

PERSPECTIVE

Bayes analysis supports null hypothesis of anti-amyloid beta therapy in Alzheimer's disease

Edo Richard^{1,2} | Melina G.H.E. den Brok^{1,3} | Willem A. van Gool²

¹ Department of Neurology, Radboud University Medical Center, Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, the Netherlands

² Department of Public and Occupational Health, Amsterdam University Medical Centre, AMC, University of Amsterdam, the Netherlands

³ Department of Neurology, Amsterdam University Medical Centre, AMC, University of Amsterdam, the Netherlands

Correspondence

Edo Richard, MD, PhD, Department of Neurology, Radboud University Medical Center, Donders Institute for Brain, Cognition, and Behaviour, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands.

E-mail: edo.richard@radboudumc.nl

Funding information

Netherlands Organization for Health Research and Development (ZonMw, Grant/Award Number: 91718303)

[The copyright line for this article was changed on June 1, 2021 after original online publication.]

Abstract

Numerous clinical trials of anti-amyloid beta ($A\beta$) immunotherapy in Alzheimer's disease have been performed. None of these have provided convincing evidence for beneficial effects. Using traditional frequentist meta-analysis, the conclusion is that there is *absence of evidence* for a therapeutic effect, with a point estimate effect size of 0.05 (95% confidence interval -0.00 to 0.10 , $P = .055$). In addition, this non-significant effect equates to 0.4 points per year on the cognitive subscale of the Alzheimer's Disease Assessment Scale. This is well below the minimally clinically important difference. Bayesian meta-analysis of these trial data provides strong *evidence of absence* of a therapeutic effect, with a Bayes factor of 11.27 in favor of the null hypothesis, opposed to a Bayes factor of 0.09 in favor of a treatment effect. Bayesian analysis is particularly valuable in this context of repeatedly reported small, non-significant effect sizes in individual trials. Mechanisms other than removal of $A\beta$ from the brain may be probed to slow progression of Alzheimer's disease.

KEYWORDS

Alzheimer's disease, anti-amyloid beta immunotherapy, Bayesian analysis, dementia, evidence of absence, meta-analysis, randomized controlled trial

1 | INTRODUCTION

Since the turn of the century numerous therapeutic studies in Alzheimer's disease (AD) have targeted the cerebral amyloid beta ($A\beta$) deposits, one of the defining characteristics of this condition. Thousands of patients have participated in randomized controlled trials (RCTs) of anti- $A\beta$ (AAB) immunotherapy, costing billions of dollars. So far, results have been disappointing, with none of these trials reporting statistically significant beneficial effects. Despite being sufficiently powered, well performed, and published in high-ranking journals, these studies have been characterized as "negative," "failures," or "unsatisfactory." This may be unjustified and can be understood best by the characteristics of classical frequentist statistical null hypothesis testing for significance used in all RCTs. As recently reviewed in depth by Keyzers et al.,¹ non-significant study results of well-designed

RCTs are open to several explanations according to this approach: (1) trialists may simply have had "bad luck" due to sampling variability; (2) the sample size may have been too small, the target population not sufficiently well-defined, the follow-up too short, or the outcome parameters too insensitive to detect subtle effects; or (3) the true effect of a therapeutic approach is indeed absent. Both the derogatory qualifications of recent AAB immunotherapy and the ever-continuing research into this approach, suggest that the first two interpretations, "bad luck" and "too small, short, or insensitive," dominate the AD field, whereas the third option, "absence of a true effect" appears not to be considered seriously. This may have a simple explanation that is related to the traditional frequentist analyses that were used. A Bayesian analysis may offer a constructive perspective on these recent AAB trial results and allow for quantifying the null hypothesis that AAB immunotherapy does not slow down cognitive decline in AD.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association

