|  |
| --- |
| TITLE **The Prevalence and Impact of Skipping Direct Evaluation of Efficacy Before Phase 3 Trials in Cancer** |
| INTRODUCTIONDrug development typically follows a regimented process whereby the prospect of efficacy is examined in early phase “exploratory” trials, and if signal of efficacy is obtained, re-tested in more demanding, “confirmatory” trials. Yet occasionally, this timeline is compressed, with “confirmatory” phase 3 trials launched on the backs of equivocal, negative or absence of direct exploratory evidence (i.e. no earlier phase testing, earlier phase testing with ambiguous results, or an earlier phase trial that is negative on its primary endpoint).1Several reports have investigated the relationship between phase 2 evidence and phase 3 trial outcomes. For example, Addeo et al. took a sample of 67 phase 3 trials that were non-positive on their primary endpoint and published in top tier journals.2 For 42% of these trials, no matched phase 2 trial was found, and for another 12%, the matched phase 2 trial was negative on its primary endpoint. These data – though derived from a non-random sample—suggest that many phase 3 trials are launched absent supporting phase 2 evidence. Gomley et al. note that many trials testing checkpoint inhibitor combinations for myeloma were launched absent phase 2 trial evidence; these phase 3 trials were consistently negative.3 Balasubramanian found that 65% of glioma phase 3 trials were launched after equivocal or absent phase 2 trials.4 Chan et al. built a sample of over 300 phase 2/phase 3 dyads involving targeted treatments, and probed characteristics of phase 2 trials that were predictive of a positive outcome in a phase 3 trial. In this study, phase 2 negativity was predictive of phase 3 trials being negative.5 Finally, Liang et al. examined the relationship of effect sizes in matched phase 2 and 3 trials; in this study, authors restricted their sample to dyads in which the phase 2 trial was randomized. Authors observed a regression of effect sizes as drugs advanced to phase 3 trials; they also observed a greater probability of negative phase 3 studies if they were launched based on subgroup analyses in otherwise negative phase 2 trials. In this study, a quarter of phase 3 trials in the dyads were launched based on non-positive randomized phase 2 trials.6These studies all suggest it is common for phase 3 studies in cancer to be launched without direct and consistent supporting phase 2 evidence. However, none of the above studies provide a clear estimate of the prevalence of launching phase 3 trials absent direct supporting phase 2 evidence in cancer. Nor do they probe the most important moral dimensions of this widespread practice- the risk/benefit implications for volunteers and society. The present study will assess the consequences for patients and research systems of running phase 3 studies lacking direct supporting phase 2 trial evidence (or phase 2-like evidence, like expansion cohorts in phase 1- throughout we use “phase 2” as short hand for all first studies where efficacy is evaluated and endpoints prespecified). We begin this study articulating two hypotheses that might justify skipping phase 2 trials. First, phase 2 trials might justifiably be skipped when there is unusually compelling evidence from phase 1 or other research to overcome an otherwise deficit in evidence supporting launch of a phase 3. If this conjecture is correct, we should observe that primary outcome attainment and pooled effect sizes will not be significantly worse for phase 3 trials that skipped phase 2 as compared with those that did. Second and alternatively phase 3 trials lacking direct supporting evidence might be justifiably launched in cases where the deficit in direct supporting efficacy evidence is compensated by a major safety advantage for the new treatment. If this hypothesis is correct, phase 3 trials skipping phase 2 trials should show significantly greater safety as compared with those where an intervention was supported by phase 2 trials. The present study- in addition to testing the above hypotheses- also seeks to overcome methodological and sampling limitations in the existing literature. For example, almost all studies canvassed above build their phase 2 / phase 3 dyads using publications, rather than registration records. However, publication and reporting biases are likely to confound results. The present study will build its sample using sampling based on registration records. Many studies described above are opaque as to how phase 2 and 3 trials were matched with each other, or offered poorly justified matching criteria (e.g., based on what phase 3 studies cite, as opposed to what is in the literature). To our knowledge, none of these studies expressly investigated what happens in phase 3 trials when they lack any prior phase 2 study. Many did not investigate the effects of stringent vs. less stringent matching. And all of the above studies focus on subsets of phase 3 trials (e.g., glioma studies, randomized studies, studies involving targeted treatments) rather than all contemporary phase 3 trials. In what follows, we will create a sample of recent phase 3 cancer trials to investigate 1) the prevalence of launching phase 3 trials absent supporting earlier phase evidence, and 2) the relationship of supporting earlier phase evidence to risk-benefit for patients enrolled in phase 3 trials. Our study is intended to evaluate contemporary practices (e.g., if things are working as they should, studies launched absent supporting earlier phase evidence should show greater safety for the study intervention than that in phase 3 studies supported by prior earlier phase trials). Our study is also intended to inform judgments as to whether it is scientifically and ethically appropriate to skip earlier phase trials. Finally, our study will help provide a picture of the ethical tradeoffs associated with contemporary research practices against an alternate world in which all phase 3 trials were preceded by positive earlier phase trials. |
