4) the relatively frequent re-consent requests in these trials, as reported in the
16 main TEMPER report.¹⁷ We believe (with the support of our consultation) all our findings
17 relating to initial informed consent forms are detectable centrally, at the point of
18 randomisation, thereby preventing patients starting treatment if there are any issues. By
19 the time of TEMPER's main publication in 2018, a number of RCTs at the University College
20 London Institute of Clinical Trials and Methodology had started to carry out routine,
21 pre-randomisation central consent form monitoring, whereby a copy of the completed form
22 is checked at the trials unit before being destroyed securely. This may be less feasible in
23 trials with short lead-up times, such as some trials in emergency medicine. Future

1 developments in electronically documenting consent^{26,27} may be useful in resolving many
2 issues we currently face with paper-based forms.

3 Of the remaining findings, a large proportion related to patient eligibility (recognised as a
4 problem in other trials²⁸), unreported Serious Adverse Events and unreported time-to-event
5 endpoints (in these RCTs, death or cancer progression). Processes to check eligibility prior to
6 randomisation include collection of more detailed data on trial forms to verify eligibility (e.g.
7 blood results) rather than just tick-box confirmation that a patient meets each eligibility
8 criterion, and collecting pseudonymised copies of source data to verify key aspects of
9 eligibility. Our consultation confirmed that both of these are possible, although perhaps not
10 practical in all trials. Some consultation respondents also voiced concern about the
11 availability of clinical expertise required to review medical data before randomisation; it
12 may therefore be best to limit this to objective assessments (e.g. checking blood results are
13 within range) rather than more complex assessments of, for example, scan reports.

14 Timely reporting of Serious Adverse Events is fundamental to adhering to regulatory
15 requirements for RCTs, and sponsors must have processes in place to ensure all events are
16 received from trial centres. Our consultation exercise reported uncertainty about the
17 current possibility and timeliness of using routinely-collected health data in the UK (e.g.
18 Health Episode Statistics data²⁹) to identify unreported events. For now, our best suggestion
19 is therefore to collect regular data to help ascertain if any Serious Adverse Events may have
20 taken place, such as whether there have been any in-patient hospitalisations or additional
21 treatments since the last follow-up visit. We also suggest that better informing trial
22 participants about safety reporting requirements, particularly where emergency admissions
23 may take place at a non-trial hospital, may improve reporting rates.

1 Unreported deaths in trials with survival-based outcome measures can be relatively simple
2 to detect centrally by, for example, closely following-up cases where there is no data at the
3 trials unit about a given patient for a long time. It is also feasible to obtain data on patient
4 deaths through national registries, though challenges may remain regarding associated
5 costs and the timeliness of available data. Unreported disease progression can be more
6 difficult to centrally detect, depending on how much additional indicative data is collected.
7 There may be scope for using routinely-collected health data to look for changes in patient
8 treatment that may indicate disease progression, however our consultation respondents felt
9 we could not say this was feasible in the UK at present.

10 In the list of processes considered not feasible in our consultation, processes involving use
11 of routinely-collected health data were among those with the largest *potential* impact on
12 monitoring. Although there are some reported examples of remote access to individual
13 patient records for source data verification,³⁰⁻³² there are currently substantial legal and
14 information governance barriers to this becoming routine, certainly in the UK. The
15 feasibility of regular, timely access to linked electronic health record data for research
16 purposes is easier to envisage, and could facilitate identification of unreported (serious)
17 adverse events, trial outcome data and verification of, for example, health economics data.
18 The suitability and availability of these data for complex, time-sensitive purposes is yet to be
19 proven, however.^{33,34}

20 We acknowledge several limitations not already mentioned. The central monitoring
21 processes we propose may be most suitable for trials like those included in TEMPER, which
22 were late-phase trials with already-licensed IMPs posing moderate risk to trial participants.
23 A higher degree of reliance on central monitoring may be less appropriate for higher-risk

1 trials. The feasibility of our monitoring and prevention processes was confirmed not through
2 a broad survey of trialists, but through a convenience sample of staff and collaborators,
3 mainly working on trials of cancer treatments, at the University College London Institute of
4 Clinical Trials and Methodology in the UK. The consultation exercise had a relatively low
5 response rate. Our results may therefore not be generalizable to different settings.

