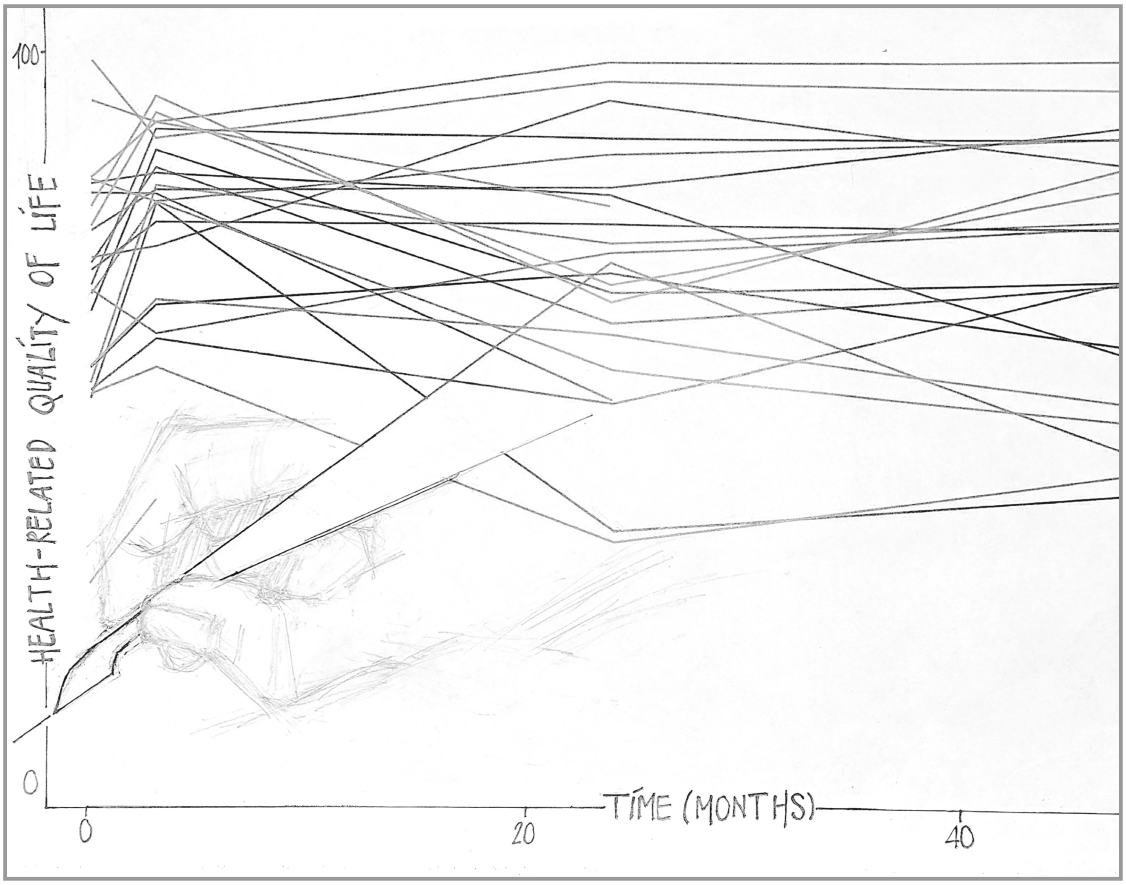
Longitudinal Studies in Pediatric Gastrointestinal Surgery: Theory and Practice



Rebecca Kirsten Stellato

Longitudinal Studies in Pediatric Gastrointestinal Surgery: Theory and Practice

Rebecca Kirsten Stellato

Longitudinale studies in gastro-intestinale kinderchirurgie: theorie en praktijk

Author: Rebecca Kirsten Stellato
Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Jacolijn de Krom, persoonlijkproefschrift.nl
Cover art: Arja Verduyn

Longitudinal Studies in Pediatric Gastrointestinal Surgery: Theory and Practice

Longitudinale studies in gastro-intestinale kinderchirurgie: theorie en praktijk
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 11 december 2023 des ochtends te 10.15 uur

door

REBECCA KIRSTEN STELLATO

geboren op 24 december 1968
te New Orleans, Verenigde Staten

PROMOTOREN:

Dr. M.Y.A. Lindeboom
Prof. dr. M.J.C. Eijkemans †

COPROMOTOREN:

Dr. S.H.A.J. Tytgat
Dr. M. Schipper

BEOORDELINGSCOMMISSIE:

Prof. dr. M.A. Boermeester
Prof. dr. H.C. Boshuizen
Prof. dr. M.M. Rovers
Dr. J. Ruiterkamp
Prof. dr. M.C.J.M. Sturkenboom

LIST OF CONTENTS

CHAPTER 1

General Introduction and Outline of Thesis

Much has been written in the past four decades about the poor methodological quality of medical research,[1-3] both in clinical trials[4,5] and in observational research.[6,7] Poorly designed medical studies waste patients, time and resources, and contribute little to evidence-based medicine.[5]

The term "evidence-based medicine" was first coined in 1991, though calls for the use of evidence in clinical practice had been heard much longer.[8] Evidence-based practice has been defined as the incorporation of the best available external evidence into clinical practice, complementing the expertise of the clinician and the preferences of the patient.[9]

More recently, attention has been focused on evidence-based surgery (EBS).[10-12] In 2009 the Lancet published a series of articles written by the IDEAL collaborative, which investigated the challenges to obtaining high-level evidence in surgery,[13] and made recommendations for improving study design and reporting in the field of surgery.[14]

Despite these developments, the adoption of evidence-based surgery in pediatric surgery has been comparatively slow.[15-18] Numerous obstacles hindering the collection of high-quality general and pediatric surgical evidence have been cited in the literature,[15,18,19] and there is still ample room for improvement in the quality of the available studies.[17-21] The challenges and some potential solutions are described in the following sections.

CHALLENGES IN EVIDENCE-BASED PEDIATRIC SURGERY

Study design

Relatively few published surgical articles are randomized, controlled trials (RCT),[10,11] though that design is generally accepted as higher-quality evidence in evidence-based medicine. Designing good surgery RCTs is particularly difficult due to the nature of surgical interventions. Double blinding and the use of a placebo control are more challenging than in a trial examining the effect of a pharmacological intervention.[11,12,22,23] It can be difficult to standardize surgical procedures due to experience and personal preferences of the surgeon, and to the continual evolution of surgical procedures.[11,12] The complicated nature of surgical interventions (including not only the surgical procedure itself but also the anesthesiology and pre- and postoperative care) also makes standardization of protocols challenging.[24] Furthermore, surgical studies are generally not as well-funded as studies comparing two medications, or one medication to a placebo.[11,18,22]

Surgical studies in children are additionally afflicted by the small number of available patients, often leading to underpowered studies.[19,20] Other obstacles in pediatric surgical studies include the vulnerability of the patient group, the need to obtain consent by

proxy, the often urgent nature of the surgery,[19] a lack of equipoise,[13,21] and the influence of comorbidities.

Reviews have found that 0.04% to 1.9% of reported pediatric surgery studies were randomized, controlled trials.[15,18,20,25] Because a randomized trial is not always feasible or ethical in pediatric surgery, most pediatric surgical studies are observational in nature. Observational studies, when performed well, can be accepted as high-quality evidence.[8,14] However, reviews indicate that most studies in surgery, and especially pediatric surgery, are case series.[11,20] A case series can be a quick way to report on a new technique. However, a well-designed cohort study, in which the patient population and study period are defined and data collection is standardized, is preferable in terms of level of evidence.[13,26]

Reporting of methodology

Among published surgical studies, several methodological problems have been described.

Systematic reviews have demonstrated that criteria essential to the evaluation of design and methodology are often not, or inadequately, reported.[27-29] The few reviews of methodology in pediatric surgery indicate that adherence to guidelines on reporting is also sub-standard.[17,19-21,30] For instance, sample size estimation is often either not reported at all, or not presented reproducibly.[29,31] If the reporting of the methodology of a study is poor, the quality of the study cannot be determined and the results may be misleading.[1,5,6]

Furthermore, even when good-quality data has been collected, if it is not analyzed correctly, results can be incorrect.[32,33] While the medical literature is rich with reviews of poor reporting of methodology, very few reviews can be found related to incorrect use of statistical analysis. One review of surgical literature found that 27% of studies used either incorrect statistical methods or reported results incorrectly.[32]

ADDRESSING THE CHALLENGES

Study design

The IDEAL Collaboration first published their recommendations for evaluating surgical innovation in the Lancet in 2009.[14] While recognizing the challenges faced in designing clinical trials in surgery, they emphasized the need for better study designs and better reporting of surgical research. When feasible, RCT's are preferred. When not feasible, well-designed observational studies such as prospective cohort studies and long-term registry studies were recommended as the most appropriate designs to assess surgical interventions.

Because the vast majority of the surgical literature is observational in nature, it is important that those studies be designed and analyzed with care in order to avoid problems with bias and confounding.[26] Prospective cohort studies provide more evidence for etiology and prognosis than retrospective cohorts or case series.[7,26]

Collection and analysis of repeated measures

In addition to providing stronger medical evidence, prospectively designed studies allow for the collection of repeated measurements on outcomes. These repeated measures can assist in examining effects of treatment over time, or changes in clinical parameters. An additional benefit to collecting repeated measures is the increase in power often resulting from such a design. Since pediatric surgery populations are often small, collecting more data on the same children can provide more information, and statistical power, than a single measurement.[34]

One measure that can be collected without additional clinical visits are questionnaire-based patient-reported outcomes measures (PROMs), such as health-related quality of life (HRQoL). PROMs are increasingly used in the assessment of effectiveness of medical interventions, and are especially important in patient-centered care.[35,36] PROMs should be collected at several moments in time in order to acquire information about trajectories over time and reduce potential bias.[37]

Reporting guidelines

In addition to improvements in study design, reporting of methodology in pediatric surgical studies is important. Numerous guidelines for reporting of medical studies have been developed, including for clinical trials (the CONSORT statement[2]) and for observational studies (STROBE[38]). Adherence to these guidelines assists the readers of published medical studies to assess the quality of the study and level of evidence provided.

AIMS AND OUTLINE OF THE THESIS

This thesis describes the current state of longitudinal data analysis in the pediatric gastrointestinal (GI) surgical literature, and two attempts to gather the best available evidence for minimally invasive upper GI pediatric surgery. The focus is on performing these studies using the best possible design and analysis given the constraints.

The two surgical procedures described in this thesis were performed via minimally invasive surgery (MIS). MIS causes less trauma to the patient, decreasing both recovery time and medical costs. The studies were both observational in nature. In each case, a lack of equipoise dictated that an RCT in children would be unethical. However, the current state of knowledge about short- and long-term effects of the procedures in a pediatric population was lacking. Both studies were set up as one-arm trials and

registered their protocols in advance. They also used validated instruments for outcomes where possible, and collected short- and long-term follow-up on HRQoL and gastroesophageal reflux disease symptoms.

This thesis comprises three parts:

Part I: Longitudinal data analysis in pediatric gastrointestinal surgical literature

The thesis begins with an investigation of reporting on, and analysis of, longitudinal data in the pediatric GI surgical literature. **Chapter two** reports the results of a systematic review of the literature for the years 2010-2019. All scientific articles that reported on repeated measures in a pediatric GI surgical study in this time period were included. Criteria related to the reporting of methodology and statistical analysis were examined, and the type of longitudinal data analysis is described and evaluated. In order to make use of the additional information afforded by repeated measures, proper statistical analysis is important. Longitudinal designs and appropriate analysis methods for repeated measures have become more common in epidemiologic research. However, based on our knowledge of the pediatric gastrointestinal (GI) surgical literature, our hypothesis was that these designs and methods had not yet been fully integrated into that field of research.

Chapter three presents a simulation study that demonstrates the potential problems with suboptimal longitudinal data analysis methods in a simulation of a realistic pediatric surgical setting. Data were simulated on the basis of results from several reports on antireflux surgery (ARS) in neurologically impaired and neurologically normal children. Different levels of missing data were used, and two type of missingness ("missing completely at random" and "missing at random"[39,40]) were examined. Bias, coverage of the nominal 95% confidence intervals, and power were examined for several types of longitudinal data analysis. The "traditional" methods (paired and/or independent *t*-tests and repeated measures ANOVA) were contrasted with more "modern" methods (generalized estimating equations, covariance pattern models, and linear mixed effects models).

Part II: Short- and long-term effects of laparoscopic anti-reflux surgery

Gastroesophageal reflux (GER) is common in infants[41] and usually resolves within the first year of life.[42] When reflux causes troublesome symptoms or complications, it is diagnosed as gastroesophageal reflux disease (GERD). Conservative management of GERD includes lifestyle changes[43] and medical treatment with one or more drugs, the most powerful of which are proton pump inhibitors, which have been found to be safe and effective in children.[44] When conservative management fails, ARS can be used to treat severe GERD. ARS involves fundoplication, or wrapping the fundus of the stomach around the esophagus.

Several techniques are used for ARS: a complete (360°) fundoplication developed by Nissen[45] (Figure 1a), or a partial (270°) fundoplication, either posterior (Toupet,[46] Figure 1b) or anterior (Thal,[47] Figure 1c). Until fairly recently, complete fundoplication was thought to control reflux better than partial; however, Nissen fundoplication has been associated with more severe dysphagia than Thal.[48,49] A systematic review found similar rates of subjectively reported reflux recurrence both short (<6 months) and longer term (>12 months) for complete and partial fundoplication. After partial fundoplication, lower rates of dilatations for dysphagia were observed, along with lower levels of postoperative dysphagia, though the latter was not statistically significant.[50] These results were primarily based on retrospectively collected data. Information on long-term results of ARS in children has been lacking, especially from prospectively designed studies.

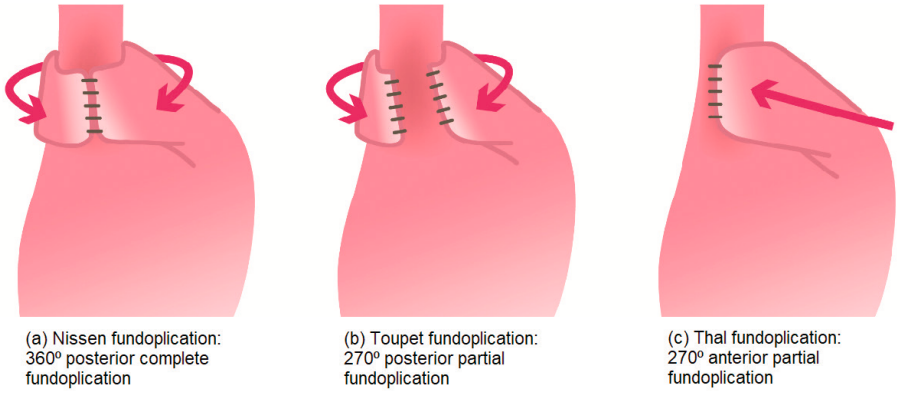


Figure 1. Fundoplication procedures[51]

In the second part of this thesis we examine the short-, middle-, and long-term effects of laparoscopic anti-reflux surgery in a group of 25 children who underwent laparoscopic anti-reflux surgery (LARS) with either a Thal or Nissen fundoplication. A prospective, three-center study was designed to examine both the short-term effects via more objective assessment tests,[52] reflux symptoms and PROMs (health-related quality of life questionnaires). In **chapter four**, the short-term effects of LARS are described. Questionnaires on self- or proxy-reported GERD symptoms and HRQoL were obtained before and approximately 3 months after LARS. Changes in these repeated measures were examined, and an attempt was made to identify predictors of HRQoL in children undergoing LARS. **Chapter five** reports on the middle-term (one and two years after LARS) self-reported HRQoL and GERD symptoms, and in **chapter six** the long-term (i.e. five-year) effects of LARS on HRQoL and GERD symptoms are examined.

Part III: Short- and long-term effects of laparoscopic gastrostomy placement

The second procedure we investigated, to describe its effect over time, was the laparoscopic gastrostomy placement (GP). An infra-umbilical 6 mm trocar was introduced for the camera. Between the umbilicus and the costal margin, a small incision was made through which a Babcock clamp was introduced to grasp the lateral wall of the corpus under direct laparoscopic view. This part of the stomach was then sutured to the fascia of the abdominal wall with Vicryl sutures in four directions. The stomach was insufflated by the anesthesiologist. With clear laparoscopic view a needle was inserted into the stomach. A peel-away dilator was placed using the Seldinger technique followed by introduction of a gastrostomy catheter. Finally, the balloon of the catheter was inflated with sterile water. (Figure 2). GP is a frequently performed procedure that provides long-term enteral tube feeding in children with swallowing or other feeding difficulties.[53,54] The majority of these patients have significant neurologic impairment or congenital heart disease. Other indications for GP include inadequate caloric intake in children with chronic medical diseases e.g. cystic fibrosis, and chronic lung, renal or metabolic disease.[55,56] Parents choose GP primarily for the improvement in nutritional status and global condition.[57] Most GP research has focused on the physical results, such as gain in height and weight.[58-60] Little is known about short- or long-term effects on important PROMs, such as HRQoL.

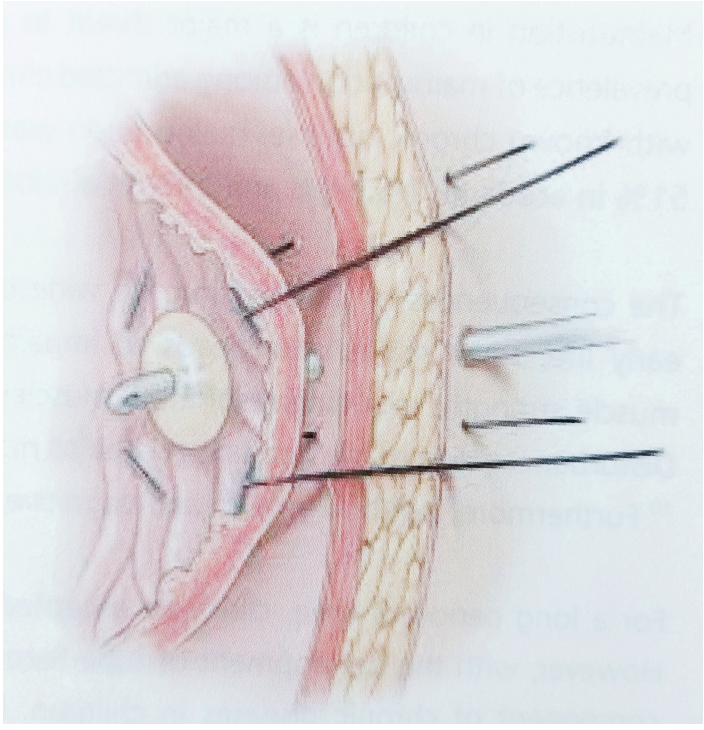


Figure 2. Laparoscopic gastrostomy placement procedure[61]

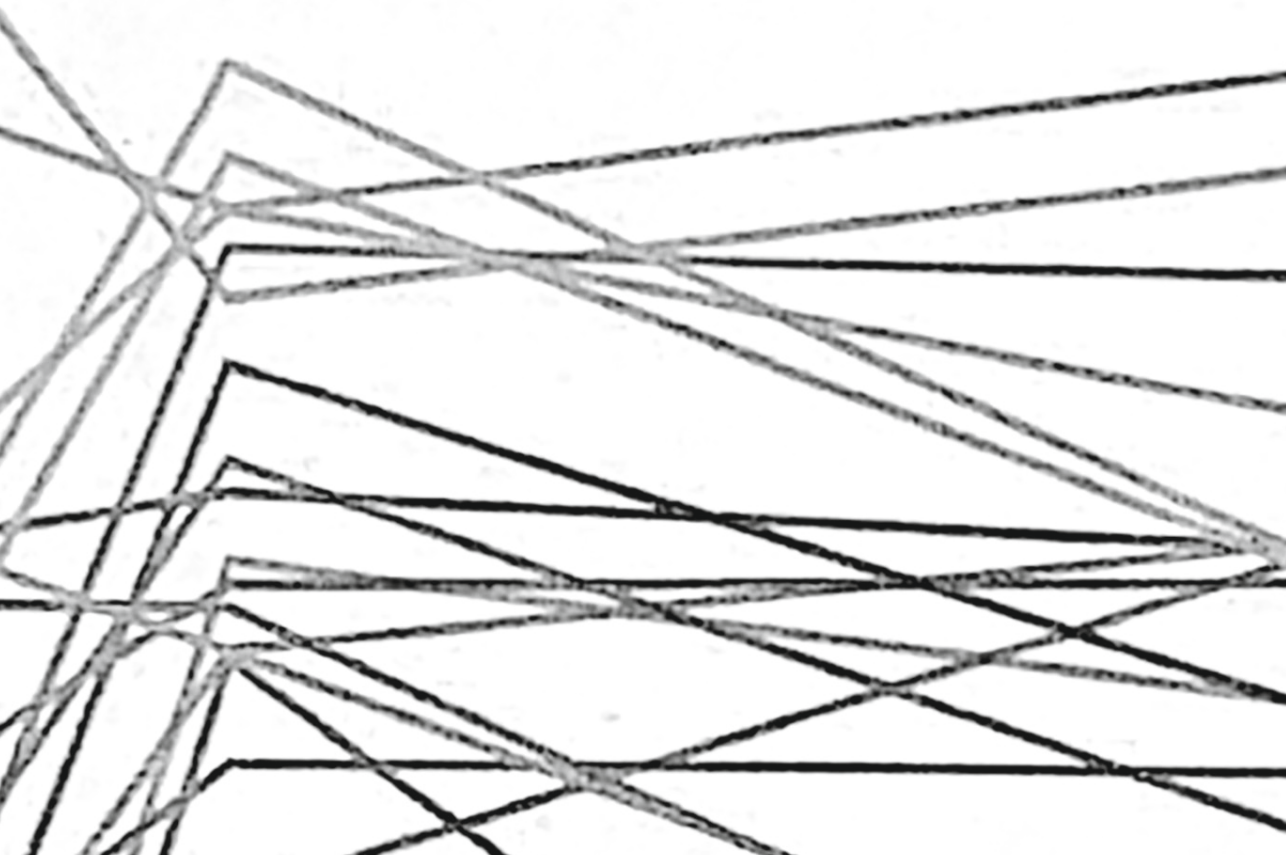
The third part of this thesis describes the results of a prospective study examining the effects of laparoscopic GP in a group of 50 children who underwent the surgery. **Chapter seven** examines the short-term (three to six months after surgery) effects of GP on health-related quality of life and self-reported reflux symptoms. Five years later, the same children (or their parents/caregivers) were asked to repeat the questionnaires; **Chapter eight** reports the findings of the long-term patterns of quality of life and symptoms.

REFERENCES

1. Altman DG. The scandal of poor medical research. *BMJ*. 1994;29(308):283-4.
2. Altman D, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-94.
3. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med*. Feb 2016;109(2):67-77.
4. Pocock S, Hughes M, Lee R. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317(7):426-32.
5. Pirosca S, Shiely F, Clarke M, Treweek S. Tolerating bad health research: the continuing scandal. *Trials*. Jun 2 2022;23(1):458.
6. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. Oct 16 2004;329(7471):883.
7. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. Feb 1 2005;161(3):280-8.
8. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. Jul 22 2017;390(10092):415-423.
9. Sackett DL, Rosenberg WMC, Muir Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-2.
10. Horton R. Surgical research or comic opera: questions but few answers. *The Lancet*. 1996;347:984-5.
11. Wente MN, Seiler CM, Uhl W, Buchler MW. Perspectives of evidence-based surgery. *Dig Surg*. 2003;20(4):263-9.
12. Blencowe NS, Cook JA, Pinkney T, Rogers C, Reeves BC, Blazeby JM. Delivering successful randomized controlled trials in surgery: Methods to optimize collaboration and study design. *Clin Trials*. Apr 2017;14(2):211-218.
13. Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet*. 2009;374:1097–104.
14. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374:1105–12.
15. Hardin W, Stylianos S, Lally KP. Evidence-based practice in pediatric surgery. *J Ped Surg*. 1999;34(5):908-12.
16. Moss RL, Henry MC, Dimmitt RA, Rangel S, Geraghty N, Skarsgard ED. The role of prospective randomized clinical trials in pediatric surgery: state of the art? *J Pediatr Surg*. Aug 2001;36(8):1182-6.
17. Rangel SJ, Kelsey J, Henry MCW, Moss RL. Critical analysis of clinical research reporting in pediatric surgery: justifying the need for a new standard. *J Ped Surg*. 2003;38(12):1739-1743.
18. Ostlie DJ, St Peter SD. The current state of evidence-based pediatric surgery. *Journal of Pediatric Surgery*. 2010;45(10):1940-1946.
19. St-Louis E, Oosenbrug M, Landry T, Baird R. Enrollment and reporting practices in pediatric general surgical randomized clinical trials: A systematic review and observational analysis. *J Pediatr Surg*. May 2018;53(5):879-884.
20. Thakur A, Wang EC, Chiu TT, et al. Methodology standards associated with quality reporting in clinical studies in pediatric surgery journals. *J Pediatr Surg*. Aug 2001;36(8):1160-4.
21. Blakely ML, Kao LS, Tsao K, et al. Adherence of randomized trials within children's surgical specialties published during 2000 to 2009 to standard reporting guidelines. *J Am Coll Surg*. Sep 2013;217(3):394-399 e7.
22. Das AK. Randomised clinical trials in surgery: a look at the ethical and practical issues. *Indian J Surg*. Aug 2011;73(4):245-50.

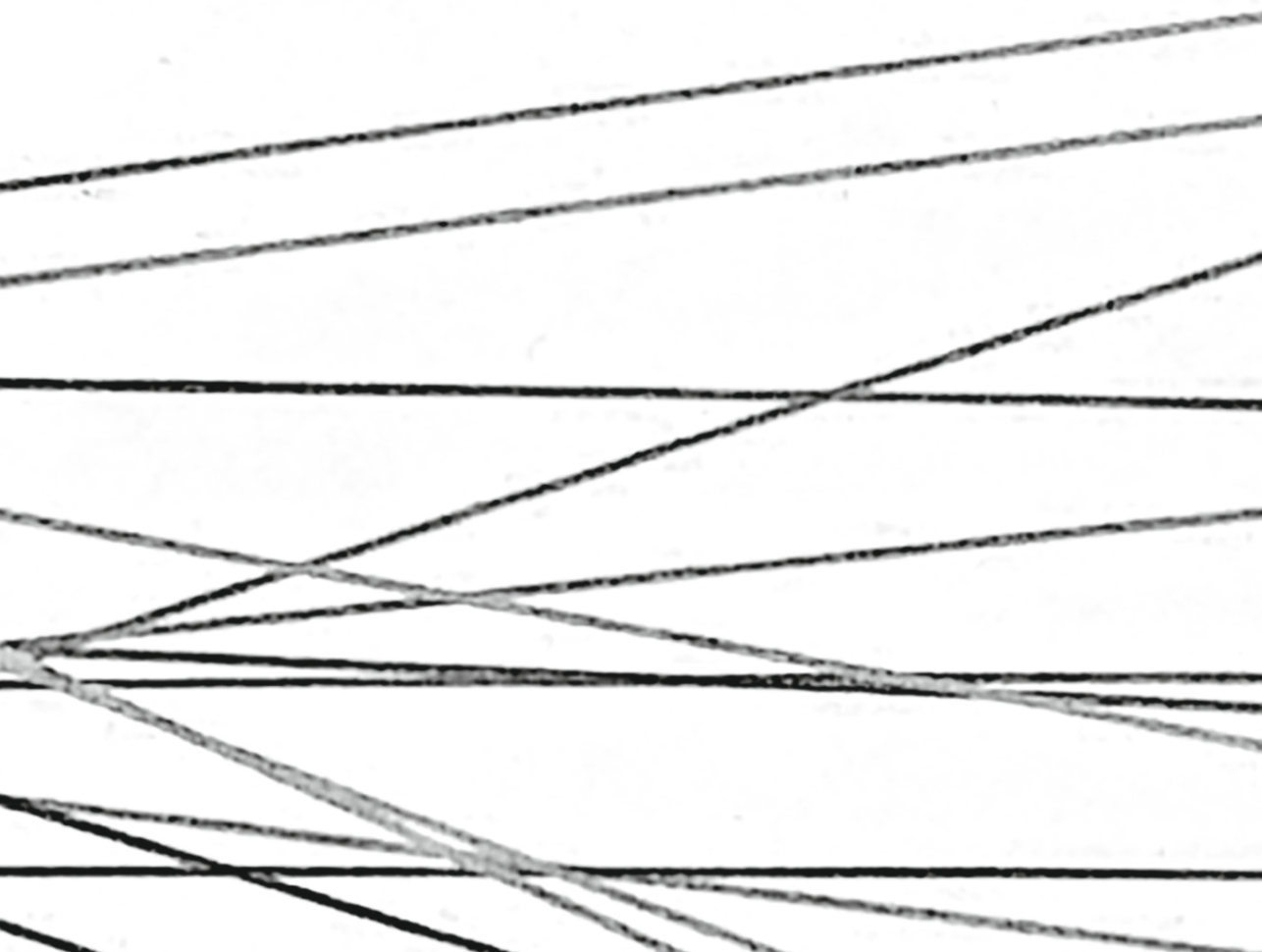
23. Slim K. Limits of evidence-based surgery. *World J Surg*. May 2005;29(5):606-9.
24. Meakins JL. The rules of evidence-based medicine: can they be generalized to improve GI surgical practice? *J Gastrointest Surg*. Apr 2008;12(4):620-3.
25. Allin B, Aveyard N, Campion-Smith T, et al. What Evidence Underlies Clinical Practice in Paediatric Surgery? A Systematic Review Assessing Choice of Study Design. *PLoS One*. 2016;11(3):e0150864.
26. Hartz A, Marsh JL. Methodologic issues in observational studies. *Clin Orthop Relat Res*. Aug 2003;(413):33-42.
27. Hall JC, Platell C, Hall JL. Surgery on trial: an account of clinical trials evaluating operations. *Surgery*. 1998;124:22-27.
28. Ellis C, Hall JL, Khalil A, Hall JC. Evolution of methodological standards in surgical trials. *ANZ J Surg*. 2005;75:874–877.
29. Wenner DM, Brody BA, Jarman AF, Kolman JM, Wray NP, Ashton CM. Do surgical trials meet the scientific standards for clinical trials? *J Am Coll Surg*. Nov 2012;215(5):722-30.
30. Curry JI, Reeves B, Stringer MD. Randomized controlled trials in pediatric surgery: could we do better? *J Pediatr Surg*. Apr 2003;38(4):556-9.
31. Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. *BMJ*. Mar 21 2013;346:f1135.
32. Kurichi JE, Sonnad SS. Statistical methods in the surgical literature. *J Am Coll Surg*. Mar 2006;202(3):476-84.
33. Thiese MS, Ronna B, Robbins RB. Misuse of statistics in surgical literature. *J Thorac Dis*. Aug 2016;8(8):E726-30.
34. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med Res Methodol*. 2003;3(22)
35. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient reported outcomes in randomized trials. *Journal of the American Medical Association*. 2013;309(8):814-822.
36. US Department of Health and Human Services Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2022. Accessed 16 September 2022, 2022.
37. Kluzek S, Dean B, Wartolowska K. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. EBM opinion and debate. *British Medical Journal Evidence-Based Medicine*. 2022;27(3):153-155.
38. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. Dec 2014;12(12):1500-24.
39. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Second ed. John Wiley & Sons; 2002.
40. van Buuren S. *Flexible Imputation of Missing Data*. Second ed. Chapman and Hall/CRC; 2018:444.
41. Nelson SP, Chen EH, Syniar GM, Christoffel KK. Prevalence of symptoms of gastroesophageal reflux during infancy. A pediatric practice-based survey. Pediatric Practice Research Group. *Arch Pediatr Adolesc Med*. Jun 1997;151(6):569-72.
42. Nelson SP, Chen EH, Syniar GM, Christoffel KK. Prevalence of symptoms of gastroesophageal reflux during childhood: a pediatric practice-based survey. Pediatric Practice Research Group. *Arch Pediatr Adolesc Med*. Feb 2000;154(2):150-4.
43. Orenstein SR, McGowan JD. Efficacy of conservative therapy as taught in the primary care setting for symptoms suggesting infant gastroesophageal reflux. *J Pediatr*. Mar 2008;152(3):310-4.