| DEFINITIONS:“Dyad” = phase 3 trial paired with its supporting evidence. Note: if there is no prior earlier phase trial testing the same drug in the same indication, then the dyad is effectively a monad (e.g. no earlier phase evidence paired with a phase 3 trial).“Matched dyad” = earlier phase trial and phase 3 trial using identical conditions. Note- unless otherwise specified, assume “stringent” matching criteria. Elsewhere in the protocol we will use “relaxed” matching criteria (e.g. non-stringent matching).“Supporting” evidence = Earlier phase trial that is (+) on a pre-specified primary endpoint within a matched dyad.“Non-supporting” evidence = Earlier phase trial that is not (+) on a pre-specified primary endpoint within a matched dyad.(+) trial = a clinical trial that results in a statistically significant outcome in the right direction for a new drug, for all pre-specified primary endpoints.7(-) trial = a clinical trial that results in a statistically significant outcome in the wrong direction for a new drug, based on a pre-specified primary endpoint.Non-positive trial = a clinical trial that does not result in a statistically significant outcome in the right direction for a new drug, based on a pre-specified primary endpoint. |
| PRIMARY OBJECTIVE1. To assess the prevalence of cancer phase 3 clinical trials launched absent matched supporting earlier phase efficacy evaluation.
2. To compare the risk/benefit of phase 3 trials launched absent stringently matched supporting earlier phase trial evidence with that for phase 3 trials launched with stringently matched supporting earlier phase trial evidence.
 |
| SECONDARY OBJECTIVES1. To explore conditions under which phase 3 trials are launched absent matched supporting earlier phase evidence.
2. To compare the risk/benefit of phase 3 trials launched absent permissive matched supporting earlier phase trial evidence with that for phase 3 trials launched with permissively matched supporting earlier phase trial evidence.
3. To explore whether the use of a phase 1 expansion cohort is equivalent to an earlier matched phase 2 trial in terms of providing prior evidence for phase 3 trials.
4. To compare the risk and benefit (using a risk-adjusted efficacy measure) of phase 3 trials launched absent matched supporting earlier phase trial evidence with that for phase 3 trials launched with supporting earlier phase trial evidence.
5. To explore whether there are predictors of skipped earlier phase efficacy evaluations.
 |
| METHOD OVERVIEWWe will create a sample of phase 3 trials. We will then search backwards for matched earlier phase trials. To estimate prevalence, we will estimate the proportion of phase 3 trials that are launched based on a positive earlier phase trial result. We will then assess whether confirmatory trials launched on earlier phase evidence have significantly different risk and benefit profiles. |
| SAMPLE AND SAMPLING METHODSPhase 3 Trial Sampling: All phase 3 solid tumor cancer trials registered as closed or terminated in trials with actual primary completion dates from 2013-2018 will be downloaded from ClinicalTrials.gov for screening and every trial matching our criteria will be included in our sample working backwards through time until we reach our target sample of 170 phase 3 trials. This strategy of inclusion was chosen to provide the most recent sample of trials while still allowing at least 3 years for results data to become available.Eligibility criteriaEligibility: a) randomized; b) must use a comparator that is either placebo or another treatment (as opposed to another dose of same drug); c) at least one primary endpoint is efficacy; d) must test an anti-cancer FDA-regulated product; e) enrollment exceeding 100 patients; f) results for final analysis available; g) must be the first phase 3 trial to close and is registered on clinicaltrials.gov investigating this treatment in the specific indication:8 h) has PFS as a registered outcome; i) at least one US,CAD or EU (+UK) research site. Exclusion criteria: a) adjuvant, radiotherapy, prevention, or surgery trial; b) cell therapies where the exact identity of the product cannot be ascertained c) treating side effects of cancer or treatment d) head-to-head trials of standard of care interventions as the primary analysis or where it is unclear which is the experimental intervention (determined by an oncologist); f) Hematological Malignancies; e) trials testing drug cocktails or variations thereof (e.g. FOLFORI) f) studies for which matching is going to be extremely complicated due to intervention/indication conditions (see Appendix). Trial outcomes will be sought via searches of clinicaltrials.gov, OVID (MEDLINE and EMBASE), Google, and if necessary, principal investigator queries. Earlier Phase Trial Matching:Two trained researchers will search clinicaltrials.gov, EU Clinical Trials Register (for trials with only EU sites), OVID, and citations within phase 3 publications/protocols where available for earlier phase trials preceding the phase 3. Only trials registered as testing efficacy (e.g., response rate, time-to-event) as a primary endpoint or phase I with expansion cohorts or basket / umbrella -type trials containing pre-specified efficacy hypotheses and investigating the same drug-indication- pairing will be included as preceding exploratory trials. Matched exploratory evidence will be classified as follows:1. Stringent match: Must match for all of the following criteria (see Appendix for explanation of choices)
	1. Timing: Due to the inconsistency of trial completion date reporting, earlier phase data will have been considered available at the start of the phase 3 trial when the earlier phase trial started at least 12 months before the start of the phase 3 trial of interest.5
	2. Intervention: at least one arm must match each other.
		1. Exact same drug/combination
		2. Same dose (defined as phase 3 dose falling within 50% and 200% of dose used in phase 2 trial;7
		3. Same schedule (as determined by an oncologist);
	3. Population: Populations under investigation match or are included in a broad population, based on histology and marker status and an appropriate analysis\* was conducted.
		1. Earlier phase trials investigating more than one histology or biomarker must separately analyze the specific histology or biomarker used in the phase 3 per an acceptable analysis.
		2. Acceptable matches include

