The SPARTACUS Trial: Controversies and Unresolved Issues

Authors
Felix Beuschlein1, 2, Paolo Mulatero3, Evelyn Asbach1, Silvia Monticone3, Cristiana Catena4, Leonardo A. Sechi4, Michael Stowasser5

Affiliations
1 Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, LMU, Munich, Germany
2 Klinik für Endokrinologie, Diabetologie und Klinische Ernährung, Universitätsklinikum Zürich, Zürich, Switzerland
3 Division of Internal Medicine and Hypertension, Department of Medical Sciences, University of Torino, Torino, Italy
4 Clinica Medica and Chair of Internal Medicine, Department of Medicine, University of Udine, Udine, Italy
5 Endocrine Hypertension Research Centre, University of Queensland Diamantina Institute, Greenslopes and Princess Alexandra Hospitals, Brisbane, Australia

ABSTRACT
Adrenal vein sampling (AVS) is considered the gold standard for the differential diagnosis in patients with primary aldosteronism (PA). The distinction between unilateral and bilateral disease dictates the targeted therapeutic approach with surgery for aldosterone-producing adenomas and medical therapy for patients with bilateral hyperplasia. Thereby, this diagnostic step is crucial in clinical care. As AVS is an invasive, not well-standardized procedure that is restricted to few specialized centers, several attempts have been made to simplify diagnostic algorithms. In this clinical scenario, the recently published SPARTACUS trial aimed at answering the question whether AVS in fact is superior for differential diagnosis in comparison to imaging of the adrenal glands. In this multicenter study, patients were randomized to be treated according to AVS results or based on abdominal imaging only. Clinical outcome in both patient groups after one year was reported as not different. While the study results found broad interest, it also stirred considerable controversies. This review provides an overview on the different views regarding the outline of the SPARTACUS trial and the interpretation of its results.

Introduction
Primary aldosteronism (PA) represents the most frequent curable form of arterial hypertension [1–3]. Its relevant cardiovascular comorbidities, with potential reversal or improvement after early diagnosis and therapy, justify a screening in high-risk populations [4–10]. These include patients with resistant hypertension, spontaneous or diuretic-induced hypokalemia, adrenal incidentaloma, or sleep apnea [11]. The two major forms of PA are unilateral aldosterone-producing adenoma (APA), which is treated by surgery, and the more common bilateral adrenal hyperplasia, which is addressed by mineralocorticoid receptor antagonist (MRA) treatment [12, 13].

As therapeutic approaches diverge substantially between the two subtypes, differential diagnosis is essential for patient care [14, 15]. Following the establishment of the diagnosis of PA, current guidelines suggest the performance of adrenal computed tomography (CT), which allows to exclude adenocortical carcinoma as a rare cause of aldosterone excess [11, 16]. While CT scans have a high sensitivity in detecting larger adrenal tumors, they only display a limited specificity for endocrine active lesions due to the high prevalence.
of adrenal incidentalomas with increasing age [17]. Accordingly, bilateral adrenal vein sampling (AVS) is currently recommended as the gold standard in differential diagnosis in PA patients willing to undergo adrenalectomy (ADX) in case of unilateral disease. The rationale of this procedure is to measure aldosterone directly at its off-spring in the effluents of the adrenal veins. A systematic review had pointed out a discordance between CT- and AVS-based diagnosis in approximately 40% of the cases [17]. While being safe in experienced hands, AVS has been judged as invasive, technically demanding, and relatively expensive [18, 19]. In many centers, success rates are low [20]. A relevant proportion of patients not treated in specialized centers is deprived of this diagnostic tool due to lacking resources. Additionally, AVS is not well standardized between centers regarding selectivity and lateralization indices, sequential or simultaneous catheterization and the use of cosynthire, thereby complicating comparability of diagnosis findings. Finally, it has been criticized, that prospective studies for the reliability of AVS results are lacking. As a consequence, efforts have been made to avoid AVS at least in some patients based on different algorithms and scores that had been demonstrated to provide some level of subtype prediction [21–23]. Alternative functional imaging techniques including metomidate PET-CT or the use of specific aldosterone synthase tracer are currently under investigation [24, 25].

