

The Impact of Non-Drug-Related Toxicities on the Estimation of the Maximum Tolerated Dose in Phase I Trials

Alexia Iasonos¹, Mrinal Gounder², David R. Spriggs³, John F. Gerecitano², David M. Hyman³, Sarah Zohar⁴, and John O'Quigley⁵

Abstract

The rate of observed dose-limiting toxicities (DLT) determines the maximum tolerated dose (MTD) in phase I trials. There are cases in which non-drug-related toxicities or other-cause toxicities (OCT) are flagged as DLTs, or vice versa, due to attribution errors. We aim to assess the impact of such errors on the final estimate of MTD. We compared the impact of attribution errors using 2 trial designs—the "3+3" dose-escalation scheme and the continual reassessment method (CRM). Two attribution errors are considered: when a DLT is classified as an OCT (type A error) and when an OCT is misclassified as a DLT (type B error). The impact of these errors on accuracy, patient safety, sample size, and study duration was evaluated by varying the probability of occurrence of each error through simulated trials. Under no errors, CRM is on average 35% more accurate than 3+3 in finding the true MTD. This improved accuracy is maintained in the presence of errors. At a 15% type B error rate, CRM recommends a dose within 2 levels of the true MTD 68% of the time, compared with 17% of the time using the 3+3 method. A DLT must be attributed as an OCT 30% of the time to increase the accuracy of 3+3; otherwise the method recommends a wrong dose approximately 75% of the time. CRM is more robust to toxicity attribution errors compared with the 3+3 as it uses information from all treated patients, leading to a more accurate MTD estimation at the frequency of attribution errors anticipated in phase I clinical trials. *Clin Cancer Res*; 18(19); 5179–87. ©2012 AACR.

Introduction

The objective of phase I studies is to establish the safety, dose, and schedule of a new drug or regimen for further clinical development. In general, it is assumed that for cytotoxic agents, higher drug exposure correlates with improved efficacy as measured by tumor responses. Higher drug exposure is also associated with increasing severity (grade) of adverse events (AE), which are typically measured by standardized criteria such as Common Terminology Criteria for Adverse Events v 4.0. When patients experience a prespecified, unacceptable rate of dose-limiting toxicities (DLT) at a particular dose level, typically 33% or higher, the trial design declares the level below as the maximum tolerated dose (MTD), which becomes the recommended phase II dose (RP2D) for further clinical development.

Serious toxicities that otherwise would be counted as DLTs, if deemed non-drug-related, do not contribute equally in the determination of the MTD. Because the endpoint of phase I trials is the number of DLTs related to the experimental drug, it is essential that clinicians minimize errors when attributing toxicities.

Accurate attribution of toxicities to experimental drug versus other competing factors such as concomitant chemotherapy or medications, cumulative toxicities from previous treatment, disease progression, comorbidities, and/or intercurrent illness is not always clear. The challenge arises as phase I trials are typically the first in-human studies of novel agents, and the investigators rely on the drug's proposed mechanism of action, the toxicities observed in animal studies, and temporal associations. Clinical experience shows that it is sometimes beyond the clinician's ability to definitively determine if a given toxicity is due to the study drug, other cause, or combination (1–4). In these situations, physicians may be inclined to attribute DLTs as possibly related to the study drug as a precaution, to reduce potential patient harm. This may bias DLT attribution against study drugs and lead to underestimation of the MTD. A recent report evaluated causality attribution for serious AEs (SAE) during phase I oncology studies and concluded that a new causality assessment tool is needed (4). Moreover, other reports suggest that there might have been overreporting of SAEs (5) as until recently there were no specific guidelines for drug causality (6). For this reason, the U.S. Food and Drug Administration (FDA) recently

Authors' Affiliations: ¹Department of Epidemiology and Biostatistics, ²Developmental Therapeutic Center, and ³Gynecologic Medical Oncology Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York; ⁴INSERM, U717, Department Biostatistics; and ⁵INSERM, Université Paris VI, Place Jussieu, 75005 Paris, France

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Alexia Iasonos, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street, 3rd Floor, New York, NY 10065. Phone: 646-735-8115; Fax: 646-735-0010; E-mail: iasonosa@mskcc.org

doi: 10.1158/1078-0432.CCR-12-0726

©2012 American Association for Cancer Research.

issued a new regulation that clarifies the definition of drug-related AEs (7).

In this article, we estimate the impact of attribution errors on the outcome of phase I trials. Given the potential for attribution errors in toxicity assessment and the limited number of patients treated in phase I studies, the established MTD might not necessarily be the same as the true MTD, which corresponds to an acceptable toxicity rate. We address the questions of how frequently attribution errors can occur without establishing an MTD that is significantly above or below the actual MTD and what are the implications of attribution errors on drug development. There are 2 types of errors that might occur in toxicity attribution: (i) when an investigator incorrectly attributes a DLT as an event due to other causes when, in fact, it is related to the experimental drug (type A error), and (ii) when an investigator incorrectly attributes toxicity to an experimental drug when in fact it is related to other causes (type B error). A type A error can result in additional patients being enrolled to higher dose levels and being exposed to toxic levels of the drug, thus adversely affecting patient morbidity and mortality. On the other hand, a type B error will result in early termination of accrual and denote all levels above as unsafe, consequently recommending a subtherapeutic dose for further studies. This could ultimately result in abandonment of otherwise effective therapies. A type B error can also lead to unnecessary dose expansion, resulting in an increase in trial duration and cost.