6 We asked our consultation respondents to comment on each suggested process in isolation,
7 and to implement all processes in combination would be a significant undertaking. This
8 again highlights the importance of properly resourcing all central monitoring activity. Finally,
9 we should acknowledge that we could implement all these suggested processes in a
10 subsequent trial in a different setting and find different, important findings at trial centres.
11 Comparison with our earlier work²¹ supports this to some extent.

12 We recommend that trialists review the most common errors in their trials and give due
13 consideration to implementing the central processes – including, but not limited to, those
14 reported in this work – that could detect or prevent the majority of serious issues, alongside
15 processes to address additional, trial-specific risks. There is scope for more formal
16 evaluation of our resulting monitoring plan, and generation of additional data on efficacy
17 and costs, both direct and indirect.

18 In conclusion, we recommend the process we have explored here of systematically using the
19 results of monitoring to induce continual improvement of trial processes should be a
20 routine, ongoing exercise, as it will ultimately lead to more robust data, safer participants
21 and better trials. Standardised, systematic recording of data about clinical trial monitoring
22 may be a necessary precursor to this, to facilitate review and assessment of trends within
23 and across trials.

1

1 **Acknowledgements**

2 We would like to acknowledge the support of those who helped this work by participating in
3 the consultation exercise, and of Matthew R Sydes, Sharon B Love and Mahesh KB Parmar at
4 Medical Research Council Clinical Trials Unit at University College London for their reviews
5 of draft manuscripts. We also thank the wider TEMPER study team for their support, the
6 trial teams whose trials took part in TEMPER and the site staff whose visits were included in
7 TEMPER.

1 **Conflicts of interest**

2 The Authors declare that there are no conflicts of interest.

3

1 **Funding**

2 The TEMPER study was funded by a grant from Cancer Research UK (grant C1495/A13305)
3 from the Population Research Committee); additional support for the present work was
4 provided by the Medical Research Council London Hub for Trial Methodology Research
5 (MC_UU_12023/24).

