Continuous intrajejunal levodopa INfusion VErsus deep brain STimulation in advanced Parkinson's disease (INVEST): statistical and health economic analysis plans

D. van Poppelen[1,2], R.M.A. de Bie,[1,2], M. van Barreveld[3], P.R. Schuurman[4], F. van Delft[5], G.J. Geurtsen[2,6], R.J. de Haan[7], M.G.W. Dijkgraaf[3] and J.M. Dijk[1,2] on behalf of the INVEST study group

[1]Amsterdam UMC, University of Amsterdam, Department of Neurology, Meibergdreef 9, Amsterdam, the Netherlands.
[2]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands.
[3]Amsterdam UMC, University of Amsterdam, Department of Epidemiology and Data Science, Amsterdam Public Health, Meibergdreef 9, Amsterdam, the Netherlands.
[4]Amsterdam UMC, University of Amsterdam, Department of Neurosurgery, Meibergdreef 9, Amsterdam, the Netherlands.
[5]Radboud UMC, Radboud University, Department of Gastroenterology, Geert Grooteplein Zuid 10, Nijmegen, the Netherlands.
[6]Amsterdam UMC, University of Amsterdam, Department of Medical Psychology, Meibergdreef 9, Amsterdam, the Netherlands.
[7]Clinical Research Unit, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, the Netherlands.

Preface

This statistical analysis plan (SAP) and consecutive health economics analysis plan (HEAP) describe the planned main analyses of the Infusion Versus Stimulation (INVEST) study, in which continuous intrajejunal levodopa infusion and deep brain stimulation are compared in patients with Parkinson's disease and motor response fluctuations.

The timeframe described in this SAP and HEAP was the original 12-month follow-up point. The SAP and HEAP are written, finalized, sent to the medical ethics board that approved the study protocol (METC AMC Amsterdam), and uploaded to www.clinicaltrials.gov (timestamped) prior to locking the INVEST database and prior to initiation of all analyses. Therefore, this SAP and HEAP were drawn up without insight in collected patient data whilst the course of the study (*e.g.*, number of included patients per study arm, number of patients preliminary terminating study participation, missing visits, and loss-to-follow-up) was taken into account.

The planned analyses of clinical and economic outcomes are reported consecutively in this document. Although the primary outcomes were economic, these are dependent on clinical outcomes. Therefore, the latter are reported first. The sections considering the health economics analysis contain references to the clinical paragraphs. The intention is to publish the results of the clinical and economic analyses separate.

For consistency, this SAP and HEAP are written in the past tense, also regarding matters that still have to take place.

Index

Statistical Analysis Plan

Version 1.0. Date March 18, 2022.

Dutch Trial Register:	Identifier 4753, registered November 3, 2014
EudraCT:	Number 2014–001501-32
Clinicaltrials.gov:	NCT02480803
Sponsor protocol number:	NL51240.018.14 INVEST study Version 7.0, October 23, 2019
Primary source of funding:	ZonMw, The Netherlands Organization for Health Research and Development.
Co-financier:	Medtronic Europe
Trial protocol version:	BMC Neurology 2020;20:40, PMID 32005175

The funding parties had no role in the design of the study, the SAP, analyses, interpretation of data, or decision to submit results.

Signatures

Author	Principal Investigator	Chief Investigator
		
D. van Poppelen MD	Prof. R.M.A. de Bie PhD MD	J.M. Dijk PhD MD
Date: March 18, 2022	Date: March 18, 2022	Date: March 18, 2022

Introduction

Continuous intrajejunal Levodopa Infusion (CLI) and Deep Brain Stimulation (DBS) are accepted therapies for the treatment of advanced Parkinson's disease (PD). CLI has a lower level of evidence compared to DBS and is probably more expensive. Whether or not such a cost difference is justified by a difference in effectiveness is unknown because no head-to-head comparison of CLI and DBS has been performed. This statistical analysis plan (SAP) and health economics analysis plan (HEAP, from page 18 onwards) describe the major analyses for the ongoing INfusion VErsus STimulation trial regarding 12 months follow-up.

Synopsis of the original study protocol

The INVEST study is an open label multicenter randomized controlled trial (RCT) comparing bilateral subthalamic nucleus CLI and DBS in patients with advanced PD. The study protocol was extensively described previously.[1] In summary, eligible patients had severe response fluctuations, bradykinesia, dyskinesia, or painful dystonia despite optimal pharmacological treatment. Exclusion criteria included previous neurosurgery for PD, contra-indications for CLI or DBS, and Hoehn and Yahr stage 5 in on-drug state (on-drug state: the typical functional state when patients are receiving medication and have a good response).[2]

The hypothesis was that treatment with CLI and DBS would both be effective for PD, but that CLI would be more expensive than DBS and that the surplus in costs would not be sufficiently compensated for by more improvement of quality of life compared to DBS. This would implicate that CLI is not a cost-effective treatment.

This design of the study was an RCT with ancillary patient preference observational arms which is also known as a "patient preference trial" or "comprehensive cohort study".[3] In the RCT, patients with PD who met the inclusion criteria and who were willing to be randomized between CLI and DBS treatment were assessed for eligibility for both treatments. Eligible patients were randomized using online computer software (ALEA, Amsterdam UMC). Quality of life, costs, disability, PD specific motor and non-motor symptoms, neuropsychiatric comorbidities and caregiver burden were assessed 12, 24 and 36 months after initiation of advanced therapy (*i.e.*, CLI and DBS). Furthermore, at three, six, and nine months follow-up, additional assessments regarding costs, utilities and adverse events were planned.

It was expected that only a proportion of patients would be willing to be randomized between the two freely available therapies. Patients who were eligible to participate in the RCT but were not willing to be randomized, were asked to participate in the ancillary patient preference observational study. The observational study arms served primarily to determine the extent of the external validity of the RCT-results. Furthermore, data from RCT patients were pooled with data from the patients participating in the ancillary patient preference observational study, to compare specific clinical outcomes in a larger population. We aimed to randomize a total of 66 patients. There was no minimum of participants in the ancillary patient preference observational arms. Although an open-label study, data collection was performed by investigators who had no therapeutic relationship with study participants. Study data was not accessed by treating physicians. Treating physicians were aware of patients' study participation and allocated treatment if appropriate. Patients were treated in accordance with the center specific usual care regarding the procedures (*e.g.*, planning of DBS approach, usage of general anesthetics during placement of the DBS electrodes, possible pre-CLI medication test phase, and dose or device settings).

Synopsis of study protocol development

The first patient was included in December 2014. During the study, new insights and potential weaknesses in the original design were addressed in (substantial) amendments of the study protocol. Major adjustments are presented below. If applicable, this is also addressed in the relevant following paragraphs.

In the initial design of the study, patient randomization was stratified according to both experience of the including center and the response to a supra-threshold levodopa-dose. A center was considered experienced in CLI or DBS treatment if in the previous two years at least five patients per year were started with CLI or DBS respectively in that center. Centers were able to change in stratum in case experience grew. Yet, four different strata based on experience of the including center (*i.e.*, little and much experience for both CLI and DBS) with two strata of levodopa response, would have led to a total of eight strata. This would lead to too few patients in at least some strata, considering the planned number of patients to be randomized. To reduce the number of strata, the protocol was amended in 2016 to include only stratification according to experience with both therapies.

In the original study protocol, a follow-up period of 12 months was described, but it was already stated that a longer follow-up period was desirable for a more reliable assessment and prediction of long-term effects and costs. After acquisition of additional funding, 24 and 36 months follow-up visits were added to the study protocol in 2017. All patients included were offered to participate in the extended follow-up. This statistical analysis plan describes the statistical analysis of the major outcomes of the first 12 months follow-up of the INVEST study, and not the ongoing 24- and 36-month follow-ups.

After it proved challenging to include patients in other centers than the primary including hospital (Amsterdam UMC), the recruitment and consent procedure was adjusted twice to enhance inclusion in the ancillary patient preference observational arms in other participating centers. First, it was no longer obligatory to sign the informed consent form face-to-face with the researcher for eligible patients who did not consider participation in the RCT but were willing to be included in the ancillary patient preference observational arms (2016). Patients could return the informed consent form by mail as of then. Second, treating physicians from non-participating centers were also allowed to register patients for the ancillary patient preference observational arms, after which the research team contacted the patient and enrolled them if eligible (2019). Furthermore, to enhance recruitment in the RCT in other centers, treating physicians were contacted frequently and patients were informed through the Dutch Parkinson Society. This did not lead to protocol changes.

Concerning the primary outcome, the usage of the Parkinson's Disease Questionnaire-39 (PDQ-39) was operationalized as follows: the difference between the two treatment groups (*i.e.*, DBS and CLI) in change from baseline to 12 months on the PDQ-39 summary index score. The change (*i.e.*, from baseline to 12 months) on total PDQ-39 was chosen as this score displays alteration in disease related quality of life after instalment of the allocated treatment and therefore is largely independent of possible differences in baseline values in the relatively small populations. The PDQ-39 has been a frequently used outcome measure in clinical trials for PD.[4–6] Use of repeated PDQ-39 measures was planned in the original design; a PDQ-39 assessment was planned at the nine months follow-up time-point to be able to pool data with a designed similar trial by another study group. Unfortunately, that study and planned international pooling of data never came into existence. Because there was a limited number of repeated PDQ-39 measurements (*i.e.*, at baseline, nine months, and 12 months follow-up), no major differences between both follow-up points were

expected. Also, up to 2016 the nine months visits proved incomplete with regard to the PDQ-39 assessments. Therefore, it was decided not to perform a repeated measurement assessment.