Attribution errors will affect the operational characteristics of various designs in different ways, as different designs do not react in the same way in the presence of DLTs. Simulation studies allow us to compare the estimated MTD (RP2D) with the true MTD; thus, we simulated hypothetical rates of DLTs in the presence of attribution errors and evaluated the effect of errors on dose escalation and the RP2D. We compared the impact of attribution errors on 2 trial designs—the standard design ("3+3") (8) and the continual reassessment method (CRM; ref. 9). We hypothesize that adaptive designs, which use the accumulated data from all patients, are able to reduce the impact of attribution errors when they occur.

Methods

Data collection

In practice, the true underlying cause of an AE is unknown; hence, investigators cannot know for sure whether the number of DLTs observed in a phase I trial truly corresponds to drug-related SAEs (4). For this reason, a comparison of different phase I designs must be carried out with simulated hypothetical data that are generated under the same circumstances. We followed 2 prospective dose-escalation algorithms as described below. Regardless of the design, each patient has the same probability of experiencing a DLT at different dose levels; therefore, the dose-escalation algorithm alone determines the trial enrollment and the final MTD (10).

In this comparison, we included the 3+3 dose escalation method because it is the most commonly used phase I

design (11) and CRM, which is a model-based design. CRM determines whether to escalate, deescalate, or retain the level based on the accumulated data on all cohorts simultaneously and the best estimate of the MTD. Because CRM is an adaptive design, the dose-escalation rules are not known in advance but instead depend on the history of the trial. The design estimates *a priori* the toxicity rates for each dose level and refines these rates as the trial progresses and more patients are accrued (9).

Dose-escalation algorithms

The rules that determine dose escalation for the 3+3 design are shown in Supplementary Table S1. The rules that determine dose escalation for CRM (Fig. 1) are as follows:

- Before the start of the trial, 6 dose levels are chosen and each dose level is preassigned a toxicity rate. For example, investigators assign a toxicity rate of 5%, 10%, 20%, 30%, 40%, and 50% for dose levels 1 through 6, respectively. As the trial progresses and data regarding DLTs are accumulated, the preassigned toxicity rates are refined by sequential estimation such that a dose level associated with a 33% DLT rate is identified (12).
- The first dose is the lowest dose.
- DLT is evaluated only during the first cycle, which is 21 days in duration.
- Skipping dose levels is not allowed in dose escalation, but deescalation by more than one dose level is permitted.
- The trial stops when a prespecified number of patients have been accrued. Here, the sample size is 20 patients.

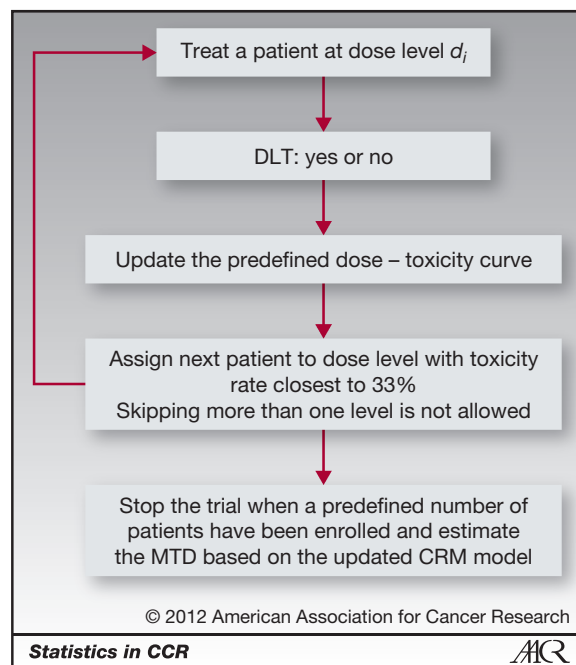


Figure 1. Dose-escalation algorithm for the CRM.