6

7

1 Supplementary information

2 - Supplementary tables S1a-c

3 - Consultation exercise text

1 References

2

- 3 1. Meeker-O'Connell A, Glessner C, Behm M, et al. Enhancing clinical evidence by
4 proactively building quality into clinical trials. *Clin. Trials* 2016; 13, 439–444.
- 5 2. Baigent C, Harrell FE, Buyse M, et al. Ensuring trial validity by data quality assurance
6 and diversification of monitoring methods. *Clin. Trials* 2008; 5, 49–55.
- 7 3. Macefield RC, Beswick AD, Blazeby JM, et al. A systematic review of on-site
8 monitoring methods for health-care randomised controlled trials. *Clin. Trials* 2013;
9 10, 104–124.
- 10 4. MHRA. *Good Clinical Practice Guide*. 2012;
- 11 5. Yusuf S, Bosch J, Devereaux PJ, et al. Sensible guidelines for the conduct of large
12 randomized trials. *Clin Trials* 2008; 5, 38–39.
- 13 6. Duley L, Antman K, Arena J, et al. Specific barriers to the conduct of randomized trials.
14 *Clin. Trials* 2008; 5, 40–48.
- 15 7. Sertkaya A, Wong HH, Jessup A, et al. Key cost drivers of pharmaceutical clinical trials
16 in the United States. *Clin Trials* 2016; 13, 117–126.
- 17 8. Olsen R, Bihlet AR, Kalakou F, et al. The impact of clinical trial monitoring approaches
18 on data integrity and cost—a review of current literature. *Eur. J. Clin. Pharmacol.*
19 2016; 72, 399–412.
- 20 9. Tudur Smith C, Stocken DD, Dunn J, et al. The value of source data verification in a
21 cancer clinical trial. *PLoS One* 2012; 7, 12.
- 22 10. Buyse M, Squifflet P, Coart E, et al. The impact of data errors on the outcome of
23 randomized clinical trials. *Clin. Trials* 2017; 174077451771615.
24 doi:10.1177/1740774517716158
- 25 11. MRC/DH/MHRA Joint Project. Risk-adapted Approaches to the Management of
26 Clinical Trials of Investigational Medicinal Products. 2011; Available at:
27 [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/34](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343677/Risk-adapted_approaches_to_the_management_of_clinical_trials_of_investigational_medicinal_products.pdf)
28 [3677/Risk-](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343677/Risk-adapted_approaches_to_the_management_of_clinical_trials_of_investigational_medicinal_products.pdf)
29 [adapted_approaches_to_the_management_of_clinical_trials_of_investigational_med](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343677/Risk-adapted_approaches_to_the_management_of_clinical_trials_of_investigational_medicinal_products.pdf)
30 [icinal_products.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343677/Risk-adapted_approaches_to_the_management_of_clinical_trials_of_investigational_medicinal_products.pdf). (Accessed: 8th September 2017)
- 31 12. Food and Drug Administration. Guidance for Industry: Oversight of Clinical
32 Investigations — A Risk-Based Approach to Monitoring. 2013; Available at:
33 <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM269919.pdf>. (Accessed: 8th
34 September 2017)
- 35 13. European Medicines Agency. Reflection paper on risk based quality management in
36 clinical trials. 2013; Available at:
37 [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/201](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf)
38 [3/11/WC500155491.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf). (Accessed: 16th January 2017)
- 39 14. International Conference on Harmonisation of Technical Requirements for
40 Pharmaceuticals for Human Use (ICH). Integrated Addendum to ICH E6(R1): guideline

- 1 for good clinical practice E6(R2). 2016; Available at:
2 https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2_Step_4_2016_1109.pdf. (Accessed: 25th February 2017)
- 3
- 4 15. Brosteanu O, Schwarz G, Houben P, et al. Risk-adapted monitoring is not inferior to
5 extensive on-site monitoring: Results of the ADAMON cluster-randomised study. *Clin.*
6 *Trials* 2017; 14,
- 7 16. Bakobaki J, Joffe N, Burdett S, et al. A systematic search for reports of site monitoring
8 technique comparisons in clinical trials. *Clin. Trials* 2012; 9, 777–780.
- 9 17. Stenning SP, Cragg WJ, Joffe N, et al. Triggered or routine site monitoring visits for
10 randomised controlled trials: results of TEMPER, a prospective, matched-pair study.
11 *Clin. Trials* 2018; 174077451879337. doi:10.1177/1740774518793379
- 12 18. Medicines and Healthcare products Regulatory Agency. Good clinical practice for
13 clinical trials - GOV.UK. Available at: <https://www.gov.uk/guidance/good-clinical-practice-for-clinical-trials#inspection-outcomes>. (Accessed: 16th December 2018)
- 14
- 15 19. ObjectPlanet. Conduct Online Surveys using Opinio. Available at:
16 <http://www.objectplanet.com/opinio/>. (Accessed: 16th May 2018)
- 17 20. Health Research Authority. Do I need NHS Ethics approval? Available at:
18 <http://www.hra-decisiontools.org.uk/ethics/index.html>. (Accessed: 3rd July 2018)
- 19 21. Bakobaki JM, Rauchenberger M, Joffe N, et al. The potential for central monitoring
20 techniques to replace on-site monitoring: findings from an international multi-centre
21 clinical trial. *Clin. Trials* 2012; 9, 257–264.
- 22 22. Bertolet M, Brooks MM & Carson JL. Commentary on Engen et al: Risk-based,
23 dynamic, process-oriented monitoring strategies and their burden. *Clin. Trials* 2019;
24 174077451988161. doi:10.1177/1740774519881617
- 25 23. Wilson B, T. P, J. G, et al. Defining a Central Monitoring Capability: Sharing the
26 Experience of TransCelerate BioPharma’s Approach, Part 1. *Ther. Innov. Regul. Sci.*
27 2014; 48, 529–535.
- 28 24. Journot V, Pérusat-Villetorte S, Bouyssou C, et al. Remote preenrollment checking of
29 consent forms to reduce nonconformity. *Clin. Trials* 2013; 10, 449–59.
- 30 25. Wyman Engen N, Huppler Hullsiek K, Belloso WH, et al. A randomized evaluation of
31 on-site monitoring nested in a multinational randomized trial. *Clin. Trials* 2019;
32 174077451988161. doi:10.1177/1740774519881616
- 33 26. Stevens N, Edwards L, Balayah Z, et al. Risk based survey evidence supports electronic
34 informed consent as a recruitment method for UK clinical trials. *J. Clin. Epidemiol.*
35 2016; 77, 134–136.
- 36 27. Health Research Authority & Medicines and Healthcare products Regulatory Agency.
37 *Joint statement on seeking consent by electronic methods*. 2018;
- 38 28. Yelland LN, Kahan BC, Dent E, et al. Prevalence and reporting of recruitment,
39 randomisation and treatment errors in clinical trials: A systematic review. *Clin. Trials*
40 2018; 174077451876162. doi:10.1177/1740774518761627
- 41 29. NHS Digital. Hospital Episode Statistics (HES) - NHS Digital. Available at:

