**Supplement Page #**

**Supplemental Methods………...…………………………………….……………………..……….……….……….x**

**Supplemental Statistics…………….……………….………………………………………………………..…….....x**

**Protocol Deviations……………………………………………………………………………….………….…….....x**

**Supplemental Results……………………………………………………………...…………………………,………x**

**eTable 1. Phase 2 bypass and the proportion of phase 3 that are positive within conditions**

**eFigure 1. Phase 2 bypass and the pooled RRs for WdAEs**

**Supplemental References……………………………………………………………………………..………..….....x**

**Supplemental Methods**

Phase 3 Sample Creation

ClinicalTrials.gov search parameters for phase 3 trials:

1. **Condition or disease (including synonyms built into ClinicalTrials.gov:** 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 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.
   2. Semi-automatic screening (using excel filters) for phase 3 trials:
7. **Primary completion date**: checked that type is “Actual” and not “Anticipated”
   1. Excluded, \*unless\* trial had an “Actual” overall completion date;
8. **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;
9. **Trial size:** <30
10. **Trial status:** exclude if the trial recruitment status was:
    1. Withdrawn (i.e. no patients enrolled);
11. **Indication:** excluded if primary purpose is
    1. Diagnostic;
    2. Screening;
    3. Basic Science
12. **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;
13. **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 secondary condition in patients with included conditions (ie infection in PD patients and immune responses to vaccines in MS patients)
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 trials 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. Relapsing-remitting MS
      2. Trials investigating treatment for CIS only were excluded
   6. Progressive Multiple sclerosis,
      1. Primary progressive MS and secondary progressive MS
      2. Trials investigating treatment for CIS only were excluded
   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 TrialViewer2 to search ClinicalTrials.gov for all earlier phase 3 trials of our experimental drug-of-interest. In addition, we searched for earlier phase 3 trials in our phase 3 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 the same 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 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 mostly reliant on phase 3 trial publication citations indicating that the drug was a new variation on an old drug.
   2. RRMS and PPMS 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 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 in ClinicalTrials.gov records.

Phase 3 results

We searched for Phase 3 trial publications on Google Scholar using NCT ID, Title (top-line & official), varying combinations of drug names, indication, and sponsor & investigator last name. We then searched OVID using MEDLINE and EMBASE using 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 prioritized 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 ClinicalTrials.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 were included in the prevalence results but not in the positivity analysis (unless they were terminated at DSMB review-which we classified as nonpositive).

Matching phase 2 trials to phase 3 trials

We searched for phase 2 matches in phase 3 trial publications, Clinicaltrials.gov, FDA approval documents (Drugs@FDA), and author solicitation.

* For a phase 2 trial to be eligible to be a match, it had to have a primary start date that was year or more before the primary start date of the phase 3 study in our sample as indicated by ClinicalTrials.gov (or the recruitment start date in the publication if registration date was unavailable). If the date that the phase 2 trial started is unclear, publication within/before the year that the phase 3 trial started was accepted. Expanded access trials, extension studies, non-prospective trials, and trials without any accessible results were not considered.
* If a phase 2 trial passed these criteria, phase 2 trials also had to match on:
  + Indication
    - To ensure our approach for matching phase 2 and 3 trials was standardized and reproducible, we allowed phase 2 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.
    - Relapsing-remitting MS and primary 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.
  + Intervention
    - To determine whether phase 2 trials investigated the same drug or biologic, we used the following rules:
      * A trial that investigated a drug/biologic as a monotherapy could not be used as prior evidence for a trial that was 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 phase 3 trials in our sample. Two adjuvant trials with different background drugs were also accepted as matches.
        + Adjuvant trials were identified by the terms “adjuvant” or “add on”
      * 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 phase 3 trials or approval in which case the trial in our sample was excluded (see exclusion criteria)).
* Phase determination:
  + We used the phase status on ClinicalTrials.gov unless they are identified as a different phase in the publication. The phase of an earlier trial was occasionally undefined and we used the following rules to classify them (although we are aware that not all trials followed these rules, they are useful when we were forced to categorize):
    - P1-The trial was not randomized and there was no efficacy endpoint
    - phase 2—These trials could be randomized or not, could have an efficacy endpoint. In cases where trials were randomized and had a primary endpoint, we decided to call trials phase 2 (rather than phase 3) if they involved <300 patients. When the trial publication said called the trial dose-ranging or proof of concept, we put them in this category.
    - phase 3- The trial was controlled and had a primary efficacy endpoint and involved >300 patients. If we found that an earlier trial fell into this category, we excluded the relevant phase 3 trial in our sample.
  + Sample size was the deciding factor in eight cases. We decided to use FDA guidelines that indicated phase 3 trials averaged more than 300 patients.3 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 phase 2:
    - * TBI- avg p3 in our sample was 966
      * Headache- avg p3 in our sample was 1052
      * Stroke- avg p3 in our sample was 1115
      * HD- avg p3 in our sample was 695

Classification

Once we determined that a phase 2 trial was an eligible match, we extracted its positivity status and classified the associated phase 3 trial. If any p3 trial had more than one prior trial, the one closest to preceded in the order they are described above took priority.