Originally, an intention-to-treat analysis (ITT) was planned as the primary analysis. Nevertheless, during the study it was observed that after allocation, several patients immediately refused CLI treatment and either chose to be treated with DBS or quitted study participation. Analyzing these patients in an ITT could have introduced a significant bias, as their apparent preference for DBS made them not representative of a population that considered CLI a suitable treatment for advanced PD. Furthermore, for the patients for whom post-baseline data were not available, all these data would have had to be imputed, contributing to the risk of bias. A modified intention-to-treat (mITT) population was introduced as the primary analysis, because analysis in the ITT population including all randomized individuals was considered less valid.[7] The mITT population included all randomized patients who received the allocated treatment and continued in the RCT after randomization. Additionally, in line with the proposed strategy for ITT analyses for trials with incomplete outcome data as proposed by White et al.,[7] sensitivity analyses were performed to explore the robustness of the main analyses in the mITT population and included analyses in all randomized patients (the ITT population). Please see "Sensitivity analysis" for details on page 12.

While conducting the study, it was observed that comorbidities had a negative impact on completeness of the follow-up data. To assess for a possible bias towards a selective data collection of healthier patients, treating physicians of patients that had dropped out of the study were contacted to obtain additional information on the general health status and level of functioning of these patients. Please see also "Patient replacement and missing data" on page 10.

Study methods

Outcomes - randomized controlled trial

Primary outcomes

The main clinical outcome was the change from baseline to 12 months after initiation of treatment in PDQ-39 (score 0-100).[8]
The primary health economic outcomes are elaborated on page 19.

Secondary outcomes

- PD motor symptoms -- Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) part III in on-drug state (score range 0-132) [2];
- Motor experiences of daily living-- MDS-UPDRS part II in on-drug phase (score range 0-52) [2];
- Dyskinesia -- Clinical Dyskinesia Rating Scale (CDRS, score range 0-28) in on-drug state [9];
- On-drug state time without troublesome dyskinesia -- mean hours per day in on-drug state without troublesome dyskinesia assessed with a 3-day motor symptom diary;
- Medication -- PD medication expressed in Levodopa Equivalent Daily Dose (LEDD, mg per day, including the continuous intrajejunal delivered levodopa)[10];
- Functional health status -- Amsterdam Linear Disability Scale (ALDS) in on-drug state (29 items, score range 0-100) [11];
- Cognitive performance -- Parkinson's Disease Cognition Rating Scale (PD-CRS, score range 0-134) [12];
- Apathy -- Starkstein Apathy Scale (SAS, score range 0-42).[13] Development of apathy was defined as newly scored apathy at 12-month follow-up (sum score ≥14 at follow-up and <14 at baseline) and minimally 2 points increase in SAS score;

- Impulse-compulsive disorders and other compulsive behaviors -- Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease (QUIP, utilizing established thresholds)[14]. Patients were considered to have developed an impulsive compulsive disorder, other compulsive behavior or compulsive use of medication if the QUIP criteria were met at 12-month follow-up and not at baseline. Impulsive compulsive disorder, other compulsive behavior or compulsive use of medication were considered to be resolved in case the QUIP criteria were not met at 12 months whilst they were fulfilled at baseline;
- Psychiatric diagnoses -- current and previous psychiatric diagnoses were assessed with the following subsets of the Mini International Neuropsychiatric Interview (MINI): major depressive episode, (hypo)manic episode, panic disorder, agoraphobia, social anxiety disorder, obsessive-compulsive disorder, alcohol use disorder, substance use disorder, and psychotic disorder.[15] A new psychiatric diagnosis was defined as a current psychiatric diagnosis at 12-month follow-up, not being present at baseline. A psychiatric diagnosis was considered to be resolved in case it was present at baseline and not at 12-month follow-up;
- Harms -- (serious) adverse events ((S)AEs);
- Treatment course -- stopping the allocated treatment and switch to the alternative after initially receiving the allocated treatment (*i.e.*, switch from DBS to CLI or vice versa);
- Treatment satisfaction – number of patients that would recommend their treatment to others.

Certain outcomes were not part of the principal analyses. They will only be addressed in the main publication at the discretion of the reviewer and will be published separately otherwise. This applies to the following outcomes: functional health status (Hoehn & Yahr stage), non-motor symptoms (Non-Motor Symptom Checklist), laboratory measurements (levels of vitamin B6, B12 and folic acid), extensive neuropsychological assessment, depression (Hamilton Depression Scale), anxiety (Hamilton Anxiety Scale), suicidality (Columbia Suicide Severity Rating Scale), patient satisfaction based on a study tailored questionnaire, and caregiver burden based on a standardized questionnaire.

Outcomes - ancillary patient preference cohort study

Outcomes for the ancillary patient preference cohort study were the change in quality of life (PDQ-39) from baseline to 12 months after initiation of treatment; change in PD-medication expressed in LEDD from baseline to 12 months (including the continuous intrajejunal delivered levodopa); change in functional health status (ALDS) from baseline to 12 months; (S)AEs; stopping treatment; and switch to the alternative treatment after initially receiving the allocated treatment (*i.e.*, switch from DBS to CLI or vice versa).

Statistical analysis

General principles

The analyses were performed after completing the last 12-month follow-up visit, monitoring and data validation, and after submission of this SAP and HEAP to the medical ethics board and uploading of the SAP to the registration page of the study at www.clinicaltrials.gov. Analyses were performed by the investigators of the INVEST study group, independent of study sponsors, using SPSS statistics software (IBM Corp., Armonk, NY, USA) and Microsoft Excel for visualization of results.

No interim analyses were performed. Statistical uncertainty was expressed in two-sided 95% confidence intervals (CI). The main clinical outcome was viewed as significantly different between the CLI treatment RCT group and DBS treatment RCT group, if the two-sided P-value was less than 0.05. For other outcomes, no individual P-values were reported. In case of a non-normal distribution of one of the variables in a specific analysis in the RCT, that analysis was performed with bias corrected

accelerated bootstrapping, drawing 2,500 samples of the same sizes as the original samples and with replacement.

Patient flow diagram

According to the Consolidated Standards of Reporting Trials 2010 statement (CONSORT), a flow diagram illustrating the number of participants ineligible, eligible, consented, randomized, received allocated treatment, excluded, withdrawn and lost to follow-up for the primary outcome was provided. Please see the Figure SAP-1 on page 29.

Analysis populations

Modified intention-to-treat population

The primary analysis consisted of the modified intention-to-treat (mITT) population which included all randomized patients who received the allocated treatment, regardless of continuation of treatment. Analyses in the original ITT population including all randomized patients, regardless of the treatment they had actually received (*i.e.*, also patients not receiving the allocated treatment and patients quitting study participation immediately after randomization) were accounted for in sensitivity analyses (see page 12, “Sensitivity analysis”) and were presented in the supplements.

As-treated population

To provide insight in potential bias resulting from excluding patients who did not adhere to the allocated treatment or switched treatment, analyses were additionally performed in an as-treated population categorizing the randomized patients according to the actual treatment(s) they were receiving at the moment of the 12-month visit (if available).[7]

Per protocol population

The per protocol population consisted of all patients receiving the allocated treatment who completed the study protocol without protocol violations.

Ancillary patient preference population

The ancillary patient preference population consisted of all patients who signed informed consent for this element of the study, started the chosen treatment and underwent at least the baseline visit and one follow-up visit. Patients were allowed to participate in the ancillary patient preference cohort study after withdrawal of consent for the RCT prior to randomization. Please see also Appendix II for the flowchart of the Ancillary patient preference cohort.

Adherence and protocol deviations and violations

Adherence to the intervention CLI treatment was defined as having the percutaneous endoscopic gastrostomy (PEG) tube placed and having continuous administration of levodopa by means of the CLI pump. Adherence to the intervention DBS treatment was defined as having the DBS system implanted and turned on. Study patients were able to make small adjustments to their CLI or DBS settings if permitted by their healthcare professionals, major changes to treatment regimens could only have been implemented by healthcare professionals. Therefore, adherence was expected to be very good. Continuation of treatment was regularly inquired during the planned visits.

Protocol deviations were quantified and projected per study group. Protocol deviations were defined as treatment, research activities or other procedures that diverged from the study protocol were considered to have had no significant consequences for interpretation of study outcomes. A protocol violation was defined as a divergence from the protocol that might significantly have impacted the completeness, accuracy, and/or reliability of the study data, was not in line with the consent form, or

might have impacted patients' safety, welfare or rights.[16,17] In Appendix I possible protocol deviations and violations were elucidated. All protocol deviations were line-listed according to treatment group. In addition, the number and percentage of patients in each treatment group experiencing one or more protocol deviations were presented. Minor protocol deviations did not impact inclusion in all analyses. Protocol violations did not alter inclusion in the (modified) intention-to-treat analysis or the ancillary patient preference observational analyses and resulted in exclusion from the per protocol analysis.

Patient replacement and missing data

Main clinical outcome

For participants with missing values for the main clinical outcome (PDQ-39), three approaches were used:
I - In case only the 12-month follow-up PDQ-39 was missing and the 9-month PDQ-39 was present, the latter was used, minus or plus the mean difference in PDQ-39 between nine and 12 months follow-up for that treatment arm (of present data), since little change in PDQ-39 between nine and 12 months follow-up was expected.
II - In case both follow-up PDQ-39's were missing (*i.e.*, nine and 12 months) but a baseline PDQ-39 was available, multiple imputation was used by fully conditional specification using predictive mean matching with five data sets and with potentially relevant variables at baseline (*e.g.*, age, sex, PDQ-39), previously observed outcomes and adverse events as predictors. The predictive mean matching imputation was performed separately per treatment group of the mITT population and chosen to ensure that imputed values were in the appropriate range.[18] Additionally, we repeated this imputation of the PDQ-39 for the total mITT population as a scenario analysis.
III - For patients with a missing baseline PDQ-39, the imputation method as described in II was used, with follow-up PDQ-39 score(s) as covariate(s).