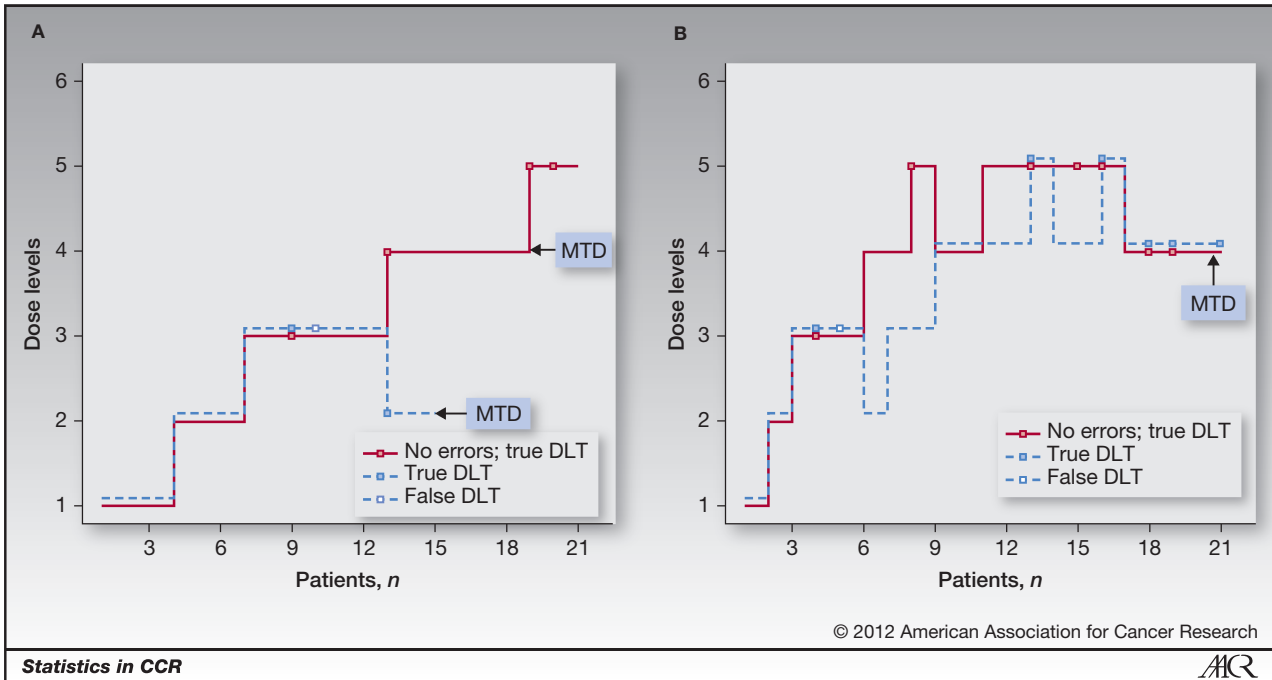


Figure 2. 3+3 (A) and CRM (B) under no error (solid line) and under type B error (dashed line) of incorrectly attributing OCT as DLT.

The MTD is the dose recommended on the basis of the updated estimates of toxicity rates based on all 20 patients' outcomes.

Statistical considerations

For both designs, we simulated 1,000 hypothetical trials that tested 6 dose levels under 6 scenarios (Table 1) that varied the location of the MTD. We varied the parameter that controls the error rates from 0% to 30% by increments of 5% in different scenarios; however, in each scenario, the error rate was constant across levels (13). For example, type A error rate is defined as a fixed probability that a true DLT is an other-cause toxicity (OCT), irrespective of dose.

We compared the performance of the above designs in terms of the following outcomes:

1. Accuracy of the final MTD by reporting the percentage of trials that found the correct MTD. If the trial recommended a wrong dose, we plotted how far away it was from the true MTD in terms of number of levels.
2. Safety (median number of DLTs) and interquartile range.
3. Trial duration (in months).
4. Sample size (fixed at 20 for CRM, varying for 3+3 by definition).

Trial duration was calculated as described by Iasonos and colleagues (14). To be concise, in the next section, we present simulation results based on 2 scenarios and accrual rates of 1 or 3 patients per month. Results with various accrual rates were similar to previous reports (14), and therefore were omitted from this article.

Results

First, we present 2 examples of hypothetical trials corresponding to the 3+3 and CRM designs in the presence of type B error for illustration. For both designs, each patient treated at dose levels 1 to 6 had a DLT rate of 10%, 17%, 22%, 30%, 45%, and 50% (scenario 1 in Table 1), respectively, and the true MTD was level 4. Figure 2A shows the trial progress with a 3+3 design in the presence of a type B error in which one OCT was incorrectly attributed as a DLT (dashed line) at level 3. Because this is the second DLT at that level, the method recommends dose 2 as the MTD. The trial is terminated early with fewer patients: 15 versus 21 if it

Table 1. True toxicity rates used to simulate the probability of a DLT at each dose level

	d ₁	d ₂	d ₃	d ₄	d ₅	d ₆
Scenario 1	0.10	0.17	0.22	0.30	0.45	0.50
Scenario 2	0.05	0.07	0.10	0.12	0.20	0.30
Scenario 3	0.05	0.20	0.35	0.45	0.50	0.55
Scenario 4	0.01	0.05	0.10	0.30	0.50	0.70
Scenario 5	0.05	0.15	0.20	0.25	0.30	0.50
Scenario 6	0.10	0.17	0.26	0.40	0.42	0.45

