Supplement

Table of Contents

Supplemental methods

**P3 Sample Creation**

CT.gov SEARCH PARAMETERS For Phase 3 Trials:

1. **Condition or disease:** Alzheimer disease OR Alzheimer's disease OR Alzheimer Dementias OR Dementia of the Alzheimer's type OR dementia alzheimers OR Dementia of Alzheimers Type OR dats OR Alzheimer Type Dementia OR Senile Dementia OR Alzheimer Syndrome OR AD OR Parkinson disease OR Parkinson's disease OR PD OR Parkinson OR Primary Parkinsonism OR Paralysis Agitans OR Shaking palsy OR ALS OR Amyotrophic lateral sclerosis OR Gehrig Disease OR Motor neurone disease OR Charcot disease OR Huntington disease OR Huntington's disease OR Huntington's chorea OR Chronic progressive hereditary chorea OR MS OR Multiple Sclerosis OR MS (Multiple Sclerosis) OR Disseminated sclerosis OR Migraine OR Cephalalgia OR Head pain OR Pain in head OR Cephalgia OR Headache OR Epilepsy OR epileptics OR seizure disorders OR epilepsia OR TBI OR Traumatic Brain Injury OR brain traumas OR Traumatic encephalopathy OR brain injuries traumatic OR traumatic brain damage OR Brain damage OR cerebral damage OR injury brain OR cerebral injury OR Stroke OR Cerebrovascular accident OR cerebral vascular accident OR Apoplexy OR Brain attack OR Brain Vascular Accident OR TIA (Transient Ischemic Attack) OR Transient Ischemic Attack OR intracerebral haemorrhage OR subarachnoid haemorrhage
2. **Study type:** “Interventional Studies (Clinical Trials)”
3. **Status of recruitment:** no restriction (looking for Actual primary completion dates, so likely mostly Completed/Terminated/Active not recruiting but completed- checked filtered results to see)
4. **Phase:** 3
5. **Study start date**: no restriction
6. **Primary completion date**: 01/01/2011-01/01/2021
   1. The end range was chosen to allow one year between primary completion and depositing results as per the Final Rule.1 Our objective was to have at least 100 phase 3 trials but we saturated the sample for a full decade of Phase 3 trials. The target minimal sample size of 100 is selected because, for a primarily descriptive study, it seems likely to deliver a reasonably robust estimate of the prevalence of phase 3 bypass. Assuming 30% trials involve phase 2 bypass, availability of 30 trials involving bypass provides a reasonable starting point for secondary objectives for a first ever exploration of the prevalence of bypass.

SEMI-AUTOMATIC SCREENING (using excel filters) For Phase 3 Trials:

1. **Primary completion date**: checked that type is “Actual” and not “Anticipated”
   1. Excluded, \*unless\* trial had an “Actual” overall completion date;
2. **Trial design**: excluded if trial was labelled as:
   1. “Non-randomized” in randomization field;
   2. “Single group assignment” in “Model” field;
   3. 1 in “Arms” field;
3. **Trial size:** <30
4. **Trial status:** exclude if the trial recruitment status was:
   1. Withdrawn (i.e. no patients enrolled);
5. **Indication:** excluded if primary purpose is
   1. Diagnostic;
   2. Screening;
   3. Basic Science
6. **Intervention/Indication:** excluded if trial:
   1. Did not include at least one intervention that was classified as a “Drug” or “Biological” “ Dietary supplement” or “Genetic” (“Other” and “combination product” were manually checked); ie exclude procedure or behavioral or device or radiation
   2. Included healthy volunteers;
7. **Trial Location**: exclude if the trial does not have a
   1. US or CAD UK, EU, Australian

MANUAL SCREENING For Phase 3 Trials:

