Diverticular Peritonitis: What Did LADIES Trial Teach Us? Analysis of Design, Application and Results

Diego Hernán Barletta
Sanatorio del Oeste, Buenos Aires, Argentina.

ABSTRACT

Background: *LADIES trial is one of the most important trials related to diverticular peritonitis. Its protocol and results were published in 2010, 2015, 2017, and 2019. Despite this trial and other published trials, the appropriate procedures for each scenario of diverticular peritonitis are still being debated, necessitating a thorough review of the methodology used in trials to validate or reject their conclusions.*

Objectives: *To analyze the methodology used in the design and the application, analysis of results and conclusions of all LADIES trial publications. Secondarily, to collaborate in the improvement of diverticular peritonitis research, and facilitate the analysis of the topic by the readers.*

Methods: *The central parts of any research were analyzed, from the research question, hypothesis development, operationalization of variables, trial design, statistical analysis and conclusions. We searched for errors, biases, and weaknesses that could challenge the study findings.*

Results: *LADIES trial was a randomized, open-label, superiority trial analyzed according to intention to treat modified in cases of non-compliance with the inclusion-exclusion criteria. Its design was generally correct, although errors, weaknesses, and biases were detected in its application. Regarding results, LOLA showed that, in Hinchey 3, laparoscopic lavage has a higher rate of early morbidity and mortality than sigmoidectomy, but with a shorter operative time. DIVA showed that, in Hinchey 3 and 4, the primary anastomosis has higher ostomy-free survival with less morbidity, combining initial surgery and ostomy closure, compared to the Hartmann procedure.*

Conclusions: *Failure to reach the calculated sample size resulted in only large differences achieving statistical significance. The low frequencies of adverse events accentuated this methodological problem. The specialization of the intervening centers and surgeons, as well as the exclusion of hemodynamically unstable patients or patients undergoing corticosteroid therapy, compromised the trial external validation.*

Keywords: *Diverticular Peritonitis: Perforated Diverticulitis; Methodology; Research; LADIES Study*

BACKGROUND

LADIES trial was an international, multicenter study that was conducted from 2010 to 2016, involving 8 academic hospitals and 34 school hospitals from the Netherlands, Belgium and Italy. Its protocol and results were published in 2012, 2015, 2017 and 2019.[1-4] It is considered one of the most relevant studies in diverticular peritonitis. Despite a large number of publications, there are still doubts about which are the most appropriate procedures for each scenario of diverticular peritonitis.

The purpose of this study was to analyze the design, application and results of LADIES trial from a methodological perspective, with the primary objective of validating or, failing that, suspecting its conclusions. And with the secondary objective of collaborating to improve the methodology used in the research of the topic in question and its analysis by the readers.

METHODS

The central parts of any research study were described. Since it is a 2-arm study, the parts that refer to all the study were analyzed together, and those that are particular, separately. The following parts of the trial were described:

- Research question and hypothesis.
- Study design.
- Operationalization of the variables.
- Statistical analysis and results.
- Conclusions.

In each of them, errors, biases and weaknesses were searched that affected the logical sequence of the research process.

RESULTS

Investigation question, hypothesis

LADIES trial aimed to answer these two questions: "First, is laparoscopic lavage (LL) in Hinchey 3 superior to sigmoidectomy in terms of morbidity, mortality, quality of life, and health costs (LOLA arm)? Second, which approach is superior for Hinchey 3 or 4 regarding ostomy-free survival, quality of life and health costs: Hartmann's procedure (HP) or primary anastomosis (PA) with or without a protective ostomy (DIVA arm)? "

In LOLA arm the alternative hypothesis was that for Hinchey 3:

- *Ha:* morbidity and mortality at 12 months of LL <

The author has no conflicts of interest.

Diego Hernán Barletta
barlettadie@gmail.com

Received: June, 2020. Accepted: July 2020

morbidity and mortality at 12 months of sigmoidectomy (S).

- *Ha:* LL quality of life score > S quality of life score.

In DIVA arm the alternative hypothesis was that for Hinchey 3 or 4:

- *Ha:* PA (with or without ostomy) ostomy-free survival > HP ostomy-free survival.
- *Ha:* PA (with or without ostomy) quality of life score > HP quality of life score.

Cost analyzes will not be addressed due to the large differences between Europe and the local environment.