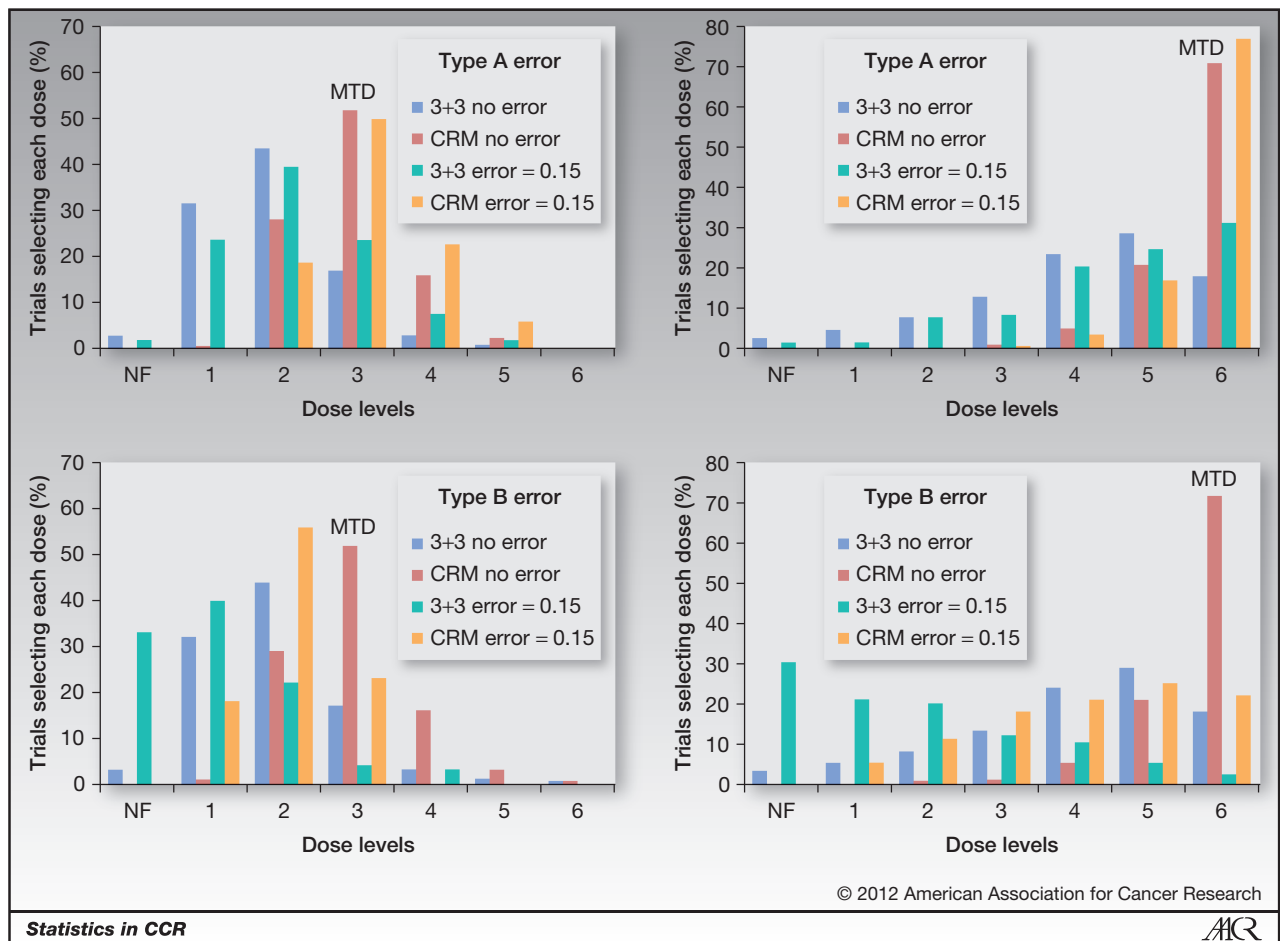


Figure 3. Percentage of trials recommending each dose level based on simulated trials comparing 3+3 with CRM under type A and B errors. In the graphs on the left, the true MTD is level 3 (scenario 3), and in the graphs on the right, the true MTD is level 6 (scenario 2). NF, dose not found because level 1 was too toxic.

had reached dose 5. In the absence of error (solid line), level 4 would be the MTD. Note that the 2 types of errors act in opposite directions, so that one reduces the effect of the other. However, we expect the type B error rate to be greater than the type A error rate in practice as investigators tend to attribute an event to the study drug when uncertain (4). Similarly, Fig. 2B shows the trial progress with a CRM design in the presence of a type B error in which one OCT is incorrectly attributed as a DLT (dashed line) at level 3. For comparative purposes, a sample size of 21 patients was used in this example for CRM. Following the dashed line, after observing 2 DLTs out of 3 patients at dose 3, CRM correctly deescalated to dose level 2. Data from subsequent patients support that level 3 is not as toxic as initially thought, and allow the method to update the estimated rates and assign patients to higher levels. It takes longer to get to the MTD, as early DLTs drop the doses to lower levels and the method needs subsequent patients without DLTs to allow experimentation to higher levels, but the final MTD is the same, which is level 4.

Figure 3 shows the summary results across 1,000 simulated trials when dose level 3 (left panel shows scenario 3) or dose 6 (right panel shows scenario 2) was assumed to be the

true MTD. In the absence of errors, the 3+3 scheme selects the correct MTD (dose 3) 17% of the time, whereas it selects dose 2 and 1 44% and 32% of the time, respectively. CRM has superior accuracy by selecting dose 3, 2, and 1 with a 52%, 29%, and 1% chance, respectively. The right panels of Fig. 3 show the scenario when the last dose is the true MTD. In the presence of type B errors, CRM selects a dose closest to the MTD 68% of the time, whereas 3+3 selects the same dose only 17% of the time.

Figure 4 shows how both methods behave as the error rates increase from 0% to 30%. We can see that the accuracy of CRM remains in the presence of type A error regardless of the location of the MTD. However, as the type B error rate increases, both methods correctly shift recommendation to lower levels below the MTD as they adapt by rejecting levels with a higher than expected number of DLTs. Simulations in which the true MTD was dose 4 and 5 confirmed these findings (data not shown). Misattributing a DLT as an OCT helps the 3+3 method by increasing its accuracy from 18% to 43% (right panel presents scenario 2) as the error rate is increased. This is because, in certain cases, it can allow the method to proceed to higher levels where activity occurs.