|  |  |
| --- | --- |
| EARLIER EVIDENCE | PHASE 3 TRIAL |
| A Phase 1 expansion cohort with a prespecified analysis of breast cancer population | Breast Cancer population  |
| A phase 2 trial investigating more than one indication with a prespecified analysis of breast cancer population | Breast Cancer population  |
| An earlier phase trial investigating the breast cancer population overall and has an analysis for HER2+ Breast Cancer population | HER2+ Breast Cancer population  |
| An earlier phase trial investigating the HER2+ Breast Cancer population | Breast Cancer Population |
| A phase 2 trial investigating NSCLC with a prespecified analysis of the squamous cell NSCLC population | Squamous Cell NSCLC |

1. Permissive match\*\*: Must match insofar as the earlier phase trial started 12 months before the start of the P3 trial (same rule as above) but can vary on on any and all of the below (note: we will record the dimensions that are invariant and caused them to fall short of stringent matching).
	* 1. Differences in intervention
			1. Slight differences in the intervention in the case of combination therapy (as determined by an oncologist)(ex. Same class of chemotherapy).
			2. Differences in the dosing of the experimental drug/s
			3. Differences in the schedule of the experimental drug/s
		2. Incomplete match for population
			1. In the following manner
				1. Earlier phase trial investigated a broad population and did not analyze in a preplanned and powered subgroup the histology or marker subgroup that was then investigated in the Phase 3 trial, but the subgroup was included.
				2. Differences between trials in regard to “after failure with x”, advanced, or metastatic status (will be flagged and confirmed with an oncologist)
2. No matched prior earlier phase trial

As a quality control on all searches, corresponding authors for each phase 3 trial will be emailed (+ 1 reminder for no responses) with the result of the matching (if any) for their trial in order to verify our classifications or to elicit evidence our searches missed. \*Note: An acceptable analysis is defined as the earlier phase trials (phase 2 or phase 1 with expansion cohorts) including at least 30 patients in the population of interest and have a pre-specified statistical analysis plan to measure efficacy for the population of interest (i.e. power calculations/specific positivity threshold) (This will be confirmed on a trial-by-trial basis with the research team.)\*\*Note: In the case that there is more than one permissively matched earlier phase trial, we will prioritize positive trials. If there is more than one permissive positive match, the one matching on the most criteria will be chosen. We will also record when there are only nonpositive permissive matched earlier phase trials but where there is a “positive” subgroup in the earlier phase trial that is identical to the population in the phase 3 trial. In the case there is a stringent match, we will not further evaluate prior trials for permissive matches.Sample size rationale / calculations:Our sample size is k=170 phase 3 trials. This sample size was justified using an *a priori* power calculation for examining a binary subgroup contrast in a meta-analysis (see PRIMARY OUTCOME ANALYSIS)12,13. The power calculation and R script can be accessed here: [link](https://github.com/RemNil/STREAM-lab/blob/main/Power_analyis_150k_missed_a_step.md). |
| EXTRACTIONPhase 3 TrialsThe following items will be extracted with the following hierarchy: a) first published report of full primary (not interim), then b) registration record: 1) all elements for matching; 2) positivity on pre-specified primary efficacy endpoint; 3) efficacy outcomes (e.g., HR for PFS; ORR for treatment arm); 4) safety outcomes (number of Serious AEs in each arm). In addition, we will collect:* date of first enrolled patient
* status of trial and if applicable, the reason for termination.
* sponsorship (large pharma firm, non-large pharma firm, non-industry)
* drug class: a) approval status of newest drug at time of enrollment (yes, no); b) combo vs. mono;
* use of biomarker enrichment design
* 5-year relative survival for indication