1.1 | Frequentist versus Bayesian analysis of trial results

To evaluate the full body of evidence of AAB immunotherapy in AD, meta-analysis of published results of RCTs can be applied. Traditional frequentist meta-analyses of published RCT results aim to provide an estimate of the true treatment effect, by calculating the average effect across studies. In most cases, a random effects model is indicated, which takes differences between studies into account, and will provide a weighted mean effect size, with a 95% confidence interval. This is particularly useful in the case of conflicting trial results. The higher number of participants exposed to the intervention or control condition generally leads to an increase in statistical power to signal a statistically significant effect. Interpretation of statistically significant trial results ($P \leq .05$) is straightforward according to the conventional frequentist approach, that is, rejection of the null hypothesis and adoption of the alternative hypothesis implying a true therapeutic effect. But the opposite finding of non-significant P values ($P > .05$) does not provide a precise, quantitative, reliable metric favoring the null hypothesis, supporting the *absence* of a therapeutic effect.¹ That is because relations among sample sizes, trial power, P values, and credibility of null hypotheses are complex: “absence of evidence is not evidence of absence.”² A frequentist approach assumes all trials to be equally meaningful (apart from the weighing based on trial characteristics), and tends to ignore the overall picture upon publication of a single, extreme finding. It lacks a strong framework allowing us to specify how trial results affect the overall odds of a specific treatment effect being plausible.

An alternative approach is using Bayesian methodology, which offers the opportunity to factor in all previous RCT results upon weighing the precise implications of new findings. This approach is based on an eighteenth-century theorem that offers a formal framework for how to update an initial belief after new observations. Following Keyzers et al. we provide here a short explanation of a Bayesian approach in the context of the interpretation of new experimental findings. In the evaluation of a new therapy, for example, absence or presence of a beneficial effect, classically reflecting the null (H_0) or alternative (H_1) hypothesis, respectively, may be equally believable initially. After obtaining the results from a first RCT, for example, in the form of an effect size, beliefs on the (in)efficacy of a new therapy can be updated, according to the degree to which the RCT result supports either hypothesis. A description for such rival accounts in a Bayesian approach can be provided in the form of odds¹:

$$\frac{p(H_0|data)}{p(H_1|data)} = \frac{p(H_0)}{p(H_1)} \times \frac{p(data|H_0)}{p(data|H_1)}$$

The posterior odds in the case of a meta-analysis can be calculated after including new data from each subsequent RCT. This is done by taking the product of the (prior) odds, based on all previously available RCT findings, and an updating factor based on the latest RCT data that have become newly available. The latter metric, the “Bayes factor,” reflects the relative performance of H_0 and H_1 in accounting for

RESEARCH IN CONTEXT

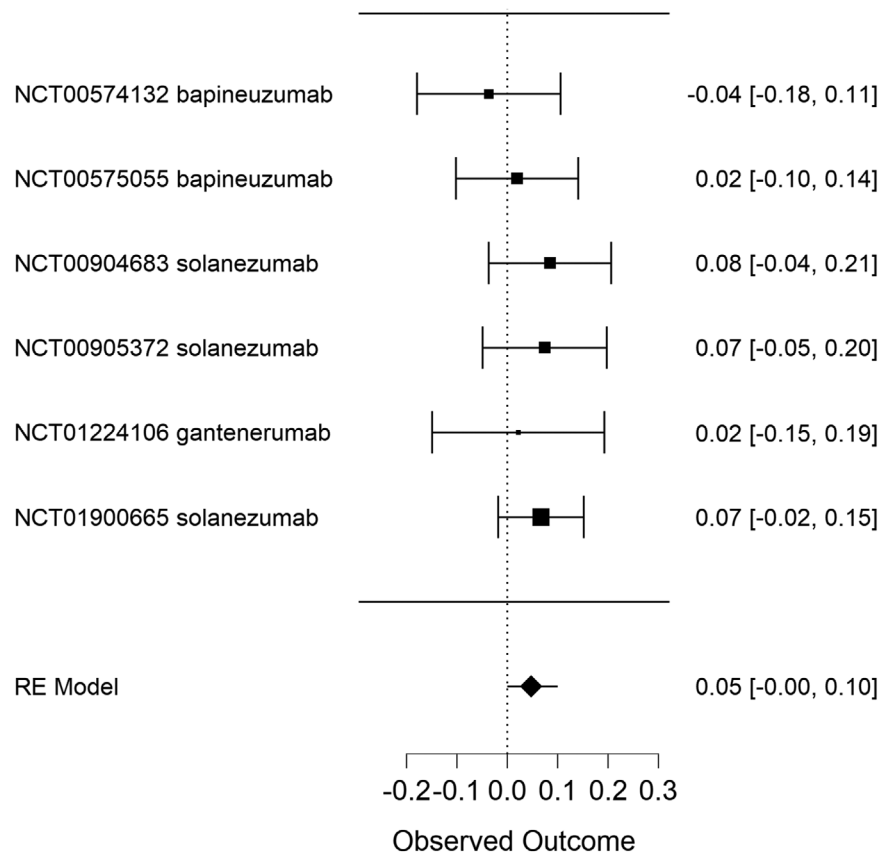
- 1. Systematic Review:** We performed a systematic review for phase 3 randomized placebo-controlled clinical trials investigating the effect of anti-amyloid beta ($A\beta$) immunotherapy in patients with Alzheimer's disease (AD). As minimal quality criterion we set a minimum of 50 patients in the highest dose arm and an intervention period of at least 12 months. This resulted in six completed trials.
- 2. Interpretation:** A conventional frequentist meta-analysis of all six published clinical trials, all with null findings, leads to the conclusion that there is *absence of evidence* for an effect of anti- $A\beta$ immunotherapy. Using a Bayesian analysis of the same data leads to the conclusion that there is strong *evidence of absence* of a therapeutic effect of anti- $A\beta$ immunotherapy in patients with AD. Results from additional trials investigating the effect of anti- $A\beta$ immunotherapy are highly unlikely to change this overall conclusion.
- 3. Future directions:** Research efforts into treatment of AD may best be diverted toward other disease mechanisms.