* Positivity of phase 2 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:
  + Sequential testing procedures were followed.
  + Trials that were stopped by DSMBs but were then positive were considered positive.
  + phase 2 futility trials were considered positive if they found that the treatment of interest was not futile.
  + When there were two primary analyses where one was positive and the other was not (inconsistent results), we used the following rules:
    - 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.
    - Multiple primaries: When researchers used multiplicity adjustment or partitioned of the alpha levels, we called inconsistent results positive
      * If they used the term “coprimaries” but adjusted the primary, we treated it as multiple primaries
      * In cases where there were 2 dose groups that were both considered primary analysis groups, we called inconsistent results positive. Therefore we did not require multiplicity adjustments for multiple dose arms.
* Each phase 3 trial was then classified into one of the following groups based on its prior evidence:

1. Preceded by a positive phase 2 trial:

* phase 3 trials were put into this category when they were preceded by one or more:
  + phase 2 trial that was positive on a clinical or a reasonably validated surrogate primary endpoint
    - Surrogate endpoints were considered reasonably validated 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 phase 3 trial of a similar drug showing clinical efficacy is associated.
      * The only surrogates that we considered to be reasonably validated were number of gadolinium-enhancing lesions and the proportion of patients with ⩾95% peripheral CD19+ B-cell depletion for multiple sclerosis trials
  + For two phase 2 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.
  + phase 2/3 are put into this category automatically.

1. Preceded by an ambiguous phase 2 trial:

* Every other phase 3 trial with a matched phase 2 trial that did not fall into the above category was put into one of the following categories:
  + Non-positive: Had a phase 2 trial that was nonpositive on their primary clinical or validated surrogate efficacy endpoint.
  + Not aimed at providing efficacy data: Had a phase 2 trial that had a primary endpoint investigating surrogate endpoints (not validated) or safety/tolerability. In addition, when the matched phase 2 trial had a primary efficacy endpoint but was not designed to evaluate significance between groups, we put the associated phase 3 trial into this category.

1. Full bypass

* phase 3 trials were put into this category when we did not find a matched phase 2 trial.
  + These were confirmed with emails to authors when emails were available.
  + When we found potential phase 2 trials but could not find any publication or results, these trials are put into the true bypass group because we could not determine if they were truly matches without information on the intervention, indication, and date.

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 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 from the study 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 was 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
      2. Post-approval = drug was approved before the primary start date
         1. Approval in other indications or with different delivery mechanisms were allowed. If the trial was looking at a new formulation for an old drug- the first formulation will be used for approval date
         2. 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.

**Supplemental Statistics**

Fisher-exact tests were performed using the “fisher.test” R function.4 Risk ratios for WdAE were pooled used the function “metabin” from the “metafor” R package.5 Paired t-tests were performed in R using “t.test”6

**Protocol deviations**

* We did not look at these variables in relationship with the prevalence of bypassing
  + Phase 2/3 vs phase 2 (these were all preceded)
  + Pediatric vs Adult vs Mixed (almost all were adult)
  + Orphan disease (all were not orphan (except maybe HD))
  + Symptoms (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 phase 2 because they we did not have a representative sample of phase 2 trials (only p2 trials that moved on to phase 3 trials).
* We did not include an analysis of phase 2 bypass and phase 3 trial benefit because there was not enough phase 3 trials reporting the same measure in more than one indication.
* We did not search OVID or PubMed for the matches due to the large sample size.

**Supplemental Results**

**eTable 1: Positivity across indications**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Indications | Positivity Rate of phase 3 | | | Termination Rate of phase 3 | | |
| Overall Positivity Rate  (N, %)1 | Type of supportive evidence | | Overall Termination Rate (N, %) | Type of supportive evidence | |
| Preceded by phase 2  (N, %) | phase 2 bypass  (N, %) | Preceded by phase 2  (N, %) | phase 2 bypass  (N, %) |
| Alzheimer's disease | 3 (10) | 2 (18) | 1 (6) | 12 (40) | 3 (27) | 9 (47) |
| Parkinson's disease | 2 (22) | 0 (0) | 2 (50) | 2 (20) | 1 (20) | 1 (20) |
| Amyotrophic lateral sclerosis | 1 (25) | 1 (50) | 0 (0) | 1 (20) | 1 (33) | 0 (0) |
| Huntington's disease | 0 (0) | 0 (0) | 0 (0) | 1 (25) | 0 (0) | 1 (33) |
| Relapsing multiple sclerosis | 14 (88) | 14 (93) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Progressive multiple sclerosis | 4 (100) | 1 (100) | 3 (100) | 0 (0) | 0 (0) | 0 (0) |
| Headache | 20 (77) | 15 (79) | 5 (71) | 2 (8) | 2 (11) | 0 (0) |
| Epilepsy | 5 (71) | 1 (50) | 4 (80) | 1 (14) | 0 (0) | 1 (20) |
| TBI | 0 (0) | 0 (0) | 0 (0) | 2 (40) | 2 (67) | 0 (0) |
| Stroke | 0 (0) | 0 (0) | 0 (0) | 3 (50) | 0 (0) | 3 (60) |
| All indications | **49 (45)** | **34 (57)** | **15 (31)** | **24 (21)** | **9 (15)** | **15 (29)** |

1Trials were only included in the positivity analysis if they had primary results available (N=108)

Chart

Description automatically generated**eFigure 1: RRs for WdAE Pooled Subgroup Analyses**

**Supplemental References**

1. FDAAA 801 and the Final Rule - ClinicalTrials.gov, https://clinicaltrials.gov/ct2/manage-recs/fdaaa (accessed 11 June 2021).

2. Clinical trials viewer, https://trials.bgcarlisle.com/ (accessed 30 March 2021).

3. Commissioner O of the. 22 Case Studies Where Phase 2 and Phase 3 Trials Had Divergent Results. *FDA*, https://www.fda.gov/about-fda/reports/22-case-studies-where-phase-2-and-phase-3-trials-had-divergent-results (2019, accessed 11 October 2020).

4. R: Proportion Test, https://search.r-project.org/CRAN/refmans/rstatix/html/prop\_test.html (accessed 21 June 2023).

5. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; 36: 1–48.

6. Tidy t-Tests with infer, https://cran.r-project.org/web/packages/infer/vignettes/t\_test.html (accessed 7 July 2023).