Earlier Efficacy TrialsThe following items will be extracted from all earlier trials evaluating efficacy:* all elements necessary for matching
* date of primary outcome completion (from registration record); date of last patient recruitment if available.
* positivity on pre-specified primary endpoint. Note: in the event there are two earlier phase trials with stringent matches, existence of one positive trial is sufficient to say that there was one positive prior earlier phase trial with a stringent match.

All items will be double coded for at least a sample to estimate error rates for each coded variable. All differences or uncertain extractions will be adjudicated in consultation with at least one oncologist. |
| PRIMARY OUTCOME ANALYSES1. Prevalence: We will report descriptive statistics (number and proportion) of phase 3 trials launched on **a)** stringent matched positive earlier phase trials vs. **b)** no prior stringent matched positive earlier phase trials, which will be reported as: i) stringent matched nonpositive earlier phase trials, ii) positive permissive matched earlier phase trials; iii) nonpositive permissive matched earlier phase trials v) absent any matched earlier phase trial evidence.
2. Meta-analyses with subgroup contrast: We will meta-analyze **a)** HR of PFS; **b)** RR of serious Adverse Events (sAEs) for phase 3 trials, stratified by phase 3 trials supported by stringently matched positive earlier phase evidence vs. those trials that are not supported by stringently matched positive phase 2.

Null hypothesis: No difference between pooled HR and RR between subgroups (Phase 3 trials preceded by stringently matched positive earlier phase trials vs. phase 3 studies not preceded by stringently matched positive earlier phase trials)Statistical test: We formally test that homogeneity exists between the sample estimates of the population parameters and that any variation is no more than expected when taking samples from the same population (“sampling error”) using the two-tailed p-value of Cochran’s Q as decision rule with a nominal significance level of 0.05.14–16To do so, we will conduct a random effects meta-analysis (inverse variance method) stratified by condition (lack vs presence of stringently matched positive earlier phase evidence) using the functions “metagen” for log-transformed HR of PFS and “metabin” for binary data, i.e., number of sAE (using R package “meta”17 and “metafor"18) . Details on meta-analytical method:* + - Inverse variance method
		- Paule-Mandel estimator for tau2 (assuming separate heterogeneity estimates for tau2 in subgroups)
		- Q-profile method for confidence interval of tau2 and tau19
		- Hartung-Knapp adjustment for random effects model

We use the Paule-Mandel estimator20 based on recommendations from Veroniki et al.21 We use the Hartung-Knapp method to adjust the variance estimator for the treatment estimate of the random effects meta-analysis.22–25 We assume separate between-study heterogeneity estimates due to the likelihood of unbalanced subgroups.26 We use variance correction if Hartung-Knapp standard error is smaller than standard error from classic random effects meta-analysis.27 |
| SECONDARY ANALYSIS 1. Prevalence: Using the same stratification (i.e., phase 3 trials supported by stringently matched positive earlier phase evidence vs. those trials that are not supported by stringently matched positive phase 2), we will also descriptively report the **a)** number and proportion of phase 3 trials positive on primary endpoint; **b)** proportion of phase 3 trials terminated for safety and/or futility. **c)** proportion of phase 3 trials funded by industry.
2. Investigation of the impact of Phase 1 expansion cohorts on the results:

To understand if the use of a phase 1 expansion cohort is equivalent to a phase 2 trial, we will reperform primary analysis #2 while only including trials that are matched to prior phase 2 stringent matches. 1. Lasagna graph: To understand where each permissive match fell short of a stringent match, we will create a lasagna graph identifying the variable/s of interest that did not match our criteria for each trial.
	1. Populations do not match with a proper analysis
		* No analysis
		* not accepted analysis
	2. Mismatch on drugs used in the intervention
	3. Dosing mismatch
	4. Schedule mismatch
	5. Line of treatment
2. Benefit-risk meta-analysis:

**a)** Based on the “Net Efficacy Adjusted for Risk (NEAR)”28, we calculate a **2×2 table combining efficacy (ORR) and safety (all cause withdrawal) results and calculate the relative risk for** “NEAR” based on the table results.

|  | ORR responders | ORR non-responders |  |
| --- | --- | --- | --- |
| Without sAE  |  a | b | Total without sAE |
| With sAE | c | d | Total suffering sAE |
|  | Total responders | Total non-responders | Total studied |