In this clinical scenario, the SPARCTACUS trial (Subtyping Prima- ry Aldosteronism: A Randomized Trial Comparing Adrenal Vein Sampling and Computed Tomography Scan) recently reported by Dekkers and colleagues in the Lancet Diabetes & Endocrinology, set out to compare AVS-based and CT-based treatment outcome in PA patients [26]. The study reported that treatment decisions based on both diagnostic tools led to similar blood pressure improvement and health-related quality of life in PA patients at one year of follow-up. Therefore, the authors postulated that the extra-costs of AVS were dispensable and that neither AVS nor CT would correctly predict PA subtypes in all cases. Due to its unexpected findings with potentially relevant and practical consequences for patient care, publication of the study triggered extensive discussions. The current review aims to outline the different points of view regarding the protocol and the results of the study.

Protocol PRO

The primary endpoint of the SPARCTACUS study was defined as blood pressure control at one year of follow-up [26]. Indeed, the aim to improve clinical outcome represents the real gold standard of diagnostic or therapeutic procedures. As AVS is being criticized to be unavailable for the majority of PA patients due to its complexity and restriction to specialized centers, simplified algorithms with improved accessibility are required. The implementation of CT scanning instead of AVS, at least in a certain proportion of PA patients, would ease the diagnostic work-up in this patient population.

For the completion of this goal, it is important to provide clinical paths that are based on the best scientific evidence, that is, randomized controlled trials. In fact, the SPARCTACUS study followed a diagnostic, randomized, controlled, multicenter design. PA was confirmed by accepted clinical standards including a salt-loading test. The sample size was determined prior to the start of the study by power calculations: 200 patients with PA were randomly assigned using a web-based algorithm to receive AVS (preceded by CT) or only CT to subtype PA. Randomization resulted in a well-balanced distribution of patients between the two groups. Patients randomized to CT were adrenalectomized in case of a unilaterally enlarged adrenal gland and a normal appearing contralateral gland. In contrast, patients with normal or bilaterally enlarged adrenal glands were treated with mineralocorticoid antagonists. For better comparison, adrenal CT was assessed by a local radiologist and reviewed by a central facility. In cases of discrepancy between these CT readings, the final decision was taken by the local center.

AVS performance is variable across centers in term of procedure and interpretation [27, 28]. In the SPARCTACUS study, AVS was performed according to accepted protocols under continuous cosyn- tropin stimulation with sequential catheterization of the adrenal veins. Based on a selectivity index of 3.0 or higher, the success rate for bilateral cannulation of the adrenal veins was reported as high as 96%. Patients with unilateral disease (based on a lateralization index of 4.0 or higher and a suppression index of 1.0 or lower) were adre- nallectomized. In cases of bilateral disease, mineralocorticoid antigen treatment was initiated. In those instances when AVS failed, patients were treated according to CT findings. Thereby, the applied criteria for the interpretation can be regarded as strict enough to ensure rigorous diagnosis. In two large studies, ACTH stimulation resulted in 1–4/46 difference in diagnosis compared with unstimulat- ed procedures [29, 30]. Therefore, it is highly improbable that this could have an impact on the final outcome of the study.

From initially 200 enrolled patients, 184 completed follow-up with even distribution in the AVS and CT group. Patients were investigated according to an intention-to-diagnose analysis for the primary endpoint at 12 months following therapy. In fact, this interval has recently been endorsed by an international expert panel as a relevant time point for re-assessment of clinical outcome in PA patients [31]. As a measure of blood pressure control, the intensity of antihypertensive treatment for obtaining target blood pressure ( < 135/85 mmHg using a semiautomatic device or < 140/90 mmHg using office measurement) was quantified as defined daily dose (DDD). This approach represents a practical end-point, as blood pressure is the parameter relevant for the patients [32]. Ambulatory blood pressure was used, which is a very objective form of monitoring. Key secondary endpoints were biochemical outcome in patients who had undergone ADX, which was analy- zed by salt-loading test. Further endpoints included physical and mental scores, the proportion of patients reaching target blood pressure, adverse events, and cost-effectiveness. The latter point is of great importance, as AVS is relatively expensive, which might not be justified if it would lack diagnostic superiority.