44. Hassall E, Kerr W, El-Serag HB. Characteristics of children receiving proton pump inhibitors continuously for up to 11 years duration. *J Pediatr.* Mar 2007;150(3):262-7, 267.e1.
45. Nissen R. [A simple operation for control of reflux esophagitis]. *Schweiz Med Wochenschr.* 1956;86(Suppl 20):590–2.
46. Toupet A. Technique d'oesophago-gastroplastie avec phréno-gastropexie appliquée dans la cure radicale des hernies hiatales et comme complément de l'opération de Heller dans les cardiospasmes. *Mem Acad Chir.* 1963;89:394–399.
47. Thal A. A unified approach to surgical problems of the esophagogastric junction. *Ann Surg*. 1968;168(3):542–50.
48. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
49. Broeders JA, Mauritz FA, Ahmed Ali U, et al. Systematic review and meta-analysis of laparoscopic Nissen (posterior total) versus Toupet (posterior partial) fundoplication for gastro-oesophageal reflux disease. *Br J Surg*. Sep 2010;97(9):1318-30.
50. Mauritz FA, Blomberg BA, Stellato RK, van der Zee DC, Siersema PD, van Herwaarden-Lindeboom MY. Complete versus partial fundoplication in children with gastroesophageal reflux disease: results of a systematic review and meta-analysis. *J Gastrointest Surg*. Oct 2013;17(10):1883-92.
51. Mauritz FA. *The Effects and Efficacy of Antireflux Surgery in Children*. Dissertation. University Medical Center Utrecht; 2015.
52. Mauritz FA, Conchillo JM, van Heurn LWE, et al. Effects and efficacy of laparoscopic fundoplication in children with GERD: a prospective, multicenter study. *Surg Endosc*. Mar 2017;31(3):1101-1110.
53. Ferluga ED, Sathe NA, Krishnaswami S, McPheeters ML. Surgical intervention for feeding and nutrition difficulties in cerebral palsy: a systematic review. *Dev Med Child Neurol*. Jan 2014;56(1):31-43.
54. Craig GM, Carr LJ, Cass H, et al. Medical, surgical, and health outcomes of gastrostomy feeding. *Dev Med Child Neurol*. May 2006;48(5):353-60.
55. Khattak IU, Kimber C, Kiely EM, Spitz L. Percutaneous endoscopic gastrostomy in paediatric practice: Complications and outcome. *Journal of Pediatric Surgery.* 1998;33(1):67-72.
56. Srinivasan R, Irvine T, Dalzell M. Indications for percutaneous endoscopic gastrostomy and procedure-related outcome. *J Pediatr Gastroenterol Nutr.* 2009;49(5):584–588.
57. Martínez-Costa C, Borraz S, Benlloch C, Lopez-Saiz A, Sanchiz V, Brines J. Early decision of gastrostomy tube insertion in children with severe developmental disability: a current dilemma. *J Hum Nutr Diet*. Apr 2011;24(2):115-21.
58. Samson-Fang L, Butler C, O'Donnell M. Effects of gastrostomy feeding in children with cerebral palsy: an AACPDM evidence report*. *Developmental Medicine & Child Neurology*. 2007;45(6):415-426.
59. Sullivan PB, Juszczak E, Bachlet AM, et al. Gastrostomy tube feeding in children with cerebral palsy: a prospective, longitudinal study. *Developmental Medicine & Child Neurology*. 2007;47(2):77-85.
60. Corwin DS, Isaacs JS, Georgeson KE, Bartolucci AA, Cloud HH, Craig CB. Weight and length increases in children after gastrostomy placement. *J Am Diet Assoc*. Sep 1996;96(9):874-9.
61. Franken J. *The Effects and Efficacy of (Laparoscopic) Gastrostomy Tube Placement in Children*. Dissertation. University Medical Center Utrecht; 2019.



PART I

LONGITUDINAL DATA ANALYSIS IN PEDIATRIC GASTROINTESTINAL SURGICAL LITERATURE



CHAPTER 2

Methodological Quality of Longitudinal Studies in Pediatric Gastrointestinal Surgery: a Systematic Review

Rebecca K. Stellato[1]
Maria Schipper[1]
G. Caroline van Baal[1]
Stefaan H.A. Tytgat[2]
Maud Y.A. Lindeboom[2¶]
Marinus J.C. Eijkemans[1¶]

[1] Department of Data Science and Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2] Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, The Netherlands

¶ These authors contributed equally to this work.

ABSTRACT

Objective
Critical reading of the pediatric gastrointestinal surgery literature indicates several important methodological problems in design, reporting and analysis. This systematic review investigates whether those problems persist in the recent literature.

Study design
A search was performed in PubMed to identify all articles published 2010 - 2019 that reported on longitudinal studies in pediatric gastrointestinal surgery. Eligible articles (with repeated measures collected in a pediatric, gastrointestinal surgical intervention) were evaluated on reporting and analysis criteria by two independent researchers.

Results
Of the 314 articles identified and screened, 22 (7%) were found to be eligible. A majority (56%) did not include repeated measures on the outcome variable(s), despite a longitudinal design. The study objective, number of participants at the beginning of the study, and exposure variable were reported most frequently (86-95%). Worst compliance was in reporting clear information on potential confounders or effect modifiers and reporting on reliability and validity of the outcome measure(s) (0-9%). Sample size justification was reported in four (18%) studies. Only eight (36%) studies used appropriate methods for analyzing repeated measures.

Conclusions
A majority of studies classified as “longitudinal” or “cohort” did not make efficient use of the design by collecting repeated measures or time-to-event data. Many recently published longitudinal studies did not report crucial information about data collection and analysis. Inefficient design and analysis likely led to reduced power; this in a field that generally reports on small study samples. Furthermore, inappropriate analysis of the collected data likely led to bias in estimated treatment effects.

INTRODUCTION

There is broad consensus that evidence-based practice in pediatric surgery is important for patients and their surgeons, as well as hospitals, insurance companies and governments.[1] New techniques and best surgical practices are constantly being evaluated, aimed at improving both efficacy and efficiency. Important research questions are investigated and data collected at no small expense in time and effort of the patients, their parents/caregivers, and their surgeons. Such valuable data must be treated with great care; poorly designed, analyzed and reported medical studies waste patients, time and resources, and contribute little to evidence-based medicine.[2]

Much has been written in the past four decades about the poor methodological quality of medical research,[3-5] both in clinical trials[2,6] and in observational research.[7,8] More recently, the quality of evidence in the field of surgery has been subjected to more careful scrutiny. Surgical studies tend to lack funding and resources needed for good clinical research[9,10] and only an estimated 4-7% of published surgical articles are randomized, controlled trials,[9,11] though that design is generally accepted as higher-quality evidence in evidence-based medicine. Among the relatively rare surgical trials several methodological issues have been identified, including poor reporting of criteria essential to the evaluation of potential bias.[12-15]

Surgical studies in children are further impeded by the need to obtain consent by proxy, the often urgent nature of the surgery,[16] and a lack of equipoise.[17] The number of available pediatric patients for most surgical procedures is generally low, often leading to underpowered studies.[16,18] The few reviews of methodology in pediatric surgery indicate that reporting in the field is also substandard.[16-20]

Since the vast majority of the surgical literature is observational in nature, it is important that those studies be designed and analyzed with care in order to avoid problems with bias and confounding.[21] Prospective cohort studies provide more evidence for etiology and prognosis than case series[8,21] and including repeated measurements of outcomes assists in examining effects of treatment over time, or changes in clinical parameters. An additional benefit to collecting repeated measures is the increase in power often resulting from such a design. However, in order to make use of the additional information afforded by repeated measures, proper statistical analysis is important.

While longitudinal designs and modern statistical methods to appropriately analyze them have become more common in epidemiologic research, our hypothesis based on knowledge of the pediatric gastrointestinal (GI) surgical literature was that these designs and methods have not yet been fully integrated into that field of research. The aims of the current review are twofold:

1) to examine the reporting and analysis of prospectively collected repeated measures data in studies of pediatric GI surgery during a ten-year period of time, offering recommendations where necessary; and
2) to demonstrate the potential bias induced by improper longitudinal data analysis using a small simulation study.

MATERIALS AND METHODS

Literature search and screening for eligibility

A search was performed on January 22, 2022 in PubMed to identify all articles published between January 1, 2010 and December 31, 2019 that reported on longitudinal studies in pediatric gastrointestinal surgery. The exact PubMed Search was:

("pediatrics"[mh] OR "infant"[mh] OR "child"[mh] OR "adolescent"[mh])
AND ("Surgical Procedures, Operative"[mh])
AND ("Gastrointestinal Tract"[mh] OR "Gastrointestinal Diseases"[mh])
AND ("follow-up studies"[mh] OR "cohort studies"[mh] OR "longitudinal studies"[mh] OR "Prospective Studies"[mh])
AND (Clinical Trial[ptyp] OR Randomized Controlled Trial[ptyp] OR Comparative Study[ptyp] OR Multicenter Study[ptyp] OR Observational Study[ptyp] OR Evaluation Study[ptyp])
AND hasabstract[text]
AND "humans"[mh]
AND English[lang]
AND "2010/01/01"[PDAT] : "2019/12/31"[PDAT]
NOT ("case series"[All Fields] OR "case"[Title] OR "case control studies"[mh] OR "Retrospective Studies"[mh] OR "Cross-Sectional Studies"[mh] OR "Systematic Review"[ptyp])
NOT ("adult"[mh] OR "middle aged"[mh])

All articles identified by the search were screened by the first author (RKS), and a random sample of 30% were screened by either the second or third author (MS, CvB). Agreement between screening was examined, and if mistakes were found in the original screening, an additional random 10% of articles would be screened by MS. In case of discrepancies, a fourth author (ML) was consulted and agreement was reached by consensus. Articles were considered eligible if they were full reports with a full text available in English, and if the studies involved only a pediatric population, a gastrointestinal surgical intervention, and repeated measures (at least 2) of a primary or secondary outcome. Transplantations were considered surgical intervention if the transplantation and/or pre-surgical variables were used in analysis, and post-surgical interventions (pain prevention, diet, antibiotics) were included only if they directly affected the surgical outcome. Exclusion criteria were: articles describing a study

protocol with no results presented; systematic reviews; case-control, cross-sectional or retrospective designs; case reports and small case series (fewer than 10 patients); non-surgical interventions (i.e. anesthesiology interventions and post-surgical interventions related to pain prevention, diet, or antibiotic use), and non-GI surgical interventions; studies in which the primary and secondary outcomes were a single event or time to a single event; and articles authored by one or more of the current authors.

Checklist and assessment of articles

The assessment form comprised 25 items (some with multiple sub-items), several of which were borrowed from a previous assessments of longitudinal studies[8] or from the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)[22] checklist, and several of which were chosen to enable the assessment of the statistical analysis.

All eligible articles were assessed on all checklist items independently by RKS and either MS or CvB. The ratings of each item were then compared; any disagreements were resolved through consensus.

For assessment of the statistical analysis of repeatedly measured (primary and/or secondary) outcomes, four categories were identified: the analysis was

- *appropriate* if the authors used mixed effects, covariance pattern (CP), or generalized estimating equations (GEE) models for >2 longitudinal measures, or used ANCOVA (adjusting for baseline) if only 2 repeated measures were collected in a trial and there were no missing data at either time point;[23]
- *correct but inefficient* if, in the absence of missing data, unpaired tests were used at one time point (usually the final time point), or paired samples t-tests within groups if only 2 repeated measures were collected;
- *inappropriate* when repeated measures ANOVA (RMA) was used, or unpaired tests (t-tests, ANOVA, Mann-Whitney or Kruskall-Wallis) between groups were applied to 2 or more time points in the presence of missing data;
- *incorrect* when the study reported unpaired tests on paired data or vice versa, or no statistical analysis at all (only descriptive statistics).

This systematic review was not registered and there was no published protocol. The data collection forms, data extracted from the included studies, data used for all analyses and analytic code can be obtained from the corresponding author.

Statistical analysis

Ineligibility criteria were tabulated for all screened articles. An article could be ineligible for more than one reason and all were tabulated. For this reason the percentages of ineligibility criteria sum to more than the total percentage of ineligible articles.

Agreement between raters for both screening and rating of articles was determined for all items; percentage of agreement is presented. The number and percentage of articles reporting each of the items on the checklist was reported. Analyses were performed in SAS/STAT software version 9.4 of the SAS System for Windows.

Simulation study

To visually demonstrate the problems with RMA and t-tests compared to linear mixed effects (LME), CP, and GEE models in studies with non-random dropout, a simulation was performed for a realistic study on health-related quality of life (HRQoL) in a small group (N=25) of neurologically impaired children undergoing anti-reflux surgery. HRQoL data at four time points in a one-year period were simulated assuming a score of 50 at baseline, and 60 at 3, 6 and 12 months, with a standard deviation of 15 at all time points. Dropout rates from 10% to 40% were applied, assuming children with lower baseline HRQoL scores were more likely to drop out. A more complete description of the simulation can be found in Appendix 1.

RESULTS

Literature search and screening for eligibility

The PubMed search identified 314 potentially eligible articles from 116 journals. Six articles (2%) were excluded due to potential authorship conflict and were not screened further, leaving 308 articles for full screening. Using the re-screening of 92 articles (30%) by either CvB or MS, no mistakes were found in the eligibility screening by RKS and no further re-screening was performed. Eligibility criteria are therefore reported from the complete screening of all articles by RKS.

Fig. 1 displays the flow chart of the 308 screened articles and reasons for exclusion. A report may be deemed ineligible for several reasons, resulting in overlap in numbers found ineligible for each reason. Of the screened articles, 48 (16%) were excluded due to design (protocol, systematic review, case-control study, case report/series, cross-sectional study, or retrospective design). Three articles were not full reports, and four others were not restricted to children.

In 129 articles (42%), the intervention being studied was not a gastrointestinal surgical intervention. In 172 reports (56%) only single measures for the outcome variables were included (i.e. post-surgical events such as success/failure or improvement/deterioration). In 20 studies (6%) the outcome was time to an event. Only 82 articles (27%) reported repeated measures of one or more outcomes. Excluding all other study designs, studies with repeated measures represent 32% of the 260 true cohort studies.

Of the 308 articles identified and screened, 22 (7%) were found to be eligible and were assessed further.

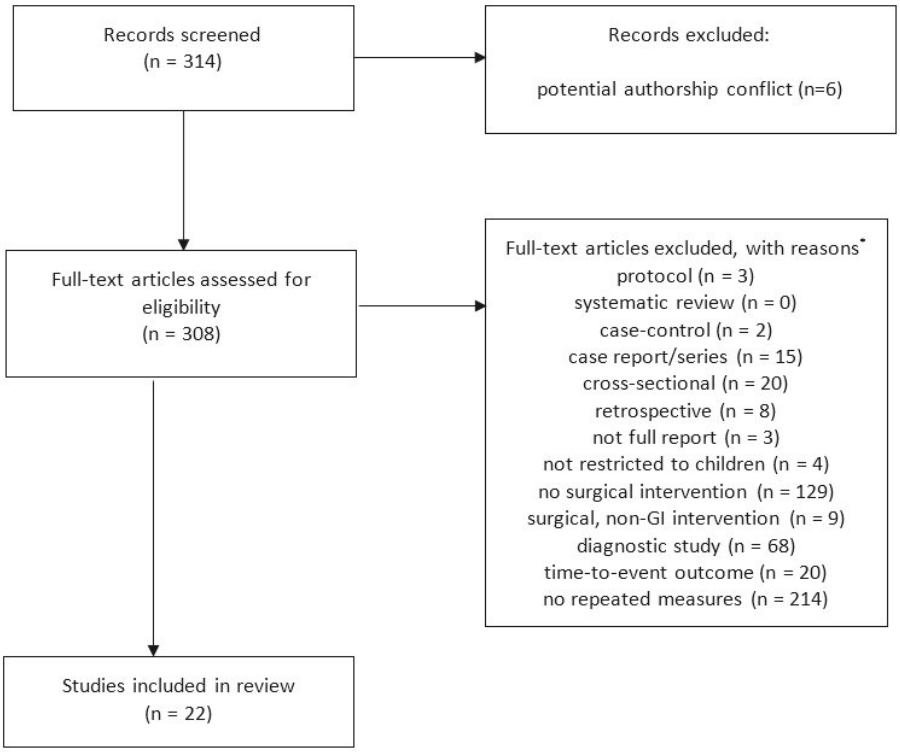


Figure 1. Flow chart screened articles

*Reports were screened for all possible ineligibility criteria; articles may be counted more than once for reasons of exclusion.

Assessment of articles

The 22 eligible articles were assessed by RKS and MS/CvB. The references for the articles can be found in Appendix 2. The median agreement between raters across all items was 83.4%, interquartile range (IQR) [80.4% - 86.3%]. All disagreements were resolved by consensus.

The median number of patients included in the studies examined was 40, IQR [27 - 61]. The percentage of articles reporting the criteria assessed ranged from 0% to 95% (Table 1), with the highest compliance for items “Number of participants at the beginning of the study stated”, “Objectives/hypotheses of the study stated”, and “Results interpreted in context of research question” (95%, 91%, and 86%, respectively) and the lowest for items “Potential confounders, effect modifiers clear”, “Reliability (repeatability) of methods mentioned”, “Validity of methods mentioned,” and “Number of participants justified” (0%, 9%, 9%, and 18%, respectively).

Table 1. Number of articles in compliance with reporting checklist items.

Criterion	N (%)	Source[a]
Objectives/hypotheses clearly stated	20 (91%)	1 (#3) 2 (#1)
Eligibility criteria, selection of participants clearly stated	14 (64%)	1 (#6)
Is the number of participants justified? Yes, justification replicable Yes, justification not replicable	2 (9%) 2 (9%)	1 (#10), 2 (#9)
Outcome variable(s) clearly stated	15 (68%)	3
Exposure(s)/predictor(s) clearly stated	17 (77%)	3
Potential confounders, effect modifiers clearly stated[b]	0 (0%)	3
Reliability (repeatability) of methods mentioned?	2 (9%)	2 (#17)
Validity of methods mentioned?	2 (9%)	2 (#19)
Was the statistical model described clearly enough to be repeatable?	9 (41%)	1 (#12), modified
Number of participants at the beginning of the study stated?	21 (95%)	2 (#15)
Number of participants at each stage/wave specified? Yes, reasons for loss to follow-up quantified Yes, reasons for loss to follow-up not quantified	6 (27%) 6 (27%)	1 (#13a), 2 (#20)
Results interpreted in context of research question(s)?	19 (86%)	1 (#18, modified)
Was there any discussion of generalizability?	6 (27%)	1 (#21), 2 (#33)

[a] 1 = STROBE statement[22] (checklist number), 2 = Tooth et al.[8] (item number), 3 = New item
[b] This item was not applicable for all articles.

Sample size was justified in only four articles. Of those four, two did not report sufficient information for a reader to repeat the sample size estimation, and the remaining two (reporting on the same cohort) reported a slightly lower required sample size than was estimated in a re-analysis using the reported information.

Six studies (27%) had an outcome variable that was measured at two time points; 17 (77%) measured outcomes at three or more time points. (Some studies had more than one primary or secondary outcome variable, potentially with different measurement schemes, and can be counted twice.) One report included an effect modifier. Most studies had some missing data, though many of these studies did not report on how missing data was handled. Only seven of the articles (32%) used appropriate methods for analyzing repeated measures, another one study (5%) used correct but inefficient methods, and 14 (64%) used inappropriate or incorrect analysis methods (Table 2).

Table 2. Types of data and analysis techniques for the 22 reviewed articles

Criterion	N (%)	Source*
Theoretical number of measurements per participant over time[b]	6 (27%)	3
2 fixed time points	17 (77%)	
3 or more fixed time points	4 (18%)	
varies per patient		
Confounders/effect modifiers accounted for	1 (8%)	2 (#28, modified)
yes	12 (92%)	
no/unclear	9	
N/A		
Missing data over time	0 (0%)	3
yes, missing data imputed	3 (14%)	
yes, missing data accounted for in analysis	11 (50%)	
yes, unclear how missing data was handled	3 (14%)	
no	5 (23%)	
unclear		
Analysis method(s) used for repeated measures	7 (32%)	3
appropriate	1 (5%)	
correct but inefficient	8 (36%)	
inappropriate, multiple testing/bias	6 (27%)	
incorrect		

[a] 1 = STROBE statement[22] (checklist number), 2 = Tooth et al.[8] (item number), 3 = New item
[b] For this item, more than one answer per article was possible, so the totals may be larger than 22.

Simulation study

The bias of three of the five methods is shown in Fig. 2 (CP and GEE models produced nearly identical results and are not displayed). For all levels of dropout, the LME model demonstrated no or only minimal bias. Using independent *t*-tests (TT) at successive time points induced a small amount of bias due to the dropout of children with lower baseline HRQoL; this effect increased with increasing levels of dropout. The bias towards higher means was much more pronounced in the RMA analyses.

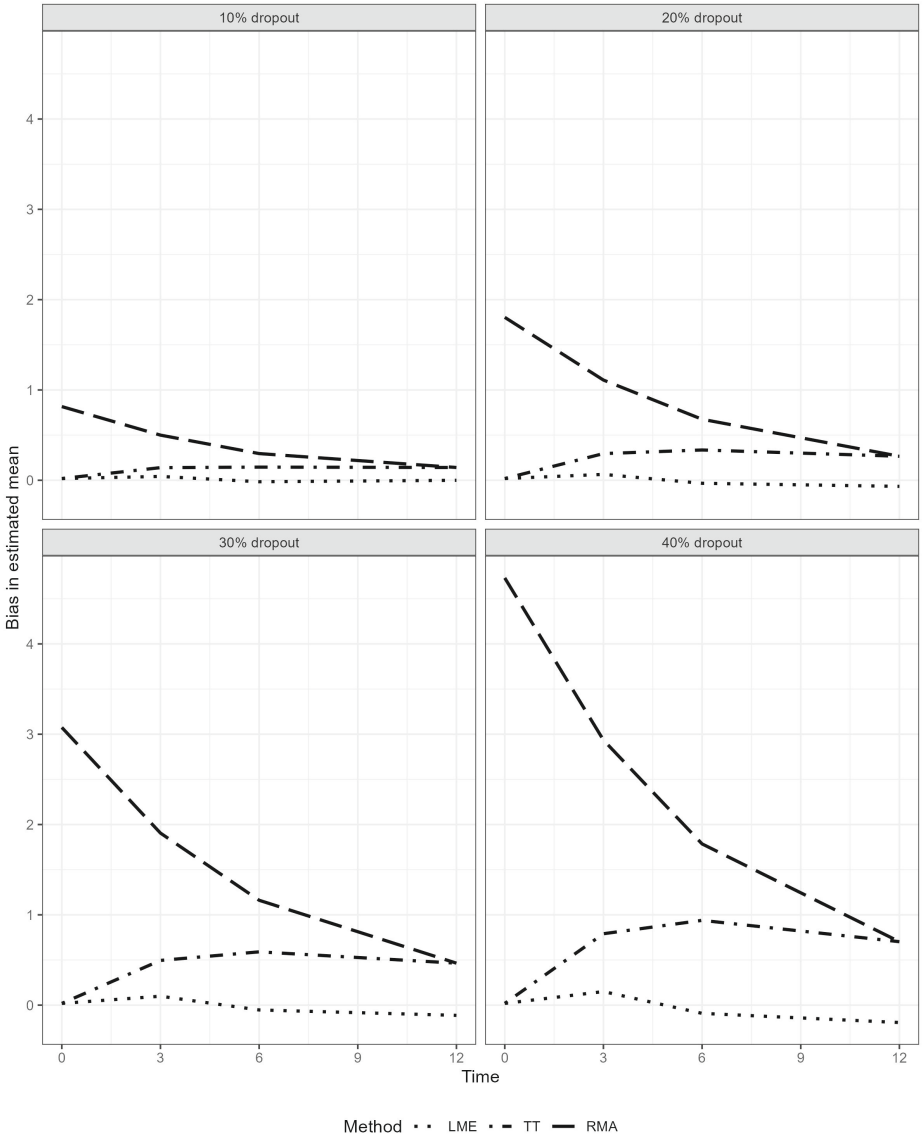


Figure 2. Potential bias in analyses of longitudinal health-related quality of life data with 10%-40% one-year cumulative dropout. On the basis of 7600 simulations using 25 patients (see Appendix A for details). Methods: linear mixed effects model (LME); repeated measures ANOVA (RMA); independent t-tests (TT)

DISCUSSION AND RECOMMENDATIONS

To our knowledge this is the first study to also systematically review the statistical methods used for longitudinal data analysis in medical journals. A few studies have reviewed either the reporting of longitudinal studies[7,8] or the general quality of statistical analysis in medical/epidemiological articles.[24-27]

A large majority of the studies examined reported on the number of included patients. Most also clearly stated the objective of the study and interpreted the results in that context. Poor adherence to reporting was found on several items that are important to the assessment of the quality of the study. In the vast majority of the articles, we observed inefficient use of design and inappropriate or incorrect analysis of longitudinal data. Although this review concentrated on pediatric GI surgery, we expect these results will also apply to other fields in which funding and resources for clinical research are limited.

Reporting Guidelines

Many recently published longitudinal pediatric GI surgery studies did not report crucial information about data collection and analysis. Sample size justification, reliability and validity of methods, statistical methods used, numbers of participants at each wave, and generalizability of results were not, or not clearly, reported in a majority of the articles examined. This in spite of the development and implementation of guidelines for reporting on trials (CONSORT statement) and observational studies (STROBE statement) in medical research.[4,22]

We reiterate the call, made by numerous authors before us, to both journal editors and authors to adhere to those guidelines. In addition to contributing to evidence-based pediatric surgery by allowing readers to more accurately assess the quality of the reported evidence, adherence to guidelines will allow for better synthesis of evidence across studies.

Sample size

The median sample size in the studies examined was 40, and sample size justification was not reported in a majority of studies. These findings are consistent with previous literature that found small samples, unclear reasons for the chosen sample size, and/or poorly performed sample size estimation in pediatric surgical studies.[16-18]

Because the pool of pediatric patients for a particular surgical intervention may be small at any one hospital or clinic, a study performed at a single center will often have difficulty achieving an adequate sample size. Multi-center trials may help overcome these difficulties, though that design requires intensive collaboration among surgeons and significant investment of resources.[1,28,29] When a multi-center trial is not feasible,

a well-designed prospective study will offer more evidence than a retrospectively collected surgical case series.[21] It is then imperative that the study is reported with enough detail to be included in a subsequent systematic review.

Efficient use of cohort design

While it was not originally a focus of this review, one result from the screening of articles is worth additional attention: more than half the pediatric GI surgery studies classified as "longitudinal" or "cohort" in PubMed did not make efficient use of that design by collecting either time-to-event data or repeated measures on one or more outcomes.

Short-term outcomes such as 30-day success rate are not adequate measures of long-term consequences of surgical interventions[29]. Furthermore, studies that use a single proportion after some period of time will, in most situations, have considerably less power to find a difference between two groups than studies that utilize time-to-event or "survival" outcomes.[30,31] Since pediatric surgery studies are generally small, the increase in statistical power afforded by the use of a time-to-event instead of a single dichotomous outcome is recommended. For instance, many surgical studies use failure of the procedure as a primary outcome; using time to failure would increase power with minimal effort.

Another way to increase power without increasing sample size is to use repeated measures on individuals over time.[32,33] One such measure that can be collected without additional clinical visits are questionnaire-based patient-reported outcomes such as health-related quality of life (HRQoL). Patient-reported outcomes are increasingly used in the assessment of effectiveness of medical interventions, and are especially important in patient-centered care,[34,35] and should be collected at several moments in time in order to acquire information about trajectories over time and reduce potential bias.[36]

Appropriate analysis of repeated measures

While repeated measures are a relatively inexpensive means to increase power, they are often subject to missing data over time.[37] Missing data in longitudinal studies presents two major problems: reduced power (leading to wider confidence intervals and larger p-values); and the potential for biased estimates.[37-39] The bias of t-tests and RMA was demonstrated in the simulation study, with considerably higher bias for the latter. This bias in RMA is induced because all data from individuals are excluded from the analysis if they have a missing outcome at any time point.

Appropriate methods for analyzing repeated measures will help considerably in correcting this bias. Commonly used statistical analysis packages have modules for (generalized) LME, CP, and GEE models, and there are multiple tutorials[39-42] and books[43-45] on appropriate methods for longitudinal data analysis. Nevertheless, these methods

do not appear to be broadly used in longitudinal pediatric GI surgery studies. Nearly two-thirds of the articles examined used inefficient, inappropriate, or incorrect methods to analyze longitudinal data, and missing data were generally not accounted for in the analyses. This may have resulted in considerable bias and/or reduced precision in the estimates produced by these studies.

Limitations of the current study

Although we carefully constructed our PubMed search with clear in- and exclusion criteria, nearly one-sixth of the study designs in the articles identified by the search were not truly cohort designs. Furthermore, more than two-fifths of the articles identified were not surgical, GI interventions. These problems highlight the difficulties of using MeSH terms to identify papers for methodological evaluation, and we cannot rule out the possibility that we missed some articles that would have been eligible for this review.

CONCLUSION

The inefficient use of the cohort/longitudinal design identified in many publications on pediatric gastrointestinal surgery has likely led to reduced power of the studies. In addition, inappropriate or incorrect analysis of repeated measures data at best make inefficient use of the available information, and at worst may have led to biased estimates of treatment effects. While these methodological issues are important in any medical study, they are especially so in studies on vulnerable (pediatric) populations. Surgical researchers are strongly encouraged to:

- use prospective cohort studies instead of case series;
- report on the studies according to the appropriate guidelines;
- collect repeated measures or a time-to-event outcome, especially when the sample is expected to be small;
- use an appropriate analysis method for the repeated measures; and
- include a statistician or methodologist in the design, analysis and interpretative phases of clinical studies.

REFERENCES

1. Ostlie DJ, St Peter SD. The current state of evidence-based pediatric surgery. *Journal of Pediatric Surgery*. 2010;45(10):1940-1946. doi:10.1016/j.jpedsurg.2010.05.008
2. Pirosca S, Shiely F, Clarke M, Treweek S. Tolerating bad health research: the continuing scandal. *Trials*. Jun 2 2022;23(1):458. doi:10.1186/s13063-022-06415-5
3. Altman DG. The scandal of poor medical research. *BMJ*. 1994;29(308):283-4.
4. Altman D, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-94.
5. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med*. Feb 2016;109(2):67-77. doi:10.1177/0141076815625599
6. Pocock S, Hughes M, Lee R. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317(7):426-32.
7. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. Oct 16 2004;329(7471):883. doi:10.1136/bmj.38250.571088.55
8. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. Feb 1 2005;161(3):280-8. doi:10.1093/aje/kwi042
9. Wente MN, Seiler CM, Uhl W, Buchler MW. Perspectives of evidence-based surgery. *Dig Surg*. 2003;20(4):263-9. doi:10.1159/000071183
10. Das AK. Randomised clinical trials in surgery: a look at the ethical and practical issues. *Indian J Surg*. Aug 2011;73(4):245-50. doi:10.1007/s12262-011-0307-5
11. Horton R. Surgical research or comic opera: questions but few answers. *The Lancet*. 1996;347:984-5.
12. Hall JC, Platell C, Hall JL. Surgery on trial: an account of clinical trials evaluating operations. *Surgery*. 1998;124:22-27.
13. Ellis C, Hall JL, Khalil A, Hall JC. Evolution of methodological standards in surgical trials. *ANZ J Surg*. 2005;75:874–877.
14. Wenner DM, Brody BA, Jarman AF, Kolman JM, Wray NP, Ashton CM. Do surgical trials meet the scientific standards for clinical trials? *J Am Coll Surg*. Nov 2012;215(5):722-30. doi:10.1016/j.jamcollsurg.2012.06.018
15. Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg*. Nov 2006;244(5):663-7. doi:10.1097/01.sla.0000217640.11224.05
16. St-Louis E, Oosenbrug M, Landry T, Baird R. Enrollment and reporting practices in pediatric general surgical randomized clinical trials: A systematic review and observational analysis. *J Pediatr Surg*. May 2018;53(5):879-884. doi:10.1016/j.jpedsurg.2018.02.009
17. Blakely ML, Kao LS, Tsao K, et al. Adherence of randomized trials within children's surgical specialties published during 2000 to 2009 to standard reporting guidelines. *J Am Coll Surg*. Sep 2013;217(3):394-399 e7. doi:10.1016/j.jamcollsurg.2013.03.032
18. Thakur A, Wang EC, Chiu TT, et al. Methodology standards associated with quality reporting in clinical studies in pediatric surgery journals. *J Pediatr Surg*. Aug 2001;36(8):1160-4. doi:10.1053/jpsu.2001.25737
19. Curry JI, Reeves B, Stringer MD. Randomized controlled trials in pediatric surgery: could we do better? *J Pediatr Surg*. Apr 2003;38(4):556-9. doi:10.1053/jpsu.2003.50121
20. Rangel SJ, Kelsey J, Henry MCW, Moss RL. Critical analysis of clinical research reporting in pediatric surgery: justifying the need for a new standard. *J Ped Surg*. 2003;38(12):1739-1743. doi:10.1016/S0022-3468(03)00631-6

21. Hartz A, Marsh JL. Methodologic issues in observational studies. *Clin Orthop Relat Res*. Aug 2003;(413):33-42. doi:10.1097/01.blo.0000079325.41006.95
22. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. Dec 2014;12(12):1500-24. doi:10.1016/j.ijsu.2014.07.014
23. Vickers AJ, G AD. Analysing controlled trials with baseline and follow up measurements. *BMJ*. 2001;323:1123–4.
24. Kanter MH, Taylor JR. Accuracy of statistical methods in Transfusion: a review of articles from July August 1992 through June 1993. *Transfusion*. 1994;34(8):697-701.
25. Cruess DF. Review of Use of Statistics in The American Journal of Tropical Medicine and Hygiene for January-December 1988. *Am J Trop Med Hyg*. 1989;41(6):619-626.
26. Olsen CH. Review of the use of statistics in infection and immunity. *Infect Immun*. Dec 2003;71(12):6689-92. doi:10.1128/IAI.71.12.6689-6692.2003
27. Karadeniz PG, Uzabaci E, Kuyuk SA, et al. Statistical errors in articles published in radiology journals. *Diagn Interv Radiol*. Mar 2019;25(2):102-108. doi:10.5152/dir.2018.18148
28. Blencowe NS, Cook JA, Pinkney T, Rogers C, Reeves BC, Blazeby JM. Delivering successful randomized controlled trials in surgery: Methods to optimize collaboration and study design. *Clin Trials*. Apr 2017;14(2):211-218. doi:10.1177/1740774516687272
29. Macrae D. Conducting clinical trials in pediatrics. *Crit Care Med*. Jan 2009;37(1 Suppl):S136-9. doi:10.1097/CCM.0b013e318192101f
30. Cuzick J. The Efficiency of the Proportions Test and the Logrank Test for Censored Survival. *Biometrics*. 1982;38(4):1033-1039.
31. Willan AR. Relative efficiency in survival analysis. *J Chron Dis*. 1983;36(2):183-192.
32. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med Res Methodol*. 2003;3(22)doi:http://www.biomedcentral.com/1471-2288/3/22
33. Tekle FB, Tan FE, Berger MP. Too many cohorts and repeated measurements are a waste of resources. *J Clin Epidemiol*. Dec 2011;64(12):1383-90. doi:10.1016/j.jclinepi.2010.11.023
34. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient reported outcomes in randomized trials. *Journal of the American Medical Association*. 2013;309(8):814-822.
35. US Department of Health and Human Services Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2022. Accessed 16 September 2022, 2022.
36. Kluzek S, Dean B, Wartolowska K. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. EBM opinion and debate. *British Medical Journal Evidence-Based Medicine*. 2022;27(3):153-155. doi:10.1136/bmjebm-2020-111573
37. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. Oct 2014;23(5):440-59. doi:10.1177/0962280213476378
38. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. *Annu Rev Clin Psychol*. 2010;6:79-107. doi:10.1146/annurev.clinpsy.032408.153550
39. Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. *Circulation*. Nov 4 2008;118(19):2005-10. doi:10.1161/CIRCULATIONAHA.107.714618
40. Albert PS. Tutorial in biostatistics: Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*. 1999;18:1707-32.
41. Schober P, Vetter TR. Repeated Measures Designs and Analysis of Longitudinal Data: If at First You Do Not Succeed-Try, Try Again. *Anesth Analg*. Aug 2018;127(2):569-575. doi:10.1213/ANE.0000000000003511

42. Fox J, Weisberg S. Mixed-Effects Models in R. *An R Companion to Applied Regression, Second Ed*. 2 ed. SAGE Publications, Inc; 2015:chap Online appendix: https://socialsciences.mcmaster.ca/jfox/Books/Companion-2E/appendix/Appendix-Mixed-Models.pdf.
43. Singer JD, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press; 2003.
44. Brown H, Prescott R. *Applied Mixed Models in Medicine*. 3rd ed. Statistics in Practice. John Wiley & Sons, Ltd.; 2015.
45. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.