We then run a meta-analysis (using the metabin R function) for relative risk for “NEAR” stratified by phase 3 trials supported by stringently matched positive earlier phase evidence vs. those trials that are not supported by stringently matched positive earlier phase trial.1. Permissive meta-analyses with subgroup contrast: We will meta-analyze **a)** HR of PFS; **b)** RR of serious Adverse Events (sAEs) for phase 3 trials, **c)** NEAR stratified by phase 3 trials supported by stringently and permissively matched positive (including permissive positivity of subgroups) earlier phase evidence vs. those trials that are not supported by stringently or permissively matched positive earlier evidence.
2. Logistic regression: to determine predictors of skipped earlier phase trials

We define a binary outcome Y (stringently matched positive phase 2: yes/no) and a set of candidate predictor variables for the phase 3 trials Xn: * 1. X1: biomarker enrichment: yes/no
		+ *greater sophistication about molecular hypothesis may drive scientific basis for skipping phase 2*
	2. X2: drug class: categorical variable (binarized/dummy coded)
		+ *certain drug classes have less need for phase 2. We make no assumptions about potential associations*
		+ *Targeted or Immunotherapy vs Cytotoxic or Other*
	3. X5: approval status: yes/no
		+ *there may be greater pressure pre-approval to skip phase 2*
	4. X6: RR of SAEs in experimental test arm: numerical
		+ *expected greater safety may reduce the need for phase 2*
	5. X7: sponsorship industry: yes/no
		+ *Industry-involvement may impact skipping phase 2. We make no assumptions about potential associations*
	6. X8: 5-year relative survival rate: numerical
		+ *refractory cancers may drive relaxation of evidentiary standards for phase 3 trial*

We use logistic regression to determine any study-level factors significantly associated with “skipping phase 2”. If a factor (e.g., X6) is significantly associated with “skipping phase 2” (while all other factors X1-X5 are accounted for in the logistic regression as well), then we interpret this as exploratory evidence that the factor makes a difference on skipping phase 2. **j)** Post-hoc analyses: In case we find strongly significant associations (p<0.01) in the logistic regression outlined in i), we will analyze the association of those strongly significant factors with our primary outcomes. If more than one factor is strongly associated, we choose only the one with the largest effect on the outcome based on standardized regression coefficients.We will select either c) and/or d) from the primary outcomes depending on whether there is a rejection of the respective null hypothesis. We then run a meta-regression for the HR of PFS (and/or RR of sAEs) for phase 3 trials. By doing so, we test on exploratory grounds the interaction between two factors: (1) presence/absence of a stringently matched earlier phase trial and (2) the largest of any strongly significant association discovered in our prior analysis i) outlined above. Meta-regression for this pre-specified interaction is calculated using the meta-for package in R.  |
| ETHICSStudy does not involve human subjects and will therefore not seek out ethical approval. |
| NOTESReferences 1. West HJ. When the Signal From earlier phase Research Should Be a Warning Sign. *JAMA Oncol*. Epub ahead of print 7 January 2021. DOI: 10.1001/jamaoncol.2020.6598.2. Addeo A, Weiss GJ, Gyawali B. Association of Industry and Academic Sponsorship With Negative Phase 3 Oncology Trials and Reported Outcomes on Participant Survival: A Pooled Analysis. *JAMA Netw Open* 2019; 2: e193684.3. Gormley NJ, Pazdur R. Immunotherapy Combinations in Multiple Myeloma — Known Unknowns. *New England Journal of Medicine* 2018; 379: 1791–1795.4. Balasubramanian A, Gunjur A, Hafeez U, et al. Inefficiencies in Phase II to Phase III Transition Impeding Successful Drug Development in Glioblastoma. *Neuro-Oncology Advances*. Epub ahead of print 22 December 2020. DOI: 10.1093/noajnl/vdaa171.5. Chan JK, Ueda SM, Sugiyama VE, et al. Analysis of Phase II Studies on Targeted Agents and Subsequent Phase III Trials: What Are the Predictors for Success? *JCO* 2008; 26: 1511–1518.6. Liang F, Wu Z, Mo M, et al. Comparison of treatment effect from randomised controlled phase II trials and subsequent phase III trials using identical regimens in the same treatment setting. *European Journal of Cancer* 2019; 121: 19–28.7. Research C for DE and. Multiple Endpoints in Clinical Trials Guidance for Industry. *U.S. Food and Drug Administration*, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry (2020, accessed 26 March 2021).8. Clinical trials viewer, https://trials.bgcarlisle.com/ (accessed 30 March 2021).9. NCCN - Evidence-Based Cancer Guidelines, Oncology Drug Compendium, Oncology Continuing Medical Education, https://www.nccn.org/ (accessed 29 March 2021).10. Zhang SX, Fergusson D, Kimmelman J. Proportion of Patients in Phase I Oncology Trials Receiving Treatments That Are Ultimately Approved. *JNCI: Journal of the National Cancer Institute* 2020; 112: 886–892.11. Rosenfeld PJ, Feuer WJ. Lessons from Recent Phase III Trial Failures: Don’t Design Phase III Trials Based on Retrospective Subgroup Analyses from Phase II Trials. *Ophthalmology* 2018; 125: 1488–1491.12. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods* 2004; 9: 426–445.13. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods* 2017; 8: 290–302.14. Sedgwick P. Meta-analyses: heterogeneity and subgroup analysis. *BMJ* 2013; 346: f4040.15. Borenstein M, Higgins JPT. Meta-analysis and subgroups. *Prev Sci* 2013; 14: 134–143.16. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004; 23: 1663–1682.17. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evidence-Based Mental Health* 2019; 22: 153–160.18. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; 36: 1–48.19. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med* 2007; 26: 37–52.20. Paule RC, Mandel J. Consensus Values, Regressions, and Weighting Factors. *J Res Natl Inst Stand Technol* 1989; 94: 197–203.21. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016; 7: 55–79.22. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med* 2001; 20: 1771–1782.23. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 2001; 20: 3875–3889.24. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 2014; 14: 25.25. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods* 2019; 10: 83–98.26. Rubio-Aparicio M, Sánchez-Meca J, López-López JA, et al. Analysis of categorical moderators in mixed-effects meta-analysis: Consequences of using pooled versus separate estimates of the residual between-studies variances. *Br J Math Stat Psychol* 2017; 70: 439–456.27. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003; 22: 2693–2710.28. Boada JN, Boada C, García-Sáiz M, et al. Net Efficacy Adjusted for Risk (NEAR): A Simple Procedure for Measuring Risk:Benefit Balance. *PLoS One*; 3. Epub ahead of print 31 October 2008. DOI: 10.1371/journal.pone.0003580.29. Jardim DL, Groves ES, Breitfeld PP, et al. Factors associated with failure of oncology drugs in late-stage clinical development: A systematic review. *Cancer Treat Rev* 2017; 52: 12–21.30. Zia MI, Siu LL, Pond GR, et al. Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. *J Clin Oncol* 2005; 23: 6982–6991.31. Monzon JG, Hay AE, McDonald GT, et al. Correlation of single arm versus randomised earlier phase oncology trial characteristics with phase 3 outcome. *Eur J Cancer* 2015; 51: 2501–2507.32. Seruga B, Ocana A, Amir E, et al. Failures in Phase III: Causes and Consequences. *Clin Cancer Res* 2015; 21: 4552–4560. |

