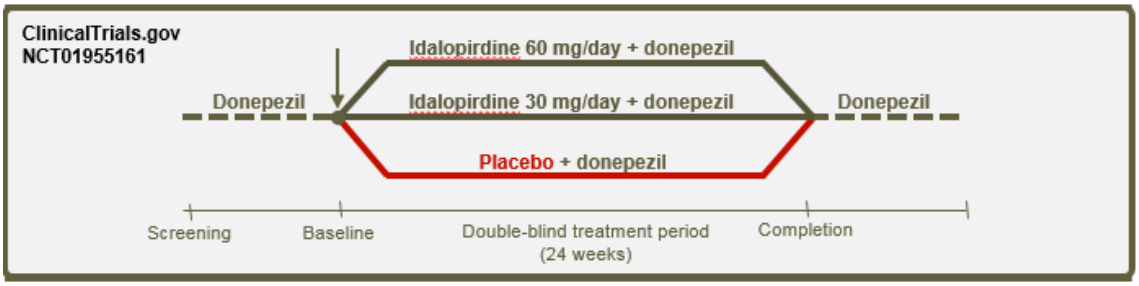
Supplementary Online Content

Atri A, Frölich L, Ballard C, et al. Effect of idalopirdine as adjunct to cholinesterase inhibitors on change in cognition in patients with Alzheimer disease: three randomized clinical trials. *JAMA*. doi:10.1001/jama.2017.20373

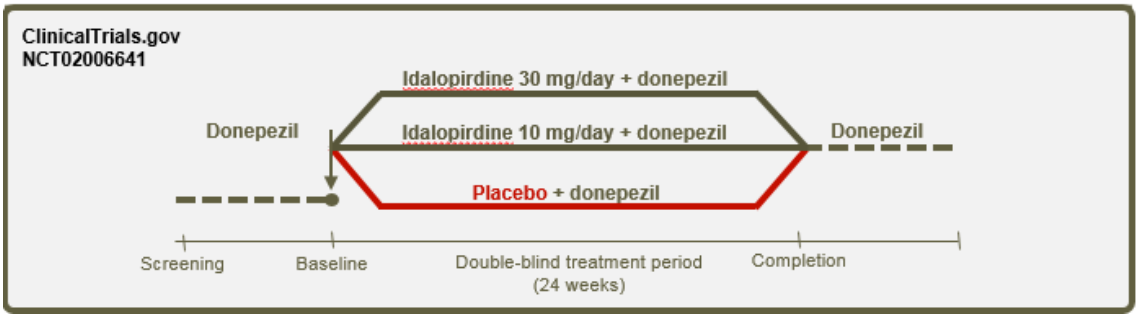
This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure 1. Idalopirdine phase III development program

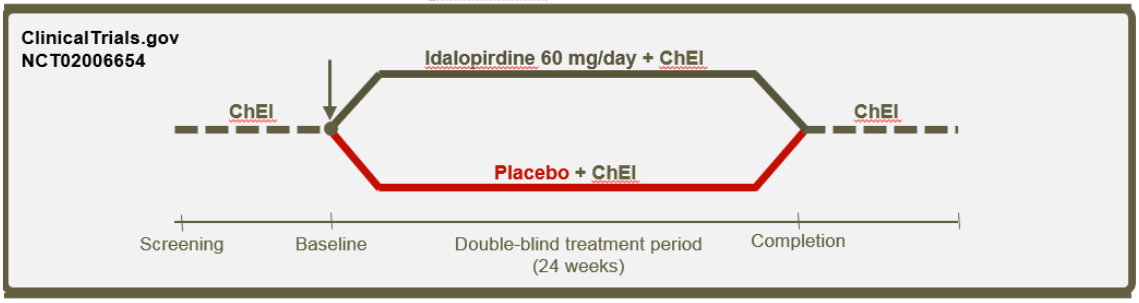
Study	Doses/ comparator	Background treatment	Treatment duration	Total population	Key efficacy endpoints
Study 1	30 mg/day 60 mg/day Placebo	Donepezil 10 mg/day	24 weeks	933	ADAS-Cog[a] ADCS-ADL$_{23}$ ADCS-CGIC



Study 2	10 mg/day 30 mg/day Placebo	Donepezil 10 mg/day	24 weeks	858	ADAS-Cog[a] ADCS-ADL$_{23}$ ADCS-CGIC