APPENDIX 1. SIMULATION OF BIAS AND PRECISION

A simulation study was performed to demonstrate the bias of methods of analysis that use complete cases (repeated measures ANOVA (RMA), paired t-tests) and methods that use complete data at one moment in time (independent t-tests between groups), compared to more appropriate longitudinal models such as generalized estimating equations (GEE) or linear mixed effects models (LME) and covariance pattern models (CPM), which use an adequately complex model for the correlation of measurements over time within patients. Based on results of several pediatric anti-reflux studies, longitudinal health-related quality of life (HRQoL) data were simulated for two groups at times 0, 3, 6 and 12, mimicking patterns that might occur after fundoplication. The results of one group is presented here.

Children were assumed to have a mean HRQoL of 50, increasing to 60 at 3 months and remaining stable until one year. A multivariate normal distribution was used, with means 50 and 60, a standard deviation of 15 points, and a month-to-month autocorrelation of 0.85 (or a correlation between baseline and one-year HRQoL of $0.85^{12} = 0.15$). The simulations occasionally led to HRQoL higher than 100; although that is not possible in practice (most QoL scores are coded as 0 - 100); in order to calculate the bias accurately we did not truncate HRQoL scores.
For the group presented here, informative dropout was generated using the exponential distribution, with percentages varying from 10% to 40%. Dropout was dependent on baseline HRQoL: children with a score less than the median were four times more likely to drop out of the study than children with a baseline score above the median.

7,600 datasets were generated and analyzed using t-tests (paired: difference in mean HRQoL 12 months vs 0 months for each group; unpaired: difference in mean HRQoL NN vs NI at all time points), repeated measures ANOVA, GEE, a CPM with a continuous autoregressive correlation structure, and a LME model with a random intercept and a random slope for time (modelled as continuous). The latter four models included (fixed) effects for group, time (modelled as categorical), and an interaction between the two. For each model, the resulting mean HRQoL for the two groups at the different time points (estimated marginal means) were compared to the true means, and the bias was calculated, averaged and plotted. For the current paper, only the results of the LME and RMA models and independent t-tests are presented here.

The simulations and analyses were performed in R version 4.0.3, using the MASS, nlme, geepack, emmeans, and ggplot2 packages.

APPENDIX 2. LIST OF REFERENCES FOR RATED ARTICLES

1. Mousavi S, Karami H, Rajabpoor A. Intractable chronic constipation in children: outcome after anorectal myectomy. Afr J Paediatr Surg. 2014;11(2):147-9.
2. McHoney M, Wade AM, Eaton S, Howard RF, Kiely EM, Drake DP, et al. Clinical outcome of a randomized controlled blinded trial of open versus laparoscopic Nissen fundoplication in infants and children. Ann Surg. 2011;254(2):209-16.
3. Bishay M, Giacomello L, Retrosi G, Thyoka M, Garriboli M, Brierley J, et al. Hypercapnia and acidosis during open and thoracoscopic repair of congenital diaphragmatic hernia and esophageal atresia: results of a pilot randomized controlled trial. Ann Surg. 2013;258(6):895-900.
4. Pacilli M, Eaton S, McHoney M, Kiely EM, Drake DP, Curry JI, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. Arch Dis Child. 2014;99(6):516-21.
5. Vowinkel T, Laukoetter M, Mennigen R, Hahnenkamp K, Gottschalk A, Boschin M, et al. A two-step multidisciplinary approach to treat recurrent esophageal strictures in children with epidermolysis bullosa dystrophica. Endoscopy. 2015;47(6):541-4.
6. Kubiak R, Eaton S, Andrews J, Grant HW. Long-term catch-up weight gain following fundoplication in children. Eur J Pediatr Surg. 2013;23(2):121-6.
7. Impellizzeri P, Simona A, Grasso M, Antonuccio P, Crisafi C, Scalfari G, et al. Surgical stress after open and transumbilical laparoscopic-assisted appendectomy in children. Eur J Pediatr Surg. 2014;24(2):174-8.
8. Dalgic N, Karadag CA, Bayraktar B, Sancar M, Kara O, Pelit S, et al. Ertapenem versus standard triple antibiotic therapy for the treatment of perforated appendicitis in pediatric patients: a prospective randomized trial. Eur J Pediatr Surg. 2014;24(5):410-8.
9. Granéli C, Kockum CC, Arnbjornsson E, Anderberg M. Outcome after Computer-Assisted (Robotic) Nissen Fundoplication in Children Measured as Pre- and Postoperative Acid Reducing and Asthma Medications Use. Eur J Pediatr Surg. 2015;25(6):532-6.
10. Wang H, Guo XN, Zhu D, Zhu T, Hu LH, Feng JX. Nursing Intervention for Outpatient Rehabilitation in Pediatric Patients with Hirschsprung Disease after Colectomy. Eur J Pediatr Surg. 2015;25(5):435-40.
11. Costerus S, Vlot J, van Rosmalen J, Wijnen R, Weber F. Effects of Neonatal Thoracoscopic Surgery on Tissue Oxygenation: A Pilot Study on (Neuro-) Monitoring and Outcomes. Eur J Pediatr Surg. 2019;29(2):166-72.
12. Golebiewski A, Anzelewicz S, Wiejek A, Lubacka D, Czauderna P. A Prospective Randomized Controlled Trial of Single-Port and Three-Port Laparoscopic Appendectomy in Children. J Laparoendosc Adv Surg Tech A. 2019;29(5):703-9.
13. Zyblewski SC, Nietert PJ, Graham EM, Taylor SN, Atz AM, Wagner CL. Randomized Clinical Trial of Preoperative Feeding to Evaluate Intestinal Barrier Function in Neonates Requiring Cardiac Surgery. J Pediatr. 2015;167(1):47-51 e1.
14. Maier S, Zahn K, Wessel LM, Schaible T, Brade J, Reinshagen K. Preventive antireflux surgery in neonates with congenital diaphragmatic hernia: a single-blinded prospective study. J Pediatr Surg. 2011;46(8):1510-5.
15. Schurman JV, Cushing CC, Garey CL, Laituri CA, St Peter SD. Quality of life assessment between laparoscopic appendectomy at presentation and interval appendectomy for perforated appendicitis with abscess: analysis of a prospective randomized trial. J Pediatr Surg. 2011;46(6):1121-5.

16. Spoel M, Meeussen CJ, Gischler SJ, Hop WC, Bax NM, Wijnen RM, et al. Respiratory morbidity and growth after open thoracotomy or thoracoscopic repair of esophageal atresia. J Pediatr Surg. 2012;47(11):1975-83.
17. Conforti A, Giliberti P, Mondi V, Valfre L, Sgro S, Picardo S, et al. Near infrared spectroscopy: experience on esophageal atresia infants. J Pediatr Surg. 2014;49(7):1064-8.
18. Miyano G, Koga H, Okawada M, Doi T, Sueyoshi R, Nakamura H, et al. Rectal mucosal dissection commencing directly on the anorectal line versus commencing above the dentate line in laparoscopy-assisted transanal pull-through for Hirschsprung's disease: Prospective medium-term follow-up. J Pediatr Surg. 2015;50(12):2041-3.
19. Turk E, Acimis N, Karaca F, Edirne Y, Tan A, Kilic C. The effect on postoperative pain of pulling the rectus muscle medially during open appendectomy surgery. Minerva Chir. 2014;69(3):141-6.
20. Knatten CK, Hviid CH, Pripp AH, Emblem R, Bjornland K. Inflammatory response after open and laparoscopic Nissen fundoplication in children: a randomized study. Pediatr Surg Int. 2014;30(1):11-7.
21. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270 degrees fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. Surg Endosc. 2010;24(6):1287-95.
22. Zeinoddini A, Heidari R, Talebpour M. Laparoscopic gastric plication in morbidly obese adolescents: a prospective study. Surg Obes Relat Dis.

CHAPTER 3

Cohort Data with Dropout: a Simulation Study Comparing Five Longitudinal Analysis Methods

Rebecca K. Stellato[1], MS
Rutger M. van den Bor[1], PhD
Maria Schipper[1], PhD
Maud Y.A. Lindeboom[2], MD, PhD
Marinus J.C. Eijkemans[1], PhD

[1] Department of Data Science and Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2] Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Submitted

ABSTRACT

Background

A simulation study was performed to visually demonstrate the problems with repeated measures ANOVA (RMA) and *t*-tests (TT) compared to linear mixed effects (LME), covariance pattern (CP) or generalized estimating equations (GEE) models in longitudinal cohort studies with dropout.

Methods

Data were generated for a realistic, observational study on health-related quality of life (HRQoL) in a small, heterogeneous sample of children undergoing anti-reflux surgery. Each generated sample comprised two groups: one with low levels (4 - 10%) of random dropout (missing completely at random, MCAR); the other with higher levels (10 - 40%), where the chance of dropout depended on lower baseline HRQoL (missing at random, MAR). Outcome data were simulated for four time points in a one-year period, assuming in both groups small but meaningful increases in HRQoL between baseline and 3 months, and thereafter constant levels to 12 months.

Five analysis methods were applied to the simulated datasets: LME; CP; GEE; RMA and independent or paired TT, as appropriate, at successive time points. The bias in estimated marginal means was examined, and the coverage and width of 95% confidence intervals for, and the power of, three within- and between-group contrasts were examined.

Results

In the group with MCAR, negligible bias was observed in all methods, coverage was close to 95%, and little difference was seen in power among methods. In the group with MAR dropout, independent and paired TT and RMA analyses displayed increasing bias and decreasing coverage and power for increasing levels of dropout. The paired TT also produced the widest confidence intervals on average, with the greatest variability. GEE displayed slightly lower coverage and higher power than LME and CP models, but bias and precision were further comparable to LME and CP. LME and CP models produced unbiased results and close to 95% coverage, even in the case of 40% MAR dropout.

Conclusions

As expected, LME and CP models performed best in terms of bias and coverage even in the case of higher levels of MAR data. Paired TT and RMA produce biased results and poor coverage and precision in the presence of MAR data.

BACKGROUND

Repeated measures on an outcome are often collected in medical studies where the primary interest lies in change in the outcome over time. When longitudinal data on subjects are collected in a study, special analysis techniques are required to properly account for the correlation of repeated measures within individuals.[1-4] An additional statistical benefit to such designs is (in most circumstances) a decrease in the standard error (SE) of the treatment estimate, thereby increasing the power to detect a treatment effect.[1]

A common problem in longitudinal studies is missing outcome measures due to dropout, missed visits or questionnaires not returned. Missing data are traditionally classified into three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR).[5,6] In data that are MCAR, the probability of missingness of the outcome is the same for all individuals. The less restrictive MAR assumes that missingness depends on data that has been observed (for instance groups of patients and/or previous outcomes). When the missingness of an outcome is dependent on variables not observed, data are MNAR.[5,6] MCAR is an often unrealistic assumption for missing data in longitudinal studies.[6] The data analysis must therefore also take more complex types of missing data into account.

Repeated measures ANOVA (RMA) is still a common way of analyzing repeated measures in some medical fields, despite several well-known problems. Because RMA uses only cases with complete outcomes at all time points, some data are discarded. This type of complete case analysis can introduce bias in estimated means and/or mean differences, since respondents with complete data are rarely representative for the original sample,[3,7,8] i.e. the data are not MCAR. In addition to introducing bias, discarding data collected on humans is also scientifically unethical for several reasons, including (but not limited to) a potential loss of statistical power due to larger SEs. RMA further assumes equal variances of the outcome measurements over time, and equal correlations between outcomes at all time points, neither of which is a realistic assumption for longitudinal data.[3]

The use of paired and unpaired *t*-tests (TT) is also still relatively common in the analysis of longitudinal data. Paired TT are used to test change from baseline, and independent samples TT are sometimes used to compare groups across time points, though this practice has long been discouraged.[9,10] Because they make use of observed data at one time point (independent samples) or complete cases (paired samples), TT will also produce biased results in the presence of MAR data. TT are also expected to have less power than models that use all data from a subject.

More appropriate methods for longitudinal data are (generalized) linear mixed effects models (LME) or covariance pattern (CP) models, and generalized estimating equations (GEE). All three models use all available data on outcome measures, hence increasing power, and provide unbiased estimates in the presence of data that is MCAR;[7,8] LME and CP models also provide unbiased estimates in the case of MAR data.[3,7,8]

Because the more technical literature on statistical theory is not generally read by non-statistical audiences, several articles/tutorials aimed at a wider audience have demonstrated the bias of RMA and GEE under MAR in real-life data.[8,11] However, because the data generating mechanism is not known in such examples, these demonstrations cannot be considered proof of which method(s) actually produced unbiased estimates.

In order to visually demonstrate the bias in RMA and TT (paired and unpaired) to pediatric surgeons, we performed a simulation study in which we varied type (MCAR and MAR) and amount of monotonic missingness (in the form of dropout) for a continuous, longitudinally measured outcome. This was done for a hypothetical, but realistic, observational study on anti-reflux surgery (ARS) in children.

METHODS

Aims

To demonstrate the bias of methods of analysis that use complete cases (RMA, paired TT) and methods that use complete data at one moment in time (independent TT between groups), compared to more appropriate longitudinal models such as LME, CPM or GEE. In addition to bias, precision and power of the methods are examined.

Data-generating mechanisms

Based on results of several pediatric ARS studies,[12-17] longitudinal health-related quality of life (HRQoL) data were simulated for 25 neurologically normal (NN) and 25 neurologically impaired (NI) children at 0, 3, 6 and 12 months, mimicking patterns that might occur after fundoplication.

Children in group 1 (NN) were assumed to have a mean HRQoL of 75 at time 0, increasing to 80 at 3 months and remaining stable to one year post-intervention. Children in group 2 (NI) were assumed to have a mean HRQoL of 50, increasing to 60 at 3 months and remaining stable to 12 months. In both groups, a multivariate normal distribution was used, with means as mentioned, a standard deviation of 15 points, and a month-to-month autocorrelation of 0.85 (or a correlation between baseline and one-year HRQoL of $0.85^{12} = 0.15$). The MASS[18] package (v7.3.58.1) for R was used to simulate the datasets. In some simulations, HRQoL for individuals exceeded 100. Although this is not possible in practice (most QoL scores are restricted between 0 - 100), in order

to demonstrate the bias due to missingness (and not due to ceiling effects) we did not truncate HRQoL scores.

Monotone missing data (in the form of dropout) was generated for both groups using the exponential distribution. In group 1 (NN), dropout was assumed to be completely at random, and percentages varied from 4% to 10% within one year. Rates of dropout for the NI group were assumed to be higher (10% to 40%), and dependent on the baseline HRQoL (MAR): children with a low score were expected to drop out more often. In the simulations, children with a HRQoL score lower than the median were assumed to be four times more likely to have missing data than children with a score above the median.

Targets of analysis

The primary target is the estimated mean HRQoL at all time points, for NN and NI separately, on the basis of estimated marginal means (EMM) for LME, CP, GEE and RMA models, and observed means for TT.

The secondary targets are three contrasts that could be of interest in such a study: the change from baseline to one year in each group, and the difference between the groups at 12 months follow-up.

Methods

Each generated dataset was analyzed in five ways:

1. TT, *t*-tests (paired *t*-tests: difference in mean HRQoL 12 months vs 0 months for each group; unpaired: Welch's *t*-test for the difference in mean HRQoL NN vs NI at 12 months)
2. RMA, using the afex[19] package (v1.2.0)
3. GEE with a first-order autoregressive working correlation matrix, using the geepack[20] package (v1.3.9)
4. CPM with a continuous first-order autoregressive correlation structure and homogeneous variances over time, using the nlme[21] package (v3.1.160)
5. LME model with a random intercept and a random slope for time (modelled as continuous), also using the nlme package

For the TT the usual estimates, SEs, and degrees of freedom (df) were used. The latter four models included an effect for group, an effect for time (modelled as categorical), and an interaction between the two. For each model, the resulting mean HRQoL for the two groups at the different time points (estimated marginal means) and their SEs and df were estimated using the emmeans[22] package (v1.8.3). For RMA and LME models, the default options in emmeans were used (denominator df for RMA and Kenward-Roger df for LME[22,23]). For CPM, approximate Satterthwaite df were used, and for the GEE the robust variance-covariance option was used.

The simulations and analyses were performed in R 4.0.3.[24] R code for the generation of the datasets and the analyses can be found in Appendices 1-3.

Performance measures

Bias was estimated by comparing the EMMs to the true population means. For precision, the empirical SE (EMM and contrasts), % gain in relative efficiency (EMM and contrasts) for all methods compared to LME, and coverage of 95% CI's for the contrasts were assessed. The power of the three contrasts was also examined. Means or percentages of these performance measures were plotted with 95% CI's based on Monte Carlo SEs (MCSE).[25] Widths of the 95% CI's for the three contrasts were also examined. Finally, any convergence problems with the LME, CPM, GEE or RMA models were described. Analysis of the simulations was performed using the rsimsum[26] (v0.11.3) and ggplot2[27] (v3.4.0) packages. The exception was power of the contrasts, which was calculated outside of rsimsum (at the time of writing the package used infinite df instead of model df for power).

Number of simulations

Since no estimate of the variance of the bias was available ahead of time, coverage was used to determine n_{sim} , the number of simulated datasets. Using the formula for the MCSE of coverage[25], 7600 datasets were generated in order to achieve a coverage MCSE of 0.25%.

RESULTS

Bias

Fig. 1 displays the bias in estimated means for the five methods in the two groups. In group 1 (MCAR) all five methods performed similarly in terms of bias: very close to 0 at all time points and for all scenarios (4% - 10% MCAR dropout). In group 2 (MAR), GEE, CPM and LME all performed similarly, with bias very close to 0 for all times and all scenarios. The independent TT displayed some bias at all times beyond baseline, with increasing bias (up to about one point) as dropout increased. The RMA analysis (based on complete cases) displays bias at all times (highest at time 0) and all scenarios, with means being estimated at 0.7 - 4.7 points above the true means in the population with 40% dropout.

Bias in the contrasts, which can also be deduced from the bias in the estimated means from Fig. 1, are presented in Fig. 2. There was no bias in the contrasts in any method for the difference between 0 and 12 months for the MCAR group, and bias of GEE, CPM and LME analyses was negligible for the within-MAR group contrast. Paired TT and RMA displayed the same amount of bias (0.7 - 4.7 points) in the MAR group. Some bias (less than one point) was seen between the groups at 12 months for RMA and independent TT for higher levels of dropout.

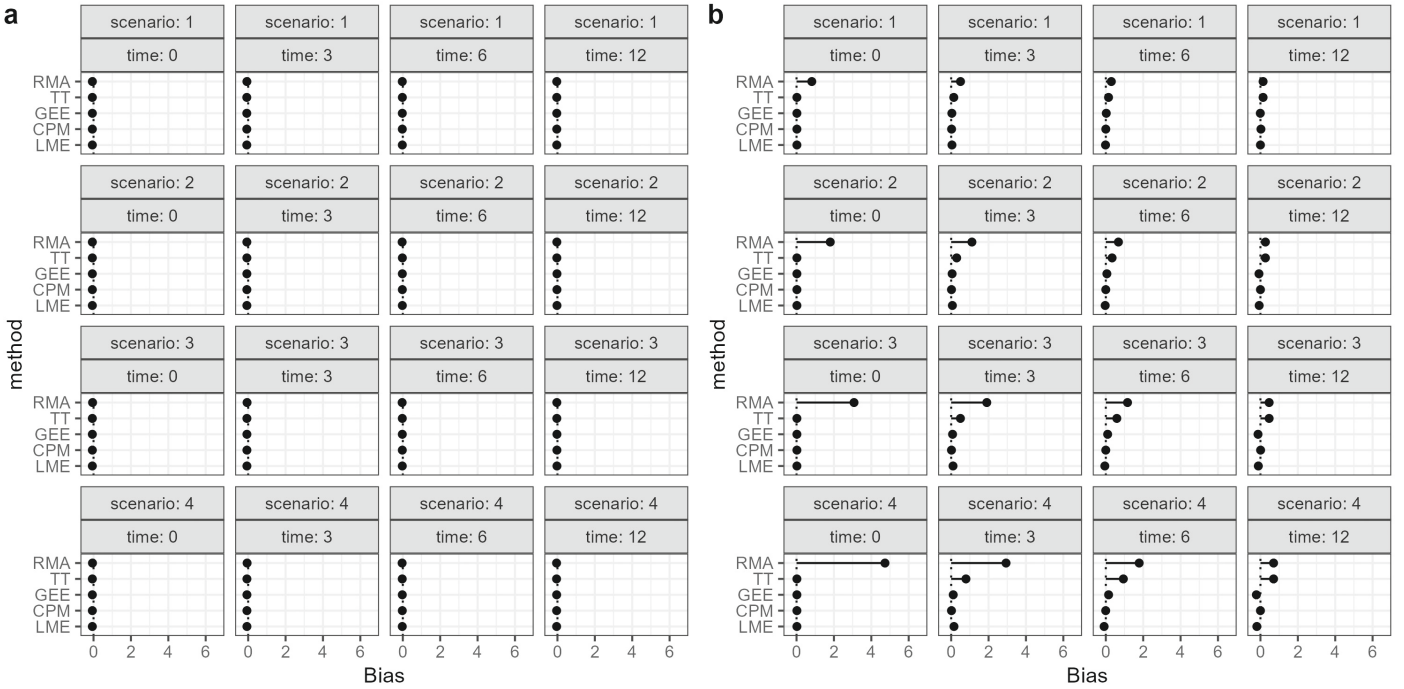


Figure. 1 Bias (with 95% Monte Carlo CI†) in estimated marginal means in four scenarios.

On the basis of 7600 simulations using 25 patients in each group.
(a) Group 1, MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. (b) Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: independent t-tests; GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
†Monte Carlo CI's not visible because width within size of point estimate.

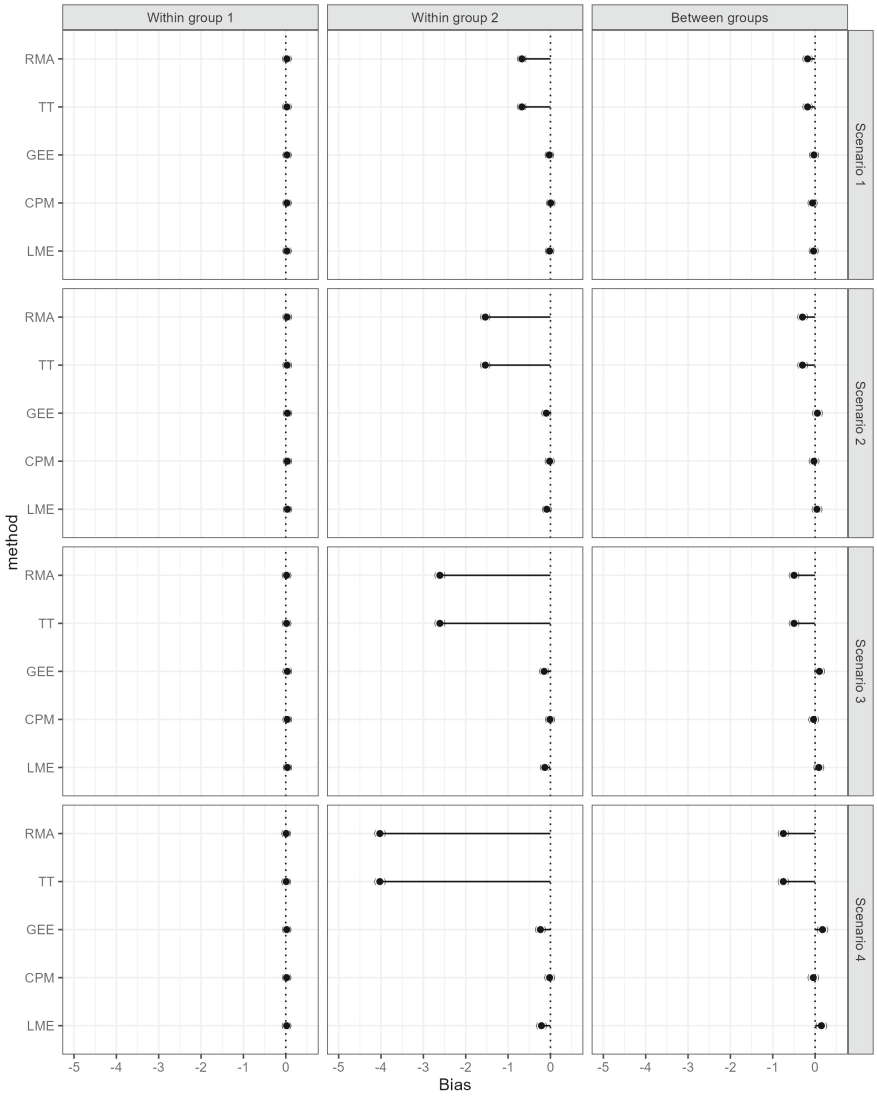


Figure 2. Bias (with 95% Monte Carlo CI†) in the within- and between-group contrasts in four scenarios.

On the basis of 7600 simulations using 25 patients in each group.
Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
†Monte Carlo CI's not visible because width within size of point estimates.

Precision

Empirical SE and relative precision

The empirical SE of the estimated means for RMA was slightly higher than other methods for the MCAR group (except at 12 months), and that pattern was more pronounced in the MAR group (Fig. 3). At 12 months, the empirical SE for LME was also slightly increased in the scenarios with 30-40% MCAR compared to other methods, and to lower levels of missingness. There was up to 10% loss in relative precision for RMA (compared to LME) in the MCAR group, and up to 40% loss in the MAR group (Fig. 4).

Fig. 5 displays the empirical SE for the three contrasts of interest. The empirical SE was larger for TT & RMA for all contrasts, though this was most pronounced in MAR group. In the between-group contrasts the difference in empirical SE's for the five methods was negligible. There was 5-16% loss in relative precision of RMA & TT compared to LME in the MAR group. GEE and CPM had a 2-3% gain in relative precision compared to LME for the within- MAR-group contrast and the between-group contrast (Fig. 6).

Coverage

The coverage of the 95% CI's for the within- and between-group contrasts is presented in Fig. 7. For all three contrasts and all scenarios, coverage was similar for CPM and LME, and very close to 95%. The coverage for GEE is consistently lower for all contrasts at all time points, though never lower than 92.3%. The coverage of the other methods was similar and very close to 95% for the 0-12 month increase in group 1 (MCAR). For the same contrast in group 2 (MAR), coverage became progressively worse for the paired TT (88.3 - 94.6%) and RMA (88.5 - 94.9%) as dropout increased. Coverage of the between-group contrast (difference in HRQoL between the two groups at 12 months) was at or just above 95% or more for all methods but GEE.

Width of 95% confidence intervals for contrasts

Widths of the 95% CI's for three contrasts are presented in Fig. 8. In all scenarios and for all contrasts, the CPM's gave, on average, the narrowest 95% CI's, and the most consistent results (smallest range in CI widths). There was very little difference in the median or range of widths of CI's for LME and RMA for the first and third contrasts; for the within-group MAR contrast, RMA has wider CI's than LME, CPM or GEE. The paired TT gave larger CI widths and the widths displayed more variation (with outliers towards wider CI's) than the others methods, especially in the group with 40% MAR dropout. GEE had relatively low median widths, but more variation in widths.

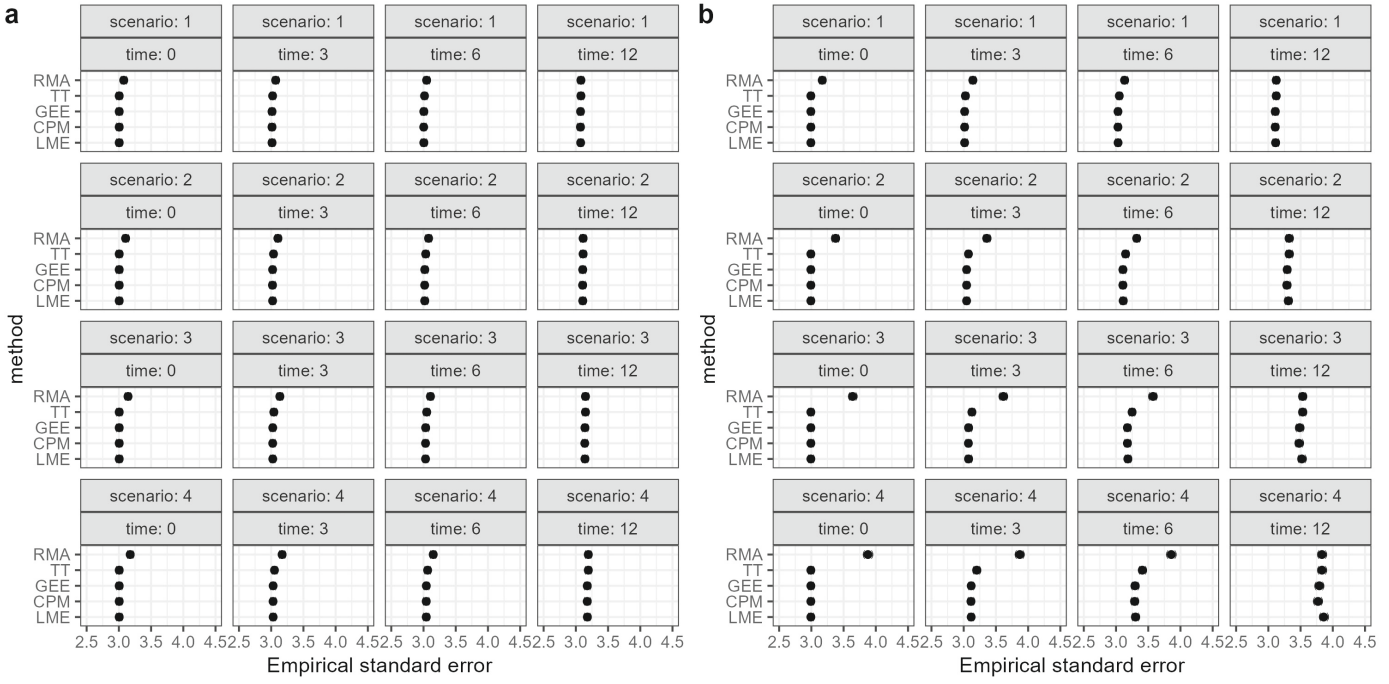


Figure 3. Empirical standard errors (with 95% Monte Carlo CI†) in estimated marginal means in four scenarios.