**Appendix**

**Matching Criteria**

A research assistant performed a literature search find previous reviews that looked at similar questions regarding to phase 2 and phase 3 transitions. We extracted the matching criteria that each review used to connect phase 3 and phase 2 trials. The graph below shows the results. This was used to understand the possible matching criteria. We decided based on the heterogeneity of this graph and discussion with an oncologist that using the same comparator drug, outcome, level of disease advancement (Such as: disease progression status (staging, advanced, metastatic, symptomatic), and line of treatment were not important for either type of match. We decided that stringent matches would need to use the same drug, have similar schedules and doses, and would need to have sufficient time between phases. Permissive matches would allow variations on these variables. We also decided to create a new criterion for population matching that allowed for subgroup analyses of the same population as long as there was an accepted analysis. Stringent vs permissive criteria for this project were determined by reviewing these materials with an experienced oncologist. 2,4–6,29–32





Screening Criteria

**CT.gov SEARCH PARAMETERS**

1. **Condition or disease**: Cancer OR cancers OR carcinoma OR carcinomas OR malignant OR malignancy OR malignancies OR tumor OR tumors OR tumour OR tumours OR neoplasm OR neoplasms OR metastatic OR lymphoma OR leukemia OR leukemias
2. **Study type**: “Interventional Studies (Clinical Trials)”
3. **Status of recruitment**: no restriction (looking for Actual primary completion dates, so likely mostly Completed/Terminated- check filtered results to see)
4. **Phase**: 3
5. **Study start date**: no restriction
6. **Primary completion date**: 01/01/2013-01/01/2019