Since missing values were expected to be missing not at random, a conformity scenario and multiple worst-case scenarios for the mITT analysis were predefined, making use of additionally gathered data if possible. If lost to follow-up, the patients' treating physicians were asked to provide information on these patients regarding their treatment and functioning at 12 months after start of the allocated treatment (*i.e.*, deceased and if so, cause of death; use of the allocated treatment; use of the alternative than allocated treatment; living at home or in a health care facility; and modified Rankin scale).[19] Loss to follow-up because of clinical improvement was deemed unlikely, therefore the conformity scenario for a patient with a missing 12-month PDQ-39 summary index score was to assign the mean PDQ-39 value in the allocated treatment arm. Possible worst-case scenarios were: (A) to assign the lowest assessed PDQ-39 in both treatment arms, and (B) scenario 1 minus ½ minimal clinically important difference (*i.e.*, ½ of -1,6 as defined in literature).[20] Based on the additional data acquired in case of loss to follow-up, the most applicable worst-case scenario was selected and justified.

In case of death, the 9-month follow-up PDQ-39 was carried forward if available. For patients alive at the moment of the 9-month follow-up, but without available PDQ-39 at that moment, the 9-month imputed PDQ-39 was carried forward. In case of death before the 9-month follow-up time point, no PDQ-39 summary index scores were used for this patient.

Secondary outcomes

For other clinical outcomes at 12 months, the last observation was carried forward if available from another visit (excluding baseline). For baseline characteristics, missing values were not imputed. For displaying baseline characteristics, the numerator and denominator were stated for dichotomous

variables (*e.g.*, male: 5/15). For continuous variables, the number of patients for whom the variable was available was stated.

Study population

Baseline characteristics

For both the RCT mITT population and for the ancillary patient preference population, baseline characteristics were presented in Table SAP1. Categorical baseline characteristics were summarized by presenting counts and percentages. Visual inspection was used for assessment of normality of data distributions. Continuous, normally distributed variables were summarized by presenting the means and standard deviations. Continuous, non-normally distributed and ordinal variables were summarized by presenting the medians and interquartile ranges. Formal statistical tests were not performed to examine baseline differences between treatment groups. No bootstrapping was used for presentation of the baseline characteristics.

Table SAP1. Baseline characteristics

	CLI-RCT (n)	DBS-RCT (n)	CLI-observational (n)	DBS-observational (n)
Age (years)				
Sex, F (% n/N)				
Time since PD diagnosis (years)				
Hoehn and Yahr stage in on-drug state				
Levodopa-equivalent daily dose (mg/day)				
MDS-UPDRS part III score - off-drug state - on-drug state - improvement off- to on-phase (%)*			*NA*	*NA*
Parkinson's disease cognitive rating score			*NA*	*NA*
Level of experience of treating center with CLI and DBS treatment (%)** - *Experienced in CLI, not in DBS* - *Experienced in DBS, not in CLI* - *Experienced in both CLI and DBS* - *Little experience in both CLI and DBS*				

CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, F: female, MDS-UPDRS: Movement Disorder Society's Unified Parkinson's Disease Rating Scale, mg: milligrams, NA: not available, RCT: randomized controlled trial
* Improvement: (score in off-drug state minus score in on-drug state) divided by score in off-drug state
**Experienced in DBS/CLI: ≥ 5 patients treated per year in the two years previous to including the patient with the respective treatment in the treating center. A center may have changed in level of experience over time.
For normally distributed data means and standards deviation were presented, for non-normally distributed data medians and interquartile ranges. For the Hoehn & Yahr stage median and range were reported.

Analyses

Randomized controlled trial

Reported outcomes were presented in Table SAP2 (part A and B). The main publication consisted of analyses in the mITT population. The analyses in ITT, as-treated and per-protocol populations were presented in supplements.

Analysis of main clinical outcome

The main analysis regarding the clinical outcome was the between-group difference in change from baseline to 12 months follow-up in total PDQ-39 summary index score in the mITT population, using a two-group *t*-test. Additionally, a multiple linear regression analysis was performed taking into account the stratifying variable (level of experience of the treatment center) and (if necessary) clinically relevant baseline imbalances. To assess whether the baseline value of the PDQ-39 served as an effect modifier (meaning that the association between intervention and outcome differed according to baseline PDQ-39) an interaction term was used in the regression analysis. In case of a significant interaction term, the analysis was additionally performed in two separate strata: one with patients with a median PDQ-39 or below and the other with patients with a PDQ-39 above median.

Sensitivity analyses

Sensitivity analyses concerning missing data as described in the paragraph "Patient replacement and missing data" were performed in the mITT population. To further assess robustness of the mITT analysis for the main clinical outcome additional sensitivity analyses were performed.

In line with the proposed strategy for ITT analyses for trials with incomplete outcome data as proposed by White et al, analyses were performed in the ITT population, including the patients who had been excluded from the mITT population (*i.e.*, patients who immediately after randomization either had chosen not to accept the allocated treatment or had quitted participation in the study).[7] For missing values for the main clinical outcome in the analyses in the ITT population, the approach for handling missing values as used for the mITT analyses as described in "Patient replacement and missing data" on page 10 was applied; first making use of the previously imputed values for the mITT population and additionally imputing missing values of the ITT-population based on the total population according to the conformity scenario and the most applicable worst case scenario chosen for the mITT analysis. In addition, a "best case" scenario was applied as for patients who immediately quitted study participation after randomization and for whom the main clinical outcome at 12 months was missing, a favorable outcome was possible. In this scenario, for patients who immediately quitted study participation after randomization, missing values for the 12-month PDQ-39 summary index score were imputed with the highest PDQ-39 in the treatment arm the patient had been allocated to. Other missing values were imputed according to the previously described conformity scenario.

Additionally, an as-treated analysis was performed to make the effect of treatment switching intelligible for interpretation of the results of the main analysis. Patients could have switched to the other than the allocated treatment either immediately after randomization, without first having had the allocated treatment (these patients were not included in the mITT analysis), or during the follow-up after initial treatment with the allocated therapy. In the analysis of the as-treated population, patients were analyzed according to, if known, the actual treatment they were receiving at the moment of the 12-month visit. In case of add-on treatment with the alternative than the allocated treatment, patients were analyzed in the group of the primary therapy. Missing data were not imputed in this analysis. To evaluate the robustness of the main clinical analysis this was considered the adequate approach while more sophisticated adjustment methods for treatment switching could have led to error in a relatively small RCT.[7]

Finally, to evaluate the effect of protocol violations on the analysis of the main clinical outcome, analysis in the per protocol population was performed.

Analysis of secondary clinical outcomes

The difference between both treatment groups in the mean change from baseline to 12 months in total MDS-UPDRS part II and III was shown with 95% CI. Similarly, the difference between groups in mean change from baseline to 12 months in CDRS was shown, as well as the difference in mean change in total time in both off-drug state and on-drug state without bothersome dyskinesias as reported in the 3-day motor diary, the difference in mean change in usage of dopaminergic medication (LEDD), and ALDS in on-drug phase, all with 95% CI. Discontinuation of the allocated treatment or switch to the other treatment option were summarized using simple descriptive statistics. The change in cognition expressed in points on PD-CRS from baseline to 12 months was reported. For apathy, the numbers and frequencies of patients with apathy at baseline and after 12 months follow-up were shown. Moreover, the numbers and frequencies of patients who had developed apathy after 12 months (*i.e.,* no apathy at baseline, but presence of apathy after 12 months), was reported for both groups, including the difference between these frequencies and 95% CI. Similarly, numbers and frequencies of impulsive-compulsive disorders and, separately, of development of impulsive-compulsive disorders were shown, with difference and 95% CI of the latter. Additionally, the numbers and frequencies of resolution of impulse control disorders were reported for both groups as well as the difference and 95% CI. Psychiatric diagnoses were reported analogues to impulse-compulsive disorders.

The number of patients that would recommend their treatment at 12 months to others was reported and the difference in frequency between the treatment groups was reported with 95% CI.

We presented line listing of safety outcomes using counts and percentages. All (S)AEs, including death were reported for each treatment arm. See Table SAP3 for an overview of (S)AEs. The treatment course per group was expressed in number of patients stopping allocated treatment and number of patients switching to the alternative treatment.

Table SAP2. Main and secondary clinical outcomes at 12 months, mITT RCT population
Part A: continuous outcomes

Outcome	Baseline				Within-treatment Change from Baseline to 12 Months				Between-Group Difference in Change from Baseline	P-value
	CLI		DBS		CLI		DBS			
	no. of patients	mean ±SE	no. of patients	mean ±SE	no. of patients	mean ±SE	no. of patients	mean ±SE	mean (95% CI)	
Main clinical outcome										
PDQ-39 summary index score										P
Secondary clinical outcomes										
MDS-UPDR-III score in on-phase										NR
CDRS in on-phase										NR
Time in on-phase without troublesome dyskinesias (hr)										NR
Time in off-phase (hr)										NR
LEDD (mg per day)										NR
ALDS score in on-phase										NR
PD-CRS score										NR

ALDS: Amsterdam Linear Disability Scale, CDRS: Clinical Dyskinesia Rating Scale, CI: confidence interval, CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, hr: hours, LEDD: Levodopa Equivalent Daily Dose, MDS-UPDRS: Movement Disorder Society's Unified Parkinson's Disease Rating Scale, mg: milligrams, NR: not reported, PD-CRS: Parkinson's Disease Cognitive Rating Scale, PDQ-39: Parkinson's Disease Questionnaire-39, SE: standard error

Table 2. Main and secondary clinical outcomes at 12 months, mITT RCT population
Part B: dichotomous outcomes