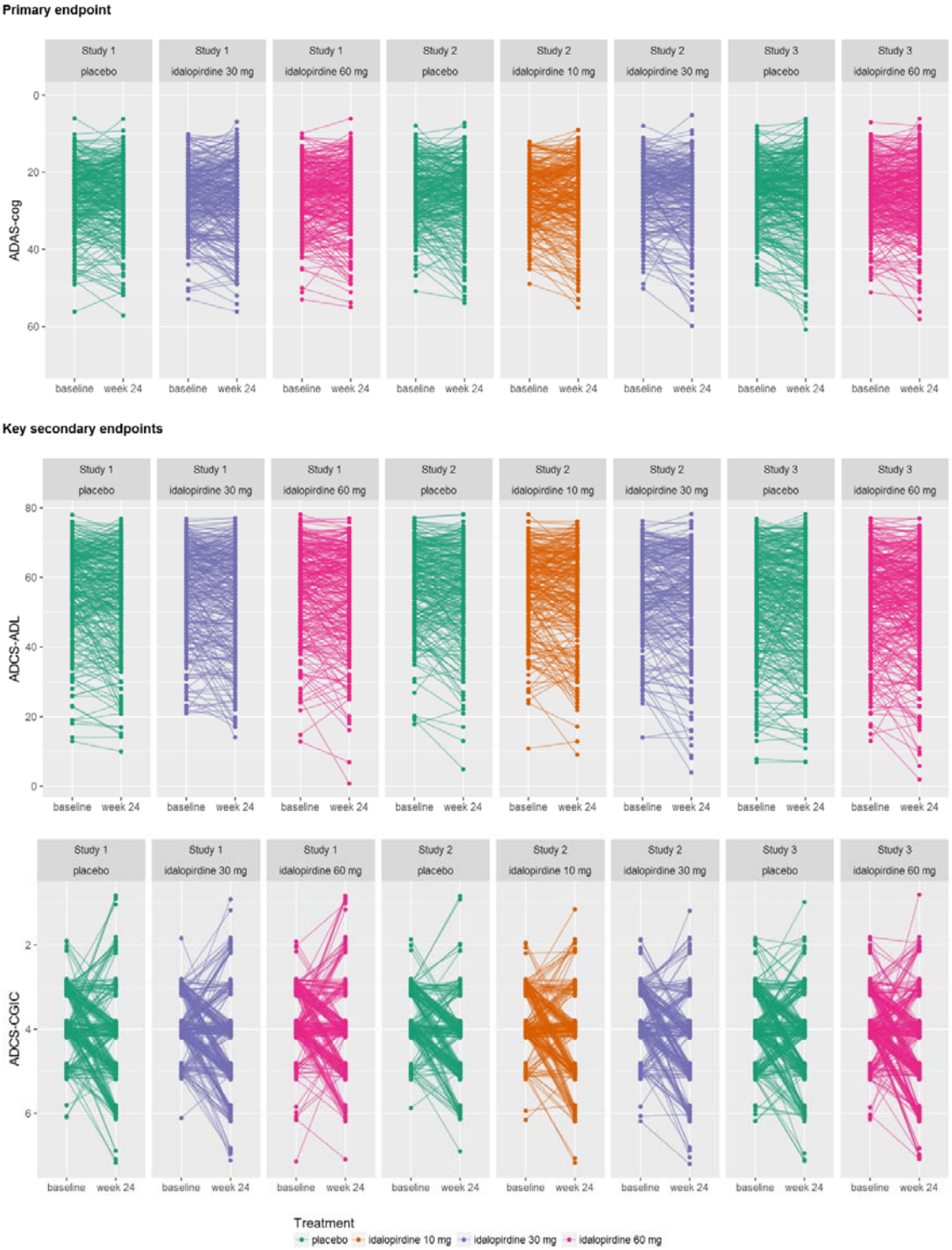
Study 3	60 (30) mg/day Placebo[b]	ChEI (donepezil, rivastigmine, or galantamine)	24 weeks	734	ADAS-Cog[a] ADCS-ADL$_{23}$ ADCS-CGIC



STAR Extension **ClinicalTrials.gov** **NCT02079246**	60 (30) mg/day[c]	Donepezil	28 weeks	1,463 (patients completing Study 1 and 2)	ADAS-Cog[a] ADCS-ADL$_{23}$ ADCS-CGIC (primary endpoint is safety & tolerability)
Extension follow-up sub-study	Memantine 20 mg/day or 28 mg/day	Idalopirdine 60 (30) mg/day + donepezil	24 weeks	101 completers of 28-week extension study	Safety & tolerability