On the basis of 7600 simulations using 25 patients in each group.
(a) Group 1, MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. (b) Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: independent t-tests; GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
†Monte Carlo CI's not visible because width within size of point estimates.

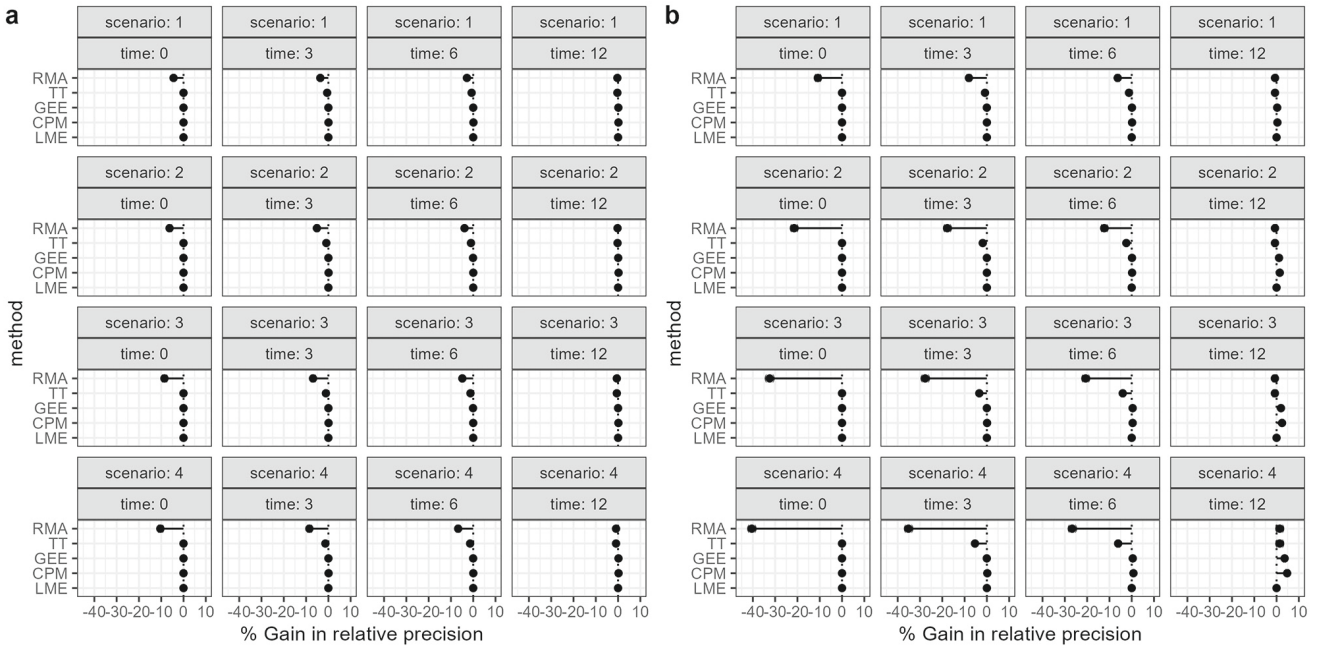


Figure 4. Percentage gain in relative precision‡ (with 95% Monte Carlo CI†) of the estimated marginal means in four scenarios.

On the basis of 7600 simulations using 25 patients in each group.
(a) Group 1, MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. (b) Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: independent t-tests; GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
‡Compared to LME
†Monte Carlo CI's not visible because width within size of point estimates.

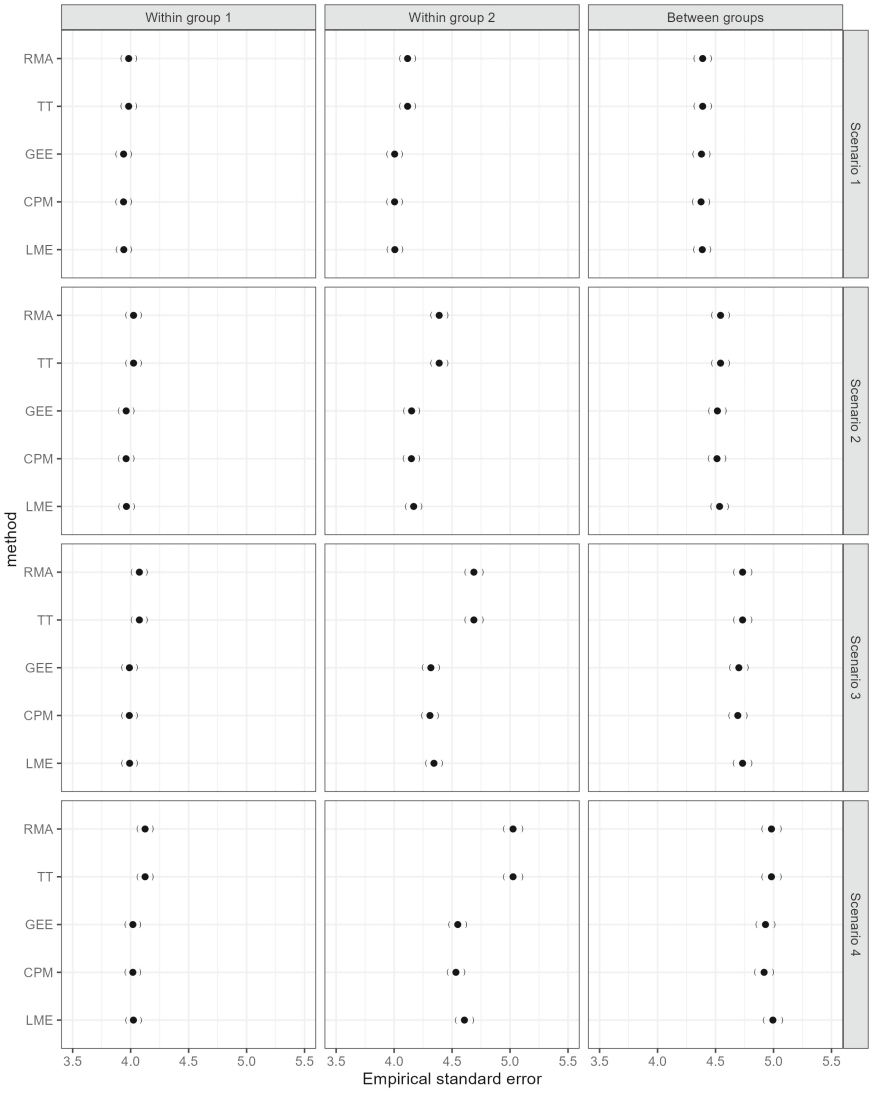


Figure 5. Empirical standard error (with 95% Monte Carlo CI) for the within- and between-group contrasts.

On the basis of 7600 simulations using 25 patients in each group.
Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.

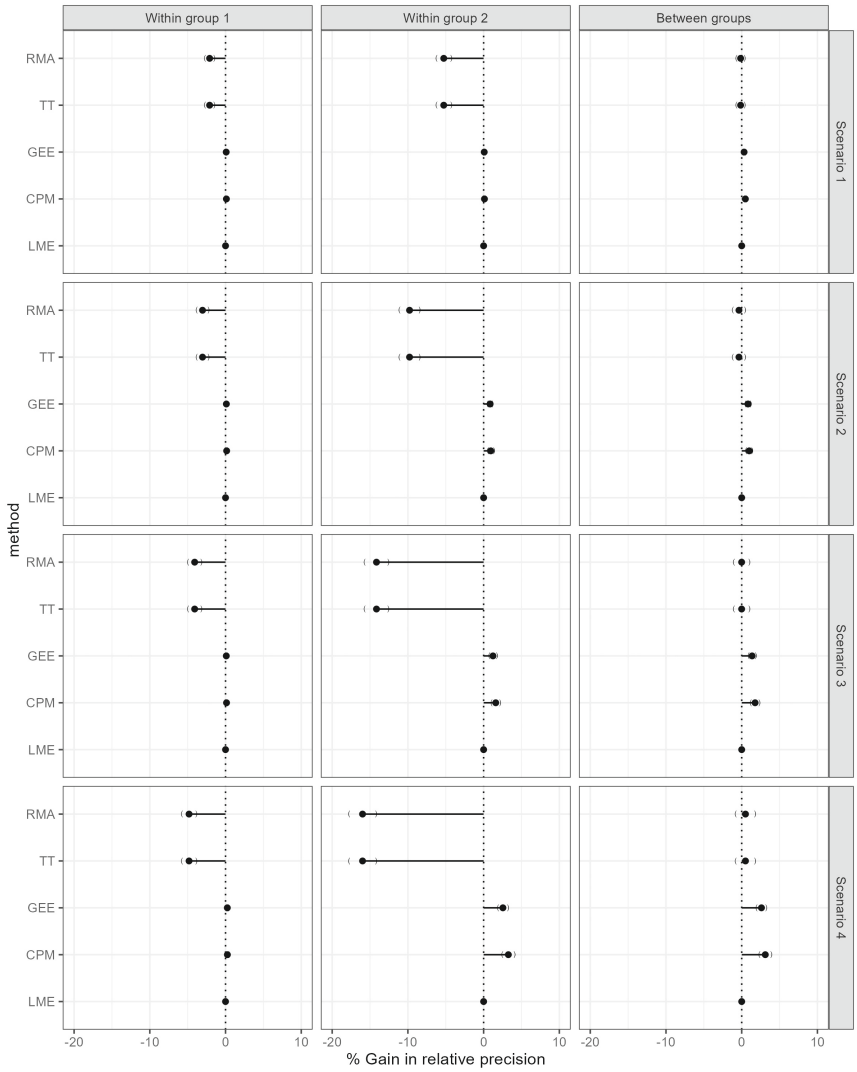


Figure 6. Percentage gain in relative precision‡ (with 95% Monte Carlo CI) in the within- and between-group contrasts in four scenarios

On the basis of 7600 simulations using 25 patients in each group.
Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
‡Compared to LME

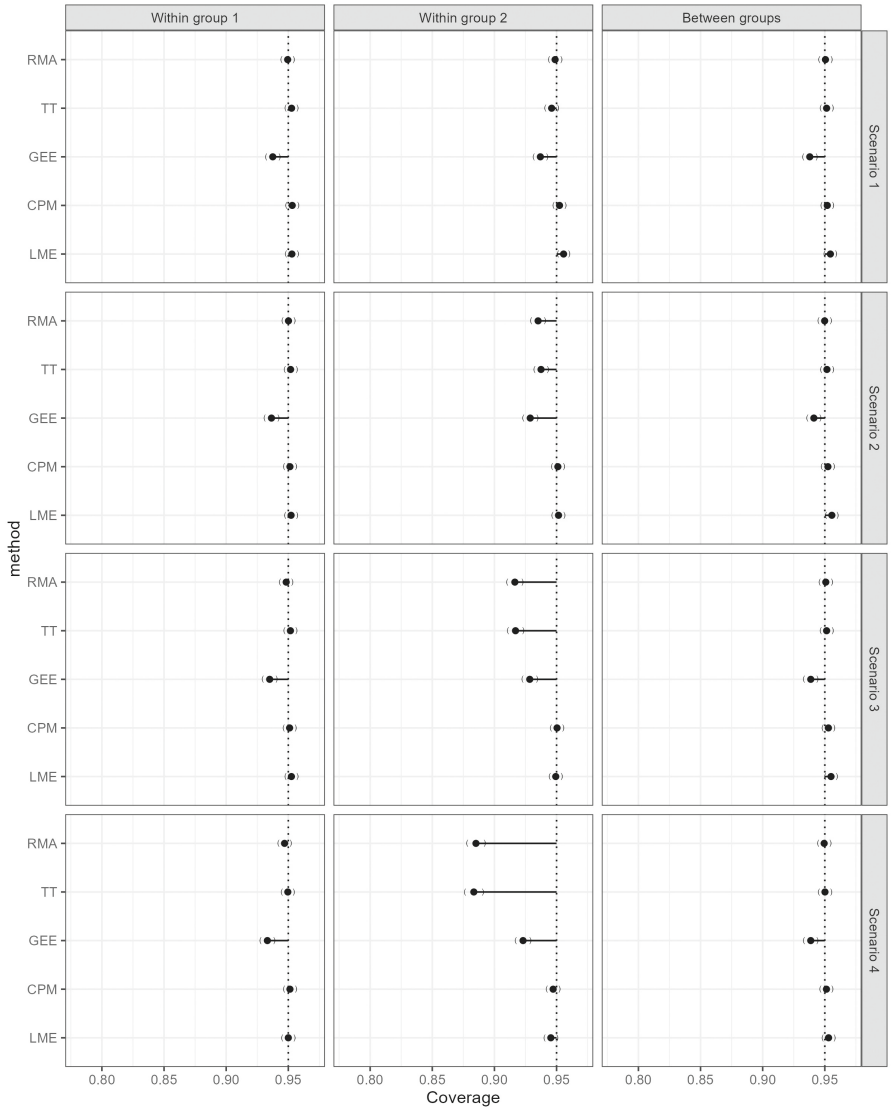


Figure 7. Coverage (with 95% Monte Carlo CI) of the 95% CI's for the within- and between-group contrasts.

On the basis of 7600 simulations using 25 patients in each group.
Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.

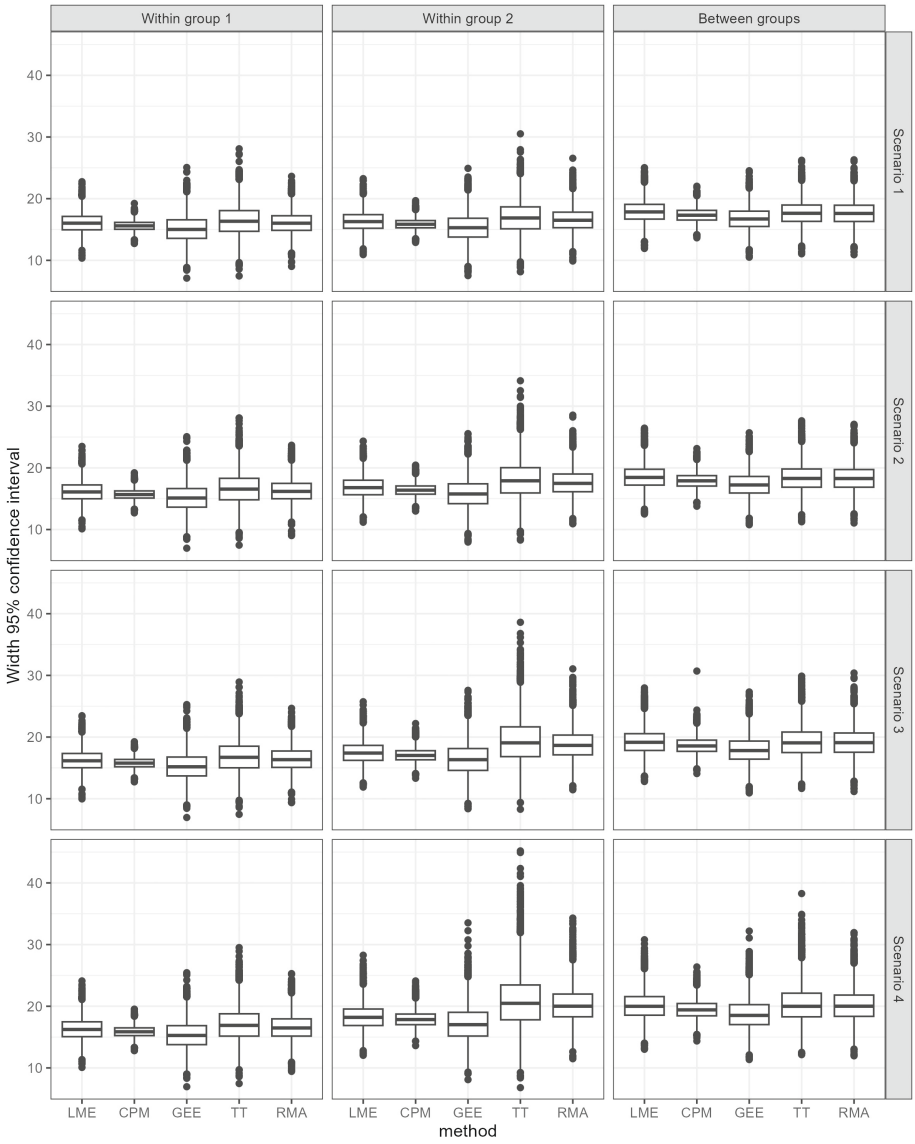


Figure 8. Width of 95% CI's for the within- and between-group contrasts.

For two of the four scenarios, on the basis of 7600 simulations using 25 patients in each group. Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.

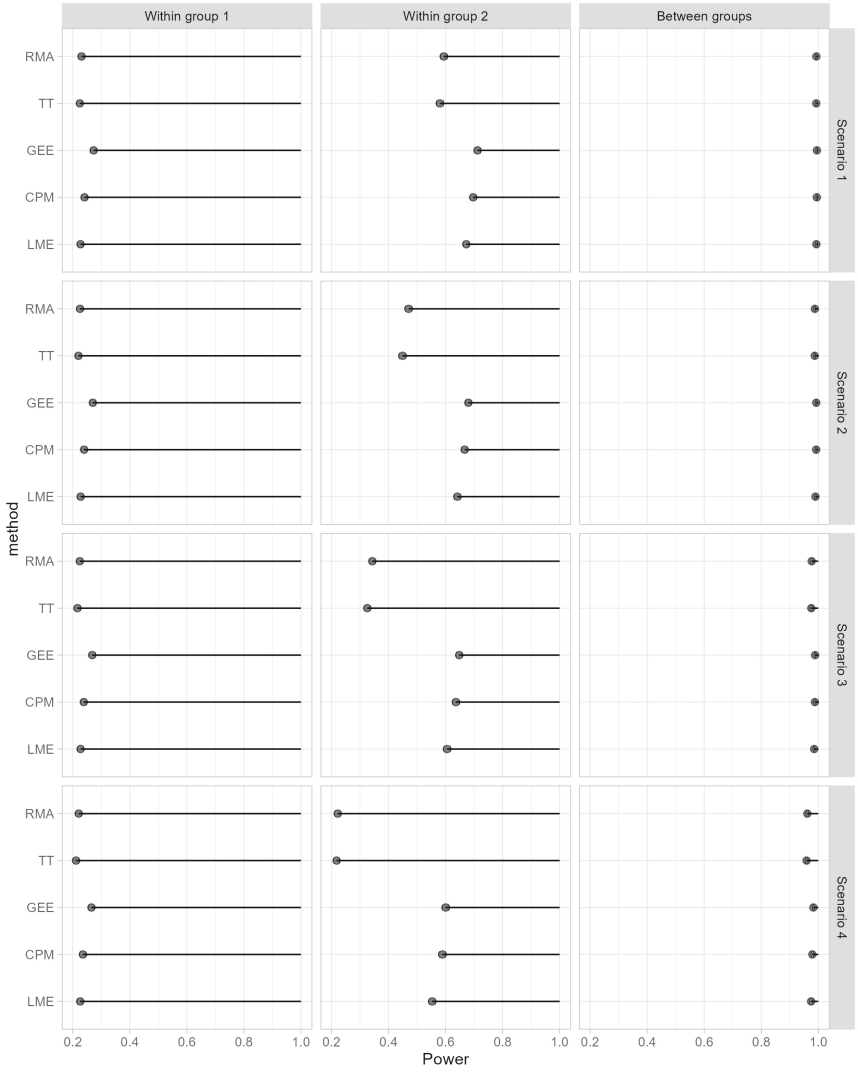


Figure 9. Power (with 95% Monte Carlo CI†) of the within- and between-group contrasts.

On the basis of 7600 simulations using 25 patients in each group.
Column 1: group 1 contrast 0 - 12 months; column 2: group 2 contrast 0 - 12 months; column 3: group 1-group 2 contrast at 12 months. Group 1 MCAR dropout scenarios 1: 4%, 2: 6%, 3: 8% and 4: 10%. Group 2, MAR dropout scenarios 1: 10%, 2: 20%, 3: 30% and 4: 40%.
RMA: repeated measures ANOVA; TT: paired sample t-tests (within-group contrasts) or independent samples t-tests (between-group contrasts); GEE: generalized estimating equations; CPM: covariance pattern model; LME: linear mixed effects model.
†Monte Carlo CI's not visible because width within size of point estimates.

Power

The power of the contrasts is presented in Fig. 9. Power for the within-group contrasts was low for both groups and lower in group 1 than in group 2, consistent with the size of the increase in HRQoL for the two groups. No large differences in power were observed for group 1 (MCAR), although GEE had the highest power (approximately 0.27) across all scenarios. In group 2 (MAR) the paired TT and RMA had considerably lower power than the other three methods. Power decreased as dropout increased for all methods, though most markedly for TT and RMA. GEE again had the highest power in all scenarios, followed by CPM and LME. The power for the between-group contrast was very close to one for all methods in all scenarios, though power of RMA and independent TT decreased slightly for the higher levels of missingness.

Convergence problems

The LME analyses produced warnings about convergence in 0.03%, 0.05%, 0.05% and 0.11%, respectively, of the simulations in dropout scenarios 1-4, though it was still possible to estimate the models and resulting EMMs and contrasts. No convergence problems were encountered with the other 4 analysis methods. However, in two simulations with higher missings (one each in scenarios 3 and 4), an error was encountered in the EMMs for CPM. In those simulations, error df had to be used instead of the approximate Satterthwaite df.

DISCUSSION

Previous articles and books have presented theoretical arguments for the use of linear mixed effect models, CPM or GEE for analyzing longitudinal data.[1,3,8] A few studies have demonstrated the bias of TT and RMA on existing datasets.[8,11] To our knowledge, this is the only study comparing more advanced methods (LME, CPM, GEE) to the traditional methods (TT, RMA) using simulated datasets.

While the independent TT for the between-group contrast did not induce much bias and had coverage close to 95% even with high levels of dropout, the precision of the paired TT for increase in HRQoL tended to be lower than for the same contrast from a LME, CPM or GEE model: TT had consistently wider CI's than the CP and GEE analyses for the group with MAR dropout. To a lesser extent this was also the case in the group with lower levels of MCAR dropout, though in that group the highest level of dropout was 10%. Increasing dropout would likely have further decreased the precision of the paired TT in the MCAR group as well.

As expected, RMA and paired TT gave biased estimates and poorer coverage for the group with MAR. This is primarily due to the fact that these methods employ listwise deletion, removing every individual with one or more missing outcomes. RMA also depends on unrealistic assumptions about correlations of repeated measures over

time.[3,7] Surprisingly, then, the 95% CI's for contrasts after RMA were relatively narrow, and were only slightly wider than the LME CI's in the within-MCAR and between-group contrasts. This could be due to the lower levels of missingness in the group with MCAR.

While the bias in GEE was very similar to LME and CPM, the coverage for all contrasts and scenarios was slightly lower, and the power slightly higher, for GEE than for LME or CP models. This is likely due to the use of z-testing in the GEE contrasts, which in these small samples resulted in narrower 95% CI's.

Results for the CP models were nearly identical to those of the LME models for both bias and coverage, with better precision for CP models. The major drawback to LME was the occasional problem with convergence. Also, some small problems were seen in precision of LME, likely because linear random effects were applied to non-linear trends. The use of linear random effects for time is common even when the overall (fixed) time trend is not completely linear. It is interesting to note that this method may produce less precise estimates than a CP model, especially at later time points.

Strengths and Limitations

The advantage of using simulations is that the data-generating mechanism is known. It is therefore possible to calculate average bias, precision, and coverage for the different methods. The current study used the ADEMP framework[25] for design and reporting.

Several limitations of the current study must be addressed. Generated HRQoL values were not truncated at 0 and100, because the purpose of the study was to demonstrate the performance of the five methods on normally distributed outcomes. If HRQoL values had been subject to a ceiling of 100 (as they are in practice), all methods discussed here would have produced biased results for the group with high levels of HRQoL. In the case of ceiling or floor effects, other models are recommended.[28,29]

Another assumption made for reasons of simplicity was that the variance of HRQoL was the same for both groups at all time points. This may not be a realistic assumption for longitudinal data collected in disparate patient groups (e.g. NN vs NI). LME, CPM and independent TT can easily be adjusted to account for differences in variances, and GEE with a robust estimator should account for heterogeneity. Future simulation studies could vary heterogeneity between groups and/or time points to compare models that are corrected for heterogeneity to those that are not.

To ground the simulation in a realistic setting, low levels of missing were used in the NN group. Since this was also the group in which we assumed MCAR, it was difficult to assess the effect of large levels of MCAR on precision. That the missingness only depended on the baseline HRQoL was also a simplification for purposes of demonstration; in

reality, missing data will depend on more factors. A study using more complicated MAR structures and comparing similar levels of MCAR and MAR data could help clarify this.

Power for the between-group contrast was quite high and nearly indistinguishable across the methods. This was primarily due to the assumption of a 20-point difference between the groups (equivalent to a difference of one standard deviation). A smaller difference between groups, though less realistic for this particular situation, would have elucidated the differences in power across the methods.

Finally, no multiple imputation method was used. RMA, GEE and TT might all have performed better, certainly in terms of bias, if missing data were first imputed. A previous study found differences in estimates from GEE before and after multiple imputation,[8] though that study analyzed an existing dataset for which the true data-generating mechanism was not known and could therefore not estimate bias of either method.

CONCLUSIONS

As expected, LME and CP models performed best in terms of bias and coverage even in the case of higher levels of MAR data. CPM slightly outperformed LME in this study, likely due to the non-linear trends over time. All methods gave fairly comparable results for low levels of MCAR data. Paired TT and RMA produce biased results and poor coverage and precision in the presence of MAR data. GEE produced unbiased results for both MCAR and MAR data, but slightly too narrow 95% CI, resulting in slightly poorer coverage and exaggerated power.

ACKNOWLEDGEMENTS

We are grateful to Dr. Peter van de Ven for his assistance in brainstorming about the data generating mechanism.

REFERENCES

1. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.
2. Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics*. 1982;38:963-974.
3. Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. *Circulation*. Nov 4 2008;118(19):2005-10.
4. Albert PS. Tutorial in biostatistics: Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*. 1999;18:1707-32.
5. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Second ed. John Wiley & Sons; 2002.
6. van Buuren S. *Flexible Imputation of Missing Data*. Second ed. Chapman and Hall/CRC; 2018:444.
7. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. *Annu Rev Clin Psychol*. 2010;6:79-107.
8. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. Oct 2014;23(5):440-59.
9. Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *Br Med J*. 1990;300:230-5.
10. Omar RZ, Wright EM, Turner RM, Thompson SG. Analyzing repeated measurements data: a practical comparison of methods. *Stat Med*. 1999;18:1587-1603.
11. Twisk JWR. Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. *Eur J Epi.* 2004;19:769–776.
12. Kawahara H, Okuyama H, Kubota A, et al. Can laparoscopic antireflux surgery improve the quality of life in children with neurologic and neuromuscular handicaps? *J Pediatr Surg*. 2004;39(12):1761-1764.
13. Srivastava R, Downey EC, Feola P, et al. Quality of life of children with neurological impairment who receive a fundoplication for gastroesophageal reflux disease. *J Hosp Med*. May 2007;2(3):165-73.
14. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270 degrees fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. *Surg Endosc*. Jun 2010;24(6):1287-95.
15. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
16. Knatten CK, Kvello M, Fyhn TJ, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. *J Pediatr Surg*. Jul 2016;51(7):1115-21.
17. Stellato RK, Mulder FVM, Tytgat SHA, et al. Two-year outcome after laparoscopic fundoplication in pediatric patients with gastroesophageal reflux disease. *J Laparoendosc Adv Surg Tech A*. 2020;in press
18. Venables WN, Ripley BD. *Modern Applied Statistics with S. Fourth edition.* Springer; 2002.
19. Singmann H, Bolker B, Westfall J, Aust F, Ben-Shachar M. afex: Analysis of Factorial Experiments. R package version 1.2-0. 2022. https://CRAN.R-project.org/package=afex
20. Højsgaard S, Halekoh U, J Y. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*. 2006;15(2):1-11.
21. Pinheiro J, Bates D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-160. 2022. https://CRAN.R-project.org/package=nlme
22. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.3. 2022. https://CRAN.R-project.org/package=emmeans

23. Luke SG. Evaluating significance in linear mixed-effects models in R. *Behav Res Methods*. Aug 2017;49(4):1494-1502.
24. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2022. https://www.R-project.org/
25. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. May 20 2019;38(11):2074-2102.
26. Gasparini A. rsimsum: Summarise results from Monte Carlo simulation studies. *Journal of Open Source Software*. 2018;(26)
27. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag 2016.
28. Twisk J, Rijmen F. Longitudinal tobit regression: a new approach to analyze outcome variables with floor or ceiling effects. *J Clin Epidemiol*. Sep 2009;62(9):953-8.
29. Wang L, Zhang Z, McArdle JJ, Salthouse TA. Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behav Res*. Jul 1 2009;43(3):476-496.