Outcome	**Baseline**		**12 Months**		**Between-group Difference at 12 months**	**Development / resolution ***		**Between-group Difference in development / resolution***
	CLI	DBS	CLI	DBS	difference in proportion (95% CI)	CLI	DBS	difference in proportion (95% CI)
	n/total no. of patients (%)	n/total no. of patients (%)	n/total no. of patients (%)	n/total no. of patients (%)		n/total no. of patients (%)	n/total no. of patients (%)	
Apathy (SAS)								
Impulsive-compulsive disorder (QUIP)								
Psychiatric diagnosis (MINI)								
Patient would recommend treatment	-	-				-	-	-

CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, MINI: Mini International Neuropsychiatric Interview, NA: Not applicable, NR: not reported, SAS: Starkstein Apathy Scale, SE: standard error, QUIP: Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease

* the first number relates to development of the outcome, the second number relates to resolution of the outcome (if applicable)

Table SAP3. (Serious) Adverse Events

	CLI- RCT (n)	DBS-RCT (n)	CLI-observational (n)	DBS-observational (n)
Serious Adverse events				
Death (n,%)				
Related to therapy (n,%)				
Other causes (specified, n,%)				
Hospital admission (n,%)				
Related to therapy (n,%)				
Other reason (specified n,%)				
Other SAEs (n,%)				
Specified (n,%)				
Adverse events				
Gastro-intestinal related (n,%)				
Specified (n,%)				
Polyneuropathy (n,%)				
Cerebral hemorrhage (n,%)				
Related to therapy (n,%)				
Epilepsy (n,%)				
Other (n,%)				
Specified (n,%)				

CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, RCT: randomized controlled trial, SAE: serious adverse event

Ancillary patient preference observational study arms

For both the observational CLI and DBS groups, changes in outcomes from baseline to 12 months after start of treatment were displayed in Table SAP4. The numbers of patients who stopped their advanced treatment and who switched between advanced treatments were reported.

To assess the extent of external validity of the mITT population, differences in baseline characteristics of all randomized patients (*i.e.*, both the CLI and DBS groups combined) and the separate observational (*i.e.*, non-randomized) CLI group and DBS group of the ancillary patient preference study were shown with 95% CI in supplementary Table SAP1-S. With regard to the main clinical outcome (the change in PDQ-39 summary index scores at 12 months relative to baseline), the difference between patients randomized to CLI treatment (mITT population) and those who preferred that treatment was reported with 95% CI. The same was done for patients randomized to DBS treatment (mITT population) and those who preferred that treatment. These figures were presented in the text.

To assess the impact of the treatment on the change in PDQ-39 at 12 months relative to baseline with more power, this was analyzed using multivariable regression in the total group of patients (*i.e.*, both the mITT randomized population and non-randomized population), taking into account significant imbalanced baseline variables and participation in the RCT or preference observational study.

Table SAP4. Ancillary patient preference observational study arms, outcomes at 12 months.

Outcome	**Baseline**				**Within-treatment Change from Baseline to 12 Months**			
	CLI		DBS		CLI		DBS	
	no. of patients	mean±SE	no. of patients	mean±SE	no. of patients	mean±SE or n (%)	no. of patients	mean±SE or n (%)
Main clinical outcome PDQ-39 summary index score								
Secondary outcomes LEDD (mg per day) ALDS score in on-phase								

ALDS: Amsterdam Linear Disability Scale, CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, LEDD: Levodopa Equivalent Daily Dose, PD-CRS: Parkinson's Disease Cognitive Rating Scale, PDQ-39: Parkinson's Disease Questionnaire-39, SE: standard error
For continuous variables, mean and Standard Error were reported, for dichotomous variables the number of patients (n) and percentage (%).

Health Economics Analysis Plan

Version 1.0. Date March 18, 2022

This health economics analysis plan (HEAP) describes the plan to analyze and report the economic evaluation alongside the INVEST-study and should be considered an extension of the statistical analysis plan for the clinical outcomes. The reporting of this economic evaluation complies with the update CHEERS reporting standards 2022.[21] Any deviations from the guideline were described and explained in full.

Dutch Trial Register:	Identifier 4753, registered November 3, 2014
EudraCT:	Number 2014–001501-32
Clinicaltrials.gov:	NCT02480803
Sponsor protocol number:	NL51240.018.14 INVEST study Version 7.0, October 23, 2019
Primary source of funding:	ZonMw, The Netherlands Organization for Health Research and Development
Co-financier:	Medtronic Europe
Trial protocol version:	BMC Neurology 2020;20:40, PMID 32005175

The funding parties had no role in the design of the study, the HEAP, analyses, interpretation of data, or decision to submit results.

This HEAP was prepared by the INVEST core study group including
J.M. Dijk PhD MD, neurologist (Chief investigator)
Prof. R.M.A. de Bie PhD MD, neurologist
D. van Poppelen MD, resident in neurology
M. van Barreveld MSc, health scientist (author)
Prof. M.G.W. Dijkgraaf PhD, health economist (guarantor)

Signatures

Author	Guarantor	Chief Investigator
		
M. van Barreveld MSc	Prof. M.G.W. Dijkgraaf PhD	J.M. Dijk PhD MD
Date: March 18, 2022	Date: March 18, 2022	Date: March 18, 2022

For trial introduction and background, see the associated Statistical Analysis Plan INVEST included in this document, version 1.0. Date: March 18, 2022.

General characteristics

Aim

The aim of the economic evaluation was to answer the question: 'Is CLI treatment in PD patients associated with sufficiently higher quality adjusted life years when compared to DBS treatment, in order to justify its higher mean yearly costs?'.

Objective

The primary objective of the economic evaluation is to estimate the cost-effectiveness and cost-utility of CLI treatment versus DBS treatment in the first year of follow-up after treatment initiation. The results of the cost-effectiveness analysis (CEA) allow for local decision making on the economically most optimal treatment approach for patients with PD, based on the clinically relevant health outcomes regarding quality of life in particular, while the cost-utility analysis (CUA) provides information for health policy decision makers to decide upon efficient and affordable health care in general, based on nationally accepted evaluation criteria regarding costs per QALY in view of patients' disease burden.

Types of economic evaluation

The economic evaluation of CLI against DBS was conducted alongside the INVEST clinical trial with individual patient-level data. The costs per unit change on the PDQ-39 summary index score at 12 months follow-up versus baseline and the costs per quality adjusted life year (QALY) were the primary economic outcomes. Incremental cost-effectiveness (ICER) and cost-utility (ICUR) ratios were calculated as the costs difference divided by the difference in effect for CLI minus DBS treatment.

Jurisdiction

The trial was conducted in the Netherlands, where health insurance is obligated for inhabitants and health care is offered by public and private providers.

Perspective, time horizon and discounting

Both health care and societal perspectives were taken into account with a time horizon of 12 months following index surgery. Because the time horizon was 12 months, no discounting was performed.

Cost components

Costs encompassed health care costs, out-of-pocket expenses by patients and family members and costs of productivity loss due to sick leave from work (absenteeism) or lowered efficiency while at work (presenteeism).

Sources

Health care costs covered the use of the following use of resources:

- screening for appropriateness of treatment indication;
- index treatments CLI with infusion devices or DBS with stimulation devices;
- emergency departments visits;
- inpatient hospital admissions at neurology, psychiatry, the intensive care, or other units;
- outpatient hospital consultations (*e.g.*, medical specialist, specialist nurse);
- diagnostic procedures (*e.g.*, imaging by MRI, CT, ultrasonography or X-ray);
- therapeutic procedures (*e.g.*, ex-plantation, revision tube or other material, medication);
- out-of-hospital consultations (general practitioner, physical therapist, psychologist, other);
- institutionalized care (*e.g.*, nursing home, rehabilitation center);

- formal home care (*e.g.*, housekeeping, personal care, nursing);
- use of medication.

Out-of-pocket expenses included the following used resources:

- over-the-counter medication;
- informal help at home;
- non-reimbursable devices;
- health-related travel;
- additional expenses by informal primary caregivers.

Productivity loss focused on the impact of CLI and DBS on lost working hours from paid work only and was addressed exploratory. Lost working hours in the subgroup of patients with a paid job were quantified by summing the:

- number of days of sick leave multiplied with patient's mean number of working hours a day;
- number of affected days at work multiplied with the factor "(10 minus self-reported efficiency)/10" with self-reported efficiency reported on a scale from 0 (did not accomplish anything at all) to 10 (worked as efficiently as always).

Lost working hours in the subgroup of primary caregivers with a paid job were estimated as supplied by those caregivers.

Data on use of health care resources were derived from hospital records, electronic clinical report forms and the Medical Consumption Questionnaire (iMCQ) completed by patients or their primary caregiver at quarterly intervals.[22] The questions on the out-of-pocket expenses presumed relevant were added to the iMCQ. To quantify health-related travel, the numbers of contacts with distinct health care providers were counted and combined with standardized distances in kilometers between patients' homes and local health care providers,[23] except for the geographically more thinly scattered DBS treatment centers in which case the mean of the actual distances in kilometers between postal codes of providers and patients were used.[24] The result per patient was multiplied by two considering the 'to and from' of traveling. Data on sick leave from work and lowered efficiency while at work from patients as well as lost working hours by the primary informal caregiver were gathered with an adjusted version of the Productivity Costs Questionnaire (iPCQ).[25]

Unit costs and costing

If opportune, unit costs of (non-)hospital resources, working hours lost, and time spent on the patient by the primary care-giver were obtained from the most recent Dutch Manual on Costing in health care research (DMC).[23] The unit costs of the (re)placement of the neurostimulator system and of the tube and pump system (re)placements were based on unit costing sheets for distinct hospital care resources from the treatment center with the highest number of inclusions and subsequently capitalized based on a 5-year depreciation period (reflecting the lengths of their life cycles) in order to allow for a robust estimate of the difference in yearly treatment costs by study arm, irrespective of the chosen study's time horizon. The same hospital unit costing sheets, existing rates for mutual services excluding VAT, or passer-by prices (*e.g.*, AMC Passantenprijslijst OZP, 2020) were used for observed diagnostic examinations and therapeutic procedures not listed in the DMC. Unit costs for medication were gathered from the Dutch Pharmaceutical Compass assuming standard prescriptions.[26,27] The unit cost of lost working hours at the mean age of the included working patient at baseline was used, non-specific for gender. In case of prolonged continuous sick leave longer than the friction period, the friction cost method was applied.[28] The length of the friction period in days was calculated as the average over the study's observation period 2015-2020, weighted for the

yearly number of included patients, and equaled 97 days.[23,29] Table EE1-S in the Supplementary Material shows unit cost of the resources used. Unit costs held for the base year 2020; unit costs originating from different calendar years were price-indexed with general yearly consumer price indices from Statistics Netherlands.[30]

Costs of resources used were calculated by multiplying the frequency of distinct resources used with their respective (depreciated) unit costs. Volumes and costs (of health care used, out-of-pocket expenses, and productivity loss) are reported separately as (subtotal and total) means per patient per study arm.