[a]Indicates primary outcome measure; [b]One decrease to 30mg/day, and a return to 60mg/day, permitted based on tolerability; [c]one permanent decrease to 30mg/day permitted based on tolerability; AD=Alzheimer disease; ADAS-Cog=Alzheimer's Disease Assessment Scale - Cognitive subscale; ADCS-ADL$_{23}$= Alzheimer's Disease Cooperative Study - activities of daily living 23-item scale; ADCS-CGIC=Alzheimer's Disease Cooperative Study - Clinical Global Impression of Change scale; ChEI=cholinesterase inhibitor; MMSE=Mini Mental State Examination

eFigure 2. Patient-level baseline to week 24 changes across efficacy primary and key-secondary endpoints



ADAS-Cog=Alzheimer's Disease Assessment Scale, cognitive subscale; ADCS-ADL$_{23}$=Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item scale; ADCS-CGIC=Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change Points have been vertically jittered to decrease overlap and the y-axes for ADAS-Cog and ADCS-CGIC have been reversed such that down corresponds to deterioration.

eMethods. Additional assessments and statistical methodology

Additional assessments not reported in this manuscript

Secondary

- Neuropsychiatric Inventory (NPI) total score[1-3]
- Clinical improvement and worsening
- Composite score for ADCS-ADL_{23} and ADAS-Cog
- Area under curve analyses
- ADL basic and instrumental domains
- Dependence Scale[4]
- Mini Mental State Examination (MMSE)[5]

Pharmacoeconomic

- Resource Utilisation in Dementia Lite (RUD Lite)[6]
- EQ-5D-3L, a measure of health-related quality of life[7]

Statistical Analysis

In all of the three studies, the joint efficacy criterion for showing efficacy of a dose required demonstration of a significant positive effect on ADAS-Cog followed by a significant positive effect on either ADCS-ADL_{23} or ADCS-CGIC. A significant effect meant that the mean difference to placebo was statistically significant at a 5% significance level. Due to the multiple endpoints and multiple active dose groups, multiple testing procedures were used to control the overall type 1 error.

In each study, the efficacy analyses were performed on the full-analysis set (FAS), defined as all randomized patients who took at least one dose of investigational medicinal product (IMP) and who had a valid baseline assessment and at least one valid post-baseline assessment of the primary endpoint.
The primary endpoint (ADAS-Cog) and key secondary endpoints (ADCS-ADL_{23} and ADCS-CGIC) were analyzed in the same way in all three studies in accordance with the identical design of visit schedules and measurements. Changes from baseline were analyzed using a restricted maximum likelihood (REML)-based mixed model for repeated measures (MMRM) approach. The model included treatment (placebo and idalopirdine doses), country, MMSE stratum (12-18; 19-22), and week of treatment (Weeks 4, 12, and 24) as fixed categorical effects, baseline score as a continuous covariate, treatment-by-week interaction, MMSE-stratum-by-week interaction, and baseline score-by-week interaction. In addition, Study 3 included base therapy stratum and base-therapy-stratum-by-week interaction in accordance with the stratification for base therapy and MMSE stratum used in that study. An unstructured covariance structure was used to model the within-patient variation. The effects of idalopirdine doses were estimated as mean differences to placebo at Week 24 based on the least squares means for the treatment-by-visit interaction in the MMRM-model.

In Study 1, a Bonferroni correction for multiple doses was used initially, testing each dose at a 2.5% significance level. For each dose, ADAS-Cog was tested first and if significant, the testing procedure continued for ADCS-ADL_{23} and ADCS-CGIC. A Hochberg procedure was applied at a 2.5% significance level to adjust for multiplicity due to the two endpoints considered. If, for one of the doses, all of the null hypotheses were rejected, the 2.5% allocated to that dose was transferred to the other dose, allowing the other dose to be tested at a nominal 5% significance level (the originally allocated 2.5% plus the transferred 2.5%). This procedure protected the overall type 1 error.

In Study 2, a hierarchical testing procedure was applied, testing the 30mg dose first followed by the 10mg dose, if the 30mg dose was shown to be effective. For the 30mg dose, ADAS-Cog was tested at a 5% significance level and if the dose separated from placebo, ADCS-ADL_{23} and ADCS-CGIC was tested at a 2.5% level by applying a Bonferroni adjustment. The significance level to be applied for the 10mg dose depended on how many of the null hypotheses for key secondary endpoints for the 30mg dose were rejected; if none were rejected, the test procedure stopped and 10mg could not be considered; if one was rejected, 10mg could be tested at a 2.5% level; and if both were rejected, 10mg could be tested at a 5% level. The testing for 10mg proceeded as for each of the doses in Study 1, testing ADAS-Cog first and then ADCS-ADL_{23} and ADCS-CGIC conditional on a significant effect on ADAS-Cog, applying a Hochberg procedure at this final step.