APPENDICES

1. Generating the datasets (R base)

```
# O:/Projects/Simulation/07 Programs/Analyze data/simulations influence LTFU on long data make datasets20221118.R
# Author: Rebecca Stellato
# Date: 18 November 2022
# Aim: simulate data from a one-arm trial with 50% NI and varying degrees of
 # missings/LTFU
# Vary % of missing data at end of follow-up for NI group (10, 20, 30 and 40%),
 # and 4, 6, 8, 10% for NN
# Dependency of missingness on baseline QoL for NI group: differing
 # percentages of missing, with 4x more missing for NI kids with baseline
 # QoL < med than for NI kids with baseline QoL > med.
# Using a file (already made, code included but commented out) of random seeds
 # for each simulation
# Note: I am not truncating QoL scores >100 or <0
# Using 7600 simulations, which should be enough for a Monte Carlo SE of
 # 0.25% on 95% coverage intervals (Morris et al 2019)

# Clear workspace
rm(list=ls())

# Set path for input
inpath <- "O:/Projects/Simulation/06 Data/Raw"

# Set path for output
outpath <- "O:/Projects/Simulation/06 Data/Raw"

## Required libraries
library(MASS)          ## For mvrnorm()
library(ggplot2)

##############################################################################################
## Data simulation function
##############################################################################################

datsim <- function(
 n1 = 25,                    ## Size of first group (NN)
 n2 = 25,                    ## Size of second group (NI)
 times = c(0, 3, 6, 12),     ## Measurement times in months
 mu1 = c(75, 80, 80, 80),    ## Assumed mean at each time point (group 1)
mu2 = c(50, 60, 60, 60),     ## Assumed mean at each time point (group 2)      sd1 = 17.5 ,
                             ## The standard deviation group 1 (assumed constant across time)
 sd2 = 15,                   ## The standard deviation group 2 (assumed constant across time)
```

```
 corr.1 = .85,                ## The assumed correlation for two measurements one month apart
 plots = T,                   ## Logical; should the correlation function and means in both groups be plotted?
 miss1 = 0.05,                ## Pct missing in group 1
 miss2 = 0.05                 ## Pct missing in group 2
){

 ## Basic input checks
 if(length(mu1) != length(mu2)) stop('length(mu1) != length(mu2)')
 m <- length(mu1)
 if(length(times) != m) stop('mu1, mu2, and times should have equal length')
 if(length(n1) != 1) stop('length(n1) != 1')
 if(length(n2) != 1) stop('length(n2) != 1')
 if(length(sd1) != 1) stop('length(sd1) != 1')
 if(length(sd2) != 1) stop('length(sd2) != 1')
 if(length(corr.1) != 1) stop('length(corr.1) != 1')
 if(length(miss1) != 1) stop('length(miss1) != 1')
 if(length(miss2) != 1) stop('length(miss2) != 1')
 if(miss1 <0 | miss1 >1) stop('miss1 must be between 0 and 1')
 if(miss2 <0 | miss2 >1) stop('miss2 must be between 0 and 1')

 ## Define covariance matrix (based on power function (= continuous AR(1))
 s <- expand.grid(times, times)
 s <- cbind(s, diff = abs(s[,1] - s[,2]))
 if(plots){
 par(mfrow = c(1, 2))
 t.seq <- seq(0, max(c(12, max(diff(times)))), length.out = 1000)
 plot(corr.1^t.seq ~ t.seq, type = 'l', xlab = 'Time difference (months)', ylab = 'Correlation', lwd = 2, ylim = c(0,
1), main = 'Assumed correlation function'); grid()
 abline(v = s$diff, col = 'red', lty = 'dashed')
 }
 s <- matrix(corr.1^s$diff, nrow = m, ncol = m)
 s1 <- s * sd1^2
 s2 <- s * sd2^2

 ## Data for group 1
 d1 <- mvrnorm(n = n1, mu = mu1, Sigma = s1)
 d1 <- data.frame(id = factor(rep(1:n1, times = m)),
 wave = factor(rep(1:length(times), each = n1)),
 t = factor(rep(times, each = n1)),
 time = rep(times, each = n2),
 yW = do.call(c, lapply(1:ncol(d1), FUN = function(i) d1[,i])))
 d1 <- d1[order(d1$id, d1$t),]
 d1$group <- 1
 # Add censoring
 lambda1 <- -log(1-miss1)/max(times)
 cens1 <- data.frame(id = factor(1:n1), ct = rexp(n1, lambda1))
```

```
d1 <- merge(d1,cens1, by="id", all = TRUE)

## Data for group 2
d2 <- mvrnorm(n = n2, mu = mu2, Sigma = s2)
d2 <- data.frame(id = factor(rep((n1+1):(n1+n2), times = m)),
wave = factor(rep(1:length(times), each = n2)),
t = factor(rep(times, each = n2)),
time = rep(times, each = n2),
yW = do.call(c, lapply(1:ncol(d2), FUN = function(i) d2[,i])))
d2 <- d2[order(d2$id, d2$t),]
d2$group <- 2

## Let censoring in NI group depend on baseline QoL
# take out baseline QoL, add censoring variable, and merge y0 and ct back
# Missingness in > med is set 5%, in < med set to remaining missing
# (2*miss2-0.05)
base.y2 <- subset(d2, t==0, select = c("id", "yW"))
names(base.y2)[names(base.y2) == 'yW'] <- 'y0'
base.y2$lambda2 <- ifelse(base.y2$y0 >= median(base.y2$y0),
-log(1-(miss2*2/5))/max(times),
-log(1-(miss2*8/5))/max(times) )

base.y2$ct <- rexp(n=length(base.y2$lambda2), rate=base.y2$lambda2)

d2 <- merge(d2, base.y2, all = TRUE, by = "id")
d2 <- subset(d2, select = -c(y0,lambda2))

## Combined data
d <- rbind(d1, d2)
d$group <- factor(as.character(d$group))

## Set all times larger than censoring time to NA
d$y <- ifelse(d$time < d$ct, d$y, NA)

## Plot of means + sds
if(plots){
mu1.hat <- sapply(times, FUN = function(i) mean(d$y[d$group == 1 &
d$t == i], na.rm=TRUE))
mu2.hat <- sapply(times, FUN = function(i) mean(d$y[d$group == 2 &
d$t == i], na.rm=TRUE))
mu.all.hat <- sapply(times, FUN = function(i) mean(d$y[d$t == i],
na.rm=TRUE))
sd1.hat <- sapply(times, FUN = function(i) sd(d$y[d$group == 1 &
d$t == i], na.rm=TRUE))
sd2.hat <- sapply(times, FUN = function(i) sd(d$y[d$group == 2 &
d$t == i], na.rm=TRUE))
```

```
sd.all.hat <- sapply(times, FUN = function(i) sd(d$y[d$group == 2 &
d$t == i], na.rm=TRUE))
plot(NULL, xlim = range(times), ylim = c(30,100),
main = 'Observed means and SDs', ylab = 'Mean (+- 1 SD)',
xlab = 'Time (months)'); grid()
points(mu1.hat ~ c(times - 0.1), pch = 16, type = 'o', col = 2)
points(mu2.hat ~ c(times + 0.1), pch = 16, type = 'o', col = 3)
points(mu.all.hat ~ c(times), pch = 16, type = 'o', col = 1)
segments(x0 = times - 0.1, y0 = mu1.hat - sd1.hat, y1 = mu1.hat +
sd1.hat, col = 2)
segments(x0 = times + 0.1, y0 = mu2.hat - sd2.hat, y1 = mu2.hat +
sd2.hat, col = 3)
segments(x0 = times, y0 = mu.all.hat - sd.all.hat, y1 = mu.all.hat +
sd.all.hat, col = 1)
legend('bottomright', c('All kids', 'Group 1 (NN)', 'Group 2(NI)'),
col = c(1,2,3), lwd = 1, pch = 16, bg = 'white')

}

## Make wide version of data frame for paired t-test & RM ANOVA
## Remove incomplete cases (listwise deletion) & make long version again
dw <- reshape(d[c(1,3,6,8)], v.names = "y", idvar = "id",
timevar = "t", direction = "wide")
dcc <- na.omit(dw)
lcc <- reshape(data=dcc, idvar="id", varying = 3:6, v.names="y",
times = times, timevar = "t", direction = "long")
lcc$t <- factor(lcc$t)

## Delete censoring time & "wild" outcome
d <- d[-c(5,7)]

## Make & return list and end function
dwl <- list(d, lcc)
return(dwl)
#return(d)
}

# File of 10,000 random seeds, for reproducible simulated datasets
# Do this only once!!
# set.seed(220822)
# write(sample(1:10000000,size=10000), file=file.path(inpath, "seeds.txt"),
# ncolumns=1)

##############################################################################################
## Simulations - settings
##############################################################################################
```

```
Number of simulations
 numsim <- 7600        ## Enough for a MC SE of 0.25%

 ## File with seeds
 seeds <- read.table(file=file.path(inpath, "seeds.txt"),header=F)$V1

##############################################################################################
## Simulations - scenario 1, n = 25 (per group), 4% missing NN, 10% NI
##############################################################################################
 scnum <- 1
 # Percent missings NN & NI
 missNN <- 0.04
 missNI <- 0.1

 for (i in 1:numsim){
 # simulate data
 set.seed(seeds[i]) #sets the seed to the ith element in the vector seeds
 dwl <- datsim(n1 = 25, n2 = 25, times = c(0, 3, 6, 12),
 mu1 = c(75, 80, 80, 80),
 mu2 = c(50, 60, 60, 60),
 sd1 = 15, sd2 = 15,
 corr.1 = .85,
 miss1 = missNN, miss2 = missNI,
 plots = F)
 d <- dwl[[1]]
 lcc <- dwl[[2]]

 # Save datasets to directory
 write.table(d, file = file.path(outpath, paste0("DS_", scnum, "_", i,
 ".txt")))
 write.table(lcc, file = file.path(outpath, paste0("CC_", scnum, "_", i,
 ".txt")))

 }

##############################################################################################
## Simulations - scenario 2, n = 25 (per group), 6% missing NN, 20% NI
##############################################################################################
scnum <- 2
 # Percent missings NN & NI
 missNN <- 0.06
 missNI <- 0.2

 for (i in 1:numsim){
 # simulate data
 set.seed(seeds[i]) #sets the seed to the ith element in the vector seeds
```

```
dwl <- datsim(n1 = 25, n2 = 25, times = c(0, 3, 6, 12),
mu1 = c(75, 80, 80, 80),
mu2 = c(50, 60, 60, 60),
sd1 = 15, sd2 = 15,
corr.1 = .85,
miss1 = missNN, miss2 = missNI,
plots = F)
d <- dwl[[1]]
lcc <- dwl[[2]]

# Save datasets to directory
write.table(d, file = file.path(outpath, paste0("DS_", scnum, "_", i,
".txt")))
write.table(lcc, file = file.path(outpath, paste0("CC_", scnum, "_", i,
".txt")))

}

##############################################################################################
## Simulations - scenario 3, n = 25 (per group), 8% missing NN, 30% NI
##############################################################################################
scnum <- 3
# Percent missings NN & NI
missNN <- 0.08
missNI <- 0.3

for (i in 1:numsim){
# simulate data
set.seed(seeds[i]) #sets the seed to the ith element in the vector seeds
dwl <- datsim(n1 = 25, n2 = 25, times = c(0, 3, 6, 12),
mu1 = c(75, 80, 80, 80),
mu2 = c(50, 60, 60, 60),
sd1 = 15, sd2 = 15,
corr.1 = .85,
miss1 = missNN, miss2 = missNI,
plots = F)
d <- dwl[[1]]
lcc <- dwl[[2]]

# Save datasets to directory
write.table(d, file = file.path(outpath, paste0("DS_", scnum, "_", i,
".txt")))
write.table(lcc, file = file.path(outpath, paste0("CC_", scnum, "_", i,
".txt")))

}
```

```
################################################################################################
## Simulations - scenario 4, n = 25 (per group), 10% missing NN, 40% NI
################################################################################################
 scnum <- 4
 # Percent missings NN & NI
 missNN <- 0.1
 missNI <- 0.4

 for (i in 1:numsim){
 # simulate data
 set.seed(seeds[i]) #sets the seed to the ith element in the vector seeds
 dwl <- datsim(n1 = 25, n2 = 25, times = c(0, 3, 6, 12),
 mu1 = c(75, 80, 80, 80),
 mu2 = c(50, 60, 60, 60),
 sd1 = 15, sd2 = 15,
 corr.1 = .85,
 miss1 = missNN, miss2 = missNI,
 plots = F)
 d <- dwl[[1]]
 lcc <- dwl[[2]]

 # Save datasets to directory
 write.table(d, file = file.path(outpath, paste0("DS_", scnum, "_", i,
 ".txt")))
 write.table(lcc, file = file.path(outpath, paste0("CC_", scnum, "_", i,
 ".txt")))

 }
```

2. Analyzing the datasets (R base)

```
# O:/Projects/Simulation/07 Programs/Make data/analyze datasets20230217.R
# Author: Rebecca Stellato
# 27 February 2023, update March and April 2023
# Aim: compare five different analysis methods applied to simulated datasets:
 # LMM (on incomplete data)
 # CPM (on incomplete data)
 # GEE (on incomplete data)
 # RM ANOVA (on complete data, listwise deletion)
 # t-tests/observed means (on complete data, listwise deletion)
# The datasets have been generated in a separate R script. Missingness
 # depends on baseline QoL for NI group: differing percentages of missing
 # data, with 4x more missing for NI kids with baseline QoL < med than
 # for NI kids with baseline QoL > med.
# Write the estimated marginal means & SE's to a file, which will then be
```

```
 # analyzed using rsimsum (for example)
# I have added a returnObject = TRUE to the LME's to cover all
 # the convergence errors, also adding a tryCatch.W.E.() to mark the
 # simulations as having given an error.

# Clear workspace
rm(list=ls())

# Set path for input
inpath <- "O:/Projects/Simulation/06 Data/Raw"

# Set path for output
outpath <- " O:/Projects/Simulation/06 Data/Raw"

## Required libraries
library(nlme) ## For LME/CPM analysis
library(afex) ## For RMA analysis
library(emmeans)
library(ggplot2)
library(geepack)
library(simsalapar)

##############################################################################################
## Analysis functions
##############################################################################################

## Analysis 1: LME with RI + RS for time
mmf <- function(d){
 mm <- lme(y ~ t*group, data=d, random = ~ 1 + time|id, na.action = "na.omit",
 control = lmeControl(opt = 'optim', returnObject = TRUE))
 # EMM
 emmeans.mm <- emmeans(mm, pairwise ~ t*group)
 emm.mm <- transform(emmeans.mm$emmeans)[, 1:5] # for estimated means, SEs&df
 # Contrasts
 emm.mm.df <- transform(emmeans.mm$contrasts) # for diff 0-12
 emm.mm.cont <- emm.mm.df[emm.mm.df$contrast == "t0 group1 - t12 group1" |
 emm.mm.df$contrast == "t0 group2 - t12 group2" |
 emm.mm.df$contrast == "t12 group1 - t12 group2", 1:4]
 # Return EMM & contrasts 0-12 & 1 vs 2 for coverage
 res.emm.mm <- cbind(method = "LME", emm.mm)
 res.contr.mm <- cbind(method = "LME", emm.mm.cont)
 results.mm <- list(res.emm.mm, res.contr.mm)
 return(results.mm)

}
```

```
## Analysis 2: CPM with cAR(1)
cpmf <- function(d){
 cpm <- gls(y ~ t*group, correlation=corCAR1(form = ~ time | id),
 data=d, na.action="na.omit", control = glsControl(opt = 'optim'))
 # EMM
 emmeans.cpm <- emmeans(cpm, pairwise ~ t*group, mode = "appx-satterthwaite")
 emm.cpm <- transform(emmeans.cpm$emmeans)[, 1:5] # for estimated means, SEs & df
 # Contrasts
 emm.cpm.df <- transform(emmeans.cpm$contrasts) # for diff 0-12
 emm.cpm.cont <- emm.cpm.df[emm.cpm.df$contrast == "t0 group1 - t12 group1" |
 emm.cpm.df$contrast == "t0 group2 - t12 group2" |
 emm.cpm.df$contrast == "t12 group1 - t12 group2", 1:4]
 # Return EMM & contrasts 0-12 & 1 vs 2 for coverage
 res.emm.cpm <- cbind(method = "CPM", emm.cpm)
 res.contr.cpm <- cbind(method = "CPM", emm.cpm.cont)
 results.cpm <- list(res.emm.cpm, res.contr.cpm)
 return(results.cpm)

}

## Analysis 3: RM ANOVA
rmaf <- function(lcc){
 rma <- afex::aov_car(y~t*group+Error(id/t), data=lcc) # this is a RM ANOVA as done in SPSS
 # EMM
 emmeans.rma <- emmeans(rma, pairwise ~ t | group )
 emmeans2.rma <- emmeans(rma, pairwise ~ group | t )
 emm.rma <- transform(emmeans.rma$emmeans)[,1:5] # for estimated means, SEs & df
 # Contrasts
 emm.rma.df <- transform(emmeans.rma$contrasts) # for diff 0-12
 emm2.rma.df <- transform(emmeans2.rma$contrasts) # for diff gr 1 v 2

 c1 <- emm.rma.df[emm.rma.df$contrast=="X0 - X12" & emm.rma.df$group==1, c(1,3:5)]
 c1$contrast <- "t0 group1 - t12 group1"
 c2 <- emm.rma.df[emm.rma.df$contrast=="X0 - X12" & emm.rma.df$group==2, c(1,3:5)]
 c2$contrast <- "t0 group2 - t12 group2"
 c3 <- emm2.rma.df[emm2.rma.df$contrast=="group1 - group2" & emm2.rma.df$t=="X12", c(1,3:5)]
 c3$contrast <- "t12 group1 - t12 group2"
 rma.cont <- rbind(c1, c2, c3)

 # Return EMM & contrasts 0-12 & 1 vs 2 for coverage
 res.emm.rma <- cbind(method = "RMA", emm.rma)
 res.contr.rma <- cbind(method = "RMA", rma.cont)
 results.rma <- list(res.emm.rma, res.contr.rma)
 return(results.rma)

}
```

```
## Analysis 4: t-tests
## 4b: Independent Welch's t-test at 12 months
ttif <- function(d){
 # Means: just the observed means in the data frame d
 emm.tti <- data.frame(aggregate(y ~ t + group, data=d, FUN=mean))
 names(emm.tti) <- c("t", "group", "emmean")
 # Add SE's and df to means
 SD.tti <- data.frame(aggregate(y ~ t + group, data=d, FUN=sd))
 n.tti <- data.frame(aggregate(y ~ t + group, data=d, FUN=length))
 SE.tti <- merge(SD.tti, n.tti, by = c("t", "group"))
 SE.tti$SE <- SE.tti$y.x/sqrt(SE.tti$y.y)
 SE.tti$df <- SE.tti$y.y - 1
 emm.tti <- merge(emm.tti, SE.tti[,c(1, 2, 5, 6)], by = c("t", "group"))

 # p-value 12 months NN vs NI (Welch's t-test),
 tti <- t.test(y~group, data=d[d$t==12,], na.action = na.omit)
 e.t12.1.tti <- as.numeric(tti$estimate[1])
 e.t12.2.tti <- as.numeric(tti$estimate[2])
 res.contr.tti <- data.frame(method = "TT", contrast = "t12 group1 - t12 group2",
 estimate = e.t12.1.tti - e.t12.2.tti,
 SE = as.numeric(tti$stderr),
 df = as.numeric(tti$parameter))

 res.emm.tti <- cbind(method = "TT", emm.tti)
 results.tti <- list(res.emm.tti, res.contr.tti)
 return(results.tti)
}

## 4b: Paired t-tests at 12 vs 0 months, NI & NN
ttpf <- function(d){
 # paired t-test time 12 vs 0, groups 1 & 2
 ttp1 <- t.test(d$y[d$t==0 & d$group==1], d$y[d$t==12 & d$group==1], na.action = na.omit,
 paired = TRUE)
 ttp2 <- t.test(d$y[d$t==0 & d$group==2], d$y[d$t==12 & d$group==2], na.action = na.omit,
 paired = TRUE)

 ttp.cont1 <- data.frame(method = "TT", contrast = "t0 group1 - t12 group1",
 estimate = as.numeric(ttp1$estimate),
 SE = as.numeric(ttp1$stderr),
 df = as.numeric(ttp1$parameter))
 ttp.cont2 <- data.frame(method = "TT", contrast = "t0 group2 - t12 group2",
 estimate = as.numeric(ttp2$estimate),
 SE = as.numeric(ttp2$stderr),
 df = as.numeric(ttp2$parameter))
 res.contr.ttp <- rbind(ttp.cont1, ttp.cont2)
```

```
 return(res.contr.ttp)

}

## Analysis 5: GEE with AR(1)
geef <- function(d){
 gee <- geese(y ~ t*group, data=d, id=id, na.action = "na.omit", corstr = "ar1")
 # EMM
 emmeans.gee <- emmeans(gee, pairwise ~ t*group, vcov.method = "robust")
 emm.gee <- transform(emmeans.gee$emmeans)[, 1:5] # for estimated means, SEs & df
 # Contrasts
 emm.gee.df <- transform(emmeans.gee$contrasts)
 emm.gee.cont <- emm.gee.df[emm.gee.df$contrast == "t0 group1 - t12 group1" |
 emm.gee.df$contrast == "t0 group2 - t12 group2" |
 emm.gee.df$contrast == "t12 group1 - t12 group2", 1:4]

 # Return EMM & contrasts 0-12 & 1 vs 2
 res.emm.gee <- cbind(method = "GEE", emm.gee)
 res.contr.gee <- cbind(method = "GEE", emm.gee.cont)
 results.gee <- list(res.emm.gee, res.contr.gee)
 return(results.gee)

}

## Get N's
getns <- function(d){
 nmm <- data.frame(aggregate(y ~ t + group, data=d, FUN=length))
 ns.MAR <- c(nmm[1,3], nmm[2,3], nmm[3,3], nmm[4,3],
 nmm[5,3], nmm[6,3], nmm[7,3], nmm[8,3])
 return(ns.MAR)

}

##############################################################################################
## Analyze the simulated datasets: all scenarios at once
##############################################################################################

emm_options(save.ref_grid = FALSE)

## Number of simulations
numsim <- 7600

for (j in 1:4) {
 for (i in 1:numsim){

 d <- read.table(file = file.path(inpath, paste0("DS_", j, "_", i, ".txt")))
```

```
d$t <- factor(d$t)
d$group <- factor(d$group)
d$id <- factor(d$id)
lcc <- read.table(file = file.path(inpath, paste0("CC_", j, "_", i, ".txt")))
lcc$t <- factor(lcc$t)
lcc$group <- factor(lcc$group)
lcc$id <- factor(lcc$id)

# put means from all five analyses in one dataframe per simulation, add sim# & %missing
emm <- rbind(mmf(d)[[1]], cpmf(d)[[1]], rmaf(lcc)[[1]], ttif(d)[[1]], geef(d)[[1]])
emms <- cbind(j, i, emm)

# Warnings/errors from LME analsis
WE <- cbind(j, i, !is.null(tryCatch.W.E(lme(y ~ t*group, data=d, random = ~ 1 + time|id, na.action = "na.omit",
control = lmeControl(opt = 'optim', returnObject = TRUE)))$warning[1]))
write.table(WE, file.path(outpath, paste0("WarnErrEqVarLME.txt")), col.names = FALSE,
row.names = FALSE, append = TRUE)

# Contrasts for coverage
contr <- rbind(mmf(d)[[2]], cpmf(d)[[2]], rmaf(lcc)[[2]], ttif(d)[[2]], ttpf(d), geef(d)[[2]])
contrs <- cbind(j, i, contr)

## write results to files
# means over time for two groups
write.table(emms, file.path(outpath, "results_emm_2groups_allscn.txt"), col.names = FALSE,
row.names = FALSE, append = TRUE)
write.table(contrs, file.path(outpath, "results_contr_2groups_allscn.txt"), col.names = FALSE,
row.names = FALSE, append = TRUE)

# Not sure if it's interesting, but just in case: N per time point
write(c(j, i, getns(d)), file.path(outpath, "NperarmpertimeEqVar.txt"),
ncol=15, append=T)

}
}
```

3. Analyzing the results of the simulations (RMarkdown)

```
---
title: "Analysis of longitudinal data simulations"
author: "Rebecca Stellato"
date: "April 10, 2023 (latest update July 12, 2023)"
output:
 html_document:
 toc: TRUE
 toc_float: true
```

```
 toc_depth: 4
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(fig.dim = c(8, 6))
options(width = 300)
```

Aim
The aim of this program it to analyze the output of the simulations of longitudinal health-related quality of life (HRQoL):

 * 2 groups
 - group 1 has low levels of dropout that is MCAR
 - group 2 has higher levels of dropout that is MAR (depends on baseline HRQoL)
 * 4 scenarios (levels of dropout)
 - group 1: 4-6-8-10% within one year
 - group 2: 10-20-30-40% within one year
 * 7600 simulations per scenario
 * 5 analysis methods

using the rsimsum package for the summary of the simulation results and some useful plots.

There are three aims:
 1. Describe the bias & precision (empirical SE, % gain in relative precision, MSE) of the 5 methods for the 2 groups in the 4 scenarios
 2. Describe the bias, precision & power of the three contrasts (estimated differences in 0-12 months for groups 1 and 2, and the estimated difference between the groups at 12 months)
 3. Tabulate the number & proportions of warnings in the 4 scenarios for the LME analyses

I've discovered that rsimsum has been using infinite degrees of freedom for all methods for both coverage and power. I can change that easily for the coverage (both EM means and contrasts). It does not yet work for the power. I have to calculate the power "by hand" (I'm comparing the true value to the confidence intervals). I have taken out the plots that rely on rsimsum for the power calculations and have added my own graph for power.

Regarding the df: in the EM means analyses I used the Kenward-Roger df for the LME's, the approximate Satterthwaite df for the CPM's (all but 2 iterations: for one simulation in scn 3 and one in scn 4 I had to revert to error.df). The t-tests and RMA have their model-based df's and the GEE has infinite df because testing in GEE is with the standard normal distribution (or the chi-square).

Data preparation
There are three files of results from the simulations:
```

1. The estimated marginal means from the 5 analysis techniques (LME, CPM, GEE, RMA & independent t-tests at each moment)
2. Contrasts: estimated differences in 0-12 months for groups 1 and 2, and the estimated difference between the groups at 12 months.
3. Warnings from the LME analyses (convergence problems)

Input datasets:

* results_emm_2groups_allscn.txt
- EMMs from which I need to remove the RMA analyses
* results_contr_2groups_allscn.txx
- contrasts from which I need to remove the RMA analyses
* WarnErr_1 - _4.txt
- Warnings/errors from LME analyses (scn 1-4)

Some basic settings:
```{r FTF, include=TRUE}
# Clear workspace
rm(list=ls())

## Set path for input data
inpath <- "O:/Projects/Simulation/06 Data/Revised/"

## Set path for output (graphs, etc.)
outpath <- "O:/Projects/Simulation/08 Output/"

library(rsimsum)
library(ggplot2)
library(cowplot)
library(psych)
```

###Data preparation EMMs:
```{r DataPrepEMM, include=TRUE}
# EMMs from all 5 analyses
allemm <- read.table(file = file.path(inpath, "results_emm_2groups_allscn.txt"), header = F)
dim(allemm)
#head(allemm)
colnames(allemm) <- c("scenario", "simnum", "method", "t", "group", "mean", "SE", "df")
head(allemm)

# time variable is weird for RMA, fix it (& check!!)
allemm$time <- ifelse(allemm$t=="X0", 0,
 ifelse(allemm$t=="X3", 3,
 ifelse(allemm$t=="X6", 6,
 ifelse(allemm$t=="X12", 12, allemm$t))))
```

```
allemm$time <- as.numeric(allemm$time)
with(allemm, table(t, time))
with(allemm, table(method, time))

## First: make group variable a factor & change order method variable
allemm$group <- factor(allemm$group)
allemm$method <- as.factor(allemm$method)
allemm$method <- factor(allemm$method, levels(allemm$method)[c(3,1,2,5,4)])

# Add a column of the true values for all means
allemm$exp <- ifelse(allemm$group == 1, ifelse(allemm$time == 0, 75, 80),
 ifelse(allemm$time == 0, 50, 60))
```

###Data preparation contrasts:

```{r DataPrepContr, include=TRUE}
## Get contrast dataset
allcontr <- read.table(file = file.path(inpath, "results_contr_2groups_allscn.txt"), header = F)
#head(allcontr)
colnames(allcontr) <- c("scenario", "simnum", "method", "contrast", "estimate", "SE", "df")
head(allcontr)
dim(allcontr)
table(allcontr$contrast)
with(allcontr, table(method, contrast))

# Change order methods
allcontr$method <- as.factor(allcontr$method)
allcontr$method <- factor(allcontr$method, levels(allcontr$method)[c(3,1,2,5,4)])
table(allcontr$method)

# Reverse two of the contrasts
allcontr$estimate2 <- ifelse(allcontr$contrast == "t0 group1 - t12 group1" |
 allcontr$contrast == "t0 group2 - t12 group2", -1*allcontr$estimate,
 allcontr$estimate)

# Calculate 95% CI's by hand, for the widths of the 95% CI's
allcontr$LL <- allcontr$estimate2 - qt(0.025, allcontr$df, lower.tail = FALSE)*allcontr$SE
allcontr$UL <- allcontr$estimate2 + qt(0.025, allcontr$df, lower.tail = FALSE)*allcontr$SE
allcontr$lengthCI <- allcontr$UL - allcontr$LL

# Add a column of the true values for all means
allcontr$exp <- ifelse(allcontr$contrast == "t0 group1 - t12 group1", 5,
 ifelse(allcontr$contrast == "t0 group2 - t12 group2", 10, 20))

# For power of the contrasts: 0 not in 95% CI (in this case: LL > 0)
allcontr$CIexcl0 <- ifelse(0 < allcontr$LL, 1, 0)
```

```
head(allcontr)
```

###Data preparation warnings LMEs:

```{r DataPrepWarn, include=TRUE}
allwarn <- read.table(file = file.path(inpath, "WarnErrEqVarLME.txt"), header = F)

head(allwarn)
colnames(allwarn) <- c("scn", "simnum", "warning")
```

Analyze & summarize simulations

EMMs for the 4 time points in the two groups

I have 3 variables for splitting the simulation results: group, scenario & time. It's a bit much, so I stratify on group (which is also type & amount of missingness).

Group 1:

```{r simsumEMMgr1, include=TRUE}
simgr1 <- simsum(data = allemm[allemm$group==1,], estvarname = "mean", true = "exp",
 se = "SE", df = "df", methodvar = "method", ref = "LME",
 by = c("scenario", "time"), x = TRUE)
print(summary(simgr1), digits = 3)

biasa <- autoplot(summary(simgr1), type = "lolly", stats = "bias") +
 ggplot2::theme_bw() + ggplot2::xlim(-0.35, 6.65) + ggplot2::xlab("Bias")

covera <- autoplot(summary(simgr1), type = "lolly", stats = "cover") +
 ggplot2::theme_bw() + ggplot2::xlim(0.7, 0.97) + ggplot2::xlab("Coverage")

empSEa <- autoplot(summary(simgr1), type = "lolly", stats = "empse") +
 ggplot2::theme_bw() + ggplot2::xlim(2.5, 4.5) +
 ggplot2::xlab("Empirical standard error")

relpreca <- autoplot(summary(simgr1), type = "lolly", stats = "relprec") +
 ggplot2::theme_bw() + ggplot2::xlim(-45, 10) +
 ggplot2::xlab("% Gain in relative precision")

MSEa <- autoplot(summary(simgr1), type = "lolly", stats = "mse") +
 ggplot2::theme_bw() + ggplot2::xlim(0, 45) + ggplot2::xlab("Mean squared error")

RelErra <- autoplot(summary(simgr1), type = "lolly", stats = "relerror") +
 ggplot2::theme_bw() + ggplot2::xlim(-10, 10) +
 ggplot2::xlab("Relative % error SE")
```

Group 2:
```{r simsumEMMgr2, include=TRUE}
simgr2 <- simsum(data = allemm[allemm$group==2,], estvarname = "mean", true = "exp",
 se = "SE", df = "df", methodvar = "method", ref = "LME",
 by = c("scenario", "time"), x = TRUE)
print(summary(simgr2), digits = 3)
biasb <- autoplot(summary(simgr2), type = "lolly", stats = "bias") + ggplot2::theme_bw() +
 ggplot2::xlim(-0.35, 6.65) + ggplot2::xlab("Bias")

coverb <- autoplot(summary(simgr2), type = "lolly", stats = "cover") +
 ggplot2::theme_bw() + ggplot2::xlim(0.7, 0.97) + ggplot2::xlab("Coverage")

empSEb <- autoplot(summary(simgr2), type = "lolly", stats = "empse") +
 ggplot2::theme_bw() + ggplot2::xlim(2.5, 4.5) + ggplot2::xlab("Empirical standard error")

relprecb <- autoplot(summary(simgr2), type = "lolly", stats = "relprec") +
 ggplot2::theme_bw() + ggplot2::xlim(-45, 10) + ggplot2::xlab("% Gain in relative precision")

MSEb <- autoplot(summary(simgr2), type = "lolly", stats = "mse") +
 ggplot2::theme_bw() + ggplot2::xlim(0, 45) + ggplot2::xlab("Mean squared error")

RelErrb <- autoplot(summary(simgr2), type = "lolly", stats = "relerror") +
 ggplot2::theme_bw() + ggplot2::xlim(-10, 10) + ggplot2::xlab("Relative % error SE")

```

Put the two bias, coverage etc plots together:
```{r EMMplots, include=TRUE}
print(bias.emm <- plot_grid(biasa, biasb, ncol=2, labels="auto"))
print(cover.emm <- plot_grid(covera, coverb, ncol=2, labels="auto"))
print(empSE.emm <- plot_grid(empSEa, empSEb, ncol=2, labels="auto"))
print(relprec.emm <- plot_grid(relpreca, relprecb, ncol=2, labels="auto"))
print(MSE.emm <- plot_grid(MSEa, MSEb, ncol=2, labels="auto"))
print(RelErr.emm <- plot_grid(RelErra, RelErrb, ncol=2, labels="auto"))
```

NOTE: the coverage plots above are of the coverage of the EMMEANs, *not* the coverage of the 95% CI's for the contrasts.

Contrasts
Note: ignore the power from rsimsum.