In summary, patients were scrutinized by a well-defined clinical protocol, allocated into diagnostic procedure groups in a rand- omized fashion and prospectively followed up. In this regard, the SPARCTACUS trial has implemented the highest standards of a clinical study design. To this point, it is the first and only in the field of clinical PA research that aims at this evidence level.

Protocol CONTRA

The SPARCTACUS trial randomized 184 subjects with PA to adrenal CT scanning or adrenal CT scanning plus AVS to establish the final sub-
type diagnosis and address the patients with a unilateral disease to surgery [26]. The authors should be commended for the performance of the first randomized trial on this disease, however the design of the SPARTACUS study has some weak points that could reduce the relevance of the results and their application in clinical practice. The most important points are: 1) DDD is not the appropriate primary end-point for this type of study; 2) the SPARTACUS cohort is not representative of the general PA population and therefore the results cannot be generalized; 3) the comparison between AVS versus CT-based MRA therapy was not necessary and reduced the power of the study; 4) the sample size of the investigated population was not adequate to prove the non-inferiority of CT scanning in comparison with AVS to determine indication for ADX.

The first consideration is that DDD is not the appropriate primary end-point. First, it does not evaluate the biochemical cure of PA, the most reliable measure of the success of ADX [31]; second, it does not take into account the concomitant presence of essential hypertension that can confound the final outcome judgment; further, the intensity of the treatment depends largely on the type of drug that is considered in the calculation. As an example, a patient with blood pressure levels of 130/80 mmHg before ADX under spironolactone 75 mg, amlopidine 5 mg, lisinopril 10 mg, and hydrochlorothiazide 12.5 mg has a DDD = 3.5; the same patient, with the same blood pressure after ADX, taking ramipril 10 mg as monotherapy, has a DDD = 4. Clinicians would unanimously consider this patient largely improved and not worsened as it appears if the DDD is used as an indicator of clinical success after unilateral ADX.

The second consideration is that the SPARTACUS cohort is not representative of the general PA population and, therefore, the results cannot be generalized. Criteria for inclusion were: hypertension requiring 3 or more antihypertensive drugs in adequate doses and/or hypertension accompanied by spontaneous or diuretic-induced hypokalemia (serum potassium < 3.5 mmol/l), which means that only patients with a severe phenotype of PA were selected. In the recent PATO (primary aldosteronism in Torino) study, performed on the general hypertensive population seen in primary care practice, the majority of PA patients would not have been included in the SPARTACUS trial because of a milder phenotype [33]. For example, in the PATO study, APA accounted for 25 % of all PA cases, whereas in the SPARTACUS their prevalence was 50 %; furthermore, hypokalemia was observed in 29 % of the patients with PA the PATO study versus 68 % in the SPARTACUS cohort. The data of the PATO study are also coherent with a retrospective evaluation of the clinical and biochemical features of patients with PA in referral centers from five continents [2]. It is conceivable that patients with a severe PA phenotype and high prevalence of APA will respond to ADX (even if AVS is not performed), but this would not be the case for patients with a mild form of the disease.

Another major flaw of the design of the study is the choice of randomizing patients to CT-based MRA therapy and AVS-based MRA therapy [26]. It is in fact well known, since the publication of the studies performed by the Cleveland Clinic group, that patients with APA respond well to MRA [34]. Similarly, a more recent study demonstrated that adequate doses of spironolactone determine a blood pressure reduction similar to that obtained with ADX [35]. Therefore, this arm of the study reduces the power of the other comparison, without providing any useful novel information.