Analyses were adjusted if deemed necessary in case of Parkinson-related or Parkinson non-related outliers in costs or QALYs. If related to Parkinson's disease, these were included and reported, if caused by unrelated comorbidities, these could have been excluded.

Health outcomes

The primary health outcome in the cost-effectiveness analysis was the change in the PDQ-39 between baseline and 12 months after randomization. The change score showed the increase or decrease in number of difficulties across eight different dimensions of health (including mobility, activities of daily living, emotional well-being, and stigma) as experienced by patients in the preceding month. Dimension scores are coded on a scale of 0 (perfect health as assessed by the measure) to 100 (worst health as assessed by the measure). The change score was defined as the mean score on the eight scales at month 12 minus the mean score at baseline, in theory ranging from -100 to +100 with a score of 0 representing 'no change'.

The primary health outcome in the cost-utility analysis was the number of QALYs. Patient's health status was assessed with the EQ-5D-3L at baseline, one week, three months, six months, nine months, and 12 months post randomization. The EQ-5D contained five items: mobility, self-care, usual activities, pain/complaints, and mood (anxiety/depression).[31] Each item had three response options: no problems, some problems, or serious problems. EQ-5D scoring profiles at successive measurements were converted into health utility scores using general population based tariffs of time trade-off ratings of health states.[32] QALYs were calculated by taking the product sum of the health utility scores during follow-up and the interval lengths in years in-between successive measurements.

Health economic analysis methods

Analysis population

The CEA and CUA were done with the modified intention-to-treat population (see page 9) with references to the CONSORT-flow diagram in Figure SAP-1 and the baseline characteristics of the two RCT-subpopulations in Table EE1. Summary statistics for health utility scores at baseline were included in the manuscript text.

Cost-effectiveness threshold

The applied cost-effectiveness threshold for the sample size calculation was €120,000 as described and substantiated in the published study protocol.[1]

Statistical decision rules

Use of resources by patients were counted by type of resource by treatment group and summarized as means per patient with 95% bias corrected and accelerated confidence intervals (95% BCa CI) after non-parametric bootstrapping, drawing 2,500 samples of the same sizes as the original samples and with replacement (see Tables EE2a on hospital resources and EE2b on out-of-hospital resources and productivity loss). The associated costs of these used resources were summarized similarly, at the same disaggregated level and reported in the Supplementary Material in Tables EE3a-S and EE3b-S. Health-related out-of-pocket expenses by patients and their families were added to EE3b-S. Mean per patient costs per treatment group at aggregated level and the corresponding group differences were reported in Table EE3 in the main manuscript, along with P-values. Differences in mean aggregated costs were assessed using two-sample *t*-tests applying the abovementioned non-parametric bootstrapping procedure for 95% BCa Cis.

In addition to the clinically reported (differences in) changes from baseline in PDQ-39, the mean QALYs per patient per treatment group and the mean difference were assessed using two-sample *t*-tests applying the abovementioned non-parametric bootstrapping procedure for 95% BCa Cis.

Data cleaning

Source validation was performed by independent monitoring and related queries were solved by DvP. Inconsistencies in patient reported data on outcomes and use of resources were solved and registered by syntax (MvB). If appropriate, skipped questionnaire items due to routing were set to a zero score, again registered by syntax (MvB). Reported counts of resource use by follow-up time were checked for crossing acceptable ceiling levels (*e.g.*, 77 hours of informal care or 77 hours of work a week, allowing for 7 * 11 hours of rest a day plus 14 hours anywhere in between to have at least one consecutive period of 36 hours rest a week).[33]

Handling of missing data

Missing PDQ-39 outcome data were handled in accordance with the clinical analyses based on separate treatment groups (see page 10). Patients who died were assumed to generate zero QALYs and zero costs beyond the time of death.

For missing health utility data over time among patients alive and for missing data on resource use during follow-up the choice of imputation approach depended on the observed missing data pattern and whether or not missingness was associated with baseline variables or previously observed outcomes. Missingness of data (completely) at random was handled with multiple imputation as described above (see page 10).

In a sensitivity analysis potential missingness of data not at random was addressed to account for the expectation by the treating clinicians that missingness of data may well depend on a worsened health status (see also page 10). The following scenarios were addressed in this sensitivity analysis by shifting the costs and health outcomes from the missingness (completely) at random imputation:

a) increasing actually imputed costs by 20% and 40%

b) increasing actually imputed PDQ-39 summary index scores changes by 20% and 40%

c) decreasing actually imputed QALYs by 20% and 40%

d) combining a) and b)

e) combining a) and c).

Findings of the clinical analyses might have suggested to explore additional scenarios for the sensitivity of the economical evaluation. The higher its impact, the more extensively the sensitivity analysis was reported in the main text, with the remainder reported in the Supplementary Material.

The handling of missing data applied recommendations by Faria et al.[34]

Cost-effectiveness

Incremental analysis

Differences in costs and differences in changes in PDQ-39 from baseline between the treatment groups were used to calculate the incremental cost-effectiveness ratio (ICER), again following non-parametric bootstrapping drawing 2,500 samples of the same sizes as the original samples and with replacement to account for sampling variability. Likewise, differences in costs and differences in QALYs were used to calculate the incremental cost-utility ratio (ICUR). Differences in costs and the associated differences in health outcomes were plotted with cost-effectiveness planes,[35] for example see Figure EE-1.

For each of the 2,500 bootstraps the net health benefits (NHB) of CLI versus DBS were calculated based on the ICERs for different levels of willingness-to-pay up to €3,000 per additional unit decrease in change in PDQ-39 ($NHB_{PDQ\text{-}39}$), respectively based on the ICURs for willingness-to-pay levels up to €160,000. Cost-effectiveness acceptability curves were plotted to represent uncertainty,[36] showing the probability of CLI being cost-effective in comparison with DBS at each level of willingness to pay, see Figure EE-2.

Subgroup analyses

Considering the limited patient numbers, potentially uneven distributions of baseline characteristics despite randomization were exploratory addressed by subgroup analyses for hypothesis generation. Also, a preplanned subgroup analysis by median age was performed to detect a potential influence of being more vulnerable at a higher age.

Additional sensitivity analyses

Apart from addressing the uncertainty of missing data structures, additional sensitivity analyses were performed for the lengths of the life cycles of the DBS and CLI equipment and corresponding depreciation period (6 to 8 years in steps of 1 year versus 5 years).

Scenario analyses

A scenario analysis for the ICUR was performed to correct for health utility scores at baseline by subtracting for each patient the baseline score from the calculated QALYs during the follow-up of 12 months. Another scenario analysis was done to report the ICUR from an international perspective by applying the time trade-off based on valuation of health states based on preferences from the UK general population to the EQ-5D-3L health status profiles and conversion of Euros into English Pound by applying purchasing power parities from the Organisation of Economic Cooperation and Development (or 0.688 divided by 0.772 in 2020).[37,38]

*Table EE2a. Mean per patient use of hospital resources by strategy**

	CLI (n=)	**DBS** (n=)	**Difference** CLI vs DBS**
	Mean (95% BCa CI)	**Mean** (95% BCa CI)	**Mean** (95% BCa CI)
Screening for suitability			
preop. outpatient consultation			
inpatient stay			
on-/off-phase evaluation			
evaluation / nasogastric tube		NA	
vitamin examinations			
other laboratory			
preop. anesthetic evaluation			
EKG			
MRI-brain	NA		
Continuous Levodopa Infusion			
CLI-pump			
CLI-cassettes			
PEG-jejunostomy placement			
tube replacement			
Deep Brain Stimulation			
(re)implantation			
electrode replacement			
extension cable replacement			
battery exchanges			
Other hospital care			
Emergency department visit			
Hospital stay (in days)			
neurology ward			
psychiatric ward			
intensive care unit			
other			
Outpatient consultation			
neurologist			
CLI/DBS-nurse			
neurosurgeon			
gastroenterologist			
PEG-nurse			
psychiatrist			
(neuro)psychologist			
Imaging diagnostics			
MRI brain			
CT brain			
X-ray DBS leads			
CT abdomen			
ultrasonography abdomen			
Therapeutics			
ex-plantation (battery, lead,…)			
related medication			

BCa CI: bias corrected and accelerated confidence interval, CLI: continuous levodopa infusion, CT: computed tomography, DBS: deep brain stimulation, EKG: electrocardiogram, MRI: magnetic resonance imaging, NA: not applicable, PEG: percutaneous endoscopic gastrostomy

*P-values not reported for use of resources

**Differences rounded to zero keep their sign

*Table EE2b. Mean per patient use of non-hospital resources and productivity loss by strategy**