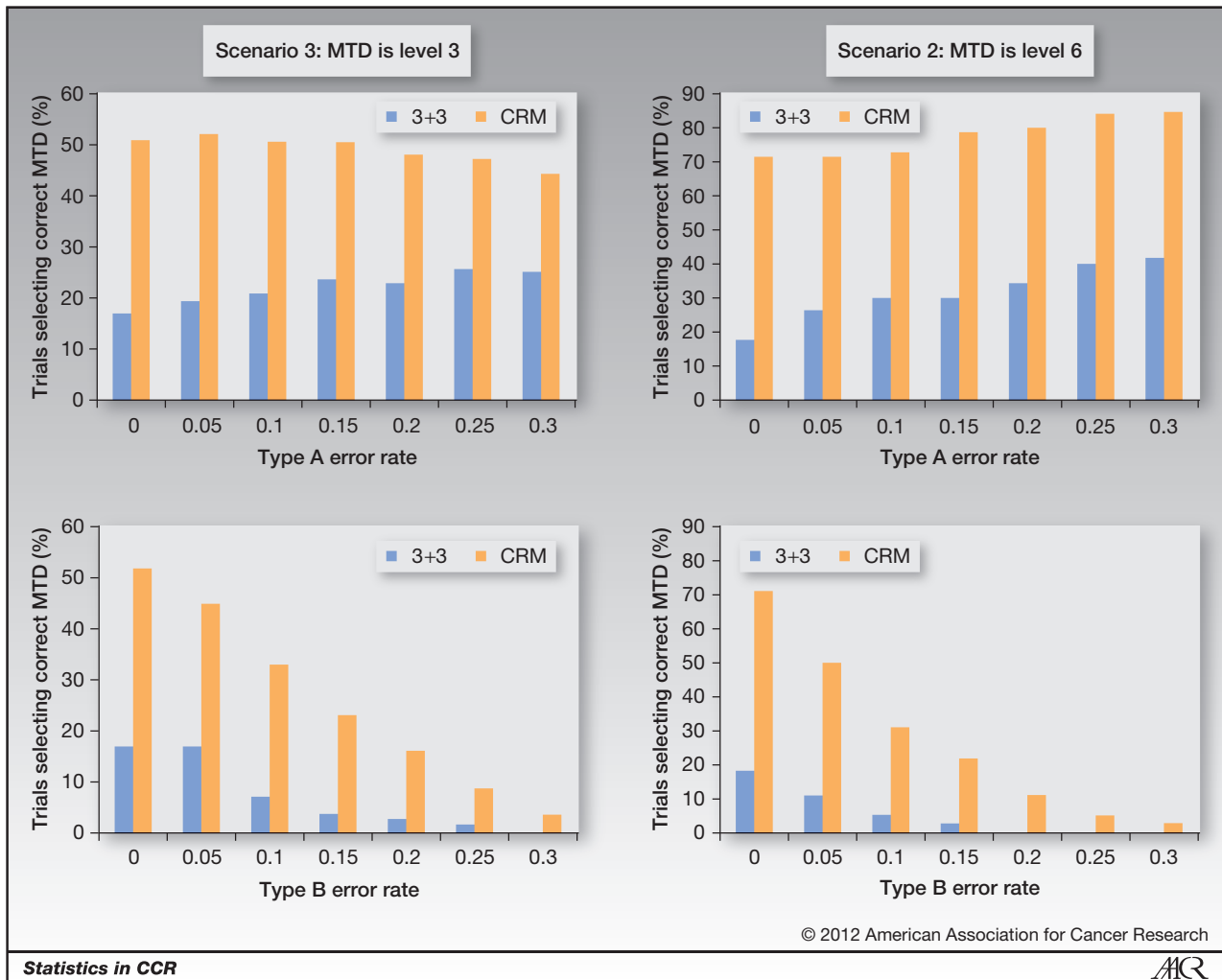


Figure 4. Percentage of trials recommending the correct phase II dose based on simulated trials. In the left, the true MTD is level 3 (Scenario 3), and in the right, the true MTD is level 6 (scenario 2). Type A error: incorrectly attributing a DLT as OCT; type B error: incorrectly attributing an OCT as a DLT.

Traditionally, the 3+3 design recommends a dose level with an observed rate of less than 33% (15, 16). By misattributing a DLT as an OCT, the number of DLTs no longer meets the cutoff of 2 of 6, and the method continues to escalate. The reduced accuracy of 3+3 is also a result of a smaller sample size and consistently treating patients at lower dose levels as measured by a smaller number of DLTs. When the MTD turns out to be among the higher levels (as in scenario 2), sample size is larger (21–24 as opposed to 20) and trial duration is longer with the 3+3 (21 vs. 20 months), whereas accuracy is much lower compared with CRM (18% vs. 73% in the absence of errors). If the MTD is among the first 3 levels, then trial duration is approximately 5 months shorter with the 3+3, as the trial can be completed with 12 to 15 patients on average (as opposed to 20 patients needed with CRM), although it recommends the correct MTD in fewer than 1 of 4 trials. Hence, this increase in sample size and resources enables the method to correct its estimate of the MTD, leading to a more accurate and robust RP2D.

So far, the CRM design was set to target a 33% rate of observed DLTs as an acceptable rate. Because the 3+3 design tends to select doses with much lower rates of toxicity, we also include simulations in which CRM targets a level with a 25% (1 in 4) rate of DLTs. Table 2 shows the percentage of trials selecting each level as well as the percentage of patients being treated at each level. The results support the previous findings, although the absolute improvement is lower. This is expected as the accuracy of any design depends on the true underlying rates and the dose–toxicity curve. The target rate is considered an external parameter, and it offers the flexibility to fine tune the design depending on the disease setting and the respective acceptable toxicity rate.

Discussion

In this article, we have investigated the impact of clinicians' errors in toxicity attribution and the choice of trial design on the estimation of the MTD in phase I trials. We

Table 2. Results of 1,000 simulated trials under true toxicity rates equal to 0.1, 0.17, 0.26, 0.4, 0.42, and 0.45 for the 6 levels, respectively, and a target acceptable toxicity rate of 25%