1. **Intervention:** Exclude if the intervention is
   1. surgery/behavioral/device/conditioning of stem cells/procedure/ biosimilar
   2. extension, discontinuation studies, phase 1/2/3
   3. head-to-head (trials pitting two approved SOC interventions against each other) or if there are more than two options for the experimental arm (ake “any anticoagulant”)
   4. treating a second condition in our conditions (ie infection in PD patients) (immune responses to vaccines)
2. **Comparator:** Exclude if the comparator is not placebo or another treatment (as opposed to another dose of same drug (no historical controls))
3. **Indication**-Must investigate treatment for the below conditions exclusively:
   1. Alzheimer's disease
      1. Excluded trjals investigating treatments for:
         1. Healthy people with AD mutations
         2. MCI without pathologic characteristics of AD
      2. Included trials investigating treatment for:
         1. Trials investigating MCI with pathologic characteristics of AD (prodromal)
         2. Mild-severe AD (however defined)
   2. Parkinson disease,
   3. Amyotrophic lateral sclerosis,
   4. Huntington's disease
   5. Relapsing Multiple sclerosis,
      1. CIS was not included
   6. Progressive Multiple sclerosis,
      1. CIS was not included
   7. Headaches,
   8. Epilepsy
   9. TBI
   10. Stroke
       1. Must be in patients who have had a stoke looking at recurrence or recovery
4. **Earlier Phase 3 trial:** Trials were excluded if they were preceded by a phase 3 or 4 trial that had at least a year of progress. We used TrialViewer to search for all earlier phase 3 trials of our experimental drug-of-interest as well as searching for earlier phase 3 trials in our P3 trial publications. We did not check for the status of the previous trial. We used the following rules when determining if earlier phase 3 trials counted as evidence for the trial in our sample (the same rules were used to match phase 3 trials to phase 2 trials):
   1. Earlier trials
      1. did not need to be exclusively in that indication
      2. could investigate the same intervention in control or experimental arm
      3. could be in any aged population
      4. could not be used if they investigated treatments in preclinical populations
         1. Example: CIS, people with AD mutation
      5. did not need to match in adjuvant status if the phase 3 in our sample was adjuvant or monotherapy. However, earlier trials for phase 3 trials in our sample investigating combination therapies also had to be testing the same combination
      6. could be investigating slight variations in the same drug such as small molecular changes or changes the delivery mechanism.
         1. If it was clear that a phase 3 trial in our sample was investigating a variation in an old drug, we checked for approval of the original drug in the same disease area and excluded the trial in our sample if the earlier drug was already approved in the same indication. This criterion was reliant on P3 trial publication citations.
   2. RRMS and Progressive MS were treated separately, and they could not be used as prior evidence for the other. If the trial was only SPMS, earlier trials in RRMS or PMS were considered prior evidence.
5. **Primary Endpoint:** Trials were only included if they had a primary endpoint that was a clinical efficacy endpoint widely used as a measure of disease modification of each disease area in phase 3 trials.
   1. Trials were excluded if they only had primary safety, tolerability surrogate primary endpoints, or primary endpoints looking only at a symptom that is not used as a measure of disease modification.
   2. Neurologist collaborators were queried: “Would you consider whether the following is a “widely used measures of disease modification in phase 3 trials for X?”
6. **Phase 3 Portion of Phase 2/3 trials**: Exclude if phase 2/3 did not progress to phase 3
   1. Trials were excluded when they were identified as phase 2 in the publication or ClinicalTrials.gov record.

**Phase 3 results**

The Google Scholar search used NCT ID, Title (top-line & official), varying combinations of drug names, indication, and sponsor & investigator last name. The OVID search using MEDLINE and EMBASE used a combination of the search terms: drug names from the experimental arm (any synonym of the drug mentioned in ClinicalTrials.gov) + the indication as listed in ClinicalTrials.gov + “Clinical trial” + “Phase 3”.

We searched for publications reporting the results of at least one primary outcome with a significance test. If we were unable to find primary publications of results, we used primary Ct.gov results. If there are no primary results on ClinicalTrials.gov, we used abstracts that reported primary results. We only used interim results if the trial was terminated. Trials without results are included in the prevalence results but not in the positivity analysis (unless they were terminated at DSMB review-which would result in a nonpositive classification).

**Matching P2 trials to P3 trials**

We searched for matches in phase 3 trial publications, clinicaltrials.gov, FDA approval documents, author solicitation, and Alzforum.