In a systematic review/meta-analysis, Kempers et al. showed that AVS and CT scanning result in a different diagnosis in around 38 % of cases [17]. In 19 % of cases, patients would have been inappropriately excluded from ADX following CT scanning, where AVS showed unilateral secretion; however, this discrepancy would not affect the results of the SPARTACUS study. Therefore, only 18.5 % of patients (14.6 % having an inappropriate ADX when AVS showed a bilateral disease plus 3.9 % having ADX on the wrong side when AVS showed aldosterone secretion on the opposite side) would be inappropriately adrenalectomized following CT scanning instead of AVS. It should be emphasized that ADX is also effective in some selected patients with bilateral PA: in fact, it resulted in hypertension cure in 15 % of cases and improvement in another 20 % [36]. Based on this assumption, no more than 6 to 9 patients are expected to have persistence of PA if a cohort of 46 patients is adrenalectomized following CT scanning alone; interestingly, the persistence of PA in the SPARTACUS study is 9/46.

In the primary aldosteronism surgery outcome (PASO) study, AVS resulted in a complete biochemical cure of PA in 94 % of the patients [31]. Using the expected rate of cure of ADX following the indication of CT scanning versus AVS, a number of 258 patients would have been necessary instead of 46, to prove the non-inferiority of CT scanning with respect to AVS for the diagnosis of unilateral PA.

In conclusion, the SPARTACUS trial conveys the strong message that in patients with PA ADX based on CT diagnosis has a similar outcome compared with ADX based on AVS findings, thereby challenging the current Endocrine Society Guideline [11]. However, the above discussed pitfalls significantly affected the results and limited its generalizability to the whole PA population.

Results PRO

Although the clinical, cardiac, and renal outcomes of PA have been shown to be comparable in patients treated with ADX or MRA [37], differentiation between unilateral or bilateral forms of this condition is still widely considered to be essential for definition of the appropriate therapeutic choice [11]. To this purpose, different approaches have been used in the past, including CT or MRI-based imaging, adrenal scintigraphy, metomidate PET-CT, and AVS. Previous retrospective investigations pointed out a substantial discordance (more than 40 %) between AVS and CT in differentiation of unilateral from bilateral adrenal disease in PA [17]. Because of the functional information provided by AVS, this was indicated as the “gold standard” for differentiation, thereby generating the preconception that AVS is almost always right, whereas CT is frequently wrong. As a consequence, AVS has been asserted as the unavoidable cross road in the diagnostic workup recommended to the majority of patients with PA [11]. However, no demonstration of this alleged superiority of AVS over other diagnostic methods for characterization of unilateral or bilateral forms of PA could be found in the medical literature and until the publication of the SPARTACUS study [26] no prospective assessment of this issue had been done. As already stated, randomized controlled trials provide the best clinical evidence for clinical decisions and the SPARTACUS trial was the first of this kind. The study compared the outcome of CT-based management with AVS-based management in an appropriately
sized sample of patients with PA who were treated with either ADX or MRA and were followed for one year. The outcome was assessed in a intention-to-diagnose analysis and both in primary (DDD and number of antihypertensive drugs used at follow-up) and most of secondary endpoints (proportion of patients reaching target blood pressure; serum potassium; plasma aldosterone levels after salt-loading post-ADX; patients with biochemical evidence of resolved PA; health-related quality of life, physical and mental; adverse events) no significant differences were observed. The only difference was found in the mean total cost of the procedure per patient that was 60% higher in those patients who underwent AVS. Notably, primary and secondary endpoints did not differ between the CT and AVS group even when patients who were treated with surgery or MRA were analyzed separately.

Although some reasonable and also some definitely questionable critiques have been raised to the SPARTACUS study, it cannot be denied that this is the only study that has approached the issue of the validity of AVS in a prospective randomized protocol, providing a clear demonstration that if CT is not foolproof for differentiation of unilateral from bilateral forms of PA, AVS is no better. These conclusions wipe out the misconception that AVS could be considered a gold standard for definition of subtypes of PA and undermine the Manichaean view that many have of it.