	CLI (n=)	**DBS** (n=)	**Difference** CLI vs DBS**
	Mean (95% BCa CI)	**Mean** (95% BCa CI)	**Mean** (95% BCa CI)
Non-hospital care			
Medication			
CLI-cassettes			
related medication			
Out-of-hospital consultation			
general practitioner			
occupational therapist			
physiotherapist			
speech therapist			
dietician			
social worker			
psychologist			
rehabilitation physician			
company physician			
Other institutional stay			
nursing home day			
rehabilitation center day			
Formal home care			
house holding			
personal			
home nursing			
Equipment			
walking devices			
other living aids			
Productivity loss in hours			
absenteeism			
presenteeism			

BCa CI: bias corrected and accelerated confidence interval, CLI: continuous levodopa infusion, DBS: deep brain stimulation

* P-values not reported for use of resources

** Differences rounded to zero keep their sign

Table EE3. Mean per patient costs by strategy

	CLI (n=)	**DBS** (n=)	**Difference* CLI vs DBS**	**P-value**
	Mean costs (95% CI)	**Mean costs** (95% CI)	**Mean costs** (95% CI)	
Hospital care				
Screening for suitability				
Continuous Levodopa Infusion				
Deep Brain Stimulation				
Emergency department visit				
Hospital in patient stay				
Outpatient consultation				
Diagnostics				
Therapeutics				
Non-hospital care				
Medication				
Out-of-hospital consultation				
Other institutional stay				
Formal home care				
Equipment				
Out-of-pocket expenses				
Productivity loss				
Total health care costs				
Total societal costs				

CI: confidence interval, CLI: continuous levodopa infusion, DBS: deep brain stimulation

*Differences rounded to zero keep their sign

References

1. Van Poppelen D, Sisodia V, De Haan RJ, et al. Protocol of a randomized open label multicentre trial comparing continuous intrajejunal levodopa infusion with deep brain stimulation in Parkinson's disease-the INfusion VErsus STimulation (INVEST) study. BMC Neurol 2020; 20(1):40.
2. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord 2008; 23(15):2129-70.
3. Torgerson D, Sibbald B. Understanding controlled trials: What is a patient preference trial? BMJ 1998;316(7128):360.
4. Deuschl G, Schade-Brittinger C, Krack P, et al. A randomized trial of deep-brain stimulation for Parkinson's disease. N Engl J Med 2006;355(9):896–908.
5. Schuepbach W M M , J Rau, K Knudsen, J Volkmann, P Krack, L Timmermann, T D Hälbig, H Hesekamp, S M Navarro, N Meier, D Falk, M Mehdorn, S Paschen, M Maarouf, M T Barbe, G R Fink, A Kupsch, D Gruber, G-H Schneider, E Seigneuret, A Kistner, P Chaynes, F Or ESG. Neurostimulation for Parkinson's Disease with Early Motor Complications. N Engl J Med 2013;368(21):2037–8.
6. Fernández-Pajarín G, Sesar Á, Ares B, et al. Continuous Subcutaneous Apomorphine Infusion before Subthalamic Deep Brain Stimulation: A Prospective, Comparative Study in 20 Patients. Mov Disord Clin Pract 2021; 8(8):1216-1224.
7. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: Lessons for intention-to-treat analysis. Clin Trials 2012; 9(4):396-407.
8. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. Qual Life Res 1995;
9. Hagell P, Widner H. Clinical rating of dyskinesias in Parkinson's disease: Use and reliability of a new rating scale. Mov Disord 1999; 14(3):448-55.
10. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. Mov Disord 2010;25(15):2649–53.
11. Weisscher N, Post B, De Haan RJ, Glas CAW, Speelman JD, Vermeulen M. The AMC Linear Disability Score in patients with newly diagnosed Parkinson disease. Neurology 2007; 69(23):2155-61.
12. Pagonabarraga J, Kulisevsky J, Llebaria G, García-Sánchez C, Pascual-Sedano B, Gironell A. Parkinson's disease-cognitive rating scale: A new cognitive scale specific for Parkinson's disease. Mov Disord 2008; 23(7):998-1005.
13. Starkstein SE, Mayberg HS, Preziosi TJ, Andrezejewski P, Leiguarda R, Robinson RG. Reliability, validity, and clinical correlates of apathy in Parkinson's disease. J Neuropsychiatry Clin Neurosci 1992; 4(2):134-9.
14. Weintraub D, Hoops S, Shea JA, et al. Validation of the questionnaire for impulsive-compulsive disorders in Parkinson's disease. Mov Disord 2009; 24(10):1461-7.
15. Sheehan D, Lecrubier Y, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. Eur Psychiatry 1997;12:224-231.
16. Bhatt A. Protocol deviation and violation. Perspect Clin Res 2012; 3(3):117.
17. European Medicines Agency. ICH guideline E3 - questions and answers (R1). 2012.
18. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol 2014; 14:75.
19. Simuni T, Luo ST, Chou KL, Fernandez H, He B, Parashos S. Rankin scale as a potential measure of global disability in early Parkinson's disease. J. Clin. Neurosci. 2013; 20(9):1200-3.
20. Peto V, Jenkinson C, Fitzpatrick RAY. Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. Age Ageing 2001; 3 0(4):299-302.
21. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. Value Heal 2022;25(1):3–9.
22. Bouwmans, C Hakkaart-van Roijen, L Koopmanschap, M Krol, M Severens, H Brouwer W. iMTA Medical Consumption Questionnaire. Rotterdam: Erasmus University Rotterdam, 2013a.
23. Hakkaart-van Roijen, L Van der Linden, N Bouwmans CK et al. Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare in Dutch: Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de

gezondheidszorg]. Rotterdam: Erasmus University Rotterdam, 2015.
24. Eenmanierom. Tool for calculating distance. [Internet]. 2022 [cited 2022 Mar. 18]. Available from: https://www.afstandberekenen.nl
25. Bouwmans, C Hakkaart-van Roijen, L Koopmanschap, M Krol, M Severens, H Brouwer W. iMTA Productivity Costs Questionnaire. Rotterdam: Erasmus University Rotterdam, 2013b.
26. Zorginstituut Nederland. Data source for unit costs of prescribed drugs. [Internet]. 2022 [cited 2022 Mar. 18]. Available from: https://www.farmacotherapeutischkompas.nl
27. Zorginstituut Nederland. Data source for yearly general consumer price. [Internet]. 2022 [cited 2022 Mar. 18]. Available from: https://www.medicijnkosten.nl
28. Koopmanschap MA, Rutten FFH. A practical guide for calculating indirect costs of disease. Pharmacoeconomics. 1996;406–66.
29. Centraal Bureau voor de Statistiek. Data source for length of friction cost period. [Internet]. 2021 [cited 2021 Dec. 28]. Available from: https://opendata.cbs.nl/statline#/CBS/nl/dataset/80472ned/table?ts=1640699996905
30. Centraal Bureau voor de Statistiek. Data source for yearly general consumer price indices. [Internet]. 2022 [cited 2022 Mar. 18]. Available from: https://www.cbs.nl/nl-nl/cijfers/detail/83131NED
31. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York) 1990;
32. Lamers LM, Stalmeier PFM, McDonnell J, Krabbe PFM, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd 2005;(1491574–8).
33. Rijksoverheid. Data source for ceiling levels work hours. [Internet]. 2022 [cited 2022 Mar. 18]. Available from: https://www.rijksoverheid.nl/onderwerpen/persoonsgebonden-budget-pgb/vraag-en-antwoord/wettelijke-regels-werktijden-en-rusttijden
34. Faria R, Gomes M, Epstein D, White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. Pharmacoeconomics 2014; 32(12):1157-70.
35. Black WC. The CE Plane: A Graphic Representation of Cost-Effectiveness. Med Decis Mak 1990; 10(3):212-4.
36. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: The role of cost-effectiveness acceptability curves. Health Econ 2001;(10):779–87.
37. Dolan P. Modeling Valuations for EuroQol Health States. Med Care 1997;(35):1095–108.
38. Organisation for Economic Co-operation and Development. Purchasing power parities (indicator). [Internet]. 2021 [cited 2021 Dec. 28]. Available from:https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