1. To determine if a P2 trial was eligible to be a match, it had to have a primary start date that is a year or more earlier than primary start date of the phase 3 study in our sample, as indicated by ClinicalTrials.gov (or the recruitment start date of the publication if registration date was unavailable). If the date that the P2 trial started is unclear, publication within/before the year that the P3 trial started was used. Expanded access trials, extension studies, non-prospective trials, and trials without any accessible results were not considered.
2. If a phase 2 trial passed the first criteria, P2 trials also had to match on:
   1. Indication
      1. To ensure our approach for matching phase 2 and 3 trials was standardized and reproducible, we allowed P2 trial in the same broad disease area to count as matches for phase 3 trials in our sample. Our broad disease areas are Alzheimer’s disease, Parkinson’s disease, amyotrophic lateral sclerosis, Huntington’s disease, relapse remitting multiple sclerosis, progressive multiple sclerosis, headache, epilepsy, traumatic brain injury and stroke.
      2. Relapse Remitting MS and Progressive MS were treated separately, and they could not cite the other as prior evidence. Secondary–progressive MS was included in progressive category but could be matched to either RRMS or PMS.
   2. Intervention
      1. To determine whether P2 trials investigated the same drug or biologic, ee used the following rules:
         1. A trial that investigated a drug/biologic as a monotherapy could not be used as prior evidence for a trial that is investigating the same drug in combination therapy (and vice-versa). Monotherapy evidence could be used for adjuvant phase 3 trials in our sample because the change may be a result of shifting populations from early line to late line patients. Adjuvant evidence could also be matched to monotherapy P3 trials in our sample. Two adjuvant trials with different background drugs were also accepted as matches.
            1. Adjuvant trials were identified by the terms “adjuvant” or “add on”: new (Experimental) + old
            2. Combo definition: new + new or old + old
         2. Slight variations on drugs were allowed to be matches such as small molecular changes or changes to the delivery mechanism. (unless the old variation of the drug preceded to P3 trials or approval in which case the trial in our sample was excluded (see exclusion criteria))
3. Positivity of P2 matches: To determine the positivity of phase 2 matches, we used the definition of positivity provided by the trial publication. We used the following rules when applicable:
   1. Sequential testing procedures were followed
   2. Trials that were stopped by DSMBs but were then positive were considered positive.
   3. In rare cases, there was no distinct primary endpoint, the coders tried to determine the objective of the study to the best of their abilities.
   4. P2 futility trials were considered positive if they found that the treatment of interest was not futile.
   5. When there were two primary analyses where one was positive and the other was not (inconsistent results), we used the following rules:
      1. Co-primaries: When they stated that all primaries had to be positive for the trial to be positive, we called inconsistent results nonpositive. We used this rule when researchers did not change adjust for multiple testing.
      2. Multiple primaries: When researchers used multiplicity adjustment or partitioned of the alpha levels, we called inconsistent results positive\*
         1. if they used the term “coprimaries” but adjusted the primary, we treated it as multiple primaries
         2. In cases where there were 2 dose groups are both considered primary analysis groups, we called inconsistent results positive, therefore not needing multiplicity adjustments for multiple dose arms.
4. Phase determination:
   1. It was occasionally unclear as to which phase an earlier trial was. We used classification used on ct.gov in the phase category unless they are identified as a different phase in the publication (prioritized) If still was not defined, we used the following rules:
      1. P1-The trial was not controlled and there was no efficacy endpoint
      2. P2—The trial publication say something like dose-ranging or proof of concept. These trials could be controlled or not, could have an efficacy endpoint or not, and involved <300 patients
      3. P3- The trial was controlled and had a primary efficacy endpoint and involved >300 patients
   2. Sample size was the deciding factor in eight cases. We decided to use FDA guidelines that indicated phase 2 trials tend to be under 300 patients. Although this undoubtedly varies by indication. We found on average phase 3 trials in relevant indications were all above 300 and it was therefore safe to use this rule to determine which trials were P2:
      * 1. TBI- avg p3 in our sample was 966
        2. Headache- avg p3 in our sample was 1052s
        3. Stroke- avg p3 in our sample was 1115
        4. HD- avg p3 in our sample was 695
           1. This was the deciding factor for four trials in our sample that were counted as phase 2 trials using these rules and were thus matches for a trial in our sample.
           2. This was the deciding factor for four trials in our sample that were counted as phase 3 trials using these rules and thus a trial in our sample.
5. Classification: Each P3 trial was classified into one of the following groups based on its prior evidence:
   1. Preceded by a positive phase 2 trial
      1. P3 trials were put into this category when they were preceded by:
         1. P2 trial that had a clinical or validated primary endpoint that was positive as defined in the trial.
            1. Surrogate endpoints were considered reasonably validated “biomarkers of disease pathophysiology”2 or type 2 biomarkers.3

If they are commonly used as a primary endpoint to evaluate efficacy in phase 2 trials in that indication because of time constraints OR make sense mechanistically and have been validated in a P3 trial of a similar drug showing clinical efficacy is associated with it