**SEMI-AUTOMATED SCREENING (using Excel filters) FOR PHASE 3** [**Studies**](https://clinicaltrials.gov/ct2/results?cond=Cancer+OR+cancers+OR+carcinoma+OR+carcinomas+OR+malignant+OR+malignancy+OR+malignancies+OR+tumor+OR+tumors+OR+tumour+OR+tumours+OR+neoplasm+OR+neoplasms+OR+metastatic+OR+lymphoma+OR+leukemia+OR+leukemias&term=&type=Intr&rslt=&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=01%2F01%2F2013&prcd_e=01%2F01%2F2019&sfpd_s=&sfpd_e=&rfpd_s=&rfpd_e=&lupd_s=&lupd_e=&sort=)

1. **Primary completion date**: check that type is “Actual” and not “Anticipated”
	1. Exclude, \*unless\* trial has an “Actual” overall completion date
2. **Intervention/Treatment**: exclude if trial:
	1. does not include at least one intervention that is classified as a “Drug” or “Biological” (“Other” is manually checked)
	2. includes healthy volunteers
3. **Trial design**: exclude if trial is labelled as:
	1. “Non-randomized” in randomization field;
	2. “Single group assignment” in “Model” field;
	3. 1 in “Arms” field
	4. Phase 2/3
4. **Trial size:** enrollment must exceed 100 patients
5. **Trial status**: exclude if the trial recruitment status is:
	1. Withdrawn (i.e. no patients enrolled);
6. **Indication:** primary purpose is
	1. Diagnostic
	2. Screening
	3. Prevention
7. **Trial Location:** exclude if the trial does not have a
	1. US or CAD or EU (+UK) trial site

**MANUAL SCREENING For Phase 3 Trials**

1. **Indication**: trial in patients with a cancer diagnosis with aim to treat cancer. Exclude if:
	1. Trial is in Hematological Malignancies
	2. Trial is in non-cancer subjects (healthy, other diseases);
	3. Trial is for symptoms related to cancer or resulting from treatment;
	4. Trial tests an intervention for a procedure related to cancer treatment;
	5. Trial is for primary prevention of cancer.
2. **Trial design**:
	1. Exclude extension, bioequivalence studies
	2. must use a comparator that is either placebo or another treatment (as opposed to another dose / schedule of same drug)
	3. Exclude if it is a non-inferiority trial
3. **Intervention**: exclude if:
	1. Intervention is a non-FDA regulated intervention (e.g. vitamin supplement);
	2. Radiation or surgery or device is given in only one of the two arms;
	3. If the intervention is an adjuvant\* therapy
		1. Check for explicit mention of “adjuvant”, maintenance, or “neoadjuvant” in the trial record. If in doubt, refer to the definitions listed at the bottom of the page. The use of as an outcome is another clue.
	4. If the intervention is a cell therapies where the exact identity of the product cannot be ascertained or drug delivery mechanisms
	5. If the trial is testing two interventions head-to-head, it must be clear which is the new treatment of interest. (e.g., exclude standard of care head-to-head trials).\*additional step to do with Joey
4. **Outcome**: exclude if primary endpoint does not include a measure of efficacy against cancer (surrogate or clinical), i.e., exclude if:
	1. Primary outcome(s) comprise measures of safety only;
	2. Primary outcomes are measure of feasibility only (e.g. MTD, pharmacodynamics, immunogenicity);
	3. Does not have PFS (radiologic) as a registered outcome
5. **Judgement of investigators:** studies for which matching is going to be extremely complicated due to intervention/indication conditions

\* **Adjuvant or neoadjuvant therapy:** treatment that is secondary to the primary treatment modality, such as surgery or radiation, to lower the risk that the cancer will come back.

\*\***Maintenance (consolidation) therapy:** treatment given to keep cancer from coming back after it has disappeared following initial therapy (i.e. to prevent a relapse).

s^ definitions from NCI dictionary / feedback from Murph

**Potential Follow-up projects**

1. **Efficacy regression**
	1. To explore relationship in efficacy effect sizes for stringently matched dyads with supportive earlier phase trials vs. permissive matched dyads
	2. For the matched dyads that we identify (stringent and permissive), we will assess the relationship between the two for a) HR of OS, b) HR of PFS, c) median OS of treatment arm, d) median PFS of treatment arm, e) ORR of treatment arm. We will note the instances when there are different comparators between the p2 and p3 trial.
2. **Protocol Investigation**
	1. When the protocols are available for phase 3 trials, we would investigate the proportion of phase 3 trials that have DSMBs and the stopping boundaries. We hypothesize that trials without prior positive evidence would be more likely to have a DSMB and would have more stringent stopping rules.
3. **Alzheimer’s Disease**
	1. See if the same trends appear in neurodegenerative diseases
4. **Negative predictive value**
5. **Looking at trajectories for pre-approval trials**
	1. Look at the fraction of trajectories where the earlier phase trial is skipped and how often do those trials advance to license vs. trajectories where the earlier phase is not skipped.
6. **Redundancy**
7. **Looking at evidence that their protocols cite**