```{r simsumContrasts, include=TRUE}
# Contrast labeller
c.labeller <- labeller(scenario = c(`1` = "Scenario 1", `2` = "Scenario 2",
 `3` = "Scenario 3", `4` = "Scenario 4"),
```

```
 contrast = c(`t0 group1 - t12 group1` = "Within group 1",
 `t0 group2 - t12 group2` = "Within group 2",
 `t12 group1 - t12 group2` = "Between groups"))
## Analysis all 3 contrasts together
contrsimsum <- simsum(data = allcontr, estvarname = "estimate2", true = "exp",
 se = "SE", df = "df", methodvar = "method", ref = "LME",
 by = c("scenario", "contrast"), x = TRUE)
print(summary(contrsimsum), digits = 3)

print(c.bias <- autoplot(summary(contrsimsum), type = "lolly", stats = "bias") +
 ggplot2::theme_bw() + ggplot2::xlim(-5, 0.5) + ggplot2::xlab("Bias") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

print(c.cover <- autoplot(summary(contrsimsum), type = "lolly", stats = "cover") +
 ggplot2::theme_bw() + ggplot2::xlim(0.78, 0.97) + ggplot2::xlab("Coverage") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

print(c.empSE <- autoplot(summary(contrsimsum), type = "lolly", stats = "empse") +
 ggplot2::theme_bw() + ggplot2::xlim(3.5, 5.5) +
 ggplot2::xlab("Empirical standard error") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

print(c.relprec <- autoplot(summary(contrsimsum), type = "lolly",
 stats = "relprec") + ggplot2::theme_bw() + ggplot2::xlim(-20, 10) +
 ggplot2::xlab("% Gain in relative precision") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

print(c.MSE <- autoplot(summary(contrsimsum), type = "lolly", stats = "mse") +
 ggplot2::theme_bw() + ggplot2::xlim(10, 45) +
 ggplot2::xlab("Mean squared error") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

print(c.RelErr <- autoplot(summary(contrsimsum), type = "lolly",
 stats = "relerror") + ggplot2::theme_bw() + ggplot2::xlim(-10, 10) +
 ggplot2::xlab("Relative % error SE") +
 ggplot2::facet_grid(vars(scenario), vars(contrast), labeller = c.labeller))

```

For the power of the contrasts, use the code & graph below:

```{r powercontr, include=TRUE}
pow <- as.data.frame(describeBy(CIexcl0 ~ contrast + method + scenario,
 data=allcontr, mat = TRUE, skew = FALSE,
 digits = 6)[,c(2:4,7)])

# Calculate the MC SE's for power
```

```
pow$SE <- round(sqrt((pow$mean*(1 - pow$mean))/7600), 3)

# Calculate the 95% CI limits
pow$lower <- pow$mean - 1.96*pow$SE
pow$upper <- pow$mean + 1.96*pow$SE

rownames(pow) <- NULL

head(pow)

# Change variable names so code below is easier
pow$method <- as.factor(pow$group2)
pow$scenario <- factor(pow$group3)
pow$contrast <- factor(pow$group1)

c.labeller <- labeller(scenario = c(`1` = “Scenario 1”, `2` = “Scenario 2”,
 `3` = “Scenario 3”, `4` = “Scenario 4”),
 contrast = c(`t0 group1 - t12 group1` = “Within group 1”,
 `t0 group2 - t12 group2` = “Within group 2”,
 `t12 group1 - t12 group2` = “Between groups”))

# Lollipop plot, as close to rsimsum as I can get
power.contr <- ggplot(pow, aes(y=method, x=mean)) +
 geom_segment( aes(y=group2, yend=group2, x=1, xend=mean)) +
 geom_point( size=2, alpha=0.6) +
 theme_light() +
 ylab(“method”) + xlab(“Power”) +
 theme(strip.background =element_rect(fill = “lightgray”))+
 theme(strip.text = element_text(colour = ‘black’)) +
 facet_grid(vars(scenario), vars(contrast), labeller = c.labeller) +
 geom_point(ggplot2::aes(x = lower, y = method), shape = 40) +
 geom_point(ggplot2::aes(x = upper, y = method), shape = 41) +
 ylim(“LME”, “CPM”, “GEE”, “TT”, “RMA”)

power.contr
```

Widths of the 95% CI’s for the 3 contrasts

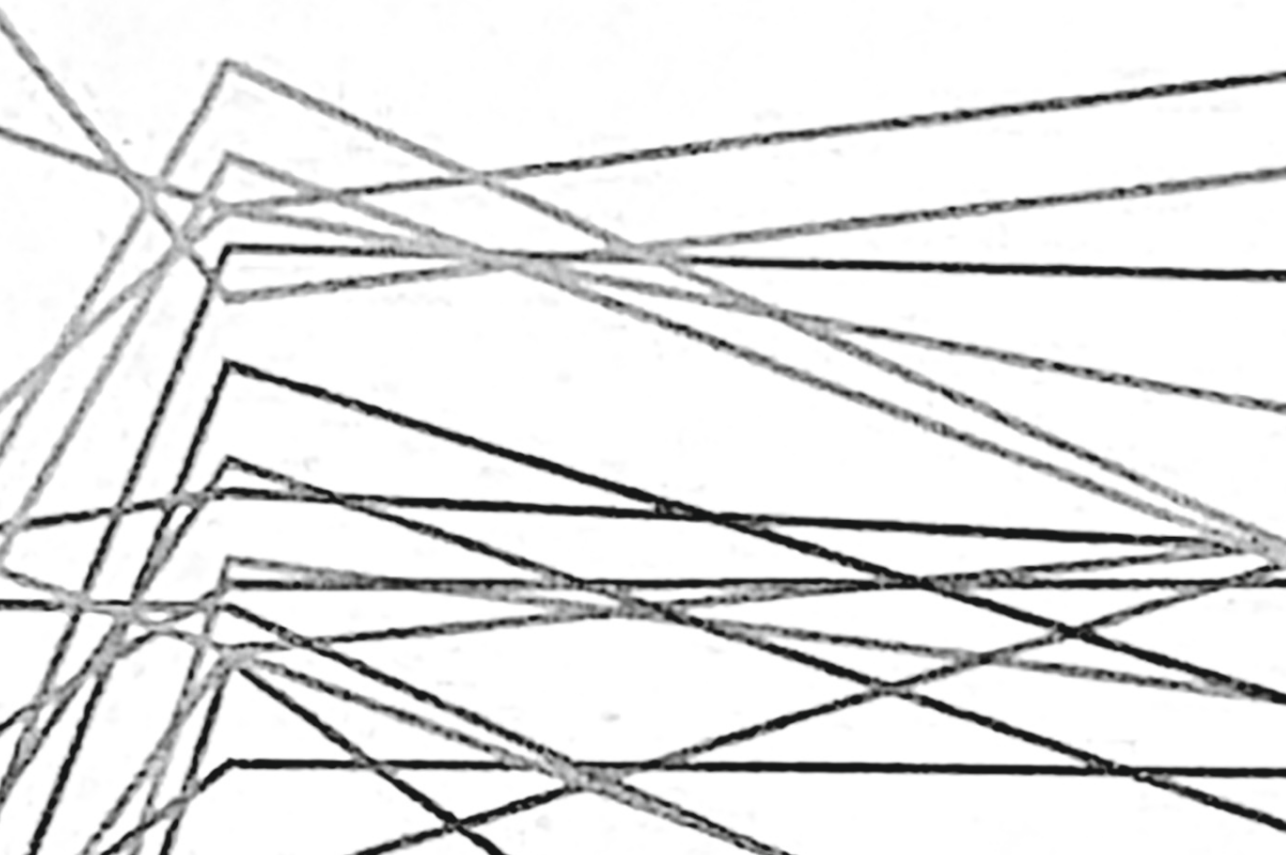
```{r WidthsCIcontr, include=TRUE}
bpCIlen <- ggplot(allcontr, aes(x=method, y=lengthCI)) +
 geom_boxplot() +
 ylab(“Width 95% confidence interval”) +
 facet_grid(vars(scenario), vars(contrast), labeller = c.labeller) +
 theme_bw()
bpCIlen
```

Warnings/errors LME models

The LME models gave some problems with convergence. I could force them to produce results anyway, but in a few cases per scenario the models were unstable for the data I had. Here I tabulate the frequencies and percentages of warnings per scenario.

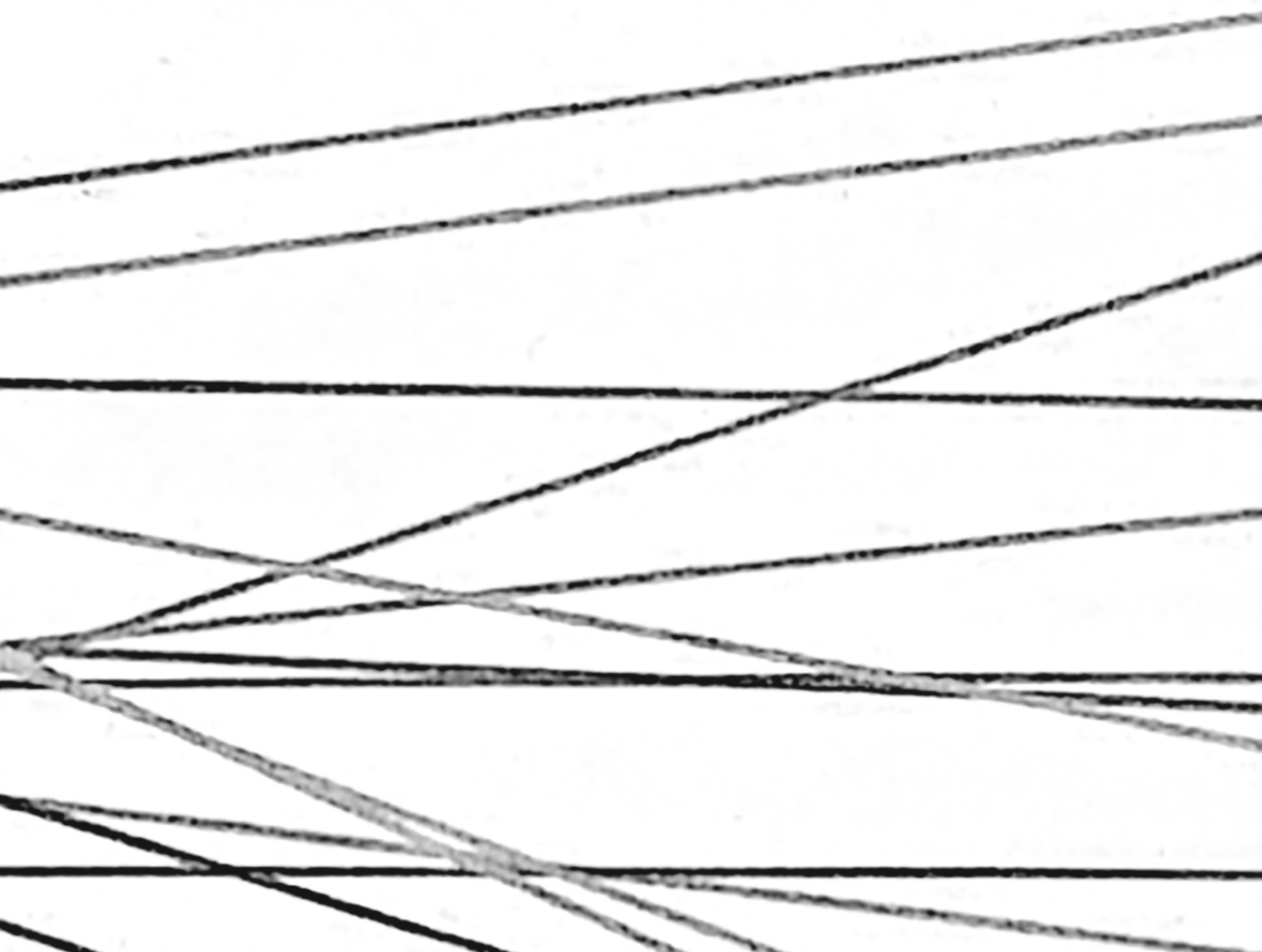
```{r Warnings, include=TRUE}
with(allwarn, table(scn, warning))
print(WE.tab <- 100*round(with(allwarn, prop.table(table(scn, warning),1)), 4))
```