Study design

This was an open-label, randomized study with analysis of superiority according to intention to treat, including patients between 18 and 85 years of age with clinical signs of diverticular peritonitis, with free air in abdominal X-ray, or CT with free air or diffuse fluid. Patients with dementia, pelvic irradiation, steroid treatment, shock, or inotropic requirements were excluded.

Randomization was in blocks of 2, 4 and 6 patients, stratified by age (over and under 60 years). In LOLA arm the distribution was 2: 1: 1 (LL, HP, PA) and in DIVA 1: 1 (HP, PA). Although technical procedures were detailed, certain decisions, such as the type of anastomosis or the decision to perform a protective ileostomy after the primary anastomosis, were left to the discretion of the surgeon.

LOLA arm design

The groups were analyzed, as corresponds to any superiority study, according to intention to treat, modified in LOLA only in 2 cases of violation of the inclusion criteria protocol.

The decision to opt for a "superiority analysis" and not to perform a “non-inferiority analysis” was due to the fact that this type of analysis requires a larger sample size to obtain a statistically significant result ($p < 0.05$). The analysis of patients "by intention to treat" is characterized by keeping the randomization of the comparative groups intact and by being more representative of reality, taking into account that it is very common for patients not to receive the assigned treatment or to receive another treatment for different reasons, and that there are losses to follow-up. This type of analysis has as a disadvantage the loss of homogeneity of the groups and therefore the loss in the expected effect of the evaluated treatment. There are authors who recommend the presentation of both analyzes, by intention to treat and by protocol, since coincident results would give greater robustness to the study.[5-7]

The authors expected a 10% morbidity and mortality for the LL group and a 25% morbidity and mortality for the S group (HP and PA), therefore, the expected difference was 15% and based on this difference the sample size was calculated.

From a population of 563 patients with perforated diverticulitis 377 were separated by inclusion and exclusion criteria, leaving 186 patients who underwent diagnostic laparoscopy, 77 of them were Hinchey 1 or 2 and were not enrolled, 19 were Hinchey 4 and enrolled in DIVA, the remaining 90 patients were included in this study as Hinchey 3. The characteristics of the excluded patients were not significantly different from those included. The groups were made up of 47 patients in the LL group (45 LL, 1 HP, 1 exclusion, and 1 loss to follow-up) and 43 in the S group (21 HP, 20 PA, 1 LL and 1 exclusion). After the exclusions, 46 remained in the LL group and 42 in the S group according to the modified intention to treat.

LOLA arm variables

The primary outcome variable chosen was a categorical variable expressed in percentages that combines mortality and morbidity greater than 12 months after initial surgery, expressed as:

$$\frac{\text{Patients deceased or with higher morbidity}}{\text{Patients treated}}$$

Measuring morbidity and mortality has the advantage of increasing the proportion of the expected event and therefore its difference, allowing reducing the sample size necessary to obtain statistical significance.

Major morbidity was defined as reoperation, evisceration, abdominal abscess requiring percutaneous drainage, acute myocardial infarction, urosepsis, kidney failure, and respiratory failure. Early morbidity and mortality was defined as that which occurred within 30 days after surgery or until discharge if hospitalization was longer than that period. Treatment failure was considered to be the presence of persistent sepsis that caused death or required reoperation. Elective sigmoidectomies in the LL group were computed as reoperations.

The accessory outcome variables were operating time, length of stay, days of life outside the hospital, late morbidity and mortality, incisional hernia, and delayed reoperations. The secondary outcome variable was quality of life measured by 3 questionnaires (SF-36, GIQLI, and EQ5D).

Statistical analysis and LOLA arm results

The statistical tests used were the usual ones and the data were reported with measures of effect, mean differences, odds ratio and 95% confidence intervals.

Seventy-three out of eighty-eight (83%) patients were

operated on by surgeons specialized in digestive surgery, the difference of this variable not being significant within the groups (37/46 in LL and 36/42 in S).

The external validation of a study refers to the applicability of the findings of the sample in the population, because the vast majority of patients with diverticular peritonitis are operated by general surgeons. The results of the study will have a significant bias regarding the applicability of their findings in daily practice.

Differences (not statistically significant) were also found in the proportion of ASA 3 and 4 patients, and in values of the POSSUM-OS scale, in both cases favoring the LL group.