No multiplicity adjustment due to multiple doses was needed in Study 3. Testing of the three endpoints proceeded as for each dose in Study 1, however, a 5% level of significance was applied for ADAS-Cog and for the Hochberg procedure that was applied for ADCS-ADL_{23} and ADCS-CGIC.

For all studies, the adjusted p-values for each endpoint were computed as the lowest significance level at which the null hypothesis of no difference to placebo would be rejected according to the testing strategy for the study. Similarly, for each dose in the studies, the overall adjusted p-value corresponding to the test of the null hypothesis of not meeting the joint efficacy criterion, was computed. That is, the lowest significance level at which the dose would be declared efficacious per the joint efficacy criterion.

A range of sensitivity analyses including pattern mixture models based analyses were performed for the primary and key secondary endpoints.

Continuous secondary endpoints were analyzed using the same structural MMRM model as described for the primary and key secondary endpoints. Binary secondary endpoints were analyzed using a Cochran-Mantel-Haenszel test for comparing each dose to placebo stratifying for country, MMSE stratum, and, in the case of Study 3, also for base treatment stratum.

Sample Size Determination

In Study 1, a total of 310 patients were randomized to each group providing a power of approximately 90% for at least one dose showing significant improvements on an overall 5% level on both ADAS-Cog and at least one of ADCS-ADL_{23} or ADCS-CGIC, assuming improvements of 2 points on both ADAS-Cog and ADCS-ADL_{23}, and 0.25 on ADCS-CGIC for both doses. The SDs are approximately 6.10, 9.15, and 1.15 for the three outcomes when adjusting for intra-patient correlation and drop-out. The SD for each endpoint was obtained from the number of patients randomized in each group (N1 and N2) and the standard error (SE) of the treatment effect estimate at 24 weeks in the Proof of Concept (PoC) phase II study, as $SD=SE/\sqrt{(1/N1+1/N2)}$. This estimate both takes into account the actual variance at 24 weeks and the loss of information due to drop-out during the study, assuming that the dropout pattern observed in PoC was representative of what was observed in this study. The estimated correlations between the endpoints are -0.27 between ADAS-Cog and ADCS-ADL_{23}, 0.38 between ADAS-Cog and ADCS-CGIC, and -0.35 between ADCS-ADL_{23} and ADCS-CGIC when adjusting for baseline scores. Multiplicity due to multiple doses and endpoints was adjusted for as explained in the Statistical Methodology. The power was evaluated by simulation from a multivariate normal distribution with the assumed mean and covariance structure described.

Study 2 and 3 sample sizes were determined similarly targeting a power of 80% in Study 2 and 90% in Study 3.

Randomization

In Study 1 and 2, symmetric randomization to the three treatment groups was stratified by MMSE stratum (12-18, 19-22). In addition, block randomization with a block size of three was used and applied within sites to balance any time trends and sites effects. The randomization was further restricted such that the first treatment in each block in the two MMSE strata within a site was different to ensure that in any analyses by country or site, the treatment effect would not be confounded with effects of country or site.

In Study 3, the symmetric randomization to the two treatment groups was stratified by MMSE stratum (12-18, 19-22) and base therapy (donepezil, rivastigmine/galantamine). The two by two combination of MMSE strata and base therapy strata generates essentially four strata for the study. In order to balance treatment allocation to strata and also take possible site effects into account the randomization was done in blocks of four based on latin squares of the following form:

1 2 1 2
2 1 2 1
1 1 2 2
2 2 1 1

The columns correspond to the four strata (MMSE strata by base therapy strata), the rows to patients and the numbers (1 and 2) to the two treatment groups. There are a total of six possible squares of this form corresponding to the six possible permutations of 1122. Random blocks of this form were generated and

assigned to sites. This procedure results in a block randomization used a block size of four balancing treatments with strata and sites.

The randomization lists were produced using computer generated random numbers applying the statistical software SAS version 9.4.

All of the three studies used an IVRS system for assigning treatments to patients as they were randomized.

Trial Completion

Studies 1, 2, and 3 stopped according to plan when the required number of patients had been randomized.