- 1 [https://digital.nhs.uk/data-and-information/data-tools-and-services/data-](https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics)
2 [services/hospital-episode-statistics](https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics). (Accessed: 17th May 2018)
- 3 30. Mealer M, Kittelson J, Thompson BT, et al. Remote source document verification in
4 two national clinical trials networks: a pilot study. *PLoS One* 2013; 8, e81890.
- 5 31. Uren SC, Kirkman MB, Dalton BS, et al. Reducing Clinical Trial Monitoring Resource
6 Allocation and Costs Through Remote Access to Electronic Medical Records. *J. Oncol.*
7 *Pract.* 2013; 9, e13–e16.
- 8 32. Weston W, Smedley J, Bennett A, et al. The Cooking and Pneumonia Study (CAPS) in
9 Malawi: Implementation of Remote Source Data Verification. *PLoS One* 2016; 11,
10 e0155966.
- 11 33. Wright-Hughes A, Graham E, Cottrell D, et al. Routine hospital data – is it good
12 enough for trials? An example using England’s Hospital Episode Statistics in the SHIFT
13 trial of Family Therapy vs. Treatment as Usual in adolescents following self-harm. *Clin.*
14 *Trials* 2018; 15, 197–206.
- 15 34. Lugg-Widger F V, Angel L, Cannings-John R, et al. Challenges in accessing routinely
16 collected data from multiple providers in the UK for primary studies: Managing the
17 morass. *Int. J. Popul. Data Sci.* 2018; 3,
18
19

1 **Tables and Figures**

2

3 **Table 1:** summary of Major and Critical findings at TEMPER monitoring visits.