The statistically significant term only refers to $p < 0.05$; it is the chance that the differences found in the study groups are due to chance. Instead, clinically significant refers to its importance or relevance in medical practice.

The study was stopped in February 2013 by the safety committee during the third protocol analysis of the data due to the high rate of early complications observed in the LL group. During the 12-month follow-up, no significant difference in morbidity and mortality was observed (30 patients in LL vs. 25 patients in S; OR 1.28, 95% CI 0.54–3.03, p = 0.5804). The mean operative time was less for the LL group: 60 vs. 120 min for S (mean difference –54.53 min, 95% CI –68.04 to –40.26, p = 0.0010).

A contradiction was observed in the presentation of data from early reoperations. Firstly, it mentions that early morbidity and mortality were higher in LL group (18 [39%] patients in LL vs. 8 [19%] patients in S group [OR 2.74, 95% CI 1.03–7.27, p = 0.0427]), which, they say, it can be explained due to the high rate of reoperations in LL vs. S group (16 and 3 patients, OR 6.3, 95% CI 1.85–26, p = 0.0041). However, later the same article mentions that sepsis was successfully controlled (patient alive and without the need for reoperations) in 35 patients of LL and 38 patients of S group. In addition, it mentions that in LL group 9 patients with persistent sepsis required reoperations and 2 more died of multi-organ failure. Similarly, we can see that LL group accounted for 2 deaths, 9 reoperations and 9 percutaneous drains.

Therefore, it is not clear if 16 patients were actually reoperated or if events were counted as patients, in which case a methodological error was incurred since the variable number of events is measured as the average of events per patient (number of total events/ number of patients and the magnitude of the effect can be expressed, for example, as Cohen´s d, but never as OR.[8]

Early mortality was 2/46 (4.3%) for LL group and 1/42 (2.4%) for S group, while mortality at 12 months was 4/46 (8.7%) for LL group and 6/42 (14.3%) for S group.

No significant differences were identified in the results of quality of life questionnaires (SF-36, GIQLI and EQ5D).

The ostomy closures were performed in 5/11 patients in LL group and in 24/35 in S group.

The LL was successful in 24/46 (52%) patients, none of them required another treatment during hospitalization, nor elective sigmoidectomy within 12 months. Thirty-six (78%) patients in the LL group and 30 (71%) in the S group were alive and ostomy-free after 12 months of follow-up (OR 1.53, 95% CI 0.55–4.30, p = 0.4193). It is mentioned that within the 88 enrolled patients, 7 (8%) cases of sigmoid carcinoma were reported, a third of the elective sigmoidectomies were due to it.

LOLA arm conclusions

The results of the study did not allow rejecting the null hypothesis.

H0: 12-month morbidity and mortality of LL ≥ 12-month morbidity and mortality S.

H0: LL quality of life score ≤ S quality of life score.

Early morbidity and mortality after initial surgery was higher in the LL group vs. the S group.

The operative time was statistically less in the LL group vs. the S group.

The authors emphasized that although there was greater early morbidity in LL, this was not reflected in higher mortality, and this could suggest that patients who failed LL could be rescued without compromising survival. These appreciations do not follow from the study since mortality was not evaluated in isolation, therefore, clinically significant differences in mortality may not have reached statistical significance due to the sample size used.

According to the authors, the failures to control sepsis in the LL group could be attributed to a misdiagnosis when mistaking fecal for purulent peritonitis, in fact, in the S group, a third of the operative specimens had perforations.

It is evident that the differentiation by virtue of the subjective classification in Hinchey 3 or 4 alone is not enough to identify those patients with persistent intraoperative colonic perforations, and could have a greater impact on LL compared to resective techniques. However, due to the study design it is not possible to know. Therefore, when a variable is considered predictor of relevant results, reproducible actions in daily practice must be incorporated into the design for their correct measurement.

DIVA arm design

After early termination of the LOLA arm, the sample size was calculated according to a 22% difference in reconstruction (72-50%), requiring 212 patients to establish this difference as statistically significant in a 2-tailed analysis with an $\alpha = 0.05$ and a power of 90% ($\beta = 0.1$). To

the 212 patients a 10% was added for possible losses to follow-up, leaving a sample of 236 patients. Like the other arm, DIVA ended prematurely due to difficulties in recruiting patients.