```{r SaveGraphs, eval = TRUE}
ggsave(bias.emm, filename=file.path(outpath, "Fig 1 EMM bias 2groups.jpg"),
 width=10, height=5, device="jpeg")
ggsave(empSE.emm, filename=file.path(outpath, "Fig 2 EMM emp SE 2 groups.jpg"), width=10, height=5,
device="jpeg")
ggsave(c.empSE, filename = file.path(outpath, "Fig 3 contrasts empirical SE.jpg"),
 width=8, height=10, device="jpeg")
ggsave(c.cover, filename = file.path(outpath, "Fig 4 contrasts coverage.jpg"),
 width=8, height=10, device="jpeg")
ggsave(bpCIlen, filename=file.path(outpath, "Fig 5 Width CIs 3 contrasts.jpg"),
 width=8, height=10, device="jpeg")
ggsave(power.contr, filename = file.path(outpath, "Fig 6 Contrasts power.jpg"),
 width=8, height=10, device="jpeg")
ggsave(c.bias, filename = file.path(outpath, "Suppl Fig 7 Contrasts bias.jpg"),
 width=8, height=10, device="jpeg")
ggsave(relprec.emm, filename=file.path(outpath, "Suppl Fig 8 EMM rel prec 2 groups.jpg"),
 width=10, height=5, device="jpeg")
ggsave(c.relprec, filename = file.path(outpath, "Suppl Fig 9 contr rel prec.jpg"),
 width=8, height=10, device="jpeg")
```



PART II

SHORT- AND LONG-TERM EFFECTS OF LAPAROSCOPIC ANTI-REFLUX SURGERY



CHAPTER 4

Laparoscopic Antireflux Surgery Increases Health-Related Quality of Life in Children with GERD

Femke A. Mauritz[1,3]
Rebecca K. Stellato[4]
L. W. Ernst van Heurn[5]
Peter D. Siersema[3]
Cornelius E. J. Sloots[6]
Roderick H. J. Houwen[2]
David C. van der Zee[1]
Maud Y. A. van Herwaarden-Lindeboom[1]

[1]Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Room: KE.04.140.5, PO Box 85090, 3508 AB Utrecht, The Netherlands

[2]Department of Pediatric Gastroenterology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

[3]Department of Gastroenterology and Hepatology, University Medical Center Utrecht, Utrecht, The Netherlands

[4]Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[5]Department of Pediatric Surgery, Maastricht University Medical Center, Maastricht, The Netherlands

[6]Department of Pediatric Surgery, Sophia Children's Hospital, Erasmus MC, Rotterdam, The Netherlands

Surgical Endoscopy 2017; 31:3122–3129

ABSTRACT

Introduction

Improving health-related quality of life (HRQoL) is increasingly recognized as an essential part of patient care outcome. Little is known about the effect of laparoscopic antireflux surgery (LARS) on the HRQoL in the pediatric patients. The aims of this study were to evaluate the effect of LARS on HRQoL in children with gastroesophageal reflux disease (GERD) and to identify predictors that influence HRQoL outcome after LARS. Methods Between 2011 and 2013, 25 patients with therapy-resistant GERD [median age 6 (2–18) years] were included prospectively. Caregivers and children with normal neurodevelopment (>4 years) were asked to fill out the validated PedsQL 4.0 Generic Core Scales before and 3–4 months after LARS.

Results

The PedsQL was completed by all caregivers (n = 25) and 12 children. HRQoL total score improved significantly after LARS, both from a parental (p = 0.009) and child's perspective (p = 0.018). The psychosocial health summary and physical health summary scores also improved significantly after LARS. HRQoL before and after LARS was significantly lower in children with impaired neurodevelopment (p \ 0.001). However, neurodevelopment did not influence the effect of LARS on HRQoL. The only significant predictor for improvement in HRQoL after LARS was age at the time of operation (p = 0.001).

Conclusions

HRQoL significantly improves after LARS. Although children with impaired neurodevelopment had lower overall HRQoL, neurodevelopment by itself does not predict inferior improvement in HRQoL after LARS. Older children have a more favorable HRQoL outcome after LARS compared to younger children. This may suggest caution when considering LARS in younger GERD patients.

INTRODUCTION

Laparoscopic antireflux surgery (LARS) is an established treatment option performed in pediatric patients with severe gastroesophageal reflux disease (GERD) resistant to medical treatment.[1,2] LARS primarily aims to decrease (acid) reflux events and to reduce reflux symptoms. However, as shown in earlier studies the effect on reflux symptoms does not always correlate to more objective assessments of success of therapy.[3,4] Furthermore, comorbidities (e.g., impaired neurodevelopment) and complications, such as dysphagia and gas-bloat syndrome,[5] may also affect success of therapy.

To better assess the impact of pediatric diseases and treatments from the perspective of the pediatric patient and their caregivers health-related quality of life (HRQoL), assessment has been increasingly recognized as an essential part of patient care outcome.[6] Effects of LARS on HRQoL have been mainly investigated in adult population. These studies almost all showed that HRQoL improves after LARS.[7-9] In the pediatric population, only few studies have focused on this outcome parameter.[10-12] HRQoL in these studies improves; however, none of these studies have used pediatric validated questionnaires. In two studies,[10,11] a questionnaire designed for adults had been modified for pediatric use and one study had only used parental proxy report to score HRQoL.[12] Furthermore, none of these studies could identify determinants that influence HRQoL outcome after LARS. The Pediatric Quality of Life Inventory (http://www.pedsql.org.) 4.0 Generic Core Scales (PedsQL) is a reliable and valid tool (also for the Dutch language) for parental proxy report and parallel child's self-report on HRQoL. It has been used to assess HRQoL in children with numerous acute and chronic health conditions, as well as in healthy populations.[6,13-17] The aim of this study was to evaluate the effect of LARS on HRQoL using the PedsQL and to identify predictors that may influence HRQoL outcome after LARS.

PATIENTS AND METHODS

We performed a prospective multicenter study in three University Medical Centers in the Netherlands performing laparoscopic fundoplication in children (Wilhelmina children's Hospital, University Medical Center Utrecht (UMCU): Sophia children's Hospital, Erasmus University Medical Center (EMC) and Maastricht University Medical Center (MUMC). From July 2011 until December 2013, we prospectively included all pediatric patients diagnosed with PPI-therapy-resistant GERD. Patients that had undergone previous esophageal or gastric surgery (except previous gastrostomy placement) and those who had structural abnormalities other than an esophageal hiatal hernia were excluded.

Patients

In total, 25 children were included. Mean age of the included patients was six (range 2–18) years at the time of fundoplication (Table 1). Impaired neurodevelopment was

present in 20% of patients (5/25 patients). Causes of impaired neurodevelopment are shown in Table 2.

Table 1. Baseline characteristics

	Median (IQR)
Age at time of operation (years)	6.0 (3.0-11.0)
Duration of hospital admission (days)	3.0 (2.0-4.5)
	n (%)
Male Gender	12 (48.0%)
Neurological Normal Development	20 (80.0%)
Gastrostomy preoperatively in situ	4 (16.0%)

Table 2. Neurological impairment (n = 5)

Charge syndrome
Mitochondrial complex II deficiency
Post hypoxic encephalopathy
Congenital rubella infection
Neurologically impairment of unknown origin with autistic behavior

Surgical procedures

All laparoscopic fundoplications were performed by experienced pediatric surgeons in pediatric laparoscopic surgery. In the UMCU the anterior, partial fundoplication according to Thal[18] was used to perform fundoplication. In the other two UMC's (EMC and MUMC) the posterior, total fundoplication according to Nissen[19] was performed. Before fundoplication, the distal esophagus was fully mobilized; the distal 3 cm of the esophagus was repositioned back into the abdomen. Both vagal nerves were identified, and after dissection of both crura the hiatus was closed routinely (UMCU and EMC). Thereafter, the fundoplication was constructed. The Thal fundoplication was performed by plicating the fundus of the stomach over 270° against the distal anterior intra-abdominal part of the esophagus and the diaphragmatic crus.[3,18] A floppy Nissen was constructed with one of the sutures of the 360° posterior wrap incorporated in the esophageal wall.[19]

Clinical assessment

Before and 3 months after laparoscopic fundoplication, clinical assessment was performed using the PedsQL 4.0 Generic Core Scale for HRQoL, a reflux-specific symptom questionnaire, 24-h multichannel intraluminal impedance-pH monitoring (MII-pH monitoring) and an 13C-labeled Na-octanoate breath test. Surgical re-interventions,

type and indication for re-intervention, endoscopic procedures, complications, and comorbidities were registered in a prospective database.

1. Health-related Quality of Life
Caregivers and children with normal neurodevelopment (>4 years) were asked to fill out the 23-items PedsQL 4.0 Generic Core Scales.[1,14,17,20,21] The scales are available for parental proxy report, subdivided in four age-adjusted questionnaires (ages: 2–4; 5–7; 8–12; and 13–18 years) and as a parallel child's self-report (ages: 5–7; 8–12; and 13–18 years). The PedsQL 4.0 Generic Core Scales comprises four domains: physical functioning (8 items), emotional functioning (5 items), social functioning (5 items), and school functioning (5 items). With the four domains, the physical health summary score, the psychosocial summary score and the total score are calculated. The physical health summary score is reflected by the physical functioning scale. The psychosocial health summary score is reflected by the mean of the other three domains (emotional, social and school functioning). Scale scores per domain were computed as the sum of the items divided by the number of items answered. Thereafter, items were reverse-scored and transformed to a 0–100 scale. Higher scale scores indicate better HRQoL.

2. Reflux-specific questionnaire
Patients and/or their parents were asked to fill out the Gastroesophageal Reflux Symptom Questionnaire.[22]

3. Ambulatory 24-h MII-pH monitoring
MII-pH monitoring was performed using an age-adjusted combined impedance-pH catheter (Unisensor AG, Attikon, Switzerland). Pathological acid exposure was defined as total acid exposure time C6, C9% in upright, and C3% in the supine body position.[23,24] The symptom index (SI) and the symptom association probability (SAP) was calculated when the patients experienced symptoms during measurement.[25,26]

4. Gastric emptying breath test
To assess gastric emptying time, we used a 13C-labeled Na-octanoate breath test.[27] Gastric emptying half time is defined as the time when the first half of the 13C-labeled substrate has been metabolized, that is, when the cumulative excretion of 13C in the breath is half the ingested amount. Gastric emptying percentiles were calculated according to the reference values obtained by van den Driessche et al.[28]

Ethical approval and trial registration

This study was registered at the start of the study in the Dutch national trial registry (www.trailregister.nl; Identifier: 2934). Ethical approval for this prospective multi-center study was obtained from the University Medical Center Utrecht Ethics Committee, and local approval was obtained by the remaining two participating centers. Prior to any

trial-related study procedure, informed consent from the patients' parents and children (≥12 years) was obtained.

Statistical analysis

Continuous variables, when symmetric, were expressed as mean ± standard error. Skewed variables were expressed as median with interquartile ranges (IQR). For statistical analysis, we used the paired sample t test or the Wilcoxon signed-ranks test. The McNemar–Bowker test was used to compare groups in case of nominal outcome measures. To assess the relationship between HRQoL and age at the time of operation, impaired neurodevelopment, reflux symptoms, acid exposure and gastric emptying, we used a linear mixed model with a random intercept per patient. A mixed model allowed us to analyze preoperative and postoperative measurements simultaneously, while taking into account correlation of measurements from the same subjects. Backwards selection was performed using the AIC. A linear regression analysis was performed to identify determinants influencing HRQoL and the effect of LARS on HRQoL. Determinant of interest included: age at the time of operation, impaired neurodevelopment, reflux symptoms, preoperative acid exposure time and preoperative gastric emptying rate. Differences with a $p < 0.05$ were considered statistically significant. All analyses were performed using IBM® 22.0.0 SPSS statistical package (IBM, Armonk, NY).

RESULTS

In total, 18 Thal and 7 Nissen fundoplications were performed. In all patients, fundoplication was completed by laparoscopy. Perioperative complications were not observed. One patient with retching due to impaired neurodevelopment developed severe recurrent reflux caused by hiatal herniation that required re-fundoplication. In six children, temporary nasogastric tube feedings were required to obtain sufficient caloric intake. This was caused by transient dysphagia (dysphagia dissolved within 3–4 months after LARS; n = 4), persistent dysphagia (>3–4 months after LARS; n = 1) or refusal of oral feedings (n = 1).

Health-related quality of life

The PedsQL was completed by all caregivers both before and after LARS for all included patients (n = 25). The HRQoL total score improved significantly after LARS from 69.8 (57.2–80.1) to 82.0 (69.3–89.2; p = 0.009; Fig. 1). Twelve children were able to fill out the parallel self-report, and their total score also improved significantly from 72.6 (7.4–82.3) to 84.6 (78.1–91.3; p = 0.018; Fig. 1).

Furthermore, the psychosocial [54.2 (69.7–77.5) to 82.5 (72.9–89.6); $p < 0.0001$] and the physical health summary [75.0 (59.4–89.1) to 92.2 (80.5–99.2); $p < 0.0001$] also significantly improved for both caregivers as well as children's self-report after LARS (Figs. 2, 3).

Patients' self-report of overall HRQoL outcomes was significantly higher (p = 0.037) than parental proxy report before LARS; self-reported and proxy-reported HRQoL scores after LARS were not different.

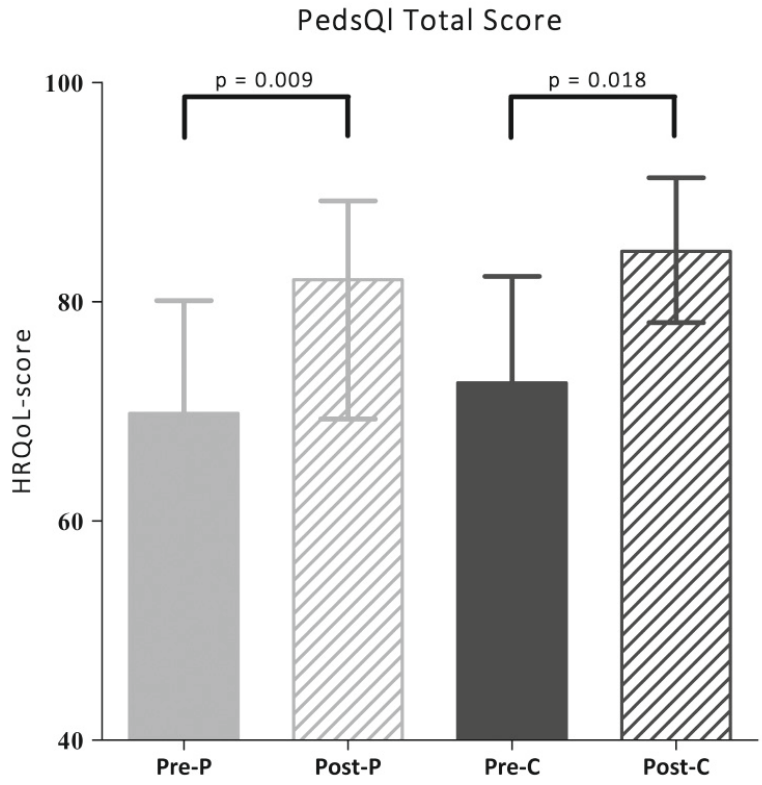


Figure 1. HRQoL assessment using the PedsQL–total score (pre = before LARS; post = after LARS; P = parental proxy report; C = child's self-report)

Reflux symptoms

Reflux symptoms significantly decreased from 16 (64%) patients with severe reflux symptoms before LARS to one (4%) patient after LARS (p = 0.001). Deterioration of symptom severity or frequency was not seen in any of the patients. Dysphagia was reported in seven (28%) patients before and in eight (32%) patients after LARS (p = 0.887). New-onset dysphagia was seen in three of these eight patients after LARS (Table 3).

Gastroesophageal functional assessment tests—total acid exposure decreased significantly from 8.5% (IQR 2.5–32.8) to 0.8% (IQR 0–21.6) after LARS (p < 0.0001). Median gastric emptying rate before LARS (percentile 75, IQR 3–99) was similar to that after LARS (70, IQR 5–99, p = 0.530).

Table 3. Symptoms (*n*=25)

	Preoperative (n; %)		3-4 months postoperative (n; %)		p-value
Reflux symptoms					
No symptoms	0	(0%)	17	(68%)	0.001
Mild reflux symptoms	2	(8%)	5	(20%)	
Moderate reflux symptoms	7	(28%)	2	(8%)	
Severe reflux symptoms	16	(64%)	1	(4%)	
Dysphagia	7	(28%)	8	(32%)	0.887

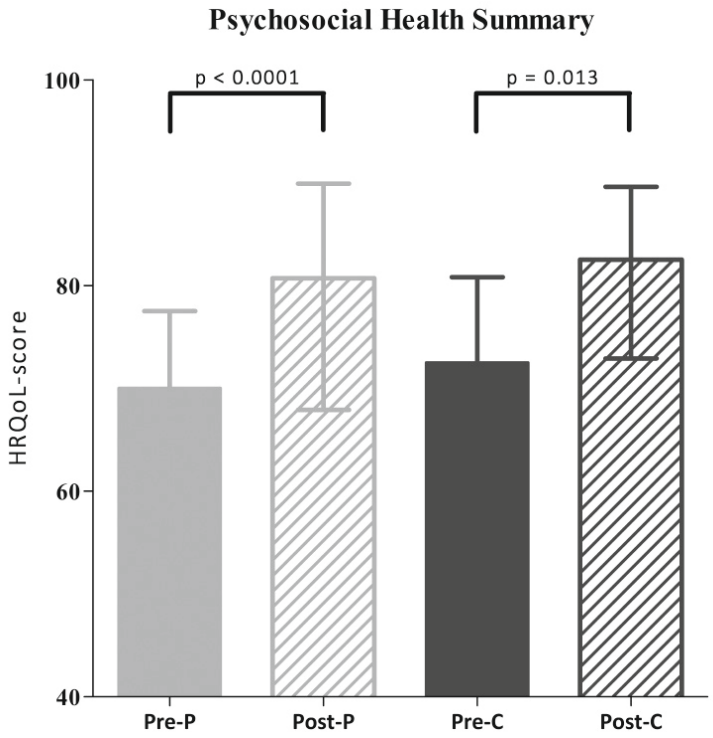


Figure 2. HRQoL assessment using the PedsQ—psychosocial health summary (pre = before LARS; post = after LARS; P = parental proxy report; C = child's self-report)

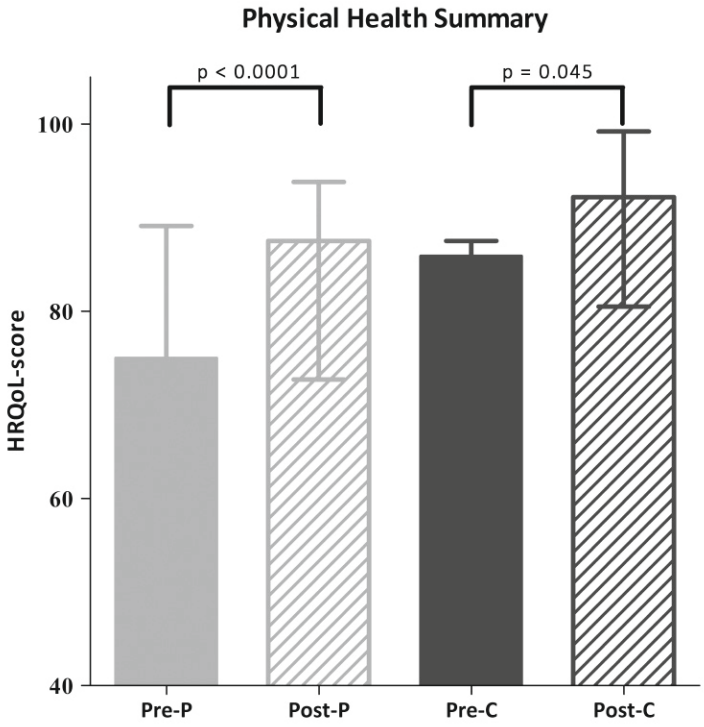


Figure 3. HRQoL assessment using the PedsQL—physical health summatabery (pre = before LARS; post = after LARS; P = parental proxy report; C = child's self-report)

Factors influencing HRQoL

Patients with impaired neurodevelopment (NI) had significantly lower HRQoL compared to patients with normal neurodevelopment (NN) (estimate 23.4; p = 0.006; (95% CI 7.2–39.5). Furthermore, reflux symptoms were also negatively associated with lower HRQoL (estimate = -4.3 l; p = 0.006; 95% CI -7.5 to -1.3). Age at the time of operation (p = 0.11), gastric emptying (p = 0.82) and total acid exposure (p = 0.75) did not significantly influence HRQoL.

Predictors for the effect of LARS on HRQoL

Linear regression analysis showed that an increase in age at the time of operation was a significant predictor for improvement in HRQoL after LARS (p = 0.001; estimate = 1.6; 95% CI 0.8–2.5; Fig. 4). Although HRQoL was significantly lower in NI children, neurodevelopment itself did not influence the change in HRQoL (p = 0.73). Preoperative gastric emptying rate, total acid exposure time and reflux symptoms also did not significantly influence the change in HRQoL (Table 4).

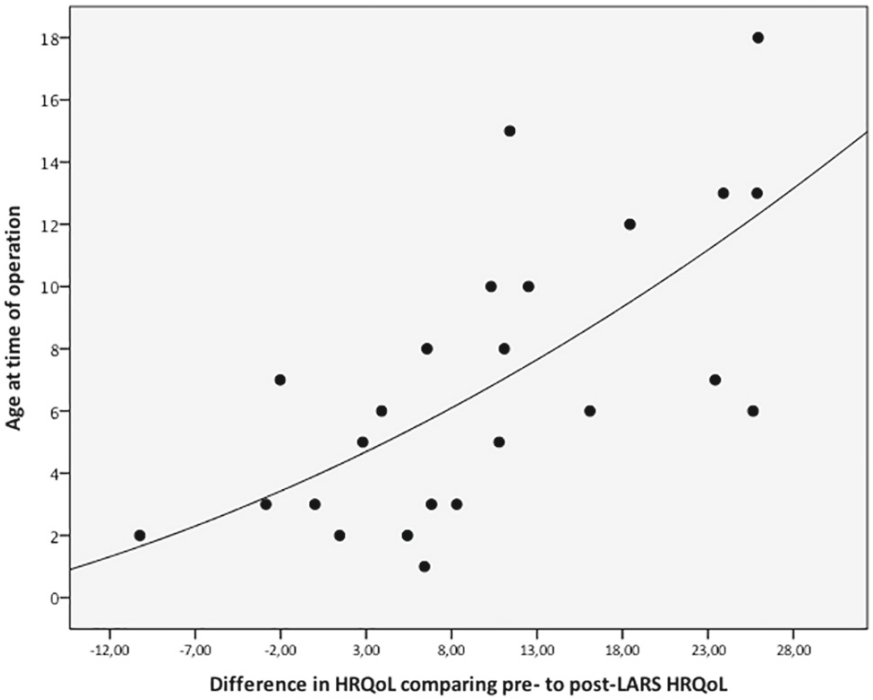


Figure 4. Scatterplot illustrating difference in HRQoL comparing pre- to post-LARS HRQoL

Table 4. Predictors for the effect of LARS on HRQoL

	B	p-value	95% CI
Age at time of operation	1.7	0.001	0.8 – 2.5
Neurological development	1.8	0.73	-9.3 – 12.9
Preoperative reflux symptoms	2.8	0.61	-8.7 – 14.4
Preoperative acid exposure time (%)	0.3	0.28	-0.3 – 0.9
Preoperative gastric emptying	-0.004	0.94	-0.1 – 0.1

Linear regression analysis (B = Beta-coefficient; 95% CI = 95% confidence interval)

DISCUSSION

This is the first study on HRQoL in children undergoing LARS that used a validated pediatric HRQoL questionnaire.[6,13-17] We demonstrate that after LARS HRQoL significantly increases and after LARS HRQoL scored were comparable to the normal HRQoL scores measured in a healthy population.[13] Furthermore, age at the time of operation is a significant predictor for improvement in HRQoL after LARS. Previous studies also showed a significant increase in HRQoL.[10-12] In two studies,[10,11] a questionnaire designed for adults had been modified for pediatric use and the third study only used parental proxy report. In contrast to these previous studies, this is the first study in pediatric LARS using a validated questionnaire for HRQoL. Using validated questionnaires in the pediatric population is important as results and questionnaires are not simply translatable for pediatric uses because pathophysiology, and patterns and symptoms of diseases may be different in children compared to adults.[29] Furthermore, this is the first study using both parental proxy reports as well as child's self-reports.

Some authors hypothesize that children with impaired neurodevelopment (NI) may not benefit to the same extent from LARS as those with normal development.[30-32] In the current study, overall HRQoL was significantly lower in NI children. However, the neurodevelopment itself did not influence the change in HRQoL. This indicates that while NI children with GERD had a lower overall HRQoL, LARS was equally effective regarding the change in HRQoL compared to children with normal neurodevelopment (NN). It is not surprising that NI children, who have more (co-)morbidity than NN children, scored lower in HRQoL; the PedsQL is able to distinguish between healthy children and pediatric patients with acute or chronic health conditions, and it is related to indicators of morbidity and illness burden.[21]

The HRQoL scores of NI children for physical, social and school functioning were scored significantly lower compared to NN children. Emotional function, however, was not scored different from NN (data not shown). As NI children have more (co-)morbidities and associated disabilities they will likely score lower in physical and social functioning. The scores in the domain school functioning may be influenced by the possibility for caregivers or patients to leave these questions open if they are not applicable. If at least 50% is filled out in a specific domain, the score over that domain can, however, still be calculated. In the domain school functioning questions regarding attitude and performance at school were generally not filled out, whereas questions regarding presence/absence due to sickness or hospital visits were almost completely filled out by the caregivers.

Before LARS, children reported a significantly higher HRQoL than their parental proxies. This difference remained when only the child's self-report was compared to the score of their parents and the scores of the parents with children aged <5 or NI children were

not taken into account. After LARS, this difference resolved as parents scored the HRQoL higher than children. It is not entirely clear why this difference in HRQoL resolves after LAR. We hypothesize that before LARS caregivers experience more burden from GERD on their child's HRQoL compared to their children's own perception. After LARS, GERD resolves in almost all children and it may therefore be possible that HRQoL assessments after LARS are therefore comparable.

Various studies on HRQoL in children indicate that information provided by caregivers does not always correspond to what children report themselves.[33,34] Pediatric patient self-report is considered to be the standard for measuring HRQoL, as it is the only genuine patient-reported outcome.[35] It can, however, be difficult to obtain self-reports in young children and children with impaired neurodevelopment. In these cases, a parental proxy report may be the only way to assess HRQoL.[36] Furthermore, it has been shown that the parents' perception of their child's HRQoL influences health care utilization more than the perception of the child itself.[37,38]

Age at the time of operation was a statistically significant predictor of improvement in HRQoL after LARS. This means that LARS has more effect on HRQoL in older children and may suggest caution when younger children are referred for therapy-resistant GERD. It has been suggested that recurrence of GERD and even the necessity for redo-fundoplication are more frequently seen in young patients. These suggestions were based on two retrospective studies both using regression analysis to identify risk factors.[39,40] Bearg et al.[39] showed that redo-fundoplication is significantly more frequent if patients are younger or have retching. Ngerncham et al.[40] reported that age less than 6 years was independently associated with increased risk of recurrence of GERD. Furthermore, it may also be possible that older children can specify their (reflux) complaints better, allowing a more precise diagnosis of therapy-resistant GERD to be made. Finally, it has been hypothesized that a young child may outgrow its fundoplication.[4]

We initially hypothesized that preoperative gastric emptying might influence the success of LARS and thereby the effect on HRQoL as this has been shown in adult literature.[41] In this study, however, we did not find an effect of gastric emptying on HRQoL.

It is plausible that reflux symptoms and acid exposure influence HRQoL assessment. In the current study, reflux symptoms were negatively associated with lower HRQoL. Remarkably, however, recurrence or persisting pathological acid exposure did not significantly influence HRQoL. It has been described before that reflux symptoms do not correlate to objective measurements of GERD, which may underscore the difficulty in symptom assessment.[3,4,42]

One of the limitations in the current study was the limited number of 25 patients included. It was therefore only possible to investigate 5 determinants in linear regression assuming that we have sufficient statistical power with 5 patients per predictor. If more patients were included in this study, we would have had more power to detect the influence of these variables, and we had been able to investigate more potential determinants of changes in HRQoL. Secondly, this limited sample size, in addition to variations in surgical technique and multiple institutions, results to various forms of potential bias, such as confounding or type 2 errors. Furthermore, not all patients were able to fill out the self-report because of impaired neurodevelopment or age and as mentioned previously not all questions could be filled out by their caregivers, because in specific domains the questions were not always suitable when considering comorbidities and patient's limitations.

In conclusion, health-related quality of life, scored by both pediatric patients and their caregivers, significantly improves after LARS. Although patients with impaired neurodevelopment have lower overall HRQoL compared to neurologically normal developed patients, neurodevelopment itself is not a predictor of inferior improvement in HRQoL after LARS. Older children have a more favorable outcome of LARS on HRQoL compared to younger children. This suggests that with the diagnosis of therapy-resistant GERD in younger children, one should possibly be cautious to perform LARS.

REFERENCES

1. Sherman PM, Hassall E, Fagundes-Neto U, et al. A Global, evidence-based consensus on the definition of gastroesophageal reflux disease in the pediatric population. *Am J Gastroenterol.* 2009;104(5):1278-1295.
2. Vandenplas Y, Rudolph CD, Di Lorenzo C, et al. Pediatric gastroesophageal reflux clinical practice guidelines: Joint recommendations of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN). *J Pediatr Gastroenterol Nutr.* 2009;49(4):498-547.
3. Mauritz FA, van Herwaarden-Lindeboom MY, Zwaveling S, Houwen RH, Siersema PD, van der Zee DC. Laparoscopic Thal fundoplication in children: a prospective 10- to 15-year follow-up study. *Ann Surg.* Feb 2014;259(2):388-93.
4. van der Zee DC, Arenas NJT, Bax NMA. The value of 24-h pH study in evaluating the results of laparoscopic antireflux surgery in children. *Surg Endosc*. 1999;13(9):918-921.
5. Humphries LA, Hernandez JM, Clark W, Luberice K, Ross SB, Rosemurgy AS. Causes of dissatisfaction after laparoscopic fundoplication: The impact of new symptoms, recurrent symptoms, and the patient experience. *Surg Endosc*. 2013;27(5):1537-1545.
6. Ingerski LM, Modi AC, Hood KK, et al. Health-Related Quality of Life Across Pediatric Chronic Conditions. *J Pediatr.* 2010;156(4):639-644.
7. Dallemagne B, Weerts J, Markiewicz S, et al. Clinical results of laparoscopic fundoplication at ten years after surgery. *Surg Endosc*. Jan 2006;20(1):159-65.
8. Slim K, Bousquet J, Kwiatkowski F, Lescure G, Pezet D, Chipponi J. Quality of life before and after laparoscopic fundoplication. *Am J Surg*. Jul 2000;180(1):41-5.
9. Velanovich V. Using quality-of-life measurements to predict patient satisfaction outcomes for antireflux surgery. *Arch Surg*. Jun 2004;139(6):621-6.
10. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270 degrees fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. *Surg Endosc*. Jun 2010;24(6):1287-95.
11. Nicolau AE, Crăciun M, Zota R, Kitkani A. Quality of life after laparoscopic fundoplication for gastroesophageal reflux disease. Preliminary study. *Chirurgia (Romania)*. 2013;108(6):788-793.
12. Srivastava R, Downey EC, Feola P, et al. Quality of life of children with neurological impairment who receive a fundoplication for gastroesophageal reflux disease. *J Hosp Med*. May 2007;2(3):165-73.
13. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL™* 4.0 as a pediatric population health measure: Feasibility, reliability, and validity. *Ambul Pediatr.* 2003;3(6):329-341.
14. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. Dec 2005;5(6):705-19.
15. Varni JW, Burwinkle TM, Seid M. The PedsQL™ 4.0 as a school population health measure: Feasibility, reliability, and validity. *Qual Life Res*. 2006;15(2):203-215.
16. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: An analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;52.
17. Varni JW, Limbers CA. The pediatric quality of life inventory: measuring pediatric health-related quality of life from the perspective of children and their parents. *Pediatr Clin North Am*. 2009;56(4):843-863.

18. Thal A. A unified approach to surgical problems of the esophagogastric junction. *Ann Surg*. 1968;168(3):542–50.
19. Nissen R. [A simple operation for control of reflux esophagitis]. *Schweiz Med Wochenschr*. 1956;86(Suppl 20):590–2.
20. Varni JW, Seid M, Rode CA. The PedsQL: Measurement model for the pediatric quality of life inventory. *Med Care*. 1999;37(2):126-139.
21. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales in Healthy and Patient Populations. *Med Care*. 2001;39(8):800-812.
22. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr.* Aug 2005;41(2):178-85.
23. Richter JE, Bradley LA, DeMeester TR, Wu WC. Normal 24-Hr ambulatory esophageal pH values - Influence of study center, pH electrode, age, and gender. *Dig Dis Sci*. 1992;37(6):849-856.
24. Smout AJPM, Breedijk M, Van Der Zouw C, Akkermans LMA. Physiological gastroesophageal reflux and esophageal motor activity studied with a new system for 24-hour recording and automated analysis. *Dig Dis Sci*. 1989;34(3):372-378.
25. Omari TI, Schwarzer A, VanWijk MP, et al. Optimisation of the reflux-symptom association statistics for use in infants being investigated by 24-hour pH impedance. *J Pediatr Gastroenterol Nutr.* 2011;52(4):408-413.
26. Weusten BLAM, Roelofs JMM, Akkermans LMA, Van Berge-Henegouwen GP, Smout APM. The symptom-association probability: An improved method for symptom analysis of 24-hour esophageal pH data. *Gastroenterology*. 1994;107(6):1741-1745.
27. Ghoos YF, Maes BD, Geypens BJ, et al. Measurement of gastric emptying rate of solids by means of a carbon-labeled octanoic acid breath test. *Gastroenterology*. 1993;104(6):1640-1647.
28. van den Driessche M, Ghoos Y, Veereman-Wauters G. Maturation of gastric emptying in healthy infants and children and reference values for the 13C octanoic acid breath test. Chapter as part of the dissertation of M. van den Driessche, University of Leuven. 2001.
29. Gold BD. Comparing GERD manifestations in children and adults. *Gastroenterol Hepatol*. 2008;4(1):40-44.
30. Capito C, Leclair MD, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. *Surg Endosc*. Apr 2008;22(4):875-80.
31. Lobe TE. The current role of laparoscopic surgery for gastroesophageal reflux disease in infants and children. *Surg Endosc*. Feb 2007;21(2):167-74.
32. Kawahara H, Okuyama H, Kubota A, et al. Can laparoscopic antireflux surgery improve the quality of life in children with neurologic and neuromuscular handicaps? *J Pediatr Surg*. 2004;39(12):1761-1764.
33. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Qual Life Res*. 2008;17(6):895-913.
34. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res*. 2001;10(4):347-357.
35. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;51.
36. Hays RM, Valentine J, Haynes G, et al. The Seattle pediatric palliative care project: Effects on family satisfaction and health-related quality of life. *J Palliat Med*. 2006;9(3):716-728.

37. Campo JV, Comer DM, Jansen-McWilliams L, Gardner W, Kelleher KJ. Recurrent pain, emotional distress, and health service use in childhood. *J Pediatr.* 2002;141(1):76-83.
38. Janicke DM, Finney JW, Riley AW. Children's health care use: A prospective investigation of factors related to care-seeking. *Med Care*. 2001;39(9):990-1001.
39. Baerg J, Thorpe D, Bultron G, et al. A multicenter study of the incidence and factors associated with redo Nissen fundoplication in children. *J Pediatr Surg*. 2013;48(6):1306-1311.
40. Ngerncham M, Barnhart DC, Haricharan RN, Roseman JM, Georgeson KE, Harmon CM. Risk factors for recurrent gastroesophageal reflux disease after fundoplication in pediatric patients: a case-control study. *J Pediatr Surg*. Sep 2007;42(9):1478-85.
41. Rebecchi F, Allaix ME, Giaccone C, Morino M. Gastric emptying as a prognostic factor for long-term results of total laparoscopic fundoplication for weakly acidic or mixed reflux. *Annals of Surgery*. 2013;258(5):831-837.
42. Draaisma WA, Rijnhart-De Jong HG, Broeders IAMJ, Smout AJPM, Furnee EJB, Gooszen HG. Five-year subjective and objective results of laparoscopic and conventional Nissen fundoplication: A randomized trial. *Ann Surg*. 2006;244(1):34-41.

CHAPTER 5

Two-Year Outcome after Laparoscopic Fundoplication in Pediatric Patients with Gastroesophageal Reflux Disease

Rebecca K. Stellato[§1], MS
Femke V.M. Mulder[§2], BSc
Stefaan H.A. Tytgat[2], MD, PhD
Tycho S. Oudman[2], BSc
David C. van der Zee[2], MD, PhD
Femke A. van de Peppel - Mauritz[#3], MD, PhD
Maud Y.A. Lindeboom[#2], MD, PhD

[§]shared first author

[#]shared last author

[1]Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, The Netherlands

[2]Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, The Netherlands

[3]Department of General Practice, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, The Netherlands

J Laparoendosc Adv Surg Tech 2020; 30:834-840

ABSTRACT

Introduction
Many studies on short-term efficacy of laparoscopic antireflux surgery (LARS) have shown good to excellent results on reflux symptom control and health-related quality of life (HRQoL). Prospective studies on the long-term efficacy, however, are scarce and indicate that the efficacy of symptom control may decline over time. The aim of this study is to assess the 2-year outcome on reflux symptoms and HRQoL after LARS.

Materials and Methods
Between 2011 and 2013, 25 children (12 males, median age 6 [2–18] years) with proton pump inhibitor resistant gastroesophageal reflux disease were included in a prospective longitudinal cohort study. To assess reflux symptoms and HRQoL, patients and/or their caregivers were asked to fill out the validated, age-appropriate gastroesophageal reflux symptom questionnaire and Pediatric Quality of Life Inventory™ before, 3 months, 1 year, and 2 years after LARS.

Results
Two years after LARS, 29% of patients had moderate to severe reflux symptoms compared with 92% (P<.001) before operation and 12% 3-4 months after operation (P=.219). The significant increase in HRQoL shortly after fundoplication (80.0 compared with 69.5 (P=.004)) is not observed after 2 years (72.0 compared to 69.5, p=.312). Correlation between the impaired HRQoL scores and the recurrence of symptoms could not be verified.

Conclusions
Although the efficacy of LARS tends to deteriorate after 2 years, LARS is still effective in controlling reflux symptoms in the majority of patients. The short-term improvement in HRQoL after LARS appears to be transient.

INTRODUCTION

Short-term follow up studies have shown that laparoscopic antireflux surgery (LARS) in children is effective in 57%-100% of children with proton pump inhibitor resistant gastroesophageal reflux disease (GERD).[1] However, prospective studies on the long-term efficacy of LARS are scarce. In the few studies with a follow-up period of >6 months, success rates varying from 63 to 96% have been reported.[1-4] Only one prospective study examined the effect of LARS in children with >5 years of follow-up, finding complete relief of symptoms in only 57% of patients 10-15 years after surgery.[5] These results indicate that the effect of LARS may deteriorate over time.

In addition to disease-specific symptoms, assessment of health related quality of life (HRQoL) may offer a better evaluation of the impact of treatment on patients.[6,7] Most longitudinal studies in adults have observed a lasting effect of LARS on long-term HRQoL.[8-14] However, the underlying etiology of GERD is different for children than for adults, and effects of LARS may also differ.[15] To date, only a few studies have assessed the effect of LARS on HRQoL in children. Three studies observed an improvement in quality of life one to six months after LARS in children.[15-17] Two studies with longer follow-up also found sustained improvement, but did not use validated questionnaires to assess HRQoL.[18,19]

Identification of patient characteristics associated with LARS success would be helpful in decision-making for both surgeons and caregivers. Several studies have examined neurological impairment (NI) as a predictor of LARS success, finding significantly higher gastrointestinal reflux (GER) recurrence in children with NI versus normal neurodevelopment (NN);[20] an elevated, but not statistically significant, risk of recurrent GERD in children with NI;[21] or no difference in overall HRQoL improvement between NI and NN children.[15] Preoperative gastric emptying (GE) was not found to be related to success.[22] To our knowledge, no other predictors of LARS success have been examined.

The aim of this prospective study is to evaluate reflux symptom control and HRQoL in a pediatric population up to 2 years after operation. Potential predictors of failure after LARS and change in HRQoL are also examined.

MATERIALS AND METHODS

Study design

A prospective, longitudinal cohort study was conducted in three University Medical Centers in The Netherlands (Wilhelmina Children's Hospital, University Medical Center Utrecht (UMCU); Sophia's Children's Hospital, Erasmus University Medical Center (Erasmus MC) and Maastricht University Medical Center (MUMC)). All pediatric patients aged 2-18 who had been referred for ARS by a pediatrician/pediatric

gastroenterologist because of therapy-resistant gastroesophageal reflux symptoms and pathological gastroesophageal reflux on pH monitoring in combination with a positive symptom association probability were eligible for inclusion. Children unable to undergo investigation, or who had undergone prior esophageal or gastric surgery (with the exception of gastrostoma placement), were excluded. No eligible patients were excluded. Patients were included between July 2011 and December 2013, and were operated between Aug 2011 and May 2014. Data on the short-term follow-up was published previously;[17,23] this article extends the follow-up of the cohort to two years post-LARS.

Surgical procedures

All laparoscopic fundoplications were performed by specialized pediatric surgeons with extensive experience in pediatric upper gastrointestinal (GI) laparoscopic surgery. In the UMCU the anterior, partial fundoplication according to Thal[24] was used to perform fundoplication. Erasmus MC and MUMC used the posterior, total fundoplication according to Nissen.25 Details regarding the surgical procedure were published previously.23

Clinical assessment

Patients were assessed before surgery and 3-4 months, 1 year and 2 years after LARS. Both children and caregivers completed a reflux specific symptom questionnaire and the Pediatric Quality of Life Inventory (PedsQL) HRQoL questionnaire. Preoperative GE half-time was obtained from a breath test with either a 375-g pancake containing 45 mg of ^{13}C-labeled Na-octanoate (children >4 years) or 100 mg of ^{13}C-labeled Na-octanoate added to a liquid formula (children <4 or unable to eat the pancake within 15 minutes); more details are provided in the previously published short-term outcome study.[23] GE half-time percentiles were calculated using reference values obtained by van den Driessche et al.[26] GE percentiles >75% were considered delayed.

Reflux specific symptom questionnaire

To assess reflux symptoms preoperatively and at 3 months and 1 and 2 years postoperatively, patients and/or their caregivers were asked to fill out the validated age-adjusted Gastroesophageal Reflux Symptom Questionnaire (GSQ).[27] All symptoms are rated for frequency and severity on a scale from 0 (none) to 7 (daily/most severe). Symptoms were defined as no symptoms (no symptoms reported), mild (mild symptoms weekly), moderate (mild symptoms daily or severe symptoms weekly), and severe (severe symptoms daily).

LARS failure

Failure after LARS was defined as either recurrence or persistence of moderate to severe GER symptoms (heartburn, regurgitation, food refusal and/or vomiting) as reported on the GSQ or the need for a redo procedure, or both.

HRQoL questionnaire

To assess the HRQoL, the Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales were used.[28] This questionnaire includes child self-report (in case of NN) for the ages 5-18 years and caregiver proxy-report for the ages 2-18 years. For both parallel reports age-adjusted versions were used (2–4, 5–7, 8–12, and 13-18 years), differing only in age-appropriate language.[28] The PedsQL scores in four different domains: physical, emotional, social and school functioning. Emotional, social and school functioning scores are summarized into the psychosocial health summary score. All four domains together represent the total score. The five-point response scale is scored from 0 to 100, resulting in a higher score indicating a better HRQoL.

Ethical approval and trial registration

This study was registered at the start of the study in the Dutch national trial registry (Identifier: 2934). Ethical approval for this prospective multicenter study was obtained from the University Medical Center Utrecht Ethics Committee, and local approval was obtained by the remaining two participating centers. Before study procedures, informed consent from the patients' caregivers and children (≥12 years) was obtained.

Statistical analysis

Continuous variables, when symmetric, were expressed as mean and standard deviation (SD). Skewed variables were expressed as median with interquartile ranges. The McNemar test was used to compare proportions of children with reflux symptoms at different time intervals.

A logistic regression model was used to identify predictors of failure 2 years after LARS. Prespecified potential predictors were preoperative GE percentile and age at the time of operation. Because two procedures were used in this study, and because previous studies identified NI as a potential predictor, the associations of failure with type of fundoplication and with NI were also examined using Fisher's exact tests.