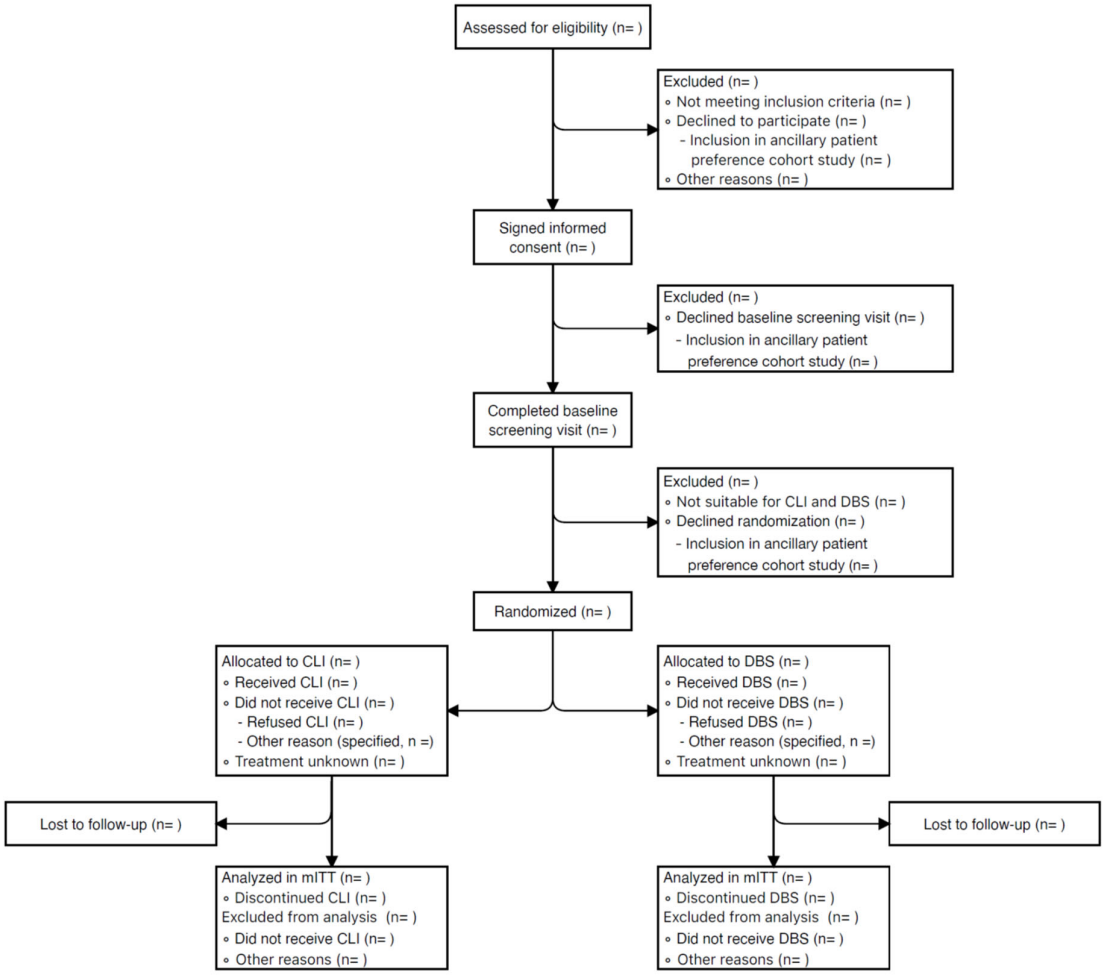


Figure SAP-1. Flowchart Randomized controlled Trial

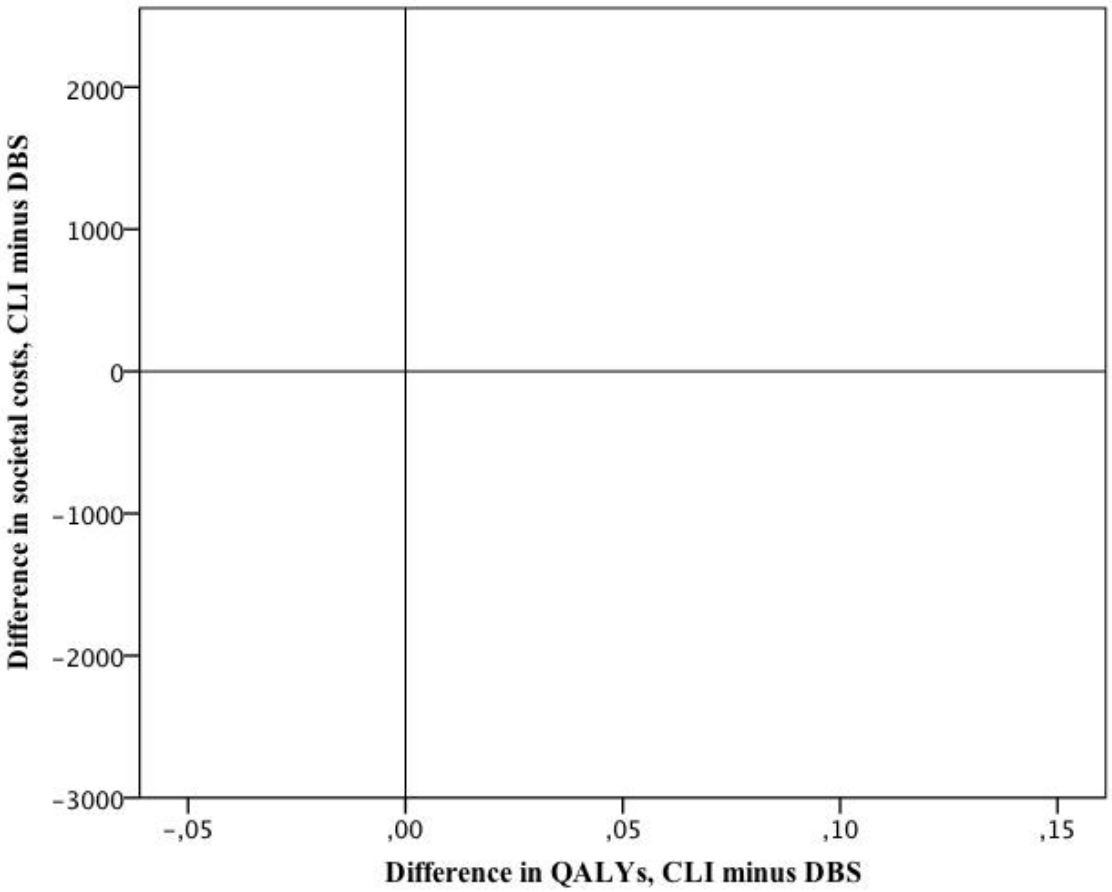


Figure EE-1. Cost-effectiveness plane for differences in costs versus differences in QALYs for CLI in comparison with DBS

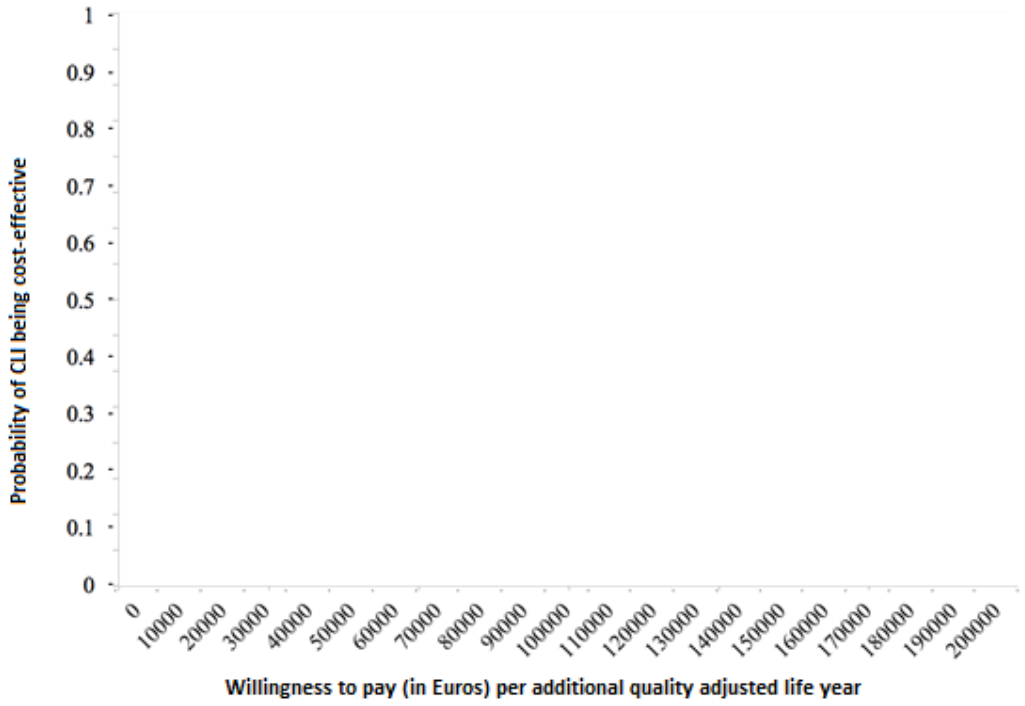


Figure EE-2. Probability of CLI being cost-effective in comparison with DBS at different levels of willingness to pay per QALY

Supplementary information

Appendix I — Possible protocol deviations and violations

Deviations

Usage of other DBS devices

e.g., usage of material from other manufacturer than Medtronic for DBS material. Although usage of material from Medtronic was described in the protocol, usage of material from other manufacturers was not expected to have a significant effect on treatment efficacy. Furthermore, the study protocol emphasized that treatments were to be given according to usual care.

Usage of other routes of CLI administration

e.g., if it proved not possible to construct a percutaneous endoscopic gastrostomy (PEG), a percutaneous radiological gastrostomy (PRG) was considered a feasible alternative by the treating physician. In line with the previous paragraph, this was deemed not to have a significant impact on treatment effect but the risk for AEs may be altered.

Missing of visits at two weeks, three, six or nine months

Since primary outcomes all were based on the assessment from the 12-month follow-up visit, missing of interim visits, although undesirable, was deemed not to have a significant impact on analyses.

Completing baseline visit after initiation of treatment (in ancillary patient preference observational arms)

During the course of the study, sometimes it proved challenging to contact patients and receive all mail-in CRFs patients had to fill in, prior to initiation of therapy in participating centers other than the primary center. If the baseline assessments were completed within reasonable time after initiation of therapy (*i.e.*, less than two weeks), it was estimated that these patients were still able to report the situation prior to therapy (as requested).

Failure to initiate follow-up after signing informed consent (in ancillary patient preference observational arms)

In case it proved impossible to plan and compete the baseline visit and subsequent visits due to logistic reasons after signing informed consent, this resulted in failure of study participation. Since the missing patients would have been at random, this probably had no influence on the integrity of analyses and was therefore marked as deviation from the protocol.

Violations

Refusal of allocated treatment

Since DBS treatment through usual care was restricted with a long waiting list, it was expected that some patients hoped to receive this treatment sooner through study participation. It was an anticipated unwanted scenario that some patients therefore would decline allocated treatment after randomization, because of a strong preference for the other treatment. This would be considered a protocol violation and made patients ineligible for the modified intention to treat population.