α and β are known as type 1 and type 2 errors, respectively They represent the possibility of making an incorrect decision regarding the null hypothesis. Type 1 error rejects a true null hypothesis (I find an effect where there is not, 5% of this difference being by chance). Type 2 error does not reject a false null hypothesis (I am not able to find the effect where there is one, I cannot find a difference between the groups, but I can be wrong 10% of the time).

Patient recruitment was carried out as follows: 93 patients were enrolled in Hinchey 3, 47 in the HP group (1 excluded, 1 PA, 45 HP; of them, 1 non-ostomy, 34 closed ostomies and 11 not closed ostomies) and 46 in the PA group (1 LL, 5 HP, 40 PA; of them, 13 non-ostomy, 29 closed ostomies and 4 not closed ostomies). Remained for analysis, according to intention to treat 92 patients, and for analysis of bowel transit reconstruction 62 patients. In Hinchey 4, 40 patients were enrolled, 21 in the HP group (1 excluded, 20 HP; of them, 10 closed ostomies and 10 not closed ostomies), and 19 in the PA group (1 excluded, 2 HP, 16 PA; of them 1 lost, 4 non-ostomy, 9 closed ostomies and 4 not closed ostomies). Remained for analysis, according to intention to treat 38 patients and for analysis of bowel transit reconstruction 19 patients.

According to the intention to treat, these 130 patients were divided into 66 for the HP group and 64 for the PA group. Sixty-five out of sixty-six (98%) patients in the HP group received the planned procedure by protocol, while only 56/64 (87.5%) patients received it in the PA group. As previously explained, the effect or results of the PA were diluted, and although this does not invalidate the study, this difference must be taken into account when drawing conclusions.

DIVA arm variables

The primary outcome variable was ostomy-free survival measured 12 months after the first surgery. It is a categorical and temporal variable.

There is a discrepancy between the secondary variables described in the 2010 protocol and the 2017 publication. In the latter, the early mortality and morbidity, the characteristics of preoperative and operative care received, and the quality of life are mentioned as secondary outcome variables.

It is not the finding of statistically significant differences in the variables between the compared groups that gives value to a study, but the logical sequence that follows from the research question, the correct elaboration of the hypotheses and variables that measure what I want to know, the correct selection of the population and sample, etc. If all this does not happen, I can only say that the differences found are statistically significant with $p < 0.05$.

Statistical analysis and DIVA arm results

The statistical tests used were the usual ones (Kaplan-Meier, Mantel-Cox and Hazard ratio for survival, Fisher's exact test for categorical variables with low frequencies and Student's test for continuous variables).

There were no statistical differences between group's characteristics. As in LOLA, 115/130 (88.5%) of DIVA patients underwent surgery by a surgeon specialized in digestive surgery, compromising the external validity of the results.

In the HP group 65/66 (98%) patients had an ostomy after the initial procedure, while in the PA group only 46/64 (73%) patients had an ostomy. Forty-four out of sixty-five (68%) patients in the HP group vs. 38/46 (83%) in the PA group underwent bowel transit reconstruction ($p = 0.085$). Those patients who did not undergo an ostomy or who were not reconstructed were excluded from the analysis of bowel transit reconstruction.

To compare the reconstruction percentages of ostomy patients within each group when it is not the main outcome variable of the study, and present it together with a $p > 0.05$, although it is statistically correct can lead to confusion. See below.

As the main result, the patients in the PA group had a statistically significantly greater ostomy-free survival than those in the HP group (94.6% [95% CI 88.7–100] vs. 71.7% [95% CI 60.1–83.3], hazard ratio [HR] 2.79 [95% CI 1.86–4.18], log-rank test $p<0.0001$.

HR is the correct way to dynamically measure survival curves because not only it tells me if the event occurred or not, but the time it takes for the event to occur. In this case it would be interpreted that the unreconstructed patients in the PA group had average 2.79 times more chances of reconstructing in the following time interval than patients in the HP group.[9-10]

No statistically significant differences were found in early postoperative results of the initial procedure. Twenty-nine out of 66 (44%) patients in the HP group and 25/64 (39%) patients in the PA group had higher or lower morbidities. The highest morbidity was observed in 8/66 (12%) patients in the HP group and in 9/64 (14%) patients in the PA group.