The results of the SPARTACUS study should not get to surprise because data on AVS previously obtained in the top referral centers performing AVS worldwide had already pointed out at its serious limitations. In the German’s Conn registry the results of 200 AVS procedures were analyzed in two phases, retrospective and prospective, after introduction of measures designed to improve the rates of successful cannulation. The rate of success in correct collection of adrenal samples was less than one third in the retrospective phase and less than two thirds in the prospective phase [20]. Also, the rate of success was extremely variable, from 80 % to less than 30 %, depending upon the stringency of the selectivity index that had been used. Similar findings were reported in Turin where the results of AVS in 64 patients with PA who had undergone the procedure twice showed an impressive disparity in the definition of successful cannulation of adrenal veins and lateralized aldosterone secretion depending upon the stringency of the criteria that were used [38]. In this study, the rate of concordance among three different criteria used for definition of lateralized secretion was 32 % and the rate of concordance between the two procedures performed in the same patient was 35 %. To notice, lateralization as detected by AVS changed from unilateral in one side to unilateral in the contralateral side in 14 % of patients. In Paris, more than 500 AVS were retrospectively reviewed comparing the different diagnostic criteria used in 4 of the top referral labs for AVS [39]. Comparison between the lab that used the most stringent criteria with the lab that used the most lenient showed a five-fold difference in the proportion of unsuccessful procedures (18 % vs. 4 %) and a two-fold difference in the proportion of lateralized aldosterone secretion (26 % vs. 60 %). Because of the lack of standardization of AVS among referral centers consensus documents have been published by expert committees [27, 28], but if one looks carefully at them many substantial differences still can be found showing that there is very little consensus even among experts. Thus, while being relatively safe in experienced hands, the procedure of AVS is invasive, technically demanding, relatively expensive, and inadequately standardized and, in the light of the findings of the SPARTACUS study, does not seem to offer any advantage over CT in the outcome of patients treated for PA.

In summary, the findings of the SPARTACUS study strongly support the concept that AVS is not a gold standard for differentiation of PA subtypes and keep wide open the possibility to define the opportunity for unilateral ADX or, alternatively, MRA treatment with diagnostic approaches, such as CT, that have a comparable level of reliability. Needless to say, this should be always done under the guidance of a balanced clinical judgment that takes into account all information on each single patients that is the first ingredient, beyond and above guidelines, for taking appropriate clinical decisions.

Results CONTRA

The investigators for the SPARTACUS Trial are to be congratulated for completing the first prospective study comparing CT- versus AVS-guided treatment of patients with PA, an impressive achievement given the large amount of planning, workload and funding support that would have been required. However, careful examination of the results reveals a number of anomalous findings that raise serious concerns about the validity and generalizability of the data and the conclusions that have been drawn. Furthermore, there are clear trends towards superiority of AVS that add weight to the argument provided above that the power of the study was insufficient to show significant differences between the two study groups in terms of treatment (and in particular, ADX) outcomes.

There are a number “odd” findings in SPARTACUS:

(1) The rate of lateralization in the CT and AVS groups was exactly the same at 50 %, which is in sharp contradiction to the reports of previous studies in which centers relying on CT-based subtype differentiation found much lower rates of detection of APA than those employing AVS [2]. One potential explanation for this is the very permissive criteria used for lateralization on CT, requiring only an enlargement (not even a mass lesion) of an adrenal, defined as a thickness of 7 mm or more in the body or limb. Surely the investigators are not suggesting that, with all we have learned through countless previous studies about the unreliability of even a mass lesion on CT, this is sufficient to warrant proceeding to surgery without any other supporting evidence of lateralization whatever? Another possible explanation for this anomalous result is selection bias towards subjects with more florid (and hence more likely unilateral) varieties of PA, leading to an over-representation of patients with larger APAs more easily detectable by CT, which is supported by other lines of evidence outlined below.

(2) A 50 % rate of lateralization is high even for AVS-based subtype differentiation when compared with most other recent reports [2, 40] with the exception of some Asian cohorts [41, 42], and centers which use very permissive lateralization criteria [38, 43] (which the SPARTACUS investigators did not). This again suggests selection bias.

(3) Moreover, the high proportion of hypokalemic patients (>60 %, where most other studies report hypokalemia in the
minority) among the SPARTACUS cohort is further evidence for selection of more severe, advanced PA, which may have been easier to localize by CT.