Level	NF	Trials selecting each dose, %						Patients treated, <i>n</i>						DLTs, <i>n</i>	Duration	Sample size, <i>n</i>	
		-1	1	2	3	4	5	6	1	2	3	4	5				6
Accrual rate: 1/mo																	
CRM																	
Error A: 0			5	24	43	21	5	2	16	28	32	17	5	2	5 (4,6)	19.9	20
0.05			3	22	41	23	8	3	14	26	33	18	7	3	5 (4,6)	19.9	20
0.10			4	20	42	23	8	4	14	26	31	18	7	3	5 (4,6)	19.8	20
0.15			2	15	38	27	12	6	13	23	30	20	9	4	5 (4,5)	20.2	20
0.20			1	15	35	28	13	7	11	23	30	21	10	6	4 (4,5)	19.8	20
0.25			1	12	32	30	15	9	11	21	29	21	12	6	4 (3,5)	20.2	20
0.30			1	7	28	32	18	#	10	19	27	23	13	8	4 (3,5)	19.8	20
3+3	1	8	25	31	27	7	1	1	31	30	21	9	2	0	3 (2,3)	15.1	15
	2	7	21	32	27	8	3	1	30	30	22	10	3	0	3 (2,3)	15.7	15
	0	6	19	30	32	9	3	1	28	29	23	12	3	1	3 (2,3)	15.6	15
	1	8	16	31	28	11	4	2	27	28	23	11	4	1	3 (2,3)	16.0	15
	1	7	15	25	31	13	7	2	27	26	22	14	5	2	3 (2,3)	16.9	15
	1	6	13	25	30	14	8	4	25	26	23	14	6	3	3 (2,3)	17.0	18
	0	5	13	24	29	17	7	6	25	25	22	15	6	3	3 (2,3)	17.6	18
Accrual rate: 3/mo																	
CRM			5	27	43	19	4	2	17	36	30	12	4	1	5 (3,6)	7.8	20
			4	24	39	25	6	2	19	36	29	13	3	1	4 (3,5)	7.9	20
			5	22	37	27	6	2	17	35	29	14	4	1	4 (3,5)	7.8	20
			3	18	39	29	8	4	16	33	30	15	5	2	4 (3,5)	7.8	20
			1	15	35	30	12	7	14	32	30	16	5	2	4 (3,5)	7.8	20
			1	11	33	33	13	8	13	31	29	18	7	3	4 (3,5)	7.8	20
			1	11	33	31	15	10	12	31	30	18	7	3	4 (3,5)	7.9	20
3+3	1	8	25	31	27	7	1	1	31	30	21	9	2	0	3 (2,3)	5.6	15
	2	7	21	32	27	8	3	1	30	30	22	10	3	0	3 (2,3)	5.8	15
	0	6	19	30	32	9	3	1	28	29	23	12	3	1	3 (2,3)	5.7	15
	1	8	16	31	28	11	4	2	28	28	23	11	4	1	3 (2,3)	5.9	15
	1	7	15	25	31	13	7	2	27	26	22	14	5	2	3 (2,3)	6.2	15
	1	6	13	25	30	14	8	4	25	26	23	14	6	2	3 (2,3)	6.3	18
	0	5	13	24	29	17	7	6	25	26	22	15	6	3	3 (2,3)	6.4	18

(Continued on the following page)

Table 2. Results of 1,000 simulated trials under true toxicity rates equal to 0.1, 0.17, 0.26, 0.4, 0.42, and 0.45 for the 6 levels, respectively, and a target acceptable toxicity rate of 25% (Cont'd)

Level	Trials selected each dose, %								Patients treated, <i>n</i>								DLTs	Duration	Sample size, <i>n</i>
	NF	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6				
Accrual rate: 1/mo																			
3+3	1	8	25	31	27	7	1	1	5	31	30	21	9	2	0	3 (2,3)	15.1	15	
	5	14	26	32	17	4	1	0	10	36	29	17	6	1	0	2 (2,3)	14.0	12	
	8	20	32	24	13	3	0	0	15	40	27	13	4	0	0	2 (2,3)	13.0	12	
	18	21	33	19	8	1	0	0	21	43	24	9	3	0	0	2 (2,3)	12.3	12	
	26	23	31	15	5	0	0	0	25	45	21	7	1	0	0	2 (2,3)	11.8	12	
	38	23	26	10	2	0	0	0	31	46	17	4	0	0	0	2 (2,3)	11.0	12	
	51	19	22	7	0	0	0	0	36	47	14	3	0	0	0	2 (2,3)	10.5	9	
Accrual rate: 3/mo																			
CRM			5	27	43	19	4	2		17	36	30	12	4	1	5 (3,6)	7.8	20	
			12	38	34	13	3	0		25	38	25	9	3	1	5 (4,6)	7.8	20	
			27	38	27	6	1	0		34	37	21	6	1	0	5 (4,6)	7.8	20	
			45	32	18	5	1	0		45	33	16	4	1	0	6 (5,7)	7.8	20	
			58	29	11	2	0	0		51	31	14	3	0	0	6 (5,8)	7.7	20	
			74	19	6	0	0	0		61	26	11	2	0	0	7 (6,8)	7.6	20	
			85	12	3	0	0	0		69	23	7	1	0	0	8 (7,9)	7.6	20	
Accrual rate: 3/mo																			
3+3	1	8	25	31	27	7	1	1	5	31	30	21	9	2	0	3 (2,3)	5.6	15	
	5	14	26	32	17	4	1	0	10	36	29	17	6	1	0	2 (2,3)	5.3	12	
	8	20	32	24	13	3	0	0	15	40	27	13	4	0	0	2 (2,3)	4.9	12	
	18	21	33	19	8	1	0	0	21	43	24	9	3	0	0	2 (2,3)	4.6	12	
	26	23	31	15	5	0	0	0	25	45	21	7	1	0	0	2 (2,3)	4.5	12	
	38	23	26	10	2	0	0	0	31	46	17	4	0	0	0	2 (2,3)	4.2	12	
	51	19	22	7	0	0	0	0	36	47	14	3	0	0	0	2 (2,3)	4.0	9	

NOTE: Each error rate varied from 0 to 0.3 by increments of 0.05. Level -1: one level below level 1; NF indicates dose not found because level 1 was too toxic. Numbers in bold indicate true MTD.

have shown that mistaking a DLT for an OCT, when it occurs less than 15% of the time, will not put patients at significant risk as estimated by the number of DLTs under either method. This is in agreement with reports that have shown that patients participating in phase I trials are not exposed to an increased risk for life-threatening events, having a less than 0.5% risk of a drug-related fatality (17–22). We refer to an individual harm when a single patient or a cohort of patients is assigned to a higher dose as a result of an error in toxicity attribution. In the 3+3 design, individual harm is minimal overall because the number of patients treated at a level higher than the MTD will not be more than 3 unless these errors occur very frequently. Under CRM, when DLTs are mistaken as OCTs, the dose may escalate to a level above the true MTD, but additional patients with DLTs will certainly result in deescalation

(23). Our simulations confirmed that overall, only a small number of patients would be exposed to levels higher than the MTD with CRM (20% and 3% at 1 and 2 levels above the MTD, respectively; Table 2).