Tables of possible outcomes to our primary analysis #2 and possible interpretations:

Efficacy

|  |  |  |
| --- | --- | --- |
| HR possible outcomes | Interpretations | Further evidence from secondaries |
| Significantly better efficacy outcomes when trial is supported | Trials that are launched without prior evidence show worse efficacy outcomes-need to watch out for launching phase 3s without phase 2s.  | Secondary #3 would find what people are using now to determine when it is ok to skip-not working.  |
| Significantly better efficacy outcomes when trial is unsupported |  earlier phase is only skipped when there is excellent mechanistic evidence supporting the drug’s effect. Or, companies somehow use easier comparators when they skip. | Secondary #3 would find some of these factors used in the decision process. If we find this, we should assess whether comparator arm is recommended for the indication in clinical practice guidelines (e.g. NCCN) with equal frequency in both supported and unsupported group. |
| Insignificant difference in efficacy outcomes when trial is supported or not | Research community is able to judge when it is necessary to have a positive earlier phase trial before launching a phase 3 trial. Good measures in place to evaluate it.  | Secondary #3 would find some of these factors used in decision process |

Safety

|  |  |  |
| --- | --- | --- |
| RR possible outcomes | Interpretations  | Further evidence from secondaries |
| Significantly better safety outcomes when trial is supported | Need to really watch out for launching phase 3s without phase 2s- the lack of observation in target group means less is known about safety in target group; they are riskier in terms of safety | Secondary #3 would find what people are using now to determine when it is ok to skip-not working |
| Significantly better safety outcomes when trial is unsupported | Aligns with hypothesis above, will skip if there is less risk of serious adverse events. Means system is approximately working well WRT ensuring a uniform risk/benefit across the field | Secondary #3 would find some of these factors used in the decision process |
| Insignificant difference in safety outcomes when trial is supported or not  | Research community is able to judge when it is necessary to have a positive earlier phase trial before launching a phase 3 trial. Good measures in place to evaluate it | Secondary #3 would find some of these factors used in decision process |

Combined

|  |  |  |
| --- | --- | --- |
| Outcome for Efficacy | Outcome for Safety | Interpretations |
| Supported has better efficacy | Supported has better safety | Need to really watch out for launching phase 3s without phase 2s- they are riskier in terms of safety and efficacy |
|  | Unsupported has better safety | Phase 3 trials that are launched without evidence may do so because it has a better safety profile but there is a trade-off in terms of efficacy  |
|  | Insignificant | Need to watch out for launching phase 3s without phase 2s- they are less efficacious and are not any safer |
| Unsupported has better efficacy | Supported has better safety | The p2 is only skipped when they are really confident in efficacy, but the safety profile is worse. This suggests risk/benefit roughly preserved across all research |
|  | Unsupported has better safety | System is working really well in interpreting safety and efficacy evidence and deciding to earlier phase trials; But system is not doing earlier phase trials well. perhaps earlier phase trials intending to support phase 3 trials should be more demanding in terms of collecting more safety and efficacy evidence |
|  | Insignificant | P2 is only skipped when they are really confident in both efficacy and safety. There is no worse safety outcomes. Good equipoise. Perhaps earlier phase studies intended to support phase 3 trials should gather greater evidence of efficacy. |
| Insignificant | Supported has better safety | Although skipping a p2 trial does not have an efficacy disadvantage, it is riskier in terms of safety. Suggests safety information is inadequate when skipping- perhaps because treatment is not tested for ssafety in p2 in the target population. |
|  | Unsupported has better safety | P3 trials that skip earlier evidence have better safety profiles and do not have an efficacy disadvantage. Suggests review systems are balancing risk and benefit well |
|  | Insignificant | There is no difference between p3 trials that are supported by prior evidence or not. Good equipoise, good research standards |

**Statistics details for each analysis and potential code:**

metagen(log(toy$HR),

 lower = log(toy$LL95CI),

 upper = log(toy$UL95CI),

 studlab = toy$COMP,

 byvar=toy$COND,

 bylab="Intervention type",

 byseparator=": ",

 sm = "HR",

 comb.random=T,

 overall = TRUE,

 overall.hetstat = TRUE,

 hakn = TRUE,

 method.tau="PM", # (Paule and Mandel, 1982)

 tau.common=FALSE,

 method.tau.ci = "QP",

 adhoc.hakn = "se", # (Knapp and Hartung, 2003)

 prediction = TRUE) # (Higgins et al., 2009))