(4) Whereas most PA cohorts show a roughly equal gender distribution, in SPARTACUS, males made up over three-quarters of PA patients. This inexplicable result again hinders generalization of SPARTACUS findings to other centers.

(5) There was a much higher localization of “APA” to the left (76% vs. right 24%) in the CT group compared to the AVS group (54% vs. 46%) and to PA cohorts in other studies. This may, at least in part, be due to nearby splenic vessels which can be mistaken for APAs. Whatever the reason, it does not bode well for the validity of CT lateralization.

(6) The overall very low rate of HT cure (14 of 92 – 15%) among the ADX patients is also unexplained. This mirrors the relatively poor outcomes of the same Netherlands and Polish centers that contributed to the recently reported PASO study when compared with almost every other center that participated [31]. Not only does this raise further concerns about generalizability, but it also would have seriously impacted on the power of the study to show differences in cure rates between the CT- and AVS-based treatment groups (see below).

(7) Given that the real “proof of the pudding” in terms of attempts at lateralization is in the response to ADX, it is uncertain why the investigators even bothered to look at responses to MRA. But even there, the surprising finding that non-lateralizing patients in the AVS group need more DDDs than those in the CT group (median 5.7 vs. 4.0; \( p = 0.05 \)) defies logical explanation and casts doubt on the effectiveness of randomization.

Notwithstanding the many anomalous findings and concerns about selection bias and generalizability raised above, there were still several important findings in SPARTACUS that argued against a CT-based approach but which were largely left unmentioned by the authors:

(1) In the CT-based treatment group, consensus could not be reached regarding lateralization in a sizable proportion of patients (11 of 98), almost all of whom were assigned to MRA, whereas AVS was unsuccessful in permitting a diagnosis in only four of 96.

(2) In keeping with an enormous body of existing data, a full 50% of the 90 patients from the AVS group who had both conclusive CT and AVS demonstrated discordant results between the two procedures.

(3) Most importantly, despite the very low hypertension cure rate observed following ADX, there was still a strong trend towards a superior cure rate among the AVS-based treatment group (22%) compared with the CT group (9%) which almost reached statistical significance (\( p = 0.08 \)). Had the study been powered to examine this (rather than DDDs) as the primary endpoint, as has been recommended by the PASO investigators [31], it is highly likely that a significant difference would have been observed. As it is, with such a low overall rate of hypertension cure, SPARTACUS was clearly seriously underpowered.

(4) Biochemical responses to ADX also tended towards superior outcomes for the AVS group, with persistent PA being observed in almost double the operated patients in the CT group compared with the AVS group (20 vs. 11%), but again with numbers too small to reach statistical significance.

In short, SPARTACUS results are non-generalizable and the study was powered to the wrong primary endpoint (rather than more meaningful ones such as cure of hypertension and PA in response to ADX). Despite this, the trends for superiority of AVS were clearly there but unfortunately ignored.

Even if we give SPARTACUS the benefit of the doubt and accept that removing the wrong gland, or inappropriately removing a gland from a patient with non-lateralizing PA, occurs in only a minority of patients in whom management is guided by CT (and was therefore not detectable in this underpowered, cohort-based study), such undesirable outcomes should be avoided wherever possible.

Conclusions

SPARTACUS has attempted to address an important clinical question in comparing AVS- with CT-based decision making in terms of the outcome of surgical and specific medical treatment in patients with PA. Its strengths include its robust protocol, the fact that it is the first randomized, prospective study in its field, the relatively strict criteria used to determine lateralization of aldosterone production on AVS, and the use of 24-h ambulatory blood pressure monitoring to assess blood pressure outcomes. However, there are also significant limitations, the most concerning of which are evidence of selection bias, anomalous results, the unusual choice of primary endpoint (DDDs), the decision to include responses to MRAs, and the low power of the study to show difference in more traditional endpoints (particularly cure of hypertension post ADX).

Hence, while both CT and AVS are clearly imperfect in predicting responses to PA treatment, it remains debatable as to whether SPARTACUS has managed to definitively refute the long-held view that AVS is still best.

Conflict of Interest

The authors declare that they have no conflict of interest.

References