The 2 trial designs do behave significantly differently when OCTs are mistakenly attributed as DLTs. Type B errors are probably more common than type A errors in phase I trials because physicians are less familiar with the side effect profile of new agents and are eager to avoid potential patient harm. A type B error rate of 30% will stall the 3+3 design at dose 1 (or dose -1 if that is permitted) with high probability (93%), as opposed to 35% when there are no errors (scenario 3). This illustrates how early DLTs cannot be over-ridden in the 3+3 design. This is consistent with the findings of other authors who have shown that the 3+3 design is conservative and tends to stop early (24), recommending an

incorrect dose on an average 75% of the time (14–16, 25, 26). This is due to the fact that the 3+3 does not use the cumulative experience of all the patients accrued in a trial, and instead it only uses the DLTs seen in the present cohort. CRM is consistently superior, on average 35% more accurate than the 3+3 in finding the true MTD, and this superior accuracy remains in the presence of attribution errors. This is a result of the adaptive nature of CRM, which allows deescalation in the presence of DLTs but subsequent reescalation if acceptable. CRM is a design with memory (27); thus, it can quickly correct the estimated rates and acceptable doses even in the presence of attribution error, as long as the error rate is less than 20%.

The above 2 errors have different implications for the process of drug development. Collective harm refers to the harm we impose on all future patients by recommending a wrong dose for future studies. A phase I trial can only estimate the MTD; the true effective dose has to be determined in a phase II efficacy study in which trials are powered both for efficacy and toxicity. Unfortunately, different phase I designs would often lead us to different MTDs, and the correct dose can be provided only through theoretical simulations. Two recent reviews (11, 28) showed that there is still reluctance among the investigators to use model-based designs, possibly because they are considered complex. The 3+3 is easy to implement, it requires very few patients if the MTD turns out to be among the first 3 levels, and in a particular setting of a 12-patient study, it can be a short trial. The major limitation, however, is that it leads to a wrong dose approximately 75% of the time, and attribution errors further increase the likelihood that the entire phase II program will be conducted with a suboptimal, possibly invalid dose. Given the narrow therapeutic window for many drugs, the result on drug development may be unrecoverable and would only be mitigated by either (i) including a consistent dose-escalation clause in phase II, which is very rarely done, and/or (ii) by using adaptive dose-finding designs. Another alternative is to develop phase I designs that do not group different attribution levels, especially groups that suggest uncertainty such as unlikely, possibly, or probably, into a dichotomous outcome of presence/absence of DLT. However, current designs use DLTs as their endpoint with the assumption that there is no misclassification of drug-related SAEs as DLTs.

We have illustrated that the impact of attribution errors depends on the magnitude of error rates and the rates of true DLTs. Our results depend on the numerical properties

of the simulated cases we studied (29, 30), and they are not based on prospective trials. The attribution error rate that occurs in practice in phase I trials is not known. Attribution errors have been noted in the phase III randomized setting (31–33); however, the reported error rates are likely less than those observed in the phase I setting (4). For this reason, we evaluated the methods under a number of different parameters and scenarios. However, error rates that change dynamically within a trial and from patient to patient are not addressed in this simulation.

Model-based designs are more accurate and more robust in the presence of clinician attribution errors. The problem of toxicity attribution becomes more relevant as we move into drug combinations of more than one novel agent with neither agent previously being tested in humans (34). In such a setting, the question of toxicity attribution has no clear answer. Although the expected clinical benefit cannot be known in such early testing, recent work suggests that phase I patients expect some benefit (35). If we were to justify their participation in a clinical trial that might be more likely to harm them than to provide benefit, then the justification must be that at least the trial will determine the correct dose for future patients. Moreover, if the drug turns out to be efficacious in later testing, then the majority of phase I patients treated under CRM will have received an efficacious dose without the need to expand accrual at the MTD (27).

Disclosure of Potential Conflicts of Interest

The authors have no conflicts of interest to disclose.

Authors' Contributions

Conception and design: A. Iasonos, M. Gounder, D.R. Spriggs, S. Zohar, J. O'Quigley

Development of methodology: A. Iasonos, S. Zohar, J. O'Quigley

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Iasonos, M. Gounder

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Iasonos, M. Gounder, D.R. Spriggs, D.M. Hyman, S. Zohar, J. O'Quigley

Writing, review, and/or revision of the manuscript: A. Iasonos, M. Gounder, D.R. Spriggs, J.F. Gerecitano, D.M. Hyman, J. O'Quigley

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Iasonos, M. Gounder, D.R. Spriggs

Study supervision: A. Iasonos, J. O'Quigley

Grant Support

This work was partially supported by the NIH (U01 CA069856; A. Iasonos and D.R. Spriggs)

Received March 1, 2012; revised June 15, 2012; accepted July 8, 2012; published OnlineFirst July 23, 2012.