the findings in a given RCT. Calculation of this Bayes factor is based on different probability distributions of effect sizes under H_0 and H_1 . Although rival hypotheses may have been equally plausible initially (i.e., $p[H_0] = p[H_1] = 0.5$), thus with a prior hypothesis odds of 1, the results from a first RCT may indicate that these findings are 10 times more likely under H_1 than under H_0 , rendering the posterior odds also 10, favoring H_1 in this example. In this case the RCT result has increased the probability of H_1 from 0.5 to $10/11 \approx 0.91$, as the posterior H_1 probability. A major advantage of Bayesian analysis as opposed to a frequentist approach is its potential to quantify the strength of evidence in both directions, not only against H_0 but also in favor of H_0 ; that is, the strength of evidence for absence of a therapeutic effect. See Keyzers et al. for an in-depth review and more examples.¹

1.2 | Meta-analysis of AAB immunotherapy trials

To compare a frequentist and Bayesian approach in the meta-analysis of AAB immunotherapy in AD, we performed a systematic literature search selecting publications on phase 3 AAB immunotherapy RCTs all using the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-Cog) as clinical outcome, to allow for comparison. As minimum quality criterion, studies had to include at least 50 AD patients in the main analysis who received active treatment in the highest dose, for at least 12 months. We chose this anti-conservative approach, that is, only analyzing the groups receiving the highest dosage, to maximally

FIGURE 1 Hypothesis testing using a frequentist approach: anti-amyloid beta immunotherapy in Alzheimer's disease. This forest plot displays six randomized controlled trials testing monoclonal antibodies and their Cohen's d effect sizes with confidence intervals based on a classical random-effects frequentists meta-analytical approach. The NCT numbers refer to the unique identification code given to clinical studies registered on ClinicalTrials.gov. The overall treatment effect was 0.048 (95% confidence interval: -0.001 to 0.096)



benefit the intervention groups. This resulted in four papers on six RCTs including a total of 6597 participants studying the effects of solanezumab, gantenerumab, or bapineuzumab, all monoclonal antibodies against A β , for 78 to 104 weeks.³⁻⁶

We first performed a traditional frequentist random-effects meta-analysis based on the published results of the primary analysis in all trials, using Cohen's d effect sizes for individual studies. This resulted in an overall treatment effect of 0.05 (95% confidence interval: -0.00 to 0.10, $P = .055$), thus not rejecting the null hypothesis (Figure 1). Subsequently, we used a Bayesian approach to investigate the plausibility of the null hypothesis being true (i.e., no treatment effect). We used the JASP software package to calculate sequential Bayes factors after adding each subsequent trial result. JASP stands for Jeffreys's Amazing Statistics Program in recognition of the pioneer of Bayesian inference Sir Harold Jeffreys. This is a free multi-platform open-source statistics package (<https://jasp-stats.org>), developed and continually updated by a group of researchers at the University of Amsterdam. Because also under a Bayesian framework certain assumptions have to be made, including that the probabilities stem from the same distribution, we used a random effects model. We used a Cauchy prior with scale $1/\sqrt{2} \approx 0.707$. This meta-analysis, based on the very same six RCTs as used for the frequentist analysis, yielded a final Bayes factor of 11.27 in favor of the null hypothesis, as opposed to a Bayes factor in favor of the alternative hypothesis of a beneficial treatment effect of 0.09 (Figure 2). The Bayes factor in favor of the null hypoth-