References

1. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*. 1994;44(12):2308-2314.
2. Cummings JL. The neuropsychianic inventory: assessing psychopathology in dementia patients. *Neurology*. 1997;48(Suppl 6):10-16.
3. Kaufer DI, Cummings JL, Christine D, Bray T, Castellon S, Masterman D, et al. Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease: the Neuropsychiatric Inventory Caregiver Distress Scale. *J Am Geriatric Soc*. 1998;46:210-215.
4. Stern Y, Albert SM, Sano M, et al. Assessing patient dependence in Alzheimer's disease. *J Gerontol*. 1994;49:M216-M222.
5. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
6. The EuroQoL Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208.
7. Wimo A, Winblad B. Resource utilisation in dementia: RUD Lite. *Brain Aging*. 2003;3:48-59.

eTable 1. Observed change from baseline in key secondary endpoints (full analysis set)

Study 1						
	Placebo		Idalopirdine 30mg		Idalopirdine 60mg	
	Baseline	Week 24 change from baseline	Baseline	Week 24 change from baseline	Baseline	Week 24 change from baseline
ADCS-ADL$_{23}$ total score	56.18 (12.96), n=304	-2.11 (7.83), n=286	55.75 (13.33), n=310	-2.32 (8.08), n=290	56.58 (12.76), n=308	-2.28 (8.60), n=282
ADCS-CGIC score	3.80 (0.75), n=303	4.31 (1.13), n=285	3.86 (0.69), n=310	4.33 (1.13), n=290	3.78 (0.72), n=307	4.11 (1.11), n=280
Study 2						
	Placebo		Idalopirdine 10mg		Idalopirdine 30mg	
	Baseline	Week 24 change from baseline	Baseline	Week 24 change from baseline	Baseline	Week 24 change from baseline
ADCS-ADL$_{23}$ total score	57.57 (11.89), n=279	-1.79 (7.06), n=259	56.57 (12.00), n=282	-1.67 (8.05), n=260	55.27 (12.94), n=275	-1.55 (7.63), n=249
ADCS-CGIC score	3.69 (0.69), n=278	4.32 (0.95), n=256	3.72 (0.68), n=282	4.26 (1.02), n=258	3.73 (0.74), n=275	4.35 (1.06), n=249
Study 3						
	Placebo		Idalopirdine 60mg			
	Baseline	Week 24 change from baseline	Baseline	Week 24 change from baseline		
ADCS-ADL$_{23}$ total score	53.93 (14.68), n=356	-1.87 (8.27), n=330	54.01 (13.27), n=361	-1.44 (8.60), n=328		
ADCS-CGIC score	3.87 (0.80), n=354	4.30 (1.05), n=324	3.85 (0.75), n=359	4.37 (1.16), n=322		

Numbers reported are mean (standard deviation), n=number of patients assessed.

ADCS-ADL$_{23}$=Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item scale (range: 0-78; higher score indicates less impairment)

ADCS-CGIC=Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change (baseline scoring ranges from 1 [normal, not at all ill] to 7 [among the most extremely ill patients], post-baseline scoring evaluates change since baseline in a similar range from 1 to 7 with 4 corresponding to no change)

eTable 2. Sensitivity analysis of primary endpoint using multiple imputation from the placebo group (full analysis set)

Study 1		
	Least-square Means (95% CI)[a]	
	Idalopirdine 30mg (n=310) versus Placebo (n=304)	**Idalopirdine 60mg (n=308) versus Placebo (n=304)**
ADAS-Cog total score	0.35 (-0.57 to 1.27)	0.09 (-0.84 to 1.01)
Study 2		
	Least-square Means (95% CI)[a]	
	Idalopirdine 10mg (n=282) versus Placebo (n=278)	**Idalopirdine 30mg (n=275) versus Placebo (n=278)**
ADAS-Cog total score	-0.04 (-1.04 to 0.96)	0.65 (-0.37 to 1.66)
Study 3		
	Least-square Means (95% CI)[a]	
	Idalopirdine 60mg (n=361) versus Placebo (n=356)	
ADAS-Cog total score	-0.49 (-1.40 to 0.42)	

Abbreviations: ADAS-Cog=Alzheimer's Disease Assessment Scale cognitive subscale (range: 0-70; lower score indicates less impairment); CI=confidence interval

[a]Least-square means and corresponding 95% confidence intervals are for the difference in change from baseline ADAS-Cog scores between the idalopirdine and placebo treatment groups at week 24. A negative value indicates a treatment effect in favor of idalopirdine. Least-square means and confidence intervals were calculated using proc mixed in SAS® version 9.4.

For further details on the analysis, see Section 10.2.4 in the statistical analysis plans for the studies.