References

1. Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009;6:430–40.
2. Ellenberg SS, Fleming TR, DeMets DL. *Data monitoring committees in clinical trials: A practical perspective*. West Sussex, England: John Wiley & Sons Ltd.; 2003.
3. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. September 2010 draft guidance for industry and investigators: safety reporting requirements for INDs and BA/BE studies. [Cited 2012 Sept 7]. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf>
4. Mukherjee SD, Coombes ME, Levine M, Cosby J, Kowaleski B, Arnold A. A qualitative study evaluating causality attribution for serious

- adverse events during early phase oncology clinical trials. *Invest New Drugs* 2011;29:1013–20.
5. Sargent DJ, George SL. Clinical trials data collection: when less is more. *J Clin Oncol* 2010;28:5019–21.
 6. FDA. Guideline for industry: clinical safety data management: definitions and standards for expedited reporting. 1995. [Cited 2012 Sept 7]. Available from: <http://www.fda.gov/downloads/Drugs/Guidance-ComplianceRegulatoryInformation/Guidances/UCM073087.pdf>
 7. FDA. Investigational new drug safety reporting requirements for human drug and biological products and safety reporting requirements for bioavailability and bioequivalence studies in humans. Final rule. *Fed Regist* 2010;75:59935–63.
 8. Storer BE. Design and analysis of Phase I clinical trials. *Biometrics* 1989;45:925–37.
 9. O'Quigley J, Pepe M, Fisher L. Continual Reassessment Method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990;46:33–48.
 10. O'Quigley J, Paoletti X, Maccario J. Non-parametric optimal design in dose finding studies. *Biostatistics* 2002;3:51–6.
 11. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *J Clin Oncol* 2007;25:4982–6.
 12. O'Quigley J, Shen L. Continual reassessment method: a likelihood approach. *Biometrics* 1996;52:673–84.
 13. Zohar S, O'Quigley J. Sensitivity of dose-finding studies to observation errors. *Contemp Clin Trials* 2009;30:523–30.
 14. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. *Clin Trials* 2008;5:465–77.
 15. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clin Trials* 2006;3:57–71.
 16. He W, Liu J, Binkowitz B, Quan H. A model-based approach in the estimation of the maximum tolerated dose in phase I cancer clinical trials. *Stat Med* 2006;25:2027–42.
 17. Horng S, Emanuel EJ, Wilfond B, Rackoff J, Martz K, Grady C. Descriptions of benefits and risks in consent forms for phase 1 oncology trials. *N Engl J Med* 2002;347:2134–40.
 18. Horstmann E, McCabe MS, Grochow L, Yamamoto S, Rubinstein L, Budd T, et al. Risks and benefits of phase 1 oncology trials, 1991 through 2002. *N Engl J Med* 2005;352:895–904.
 19. Kurzrock R, Benjamin RS. Risks and benefits of phase 1 oncology trials, revisited. *N Engl J Med* 2005;352:930–2.
 20. Muggia FM. Phase 1 clinical trials in oncology. *N Engl J Med* 2005;352:2451–3.
 21. Roberts TG Jr, Goulart BH, Squitieri L, Stallings SC, Halpern EF, Chabner BA, et al. Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials. *JAMA* 2004;292:2130–40.
 22. Sekine I, Yamamoto N, Kunitoh H, Ohe Y, Tamura T, Kodama T, et al. Relationship between objective responses in phase I trials and potential efficacy of non-specific cytotoxic investigational new drugs. *Ann Oncol* 2002;13:1300–6.
 23. Cheung Y. Coherence principles in dose-finding studies. *Biometrika* 2005;92:863–73.
 24. Hamberg P, Ratain MJ, Lesaffre E, Verweij J. Dose-escalation models for combinations phase I trials in oncology. *Eur J Cancer* 2010;46:2870–8.
 25. Zhao L, Lee J, Mody R, Braun TM. The superiority of the time-to-event continual reassessment method to the rolling six design in pediatric oncology Phase I trials. *Clin Trials* 2011;8:361–9.
 26. Reiner E, Paoletti X, O'Quigley J. Operating characteristics of the standard phase I clinical trial design. *J Comput Stat Data Analysis* 1999;30:303–15.
 27. O'Quigley J, Zohar S. Experimental designs for phase I and phase I/II dose-finding studies. *Br J Cancer* 2006;94:609–13.
 28. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *J Natl Cancer Inst* 2009;101:708–20.
 29. Conaway MR, Dunbar S, Peddada SD. Designs for single- or multiple-agent phase I trials. *Biometrics* 2004;60:661–9.
 30. O'Quigley J, Iasonos A. *Handbook of statistics in clinical oncology*. London (UK): CRC Press; 2011.
 31. Hillman SL, Mandrekar SJ, Bot B, DeMatteo RP, Perez EA, Ballman KV, et al. Evaluation of the value of attribution in the interpretation of adverse event data: a North Central Cancer Treatment Group and American College of Surgeons Oncology Group investigation. *J Clin Oncol* 2010;28:3002–7.
 32. Kaiser LD, Melemed AS, Preston AJ, Chaudri Ross HA, Niedzwiecki D, Fyfe GA, et al. Optimizing collection of adverse event data in cancer clinical trials supporting supplemental indications. *J Clin Oncol* 2010;28:5046–53.
 33. Mahoney MR, Sargent DJ, O'Connell MJ, Goldberg RM, Schaefer P, Buckner JC. Dealing with a deluge of data: an assessment of adverse event data on North Central Cancer Treatment Group trials. *J Clin Oncol* 2005;23:9275–81.
 34. Woodcock J, Griffin JP, Behrman RE. Development of novel combination therapies. *N Engl J Med* 2011;364:985–7.
 35. Locock L, Smith L. Personal benefit, or benefiting others? Deciding whether to take part in clinical trials. *Clin Trials* 2011;8:85–93.