esis should be interpreted as "strong" support, thus going beyond the notion of "not rejecting the null hypothesis" according to the frequentists approach.¹ Although the essence of a Bayesian approach is to interpret new information in light of existing information, one may wonder whether results would have been different if RCT results had appeared in a different order. Therefore, we have rerun the analyses adding the trials in a different order (i.e., arbitrarily, reverse order, and by population size). This slightly changed the steps toward the final Bayes factor after including all six trials, but did not change the overall resulting Bayes factor, as expected.

1.3 | Addition of a theoretical new RCT

Under the traditional frequentist statistical approach, addition of a new trial result that would have roughly a similar outcome as the previous six RCTs would further reduce the value of the P estimate, most probably reaching a value below .05, albeit without substantially affecting the estimate for the effect size. Under a Bayesian approach, we found that such a hypothetical, seventh study would require an unrealistic treatment effect with a Cohen's d effect size of at least 10 (standard error 0.02). Based on a typical ADAS-Cog standard deviation, such a scenario would require a study including at least 100,000 participants randomized 1:1 to active or placebo treatment and a difference between treatment groups of about 14 to 15 ADAS-Cog

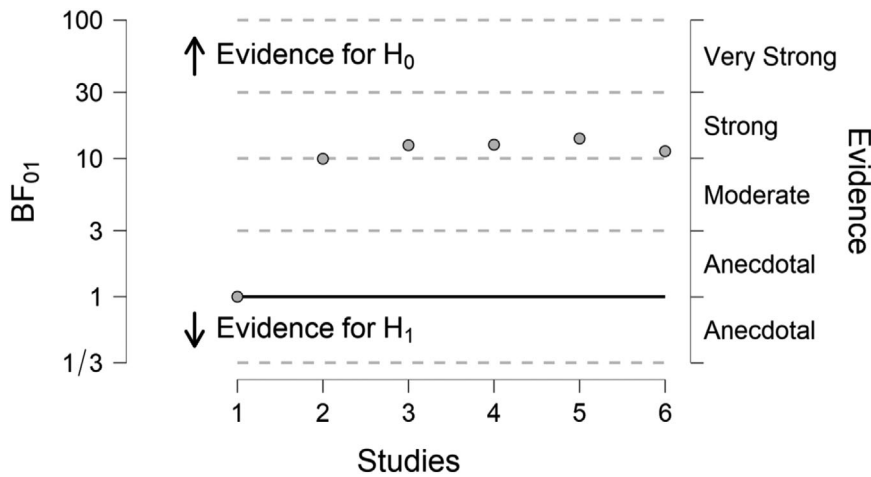


FIGURE 2 Hypothesis testing using a Bayesian approach: anti-amyloid beta immunotherapy in Alzheimer's disease. This plot shows strong evidence for H_0 . A Bayes factor (BF) between $1/3$ and 3 indicates insufficient evidence for either the null or alternative hypothesis, and a BF above 3 indicates evidence for the null hypothesis. The plot shows strong evidence supporting H_0 (no treatment effect, $BF_{01} = 11.27$) and insufficient evidence for H_1 (beneficial treatment effect, $BF_{10} = 0.09$)¹

points at endpoint. As AD patients who receive placebo deteriorate about 4 to 5 ADAS-Cog points per year, this would require a study period of at least 3 years, assuming a complete halt of cognitive decline in the active treatment group. Under such an unlikely scenario an update of the current Bayesian meta-analysis would result only in equipoise between the AAB immunotherapy null hypothesis and its alternative, thus without favoring H_1 over H_0 .