Type of finding by Monitoring Report section	Number of findings at all visits (n=94)			Individual issues ^b
	All findings	Major findings ^a	Critical findings ^a	
Investigator Site File – All	6	6	0	5
Informed consent – All	222	219	3	33
<i>Re-consent (eg failure to obtain re-consent in a timely manner)</i>	162	162	0	2
<i>Original consent (eg missing signatures, missing or in compatible signature dates, incorrect versions used)</i>	60	57	3	31
Pharmacy - All	8	6	2	8
CRF/SDV - All	76	67	9	74
<i>Unreported SAE/notable event</i>	25	25	0	25
<i>Unreported endpoint</i>	12	12	0	10
<i>Source data discrepancy (priority data)</i>	19	19	0	19
<i>Other</i>	20	11	9	20
Total Major and Critical findings	312	298	14	120

4 ^a The findings reported from the TEMPER study also included some ‘upgrades’ to Major and Critical,
5 i.e. groups of findings that collectively indicated a more serious finding. These were not considered
6 relevant outside of the context of the TEMPER study design, and were therefore excluded from this
7 exercise.

8 ^b Where a finding had occurred multiple times within or across centres, we reviewed only one of
9 each of these duplications as representative of others; these are listed here as ‘individual issues’
10 (n=120).

11

- 1 **Table 2:** summary of stakeholder groups invited to participate in the consultation exercise,
 2 completion rates, clinical trials experience and number of processes reviewed.

Stakeholder group	Site or Clinical Trials Unit (CTU)	Number distributed to	Number of responses	Completion rate	Mean # years worked in clinical trials	Number of processes reviewed
Pharmacy	Site	41	18	44%	7.5	5 ^a
Site researchers	Site	49	18	37%	5.0	28 ^b
Clinician	Site	207	12	6%	20.5	12
Operations	CTU	66	18	27%	8.5	47 ^c
Data Management	CTU	33	11	33%	10.0	31 ^d
Statistician	CTU	27	5	19%	10.0	13
Senior Trialist	CTU	6	3	50%	24.0	20
Radiologist	Site	21	2	10%	20.0	23
Total		450	87	19%	11.0	

- 3
 4 ^a Split into two surveys to same invitees, 4 in one, 1 in other (1 process missed from initial survey in error)
 5 ^b Site split into two surveys, each to half of invitees, with 14 processes in each.
 6 ^c Operations split into two surveys, each to half of invitees, with 24 and 23 processes respectively.
 7 ^d Data Management split into two surveys, each to half of invitees, with 16 and 15 processes respectively.
 8

1 **Table 3:** on-site monitoring findings judged not centrally detectable or preventable, following
 2 consensus exercise on feasibility of suggested processes.
 3

Abridged summary of finding	Number of instances	Grading	Reasons why not centrally detectable or preventable
Incorrectly graded adverse event – reported as CTCAE grade 2 instead of 4 (life-threatening); no serious adverse event reported	1	Major	<ul style="list-style-type: none"> - No simple way to detect this issue without review of source data; if serious adverse event had been submitted, could have cross-checked with those data. - Currently not possible to access and use hospital episode statistics or other electronic health records to detect this. - No obvious fool-proof way to prevent.
Unreported serious adverse event; considered ‘notable event’ in this trial and Case Report Form contained specific question about whether this type of adverse event had occurred, to which answer given had been ‘No’	2	Major	<ul style="list-style-type: none"> - Method to centrally identify cases did not work because centre misreported; therefore only way to detect is through review of source data. - Currently not possible to access and use hospital episode statistics or other electronic health records to detect this. - No obvious fool-proof way to prevent.
Unreported serious adverse event due to prolongation of hospital stay	1	Major	<ul style="list-style-type: none"> - Prolongation of hospital stay may only be clear from detailed review of medical notes. - Currently not possible to access and use hospital episode statistics or other electronic health records to detect this. - No obvious fool-proof way to prevent.
Misreporting of concomitant medication use at randomisation (key baseline data in this trial)	2	Major	<ul style="list-style-type: none"> - Data misreported; not feasible to request source data about concomitant medication to be sent to trials unit for verification, therefore only way to detect is through on-site monitoring. - No obvious fool-proof way to prevent.

4

5

Figures

Figure 1: a suggested relationship between preventative measures, monitoring and risk.