The only surrogates that we considered reasonably validated were number of gadolinium-enhancing lesions for multiple sclerosis trials and the proportion of patients with ⩾95% peripheral CD19+ B-cell depletion for multiple sclerosis trials

* + - 1. For two P2 trials, it was unclear what the primary endpoint was in a trial. We used our best judgement to determine the primary objective of the trial.
      2. P2/3 are put into this category automatically.
  1. Preceded by an ambiguous P2 trial:
     1. Every other P3 trial with a matched P2 trial that did not fall into the above category was put into one of the following categories
        1. Non-positive: Had a P2 trial that was nonpositive on their primary clinical or validated surrogate efficacy endpoint
        2. Not aimed at providing efficacy data: Had a P2 trial that had a primary endpoint investigating surrogate endpoints (not validated) or safety/tolerability. In addition, when the matched P2 trial investigated an efficacy endpoint but did not have an analysis to evaluate significance between groups, we put the associated P3 trial into this category
  2. True bypass
     1. P3 trials were put into this category when we did not find a matched P2 trial.
        1. These were confirmed with emails to authors when emails were available.
        2. When we found potential phase 2 trials but could not find pubs or results, these trials are put into the true bypass group because we could not determine if they were truly matches without information on intervention, indication, and date.
  3. If any p3 trial had more than one prior trial, the one closest to preceded in the order they are described above took priority.

**Extraction**

We extracted the following items from each phase 3 trial in our sample:

1. Termination status
   1. We extracted termination status from registration records or publications as well as whether or not it was due to futility or safety concerns.
2. Positivity status
   1. We extracted whether each trial was positive on their primary efficacy outcome. To do so, we used the definition of positivity in the statistical analysis section. The same positivity rules as above were used.
   2. If the trial was stopped by DSMB but no results were available, trials were deemed to be non-positive.
3. WdueAE in each arm
   1. We extracted the number of participants who withdrew due to adverse events from ClinicalTrials.gov or consort documents in the publications. Where there was disagreement between these sources, the publication took priority.
   2. The denominator was the number of patients at baseline randomization
   3. When there were multiple arms, we took the one that is first for hierarchical testing and the comparator arm. If there truly was not one arm with a higher priority, we took the highest dose. If one was added as an amendment, the original was taken.
4. Approval status
   1. We classified each phase 3 trial as pre or post approval depending on whether the treatment under investigation was approved at the time of trial initiation (primary start date in registration).
      1. Pre-approval = drug was approved after the primary start date or never approved  
         post-approval = drug was approved before the primary start date
   2. Approval in other indications or with different delivery mechanisms were counted. If the trial is looking at a new formulation for an old drug- the first formulation will be used for approval date
   3. If the trial was investigating a combination treatment, they both needed to be approved in that indication for the trial to be considered post-approval
5. Funding (industry vs non-industry).
   1. We extracted whether the trial was funded by a pharmaceutical company or not from publications. If no funder was available, we took the sponsor listed on ClinicalTrials.gov.
   2. When the trial was not funded by a pharmaceutical company but drug was supplied by one, we called the trial non-industry.
6. Trial sample size and duration
   1. These numbers were extracted from ClinicalTrials.gov using the following variables: Actual Enrollment, Study Start Date and Actual Primary Completion Date.

**Statistical Analyses**

Fisher-exact tests were performed using the “fisher.test” R function.17 Risk ratios for wdAE were pooled used the function “metabin” from the “metafor” R package.19 Paired t-tests were performed in R using “t.test”

PROTOCOL DEVIATIONS

* Positivity of p2 was not include times when there are multiple trials with conflicting results
* We did not look at these variables in relationship with the prevelance of bypassing
  + Phase 2/3 vs P2 (these were all preceded)
  + Pediatric vs Adult vs Mixed (almost all were adult)
  + Orphan disease (all were not orphan (except maybe HD))
  + Symptom (most were excluded)
  + Severity-too difficult to operationalize made it into degenerative
* We changed moral economy analyses to focus on Phase 3 trials rather than P2 because they were not a representative sample (only p2 trials that moved on to P3 trials)
* We did not include an analysis of bypass and P3 trial benefit because there was not enough P3 trials reporting the same measure in more than one indication
* We did not search OVID or pubmed for the matches due to huge workloads
* Primary is all indications

SUPPLEMENTAL RESULTS