2 | DISCUSSION

Based on these analyses of previously published clinical trial results, we conclude that there is strong *evidence of absence* of a therapeutic effect of AAB immunotherapy in AD and that further pursuit of this approach in AD is likely futile, even if additional studies would replicate the findings available to date, and thus decrease the P value in a traditional frequentist meta-analysis by increasing the total number of subjects in such an analysis. The overall effect size based on the frequentist meta-analysis of these six RCTs of .05 is based on the difference between a yearly average deterioration of 4.5 ADAS-Cog points in 3288 placebo-treated patients versus 4.1 points in 3309 patients with one of the three forms of AAB immunotherapy. This minute non-significant treatment effect of 0.4 ADAS-Cog points over a year, in the subgroups of trial participants subjected to the highest doses of AAB immunotherapy, invites the question how plausible a possible "true" and clinically relevant treatment effect could still be. The strong support for the null hypothesis resulting from the Bayesian analysis renders a true absence of a treatment effect much more plausible. The continued efforts of researchers and pharmaceutical companies toward persistent investment of research efforts, time, and money in AAB immunotherapy may be partially driven by the relatively narrow framework of traditional frequentist statistical methods used.

Several methodological issues should be addressed. First, it could be argued that combining different specific AAB immunotherapeutic drugs in one meta-analysis may be problematic, despite the fact that all RCTs probed essentially the same hypothesis concerning AD therapy. However, a similar approach in hypertension, epilepsy, or thromboem-

bolic disease would have resulted in clear-cut conclusions concerning the efficacy of antihypertensives, antiepileptics, or anticoagulants as drug classes. Second, it could be argued that perhaps the target populations were not optimal in the studies that were used in the present analyses, and that in earlier RCTs even some patients without cerebral amyloidosis have been recruited, in whom it was beforehand unlikely that an approach to remove $A\beta$ from the brain would be effective. However, based on both in vivo $A\beta$ imaging using positron emission tomography and ex vivo study of cerebral $A\beta$ load in trial participants who came to autopsy, there is no doubt that these AAB immunotherapies are potent drugs that can remove $A\beta$ from the brain.^{5,7} But even virtually complete removal of $A\beta$ from the brain for >10 years could not protect AD patients from progression to severe dementia.⁷ This renders the possibility of a type II error due to too-short follow-up to detect an effect unlikely.

Third, the ADAS-Cog as outcome instrument could be questioned. However, this is a widely applied cognitive instrument that has a solid track record as outcome in AD research, including these major RCTs. In essence, the exact cognitive assessment instrument used as outcome in AD RCTs may be less relevant. Even if a minute, statistically significant treatment effect could be detected using a more sensitive, detailed outcome measure, this would most likely not meet the minimally clinically important difference if this difference could not be detected by the ADAS-Cog.

There is an important caveat to this interpretation. Current approaches in AD clinical trials implicitly assume AD to be a rather homogeneous condition, for which one specific intervention, in this case AAB immunotherapy, may be beneficial. However, the heterogeneity of underlying brain pathology long has been recognized.⁸ In line with this recognition, more recent findings suggest that cerebrospinal fluid proteomics may define subtypes of AD that are potentially relevant from a pathophysiological perspective.⁹ The same may hold for additional characteristics, resulting in both clinical as well as biological heterogeneity. If corroborated, this has major implications for selection of trial participants and development of interventions in the future. Delineation of relevant AD subgroups, for instance based on a combination of age at onset, genotype, phenotype, cerebral

co-morbidity, and the specific pathophysiological processes relevant in each individual patient will most likely be necessary, to deliver the right therapeutic intervention to the right patient. Better understanding of this heterogeneity will put neutral trial results in studies based on AD as a homogenous, generic category of disease in a new perspective, and may invite new RCTs in these populations.

By contrasting the results of a traditional frequentist meta-analysis with the results from a Bayesian meta-analysis, and showing how an additional trial result would change the overall results of the latter meta-analysis, we illustrate that further pursuing AAB immunotherapy as treatment for AD as a rather homogeneous condition is likely futile. Even in the unlikely event of a new trial shifting the frequentist meta-analysis toward a significant difference, the Bayesian analysis would most likely still strongly favor the null hypothesis over the alternative hypothesis. And the chance that a new RCT would not only shift the pooled effect size toward a significant difference, but also toward a clinically relevant effect is negligible.