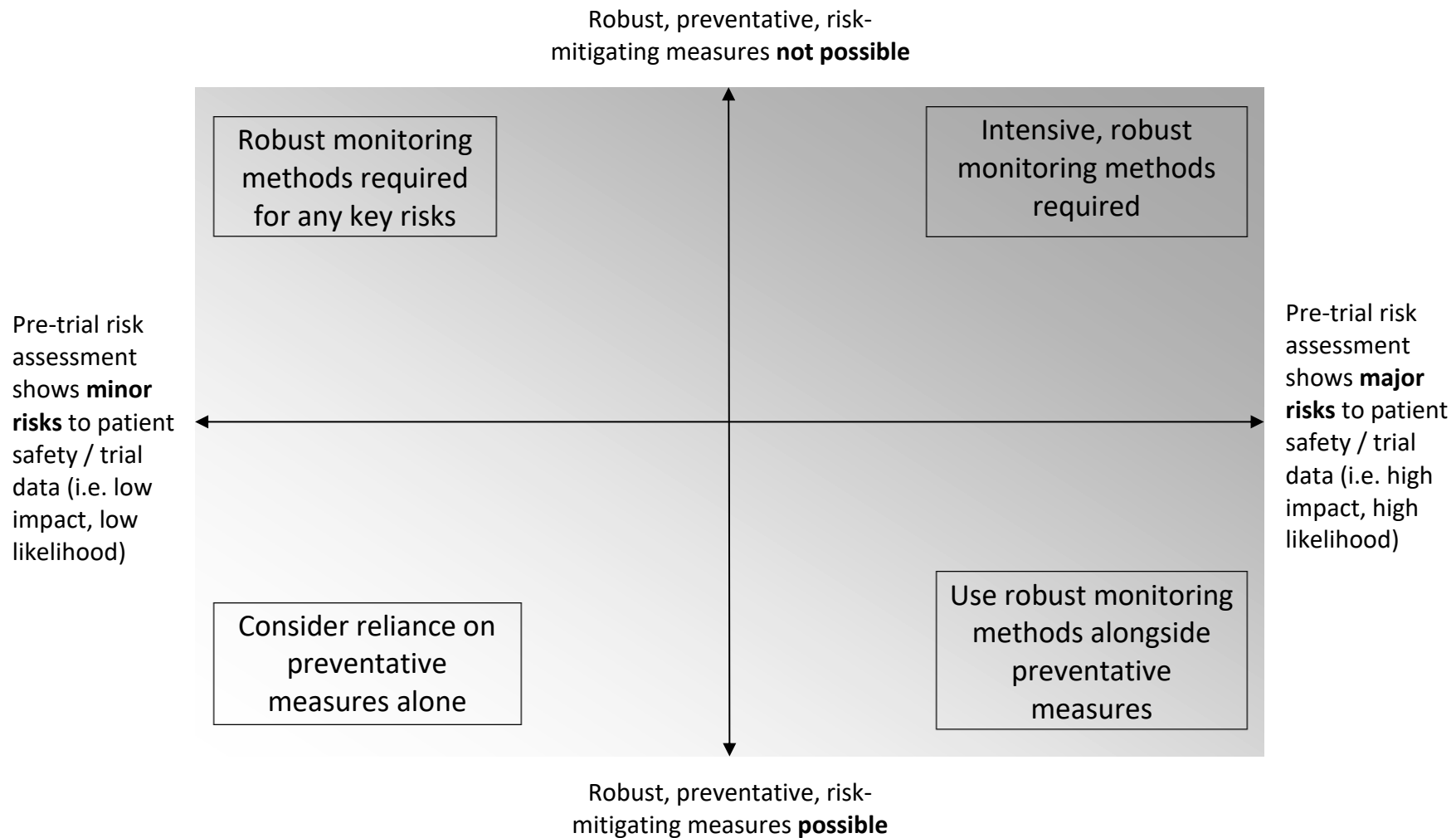


Figure 2: flow diagram describing the current study.