Mortality was not statistically different between the patients assigned to both groups (HP group 3/66 [3%] vs. PA group 4/63 [6%], $p = 0.44$). Regarding the morbidity associated with bowel transit reconstruction, this was statistically lower in PA group vs. HP group (3/38 [8%] vs.

13/66 [30%], p = 0.023). Although it did not reach statistical significance, the overall morbidity for the initial procedure and the subsequent reconstruction was lower for PA group vs. HP group (25/63 [40%] vs. 37/66 [56%], p = 0.078.

Remember that the expected effect of the PA group is diluted by the analysis according to intention to treat.

No statistically significant differences were found in the results of the life scale questionnaires in both groups.

DIVA arm conclusions

The results of the study allowed rejection of main null hypothesis. It is then accepted:

Ha: Ostomy-free survival PA (with or without ostomy) > ostomy-free survival HP.

But they did not allow rejecting the secondary null hypothesis. Then it is accepted:

H0: Quality of life score PA (with or without ostomy) ≤ quality of life score HP.

It was also observed that PA vs. HP had with statistical significance:

1. Lower early overall morbidity after bowel transit reconstruction.
2. Less average time for reconstruction.
3. Shorter postoperative hospital stay after reconstruction.

DISCUSSION AND CONCLUSIONS

This work represents the results of a first world European population with hemodynamically stable purulent or fecal peritonitis, attended by surgeons specialized in digestive surgery at third-level institutions, whose two arms were interrupted early due to the high rate of early complications in the LL group and the drop in patient recruitment in the DIVA arm. LADIES had such a small sample that only large differences could reach statistical significance. It is for this reason that the low frequencies of adverse events related to the exclusion of unstable patients or those receiving corticosteroids, as well as the specialization of the surgeons and the institutions that provided care, only increased this drawback and seriously compromised its external validity.

In conclusion, LOLA was not only unable to demonstrate the lower morbidity and mortality at 12 months after LL, but conversely showed a higher early morbidity of LL, that reached statistical significance.

LOLA also could not demonstrate a better quality of life in LL vs. S. The analysis of the operative specimens detected colonic perforations in about a third of LL patients, inadvertent during surgery, a fact that highlights the subjectivity of the Hinchey classification and the need of objectifying the intraoperative existence of a perforation for correct patient stratification.

At the same time, DIVA demonstrated greater ostomy-free survival in favor of PA vs. HP. It also showed less morbidity after PA vs. HP closure with statistical significance. DIVA did not demonstrate statistically significant differences in morbidity and mortality, or quality of life between the groups.

BIBLIOGRAPHY

1. Swank HA, Vermeulen J, Lange JF, et al. The ladies trial: Laparoscopic peritoneal lavage or resection for purulent peritonitis and Hartmann's procedure or resection with primary anastomosis for purulent or faecal peritonitis in perforated diverticulitis (NTR2037). BMC Surg 2010; 10:29-35.
2. Vennix S, Musters GD, Mulder IM. Laparoscopic peritoneal lavage or sigmoidectomy for perforated diverticulitis with purulent peritonitis: a multicenter, parallel-group, randomized, open-label trial. Lancet 2015; 386: 1269-77.
3. Vennix S, van Dieren S, Opmeer BC, Lange JF, Bemelman WA. Cost analysis of laparoscopic lavage compared with sigmoid resection for perforated diverticulitis in the LADIES trial. Br J Surg 2017;104:62-68.
4. Lambrichts DPV, Vennix S, Musters GD, et al. Hartmann's procedure versus sigmoidectomy with primary anastomosis for perforated diverticulitis with purulent or faecal peritonitis (LADIES): a multicenter, parallel-group, randomized, open-label, superiority trial. Lancet Gastroenterol Hepatol 2019; 4: 599-610.
5. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res 2011;2:109-12.
6. Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. Trials. 2011;12:58.
7. Lesaffre E. Superiority, equivalence, and non-inferiority trials. Bull NYU Hosp Jt Dis 2008; 66:150-4.
8. Kim HY. Statistical notes for clinical researchers: effect size. Restor Dent Endod 2015; 40:328-31.
9. Martínez-González MA, Alonso A, López Fidalgo J. What is a hazard ratio? Notions of survival analysis. Med Clin (Barc) 2008;131:65-72.
10. Molina Arias M. Hazard ratio: when the risk varies over time. Pediat Prim Care 2015;17:185-88.