Finally, there are important ethical and societal aspects that need to be addressed. All AAB immunotherapy probed has to be administered intravenously on a monthly basis, and most AAB immunotherapies have a considerable risk of adverse events, including amyloid-related imaging abnormalities (ARIA), which may or may not cause serious symptoms.¹⁰ This would be completely acceptable if clinically important benefits could be expected, such as, for example, in chemotherapy for different forms of malignancies. Based on the currently available trial results it is highly unlikely that AAB immunotherapy will fundamentally change the course of disease in AD in a clinically meaningful way, therefore one could question whether further exposure of large numbers of older persons to this type of experimental therapy is justified. The potential target population for these drugs is very large, and the logistical challenge of intravenous administration and the accompanying costs of the infrastructure needed and the drug itself would be astronomical. Only a clinically relevant effect beyond any doubt could justify the societal burden of this expenditure on scarce health-care resources.

3 | CONCLUSION

Our message is threefold. First, based on six well-designed RCTs there is strong *evidence for absence* of an effect of AAB immunotherapy in AD. Second, the use of Bayes factor hypothesis testing allows for quantification of the plausibility of the null hypothesis, as opposed to traditional frequentist analysis, which only allows us to conclude that there is absence of evidence. Third, whereas results from frequentist meta-analysis may invite the conclusion that additional studies to collect evidence may still be relevant, results from the Bayesian analysis strongly suggest the time has come to divert therapeutic efforts away from AAB immunotherapy. This leads to the potential to open up alternative lines of investigation that have more potential to bear fruit.

ACKNOWLEDGMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Edo Richard is supported by The Netherlands Organization for Health Research and Development (ZonMw Personal grant; VIDI grant number 91718303 to ER).

CONFLICTS OF INTEREST

ER received a personal grant from The Netherlands Organization for Health Research and Development (ZonMw). The funder did not play a role in any part (such as initiation, design, analysis, interpretation of the data, writing the report, and the decision to submit the article for publication) of this manuscript. None of the authors have any conflicts of interest.

REFERENCES

1. Keyzers C, Gazzola V, Wagenmakers EJ. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat Neurosci*. 2020;23(7):788-799. <https://doi.org/10.1038/s41593-020-0660-4>.
2. Douglas G, Altman JMB. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485. <https://doi.org/10.1080/15265161.2019.1665738>.
3. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370(4):311-321. <https://doi.org/10.1056/NEJMoa1312889>.
4. Salloway S, Sperling R, Fox NC, et al. Two phase 3 trials of Bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370(4):322-333. <https://doi.org/10.1056/NEJMoa1304839>.
5. Ostrowitzki S, Lasser RA, Dorflinger E, et al. A phase III randomized trial of gantenerumab in prodromal Alzheimer's disease. *Alzheimer's Res Ther*. 2017;9(1):1-15. <https://doi.org/10.1186/s13195-017-0318-y>.
6. Honig LS, Vellas B, Woodward M, et al. Trial of solanezumab for mild dementia due to Alzheimer's disease. *N Engl J Med*. 2018;378(4):321-330. <https://doi.org/10.1056/NEJMoa1705971>.
7. Nicoll JAR, Buckland GR, Harrison CH, et al. Persistent neuropathological effects 14 years following amyloid- β immunization in Alzheimer's disease. *Brain*. 2019;142(7):2113-2126. <https://doi.org/10.1093/brain/awz142>.
8. Van Gool WA, Eikelenboom P. The two faces of Alzheimer's disease. *J Neurol*. 2000;247(7):500-505. <https://doi.org/10.1007/s004150070147>.
9. Tijms BM, Gobom J, Reus L, et al. Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain*. 2021;143(12):3776-3792. <https://doi.org/10.1093/brain/awaa325>.
10. Sperling RA, Jack CR, Black SE, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimer's Dement*. 2011;7(4):367-385. <https://doi.org/10.1016/j.jalz.2011.05.2351>.

How to cite this article: Richard E, den Brok MGHE, van Gool WA. Bayes analysis supports null hypothesis of anti-amyloid beta therapy in Alzheimer's disease. *Alzheimer's Dement*. 2021;17:1051-1055. <https://doi.org/10.1002/alz.12379>