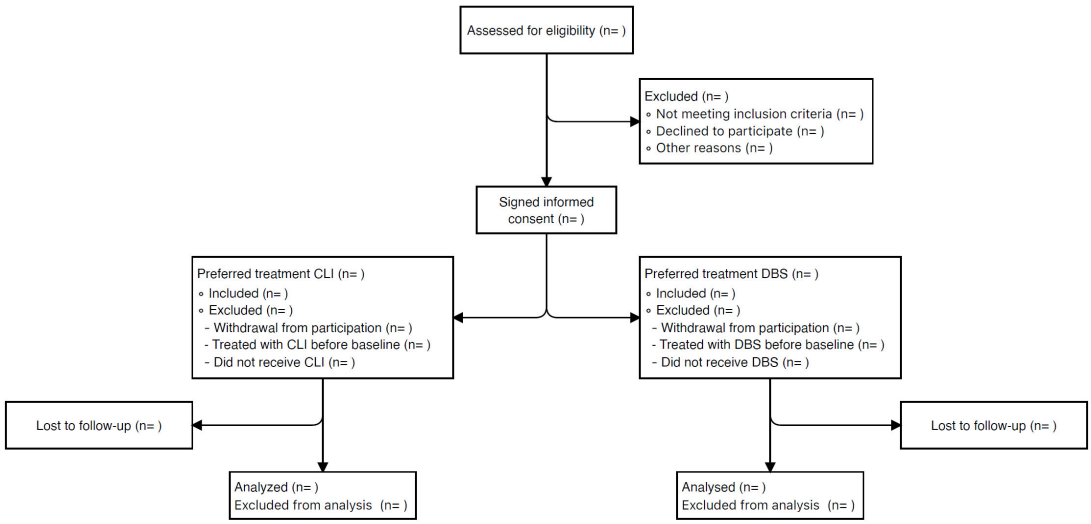
Missing of 12-month follow-up visit

Missing the 12-month visit would have resulted in significant incompleteness of data. Since this would probably not have been at random, this could have affected integrity of analyses. Therefore, this visit was deemed indispensable and missing it was considered a protocol violation.

Migration from the RCT to ancillary patient preference observational study population

If a patient found participation in the RCT too burdensome after receiving allocated treatment, continuation of study participation was encouraged. If a patient crossed-over to the ancillary patient preference arm, this was deemed a violation from the RCT protocol.

Appendix II — Flowchart ancillary patient preference observational study



Appendix III — Supplementary statistical analysis tables

Table SAP1-S. Baseline characteristics – RCT mITT population and ancillary patient preference observational study arms

	RCT all – all randomized patients (CLI or DBS)	**Observational CLI – patients chose CLI treatment**	**Observational DBS – patients chose DBS treatment**	**Difference RCT all vs observational CLI**	**Difference RCT all vs observational DBS**
	n	n	n	difference (95% CI)	difference (95% CI)
Age (years)					
Sex, F (%)					
Time since PD diagnosis (years)					
Hoehn and Yahr stage in on-drug state (%) - *Stage 1* - *Stage 1,5* - *Stage 2* - *Stage 2,5* - *Stage 3* - *Stage 4*					
Levodopa-equivalent daily dose (mg/day)					
Level of experience of treating center with CLI and DBS treatment (%)* - *Experienced in CLI, not in DBS* - *Experienced in DBS, not in CLI* - *Experienced in both CLI and DBS* - *Little experience in both CLI and DBS*					

CLI: continuous intrajejunal levodopa infusion, DBS: deep brain stimulation, F: female, MDS-UPDRS: Movement Disorder Society's Unified Parkinson's Disease Rating Scale, mg: milligrams, NA: not available, PD: Parkinson's disease, RCT: randomized controlled trial

*Experienced in CLI/DBS: ≥ 5 patients treated per year in the two years previous to including the patient with the respective treatment in the treating center. A center may have changed in level of experience over time.

For normally distributed data, mean and standard deviation were presented, for non-normally distributed data median and interquartile ranges. For dichotomous variables, the percentage stated.

Appendix IV — Supplementary health economics tables

Table EE1-S. Dutch unit costs (€) for resources used

Resource	Unit	Unit costs in 2020 euros	Source
Screening for suitability			
preop. outpatient consultation	visit		DMC
inpatient stay	day		DMC
on/off-phase evaluation	procedure		hospital ledger
evaluation / nasogastric tube	procedure		passer-by rate
vitamin examinations	procedure		passer-by rate
other laboratory	procedure		passer-by rate
preop. anesthetic evaluation	procedure		hospital ledger
EKG	procedure		passer-by rate
MRI-brain	procedure		DMC
Infusion			
CLI-pump	device*		hospital ledger
CLI-cassettes	cassette		hospital ledger
PEG-jejunostomy placement	procedure*		passer-by rate
tube replacement	procedure*		passer-by rate
Stimulation			
(re)implantation	procedure*		passer-by rate
electrode replacement	procedure		hospital ledger
extension cable replacement	procedure		passer-by rate
battery exchanges	exchange		hospital ledger
Other hospital care			
Emergency department visit	visit		DMC
Hospital stay (in days)			
neurology ward	day		DMC
psychiatric ward	day		DMC
intensive care unit	day		DMC
other	day		DMC
Outpatient consultation			
neurologist	visit		DMC
CLI/DBS-nurse	visit		hospital ledger
neurosurgeon	visit		DMC
gastroenterologist	visit		DMC
PEG-nurse	visit		hospital ledger
psychiatrist	visit		DMC
(neuro)psychologist	visit		DMC
Imaging diagnostics			
MRI brain	procedure		DMC
CT brain	procedure		DMC
X-ray DBS leads	procedure		hospital ledger
CT abdomen	procedure		passer-by rate
ultrasonography abdomen	procedure		passer-by rate
Therapeutics			
explantation (battery, lead, ...)	procedure		passer-by rate
related medication	medication day		Pharmaceutical compass

Table EE1-S continued. Dutch unit costs (€) for resources used

Resource	Unit	Unit costs in 2020 euros	Source
Non-hospital care			
Medication			
CLI-cassettes	cassettes		Pharmaceutical compass
related medication	medication days		Pharmaceutical compass
Out-of-hospital consultation			
general practitioner	visit		DMC
occupational therapist	visit		DMC
physiotherapist	visit		DMC
speech therapist	visit		DMC
dietician	visit		market price
social worker	visit		DMC
psychologist	visit		DMC
rehabilitation physician	visit		DMC
company physician	visit		DMC
Other institutional stay			
nursing home day	day		DMC
rehabilitation center day	day		DMC
Formal home care			
house holding	hour		DMC
personal	hour		DMC
home nursing	hour		DMC
Equipment			
walking devices	device		market price
other living aids	device		market price
Out-of-pocket expenses	mean monthly costs		
over-the-counter drugs	mean monthly costs		as reported
informal care	mean monthly costs		as reported
non-reimbursable devices	kilometer/mode		as reported
health related travel	transport		DMC
expenses informal caregiver	hour		DMC
Productivity loss in hours			
absenteeism	hour		DMC
presenteeism	hour		DMC

BCa CI: bias corrected and accelerated confidence interval, CLI: continuous levodopa infusion, CT: computed tomography, DBS: deep brain stimulation, DMC: Dutch Manual on Costing in health care research, EKG: electrocardiogram, MRI: magnetic resonance imaging, NA: not applicable, PEG: percutaneous endoscopic gastrostomy

*The time horizon

Table EE3a-S. Mean per patient costs of hospital resources by strategy

	CLI (n=)	**DBS** (n=)	**Difference* CLI vs DBS**
	Mean costs (95% CI)	**Mean costs** (95% CI)	**Mean costs** (95% CI)
Screening for suitability			
preop. outpatient consultation			
inpatient stay			
on/off-phase evaluation			
evaluation / nasogastric tube		NA	
vitamin examinations			
other laboratory			
preop anesthetic evaluation			
EKG			
MRI-brain	NA		
Infusion			
CLI-pump			
CLI-cassettes			
PEG-jejunostomy placement			
tube replacement			
Stimulation			
(re)implantation			
electrode replacement			
extension cable replacement			
battery exchanges			
Other hospital care			
Emergency department visit			
Hospital stay (in days)			
neurology ward			
psychiatric ward			
intensive care unit			
other			
Outpatient consultation			
neurologist			
CLI/DBS-nurse			
neurosurgeon			
gastroenterologist			
PEG-nurse			
psychiatrist			
(neuro)psychologist			
Imaging diagnostics			
MRI brain			
CT brain			
X-ray DBS leads			
CT abdomen			
ultrasonography abdomen			
Therapeutics			
explantation (battery, lead, ...)			
related medication			

CI: confidence interval, CLI: continuous levodopa infusion, CT: computed tomography, DBS: deep brain stimulation, EKG: electrocardiogram, MRI: magnetic resonance imaging, NA: not applicable, PEG: percutaneous endoscopic gastrostomy

* Differences rounded to zero keep their sign.

Table EE3b-S. Mean per patient costs of out-of-hospital resources, out-of-pocket expenses and productivity loss by strategy

	Continuous intrajejunal levodopa infusion (CLI) (N=)	**Deep brain stimulation (DBS)** (N=)	**Difference* CLI vs DBS**
	Mean costs (95% CI)	**Mean costs** (95% CI)	**Mean costs** (95% CI)
Out-of-hospital care			
Medication			
CLI-cassettes			
related medication			
Out-of-hospital consultation			
general practitioner			
occupational therapist			
physiotherapist			
speech therapist			
dietician			
social worker			
psychologist			
rehabilitation physician			
company physician			
Other institutional stay			
nursing home day			
rehabilitation center day			
Formal home care			
house holding			
personal			
home nursing			
Equipment			
walking devices			
other living aids			
Out-of-pocket expenses			
over-the-counter drugs			
informal care			
non-reimbursable devices			
health related travel			
expenses informal caregiver			
Productivity loss in hours			
absenteeism			
presenteeism			

* Differences rounded to zero keep their sign.