To identify changes over time and predictors of HRQoL, linear mixed models were used. Mixed models account for correlations of repeated measures within children, and allow for estimation of effects in the presence of missing data. In each model, the outcome was the HRQoL score (total, physical health summary, or psychosocial summary). Scores from caregiver proxy-report and self-report (when available) were analyzed in one model. Potential predictors were time since operation (modeled as categorical), gender, age, type of fundoplication, type of report (self/proxy), the presence of NI, and preoperative delayed GE. To account for correlation of self-reported and proxy-reported scores from the same child, and scores within a child over time, a random intercept per child was included and a first-order autoregressive correlation matrix[29] was added to the residuals.

Secondary analyses were performed to determine whether the observed effect of time since operation was actually an effect of the presence of reflux symptoms: "reflux symptoms" was added as a time-varying covariate to the model for the HRQoL total score. In addition, interactions of follow-up time with NI and delayed GE were examined to determine whether these groups of patients experienced a different pattern of HRQoL total score over time. These interactions were tested using a likelihood ratio test (LRT).

Differences with a P-value <.05 were considered statistically significant. All analyses were performed using SPSS version 21 (IBM Corp., Armonk, NY).

RESULTS

Between 2011 and 2013, a total of 25 children underwent LARS, of which 18 Thal and 7 Nissen fundoplications. Median age of the included patients was 6 (range 2–18) years at the time of fundoplication (Table 1). Five children (20%) had NI, with different underlying causes (CHARGE syndrome, mitochondrial complex II deficiency, posthypoxic encephalopathy, congenital rubella infection, and unknown origin). One year after surgery, 1 patient was lost to followup; 23 patients completed the GSQ, whereas only 12 patients and 13 caregivers filled out a PedsQL. Two years after LARS, all remaining 24 patients completed the GSQ, and 21 caregivers and 15 children were able to complete the PedsQL.

Table 1. Baseline characteristics of the patients

	Median (IQR)
Age at time of operation (years)	6.0 (3.0-10.0)
Duration of hospital admission (days)	3.0 (2.0-4.5)
Gastric emptying (percentile)	70.0 (12.5-85.0)
	n (%)
Male gender	12/25 (48.0%)
Impaired neurodevelopment	5/25 (20.0%)
CHARGE syndrome	1/25 (4%)
mitochondrial complex II deficiency	1/25 (4%)
posthypoxic encephalopathy	1/25 (4%)
congenital rubella infection	1/25 (4%)
unknown origin	1/25 (4%)
Type of fundoplication	18/25 (72.0%)
Thal	7/25 (28.0%)
Nissen	
Gastrostomy preoperatively in situ	4/25 (16.0%)
Preoperative delayed gastric emptying	13/24 (54.2%)

Symptom assessment

Three months after LARS, 12% (3/25) of patients reported moderate to severe reflux symptoms, a significant reduction compared with the 92% (23/25) before the operation. Two years after LARS, 29% (7/24) reported moderate to severe reflux symptoms; 1 of these was a patient who had a redo procedure after 4 months of follow-up. Although not statistically significant (P = .219), this was an increase compared with 3 months post-LARS. Reflux symptoms after LARS remain significantly better than preoperative levels (P < .001).

The total failure rate at 2 years was 8/24 (33%). In addition to patients with recurrent GER symptoms, 2 patients (8%) required a redo procedure, 1 due to hiatal herniation and 1 because of severe recurrence of symptoms.

Six patients (25%) reported use of acid suppressive medication at 2-year follow-up. Moderate to severe dysphagia was reported in 4 patients (16%) after 2 years of follow-up, of which 1 was new-onset dysphagia.

Predictors of failure

The results of the logistic regression were as follows: although there was a trend toward higher failure rates among older children (odds ratio [OR] = 1.26 per year, 95% confidence interval [CI 0.98 to 1.63]), neither age nor GE (OR = 1.02 per GE percentile, 95% CI [0.99 to 1.05]) was significantly associated with failure. The 2-year failure rates for NI (40%) and NN (32%) children did not differ significantly (P = .555), and there was no statistically significant difference for Thal versus Nissen procedures (28% and 50%, respectively, P = .362).

Health related quality of life

Figure 1 displays the HRQoL caregiver proxy scores and children's self-reported total scale scores. Most children experienced an improvement in HRQoL 3 months after LARS; 2 years after LARS this effect diminishes. This pattern is also seen in the subdomains of the HRQoL questionnaire, and both in proxy and self-reported scores (Table 2).

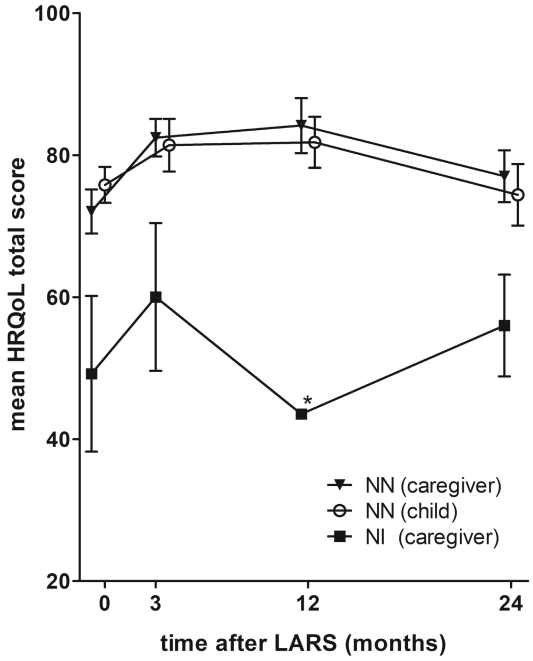


Figure. 1. Pediatric Quality of Life™ Inventory total scale scores over time for 20 children with normal neurodevelopment (NN, caregiver proxy report and child self-report) and 5 children with neurological impairment (NI, caregiver proxy report only) who underwent LARS. Points are means, bars indicate 1 SE. *It was not possible to calculate an SE for this mean because only 1 caregiver returned the questionnaire. HRQoL, health-related quality of life; LARS, laparoscopic antireflux surgery; NI, neurological impairment; NN, normal neurodevelopment; SE, standard error.

Table 2. Observed Health-Related Quality of Life Scores, as Measured by the Pediatric Quality of Life Inventory, Caregiver Proxy, and Child Self-Report

	Preoperative (mean; SD)	Postoperative 3-4 months (mean; SD)	Postoperative 1 year (mean; SD)	Postoperative 2 years (mean; SD)
Caregiver proxy	n=25	n=24	n=13	n=21
Physical	69.5 (25.4)	80.0 (20.7)	80.2 (27.3)	72.0 (29.1)
Psychosocial	66.8 (18.1)	77.0 (17.2)	81.7 (14.3)	74.6 (14.1)
Total Scale Score	67.5 (18.5)	77.8 (16.9)	81.0 (17.1)	73.0 (16.8)
Child self-report	n=12	n=12	n=12	n=15
Physical	81.3 (11.3)	88.3 (13.2)	85.9 (20.3)	77.1 (23.7)
Psychosocial	74.0 (10.4)	79.2 (14.9)	80.5 (11.7)	73.5 (17.1)
Total Scale Score	75.8 (8.8)	81.4 (12.8)	81.8 (12.5)	74.4 (16.8)

The mean total scale score (estimated by the mixed model) is the lowest before the operation and the highest 3 months after (difference in means 8.9, 95% CI [4.8 to 13.1]) (Table 3). One year post-LARS, the estimated mean total scale score is still significantly higher than preoperative (difference in means 7.7, 95% CI [2.5 to 13.0]). Two years post-LARS, the estimated mean HRQoL has nearly returned to preoperative levels and is no longer significantly different from preoperative HRQoL (difference in means 2.6, 95% CI [-2.4 to 7.7]), although it is also not significantly lower than short-term levels (difference in means 4.4, 95% CI [-2.4 to 11.2]. NI was the only factor that was significantly related to HRQoL; children with NN scored 28.8 points higher, 95% CI [16.6 to 40.9]. Girls, older children, children with Nissen fundoplication and children with delayed GE had lower total HRQoL at any given point in time; however, none of these variables was statistically significant. Caregiver proxy report scored nearly one point lower than children's self-report at the same time point; this difference was also not statistically significant.

Results of the model that included reflux symptoms were inconclusive. GER symptoms were strongly related to follow-followup time, and including both in the model resulted in unrealistic estimates and standard errors (SE), indicating multicollinearity. Descriptive statistics did not indicate a clear relation between presence of reflux symptoms and HRQoL. The patterns of HRQoL over time were similar for the physical and psychosocial health summary scores, although in the case of the physical health summary score the difference between 1 year post-LARS was not statistically significant from the preoperative level. Children with NI had significantly lower means for both scores, and children with delayed GE had a significantly lower mean physical health score (Tables 4 and 5).

Table 3. Estimated Parameters and Standard Errors from Linear Mixed Model Analysis of Caregiver Proxy and Child Self-Reported Pediatric Quality of Life Inventory Total Scale Score During Complete Follow-up Time

Parameter	Estimate	SE	95% CI	P
Intercept	47.4	9.3	28.4 - 66.5	<.0005
Postoperative 3 months	8.9	2.1	4.8 - 13.1	<.0005
Postoperative 1 year	7.7	2.7	2.5 - 13.0	.004
Postoperative 2 years	2.6	2.5	-2.5 - 7.7	.312
Female gender	-6.3	4.6	-16.0 - 3.3	.187
Normal neurodevelopment (Ref.: NI)	28.7	5.9	16.6 - 40.9	<.0005
Thal fundoplication (Ref.: Nissen)	3.9	5.6	-8.0 - 15.7	.502
Age at operation (years)	-0.7	0.6	-1.9 - 0.6	.270
Caregiver proxy (Ref.: self-report)	-0.8	2.5	-6.0 - 4.4	.761
Preoperative delayed GE (Ref.: no)	-9.1	4.4	-18.3 – 0.2	.055

Table 4. Estimated Parameters and Standard Errors from Linear Mixed Model Analysis of Caregiver Proxy and Child Self-Reported Pediatric Quality of Life Inventory Physical Health Summary Scores During Complete Follow-up Time

Parameter	Estimate	SE	95% CI	P
Intercept	29.4	11.8	5.3 - 53.5	.0187
Postoperative 3 months	9.1	3.0	3.0 - 15.1	.0037
Postoperative 1 year	6.0	3.8	-1.6 - 13.5	.1212
Postoperative 2 years	-1.8	3.6	-9.0 - 5.4	.6176
Female gender	-5.2	5.8	-17.3 - 6.8	.3754
Normal neurodevelopment (Ref.: NI)	48.5	7.5	33.1 - 63.8	<.0005
Thal fundoplication (Ref.: Nissen)	9.2	7.0	-5.6 - 23.9	.2084
Age at operation (years)	-0.8	0.7	-2.3 -0.7	.2850
Caregiver proxy (Ref.: self-report)	-1.7	3.5	-8.9 - 5.6	.6415
Preoperative delayed GE (Ref.: no)	13.7	5.5	2.1 - 25.3	.0228

Table 5. Estimated Parameters and Standard Errors from Linear Mixed Model Analysis of Caregiver Proxy and Child Self-Reported Pediatric Quality of Life Inventory PsychosocialHealth Summary Scores During Complete Follow-up Time

Parameter	Estimate	SE	95% CI	P
Intercept	54.0	10.0	33.3 - 74.7	<.0005
Postoperative 3 months	8.8	2.2	4.5 - 13.1	.0001
Postoperative 1 year	8.4	2.7	3.0 - 13.9	.0027
Postoperative 2 years	4.4	2.6	-.8 - 9.7	.0960
Female gender	-7.1	5.1	-17.7 - 3.6	.1808
Normal neurodevelopment (Ref.: NI)	21.4	6.4	8.1 - 34.7	.0028
Thal fundoplication (Ref.: Nissen)	2.1	6.2	-10.9 - 15.1	.7438
Age at operation (years)	-0.6	0.6	-1.9 - 0.8	.3709
Caregiver proxy (Ref.: self-report)	-0.5	2.6	-5.7 - 4.8	.8574
Preoperative delayed GE (Ref.: no)	7.8	4.9	-2.5 - 18.0	.1283

Predictors of postoperative changes in health related quality of life

No statistically significant interactions were found for follow-up time and NI (P value LRT .763) or time and delayed GE (P value LRT .582) on the mean total scale score. Therefore, we found no evidence that NI or preoperative delayed GE affected the progression of HRQoL.

DISCUSSION

Failure rate 2 years after LARS, based on recurrence of self-reported moderate to severe reflux symptoms and/or reoperation, was 33%. Several previously published prospective studies on longer term (2–5 years) results after LARS found lower failure rates, ranging from 10% to 22%,[3,20] although one reported a 37% failure rate after a median of 4 years follow-up.[4] However, it must be noted that there is no uniform definition of success or failure of LARS in the literature. Definition of failure ranges from recurrence of symptoms, in any severity, to solely the need for reoperation. Furthermore, subjective (questionnaires) as well as objective (pH-measurement) methods have been used to determine the outcome in various studies. It is therefore difficult to compare our findings directly with previous research. Our study categorizes a patient as “failure” when he or she reports moderate to severe reflux symptoms, that is, severe symptoms weekly or mild symptoms daily, and/or reoperation. This could lead to an overestimation of failure, since patients with a significant decrease in self-reported symptoms (from, e.g., multiple times a day before operation to less than once a day after) would still be labeled as a “failure”, whereas patients and/or their caregivers could still be satisfied with the result. Recurrent reflux was an indication for reoperation in 1 patient.

No patient characteristics were found to be significantly associated with LARS failure in this study. Previous studies specifically examining characteristics predictive of LARS failure found no association with GE[22] or NI,[15,21] although one study did find a difference in recurrence rates between NI and NN children.[20]

In this study, 12% of patients had new-onset dysphagia 3–4 months after operation; 1 of these patients (4%) still reported dysphagia after 2 years. The incidence of new-onset dysphagia is within the range of previous research.[30,31] However, these studies report only temporary dysphagia (up to half a year after surgery). The patient with persistent dysphagia also suffered from recurrence of reflux symptoms. Reflux esophagitis is associated with dysmotility of the esophagus and, therefore, dysphagia might be a manifestation of recurrent reflux.[32]

This study was the first to analyze the effect of LARS on HRQoL using validated questionnaires at multiple postoperative time points. HRQoL significantly increased 3 months after LARS.[17] After 2 years of follow-up, however, the positive effect of LARS on HRQoL decreases, although not significantly, to near-preoperative levels. This may indicate that the positive effect of LARS in children with GERD diminishes over time. To the best of our knowledge, only one previous study of fundoplication in children has published HRQoL scores at multiple postoperative time points.[19] In contrast to our findings, that study found an even stronger positive effect on HRQoL after 4 years of follow-up, compared to the short-term effect. Nearly all longitudinal HRQoL studies in adults observed a lasting effect of LARS on long-term efficacy and HRQoL.[8-14] However, in all but one[8] of these studies, there was considerable dropout over time, and none used statistical methods to correct for missing data. The observed effects, both in children and adults, may be a reflection of truly sustained HRQoL after LARS, or merely a reflection of selective dropout.

Neither delayed GE nor NI was found to be related to a different pattern of HRQoL over time. The latter is in agreement with a previous report[20] comparing NI and NN children.

The use of mixed models allowed us to analyze all available information on children (self-report or proxy) collected for the 2 years of follow-up, and allows for valid estimation in the presence of missing outcomes due to unreturned questionnaires (especially at 1 year follow-up) or data “missing by design” (no self-reports for patients <5 years, and no proxy reports for patients >18 years). The results of mixed models are less likely to be affected by selective dropout than paired analyses on complete cases.

When considering the outcomes of this study, some (methodological/statistical) limitations have to be taken into account. The sample was relatively small, and some effects, although potentially clinically relevant, were not found to be statistically significant. The small sample size is likely also the reason no determinants of failure or

increase in HRQoL could be identified. Most studies on effects of LARS, especially those in children, are quite small.

Because the type of procedure depended on the expertise of the participating center, the majority of children in this study were operated using a partial (Thal) fundoplication, whereas less than a third had a complete (Nissen) fundoplication. Despite these differences between centers, we do not expect the type of fundoplication to affect our results regarding failure. A large, multicenter study found no differences in reflux recurrence or redo procedures among children undergoing Thal, Nissen or Toupet procedures.[30] In a meta-analysis on long-term effects of complete vs. partial fundoplication, a statistically nonsignificant difference was found, with complete fundoplication resulting in slightly better reflux control.[33] A more recent study[34] and this study found slightly higher, although non-significant, failure rates among patients operated with Nissen compared with Thal.

Although we intentionally used generic HRQoL measurements, a more disease-specific questionnaire could be more sensitive to long-term improvement in HRQoL. In future research among pediatric LARS patients, the GI supplement to the PedsQL questionnaire[35] could be used.

In conclusion, LARS is an effective therapy in terms of reflux symptom control in children after 2 years, but the positive effect of LARS on HRQoL may diminish over time. We could not identify predictors of LARS failure or of postoperative improvement in HRQoL.

REFERENCES

1. Mauritz FA, van Herwaarden-Lindeboom MY, Stomp W, et al. The effects and efficacy of antireflux surgery in children with gastroesophageal reflux disease: a systematic review. *J Gastrointest Surg*. Oct 2011;15(10):1872-8.
2. Kristensen C, Avitsland T, Emblem R, Refsum S, Bjornland K. Satisfactory long-term results after Nissen fundoplication. *Acta Paediatr.* May 2007;96(5):702-5.
3. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
4. Fyhn T, Knatten C, Edwin B, et al. Randomized Controlled Trial of Laparoscopic and Open Nissen Fundoplication in Children. *Ann Surg*. 2015;261(6):1061-7.
5. Mauritz FA, van Herwaarden-Lindeboom MY, Zwaveling S, Houwen RH, Siersema PD, van der Zee DC. Laparoscopic Thal fundoplication in children: a prospective 10- to 15-year follow-up study. *Ann Surg*. Feb 2014;259(2):388-93.
6. Guyatt GH, Feeny DH, Patrick DL. Measuring Health-Related Quality of Life. *Ann Intern Med*. 1993;118(8):622-9.
7. The WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med*. 1995;41(10):1403-9.
8. Slim K, Bousquet J, Kwiatkowski F, Lescure G, Pezet D, Chipponi J. Quality of life before and after laparoscopic fundoplication. *Am J Surg*. Jul 2000;180(1):41-5.
9. Velanovich V. Using quality-of-life measurements to predict patient satisfaction outcomes for antireflux surgery. *Arch Surg*. Jun 2004;139(6):621-6.
10. Kamolz T, Granderath F, Schweiger U, Pointner R. Laparoscopic Nissen fundoplication in patients with nonerosive reflux disease. *Surg Endosc Interv Tech*. 2005;19(4):494–500.
11. Dallemagne B, Weerts J, Markiewicz S, et al. Clinical results of laparoscopic fundoplication at ten years after surgery. *Surg Endosc*. Jan 2006;20(1):159-65.
12. Pidoto RR, Fama F, Giacobbe G, Gioffre' Florio MA, Cogliandolo A. Quality of life and predictors of long-term outcome in patients undergoing open Nissen fundoplication for chronic gastroesophageal reflux. *Am J Surg*. Apr 2006;191(4):470-8.
13. Woodcock S, Watson D, Lally C, et al. Quality of life following laparoscopic anterior 90 degrees versus Nissen fundoplication: results from a multicenter randomized trial. *World J Surg*. 2006;30(10):1856-63.
14. Zéman Z, Tihanyi T. Quality of life and patient satisfaction after laparoscopic antireflux surgery using the QOLARS questionnaire. *Surg Endosc*. 2007;21(8):1418-22.
15. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270 degrees fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. *Surg Endosc*. Jun 2010;24(6):1287-95.
16. Srivastava R, Downey EC, Feola P, et al. Quality of life of children with neurological impairment who receive a fundoplication for gastroesophageal reflux disease. *J Hosp Med*. May 2007;2(3):165-73.
17. Mauritz FA, Stellato RK, van Heurn LWE, et al. Laparoscopic antireflux surgery increases health-related quality of life in children with GERD. *Surg Endosc*. Aug 2017;31(8):3122-3129.
18. O'Neill JK, O'Neill PJ, Goth-Owens T, Horn B, Cobb LM. Care-giver evaluation of anti-gastroesophageal reflux procedures in neurologically impaired children: what is the real-life outcome? *J Pediatr Surg*. Mar 1996;31(3):375-80.
19. Pacilli M, Eaton S, McHoney M, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. *Arch Dis Child*. Jun 2014;99(6):516-21.

20. Capito C, Leclair MD, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. *Surg Endosc*. Apr 2008;22(4):875-80.
21. Knatten CK, Kvello M, Fyhn TJ, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. *J Pediatr Surg*. Jul 2016;51(7):1115-21.
22. Knatten CK, Fjeld JG, Medhus AW, et al. Preoperative liquid gastric emptying rate does not predict outcome after fundoplication. *J Pediatr Surg*. Apr 2017;52(4):540-543.
23. Mauritz FA, Conchillo JM, van Heurn LWE, et al. Effects and efficacy of laparoscopic fundoplication in children with GERD: a prospective, multicenter study. *Surg Endosc*. Mar 2017;31(3):1101-1110.
24. Thal A. A unified approach to surgical problems of the esophagogastric junction. *Ann Surg*. 1968;168(3):542–50.
25. Nissen R. [A simple operation for control of reflux esophagitis]. *Schweiz Med Wochenschr*. 1956;86(Suppl 20):590–2.
26. van den Driessche M, Ghoos Y, Veereman-Wauters G. Maturation of gastric emptying in healthy infants and children and reference values for the 13C octanoic acid breath test. Chapter as part of the dissertation of M. van den Driessche, University of Leuven. 2001.
27. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr*. Aug 2005;41(2):178-85.
28. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. Dec 2005;5(6):705-19.
29. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.
30. Esposito C, Montupet P, van Der Zee D, et al. Long-term outcome of laparoscopic Nissen, Toupet, and Thal antireflux procedures for neurologically normal children with gastroesophageal reflux disease. *Surg Endosc*. Jun 2006;20(6):855-8.
31. Mathei J, Coosemans W, Nafteux P, et al. Laparoscopic Nissen fundoplication in infants and children: analysis of 106 consecutive patients with special emphasis in neurologically impaired vs. neurologically normal patients. *Surg Endosc*. 2008;22(4):1054-9.
32. Diener U, Patti M, Molena D, Fisichella P, Way L. Esophageal dysmotility and gastroesophageal reflux disease. *J Gastrointest Surg*. 2001;5(3):260–5.
33. Mauritz FA, Blomberg BA, Stellato RK, van der Zee DC, Siersema PD, van Herwaarden-Lindeboom MY. Complete versus partial fundoplication in children with gastroesophageal reflux disease: results of a systematic review and meta-analysis. *J Gastrointest Surg*. Oct 2013;17(10):1883-92.
34. Wafa TA, El-Saied A, Dessouky NM, El-Ghazaly M, Barakat T. Laparoscopic Thal versus laparoscopic Nissen fundoplication in children. *Annals of Pediatric Surgery*. 2017;13(2):74-77.
35. Varni JW, Kay MT, Limbers CA, Franciosi JP, Pohl JF. PedsQL gastrointestinal symptoms module item development: qualitative methods. *Journal of Pediatric Gastroenterology and Nutrition*. May 2012;54(5):664-71.

CHAPTER 6

Five-Year Outcome of Laparoscopic Fundoplication in Pediatric GERD Patients: A Multicenter, Prospective Cohort Study

Rebecca K. Stellato[1], MS
Nadia Colmer[2], MD
Stefaan H.A. Tytgat[2], MD, PhD
David C. van der Zee[2], MD, PhD
Femke A. van de Peppel-Mauritz[3], MD, PhD.
Maud Y.A. Lindeboom[3], MD, PhD

[1]Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

[2]Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherlands

[3]Department of General Practice, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

Journal of Gastrointestinal Surgery 2020; 25:1412-1418.

ABSTRACT

Background
Gastroesophageal reflux disease (GERD) is a common disease in children. When drug treatment fails, laparoscopic anti-reflux surgery (LARS) is considered. Short-term follow-up studies report high success rates; however, few studies report long term results. The aim of this study was to describe the long term effects of LARS in pediatric patients.

Methods
A prospective, multicenter study of 25 laparoscopic fundoplication patients was performed. At 3 months and 1, 2 and 5 years postoperatively, patients and caregivers were asked to complete the gastroesophageal reflux symptom questionnaire to assess symptoms, and the PedsQL™ to assess health related quality of life (HRQoL).

Results
Reflux symptom severity was still significantly improved 5 years after LARS compared with preoperative levels ($p < 0.0001$). However, 26% of patients reported moderate or severe reflux symptoms. Dysphagia was reported in 13% of patients 5 years after LARS, and was more common in children with neurologic impairment and children who underwent a Nissen procedure. The increase in HRQoL 3 months postoperatively appears to decline over time: 5 years after surgery, HRQoL was lower, though not significantly, than 3 months postoperatively. HRQoL at 5 years was still improved, though also not significantly, than preoperative levels. The presence of reflux symptoms after surgery was not significantly associated with lower HRQoL.

Conclusions
LARS is effective for therapy-resistant GERD in children. Five years after surgery, reflux symptoms are still improved. However, we observed a decline in symptom-free patients over time. The initial increase in HRQoL shortly after LARS appears to decline over time.

INTRODUCTION

Numerous short-term studies on anti-reflux surgery (ARS) in children have been published.[1-4] However, well-designed prospective studies and long-term follow-up data are limited. A wide range in outcome after ARS has been reported, with short-term success rates of 57-100%[1] and long-term rates of 51–96%.[5-12] This range may be caused by heterogeneous groups, different surgical techniques, and different definitions of success. For example, some studies define the presence of reflux symptoms as a failure, whereas others define only the need for redo surgery as a failure. Moreover, most studies do not use validated questionnaires to adequately assess reflux symptoms at follow-up.[1,13]

Since GERD and ARS have a substantial influence on the lives of both patients and caregivers, it is important to evaluate not only reflux symptoms, but also on the pre- and postoperative health-related quality of life (HRQoL) of patients.[3,14] Several studies have reported sustained increases in HRQoL in adults,[15-17] but because the etiology of reflux in children is different from that of adults, results from studies in adults cannot be directly generalized to a pediatric population.[3] Though the majority of pediatric studies has described reflux and reflux associated symptoms, only a few studies have reported on short-term (1 to 6 months) HRQoL of the patients.[2,3,14] Studies with longer follow-up have used interviews with, or questionnaires for, adults[5,18,19] or have reported on parent-reported "well-being".[19,20] To our knowledge, no long-term studies of HRQoL after laparoscopic anti-reflux surgery (LARS) have been reported.

The aim of this study is to report the 5-year follow-up of a prospective, multicenter cohort of 25 pediatric patients who underwent LARS. Reflux and dysphagia symptoms and HRQoL are examined longitudinally using validated questionnaires, and potential predictors of LARS failure at 5 years are investigated.

MATERIALS AND METHODS

Study Design

In a prospective, multicenter cohort study, 25 patients were included from 3 hospitals in the Netherlands: Wilhelmina Children's Hospital, University Medical Centre Utrecht (UMCU); Sophia's Children's Hospital, Erasmus University Medical Centre (EMC) and Maastricht University Medical Hospital (MUMC). Patients were included between July 2011 and December 2013, and were operated between August 2011 and May 2014. All pediatric patients aged 2-18, with proton pump inhibitor resistant GERD, were eligible for inclusion. Patients with previous esophageal or gastric surgery (except gastrostomy placement) and patients with structural abnormalities (except esophageal hiatal hernia) were excluded.

Surgical Procedure

All patients from UMCU underwent an anterior, partial (Thal) fundoplication, whereas patients from EMC and MUMC underwent a posterior, total (Nissen) fundoplication. All fundoplications were performed laparoscopically by experienced pediatric surgeons. Details on the surgical procedure have been published previously.[4]

Clinical Assessment

Patients had been assessed preoperatively and 3 months after surgery using stationary manometry, 24-h multichannel intraluminal impedance pH monitoring, and a ^{13}C-labeled Na-octanoate breath test. GE half-time percentiles were calculated using reference values from a Dutch population (unpublished chapter of the dissertation of van den Driessche, M, University of Leuven. 2001); GE percentiles higher than 75% were considered delayed. Further details of the clinical assessment have been reported previously.[4] Patients and caregivers had also been asked to complete two questionnaires preoperatively and 3 months, and 1, 2 and 5 years postoperatively.

This article extends previous research on this cohort. The short-term follow-up (3–4 months) and intermediate term follow-up (1–2 years) after ARS have been described in previous papers.[4,21]

Reflux Symptom Questionnaire

The validated, age-adjusted gastroesophageal reflux symptom questionnaire (GSQ)[22] was used to assess reflux and dysphagia symptoms; when patients were older than 18 years of age at 5-year follow-up, the reflux disease questionnaire (RDQ)[23] was used instead. With the GSQ symptoms were scored for frequency in the last 7 days and severity ranging from not at all (1) to most (7) severe. The RDQ uses slightly different symptoms and scores range from 0 (none) to 5 (daily/most severe). Reflux symptoms were scored as heartburn (two items) regurgitation (two items) and vomiting. Dysphagia was scored as swallowing problems or pain during swallowing. Reflux and dysphagia from both questionnaires were scored as no symptoms, mild (mild symptoms weekly), moderate (mild symptoms daily or severe symptoms weekly) or severe (severe symptoms daily).

ARS Failure

Failure was defined as the need for redo fundoplication and/or recurring or persistent moderate to severe reflux symptoms according to the GSQ/RDQ.

Quality of Life Questionnaire

The PedsQL 4.0 Generic Core Scales was used to assess the health related quality of life (HRQoL).[24] For patients aged 4-18 years and with normal neurological development (NN) both patients and caregivers completed a questionnaire. For patients under 4 years of age and/or neurologically impaired patients (NI), only caregivers filled in the

questionnaire, and for NN patients older than 18 years only a self-report was completed. The language used in the PedsQL is age adjusted (ages 2–4, 5–7, 8–12, 13–18). Four domains are scored with the PedsQL, i.e., physical, emotional, social and school functioning. The scores of the emotional, social and school functioning are summarized into a psychosocial score. The total score is a summary of all four domains. The scores per domain and the summary scores are converted to a scale from 0 to 100, where a higher score indicate a better HRQoL.

Ethical Approval and Trial Registration

This study was registered at the start of the study in the Dutch national trial registry (www.trialregister.nl; Identifier: 2934). Ethical approval was obtained from the University Medical Center Utrecht Ethics Committee, and local approval was obtained by the two other participating centers. Informed consent from the patients' caregivers and patients (≥12 years) was obtained prior to study procedures.

Statistical Analysis

Continuous variables were expressed as mean and standard deviation/error of the mean, categorical variables as number and percentage.

To identify predictors of failure 2 years after LARS, a logistic regression including two pre-specified variables (preoperative GE percentile and age at time of operation) was used. Because two procedures were used in this study, and because previous studies identified NI as a potential predictor,[9,20] two additional analyses were performed: the associations of failure with type of fundoplication and with NI were examined using 95% confidence intervals.

Mixed effects models were used to analyze patterns of symptoms and HRQoL over time. Mixed models account for correlation of repeated measures within a patient, and allow for estimation of effects in the presence of missing outcomes over time.[25] To test the patterns of reflux and dysphagia over time, ordinal logistic mixed models were used. In both models, fixed effects were used for time since operation (categorical), type of fundoplication (Thal or Nissen), and neurological status (NN or NI), and a random intercept per patient was used to account for repeated measures. These models estimate odds ratios (ORs) for increasing symptom severity.

Because the HRQoL scales are continuous, linear mixed models were used to examine the HRQoL over time and relate it to potential explanatory variables. These models estimate differences/changes in HRQoL for the explanatory variables. The outcome for each model was the HRQoL score (total score, physical health subscore or psychosocial subscore). Both patients' and caregivers' reports (when available) were analyzed in one model. Potential explanatory variables (fixed effects) included were gender, age at time of operation, type of fundoplication (Thal or Nissen), neurological status (NN

or NI), presence of preoperative delayed gastric emptying, presence of moderate or severe reflux symptoms, time since operation (categorical) and type of report (patient vs. caregiver). A random intercept per patient was used, together with a continuous first-order autoregressive correlation matrix for the residuals to account for repeated measures at unevenly spaced time points.

In order to determine whether patients with NI, preoperative delayed GE or presence of moderate or severe reflux symptoms have different patterns of HRQoL (total score) over time, the interaction of follow-up time with NI, preoperative delayed GE or presence of reflux symptoms was added to the mixed model and tested with a likelihood ratio test.

A p-value <0.05 was considered statistically significant. All analyses were performed using R version 3.5.1.[26]

RESULTS

LARS was performed in 25 patients between 2011 and 2013. Baseline characteristics of the participants are displayed in Table 1. The mean age at the time of surgery was 7.3 years (range 2-18). One patient was missing during the 1- and 2-year follow-up, but responded again for the 5-year follow-up. Two patients were lost to follow-up at 5 years: a neurologically impaired male who had undergone a Thal fundoplication; and a neurologically normal female who had undergone a Nissen fundoplication. At the final follow-up, 23 patients completed the GSQ (N=19) or RDQ (N=4), 19 patients completed the PedsQL, and 17 caregivers completed the PedsQL. The mean follow-up time was 5.2 years (range 4.3–6.4 years).

Table 1. Characteristics of the 25 pediatric patients who underwent LARS

Age at time of operation, years – mean ± SD	7.3 ± 4.7
Male sex	13/25 (52.0%)
Neurodevelopmental impaired	5/25 (20.0%)
Type of fundoplication Thal Nissen	18/25 (72.0%) 7/25 (28.0%)
Preoperative delayed gastric emptying	13/24 (54.2%)
Gastrostomy preoperatively	4/25 (16.0%)
Duration of follow-up, years – mean ± SD	5.2 ± 0.4

Symptoms

Reflux symptom severity was still significantly improved 5 years after LARS compared with preoperative levels (p <0.0001). Five years after LARS, 26% of the patients

reported moderate or severe reflux symptoms compared with 12% at 3 months (Fig. 1a, p = 0.0076). Type of fundoplication and neurologic impairment were not statistically significantly associated with reflux severity (p = 0.7299 and 0.1431, respectively).

Three of the 23 patients (16%) reported moderate to severe dysphagia symptoms 5 year postoperatively (Fig. 1b). Two patients with moderate dysphagia symptoms after 2 years resolved to no or mild dysphagia symptoms after 5 years, while 2 other patients without dysphagia symptoms 2 years after LARS reported new-onset dysphagia at 5 years. All 3 patients with reported dysphagia 5 years postoperatively also reported having reflux symptoms, and all 3 had undergone a Nissen fundoplication.

Dysphagia symptom severity 5 years after LARS was not significantly improved compared with severity preoperatively or 3 months postoperatively (p = 0.3488 and p = 0.8144, respectively). In fact, dysphagia severity only was significantly improved 1 and 2 years after LARS compared with preoperative levels (p = 0.0156 and p = 0.0365, respectively). Children with Nissen fundoplication had an increased chance of more severe dysphagia than children with Thal (OR = 7.1, p = 0.0014), as did NI compared with NN children (OR = 21.0, p < 0.0001).

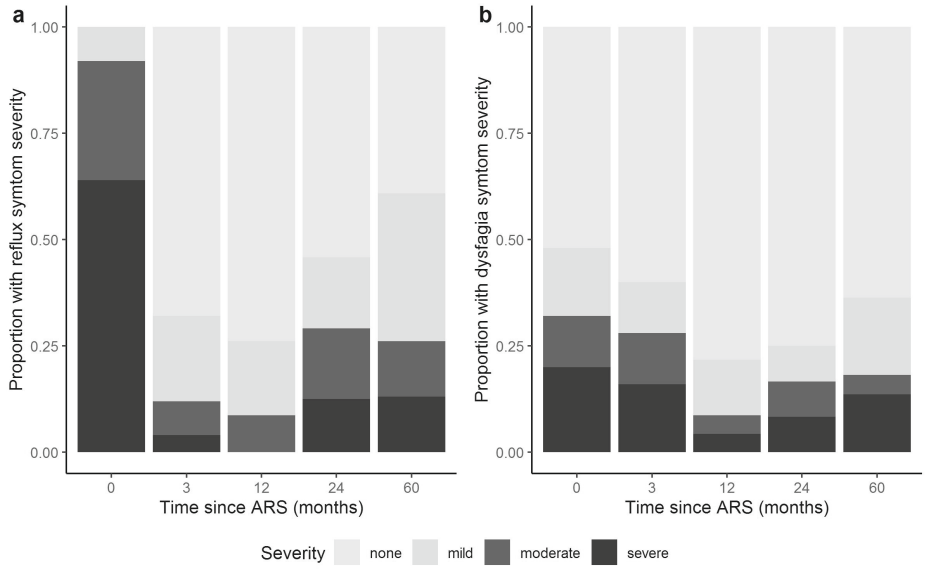


Fig. 1. Pre- and postoperative **(a)** reflux and **(b)** dysphagia symptoms for the cohort of 25 pediatric patients who underwent LARS. Symptoms are categorized as none, mild, moderate, and severe.

Failure

The failure rate after 2 years, including 2 redo procedures, was 8/24 (33%). Between 2 and 5 years postoperative no new redo procedures were necessary. At 5-year follow-up, there was a failure rate of 30%. The decrease in reflux symptoms and failure rate is due to 2 patients without a redo who reported reflux symptoms (and were therefore considered "failures") 2 years postoperatively, but reported no or mild reflux symptoms after 5 years. Including these 2 patients as failures after 5 years results in a cumulative failure rate of 39%. The 2 patients lost to follow-up at 5 years had no reported reflux symptoms after 2 years.

No variables were found to be significantly predictive of failure. The odds ratio (OR) for age was 1.02 per year (95% CI (0.98 - 1.07)), and the OR for GE was 1.00 per GE-percentile (95% CI (1.00–1.01)). NI children had a slightly higher percentage of failure than NN (40% vs. 35%, a difference of 5%, 95% CI (-30.6% to 45.6%)), and children with a Nissen procedure higher than Thal (42.9% vs 33.3%, a difference of 9.6%, 95% CI for (-25.9% to 45.9%)).

HRQoL

As reported previously, HRQOL increased significantly in this cohort 3 months after LARS, but decreased after 2 years, though not significantly.[14,21] The 5-year PedQL total score is higher, though not significantly, than preoperative levels (mean difference 4.8, 95% CI (-2.5; 12.0)) and lower (also not significantly) than 3 months postoperative levels (mean difference -4.3, 95% CI (-11.6; 2.9)).

Estimated effects of other potential predictors of HRQoL (total score and two subscores) can be found in Table 3. NI patients scored on average 25.0 (11.1; 38.8) points lower on the total score compared with NN patients (Fig. 2). Male gender, Thal fundoplication and a younger age at the time of operation resulted in higher mean HRQoL, though these differences were not statistically significant. Children with delayed GE preoperatively scored 9.7 points lower than those without delayed GE, though again this was not statistically significant, 95% CI (-20.4; 1.1). Patterns were similar for the physical and psychosocial health subscores (Table 2).

Table 3. Estimated parameters from linear mixed model analysis of caregiver proxy and child self-reported PedsQL total, physical health, and psychosocial scores during complete follow-up time.

	Total Score		Physical Health Subscore		Psychosocial Health Subscore	
Variable	Estimate [95% CI]	p-value	Estimate [95% CI]	p-value	Estimate [95% CI]	p-value
3 months postoperatively	9.1 [5.8; 12.5]	<0.0001	9.2 [4.2; 14.3]	0.0004	9.0 [5.3; 12.7]	<0.0001
1 year postoperatively	9.0 [3.7; 14.3]	0.0011	7.6 [0.1; 15.1]	0.0481	9.4 [3.8; 15.1]	0.0013
2 year postoperatively	3.5 [-1.5; 8.5]	0.1658	-0.8 [-7.8; 6.2]	0.8245	5.3 [-0.1; 10.6]	0.0532
5 year postoperatively	4.8 [-0.4; 10.0]	0.0696	1.8 [-5.4; 8.9]	0.6278	5.9 [0.5; 11.3]	0.0337
Male gender	6.4 [-4.8; 17.5]	0.2461	3.9 [-9.6; 17.3]	0.5551	7.6 [-4.1; 19.3]	0.1865
Neurologically impaired	-25.7 [-38.8; -11.1]	0.0013	-44.7 [-61.6; -27.9]	<0.0001	-17.7 [-32.2; -3.2]	0.0196
Nissen fundoplication	-3.0 [-16.7; 10.8]	0.6560	-9.1 [-25.6; 7.4]	0.2607	-0.6 [-14.9; 13.8]	0.9318
Age at operation (years)	-0.3 [-1.7; 1.1]	0.6638	-0.6 [-2.3; 1.1]	0.4729	-0.2 [-1.7; 1.3]	0.7762
Child report (vs. Caregiver)	-0.5 [-4.6; 3.6]	0.8214	0.5 [-5.0; 6.0]	0.8594	-0.8 [-5.0; 3.3]	0.6902
Preoperative delayed GE	-9.3 [-20.4; 1.1]	0.0746	-11.3 [-24.2; 1.7]	0.0836	-9.2 [-20.5; 2.0]	0.1024

CI = confidence interval; GE = gastric emptying

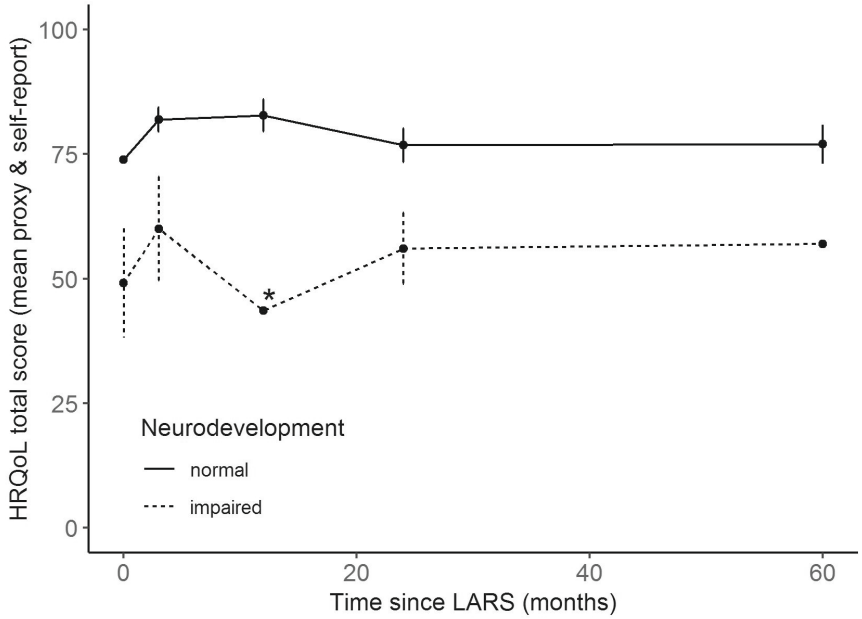


Figure 2. Mean HRQoL (PedsQL total scores) over time for the 20 children with normal neurodevelopment (mean of caregiver proxy report and child self-report) and 5 children with neurological impairment (caregiver proxy report only). Points are means; bars indicate 1 standard error. *It was not possible to calculate a standard error for this mean because only one caregiver returned the questionnaire.

Table 2. Mean (standard deviation) of pre- and postoperative HRQoL, as measured by the PedsQL (total scores, physical health and psychosocial health subscores) for the 25 pediatric patients who underwent LARS. Patient self-report (neurologically normal children aged 8 and older) and caregiver report (neurologically impaired children and all children under 18 years of age).

	Preoperatively	3 months PO	1 year PO	2 year PO	5 year PO
Patients	N=12	N=12	N=12	N=15	N=19
Tot	75.8 (8.8)	81.5 (12.8)	81.8 (12.5)	74.4 (16.8)	77.1 (17.8)
PH	81.3 (11.3)	88.3 (13.2)	85.9 (20.3)	77.1 (23.7)	80.8 (21.0)
PS	74.0 (10.4)	79.2 (14.9)	80.5 (11.7)	73.5 (17.1)	75.7 (19.1)
Caregivers	N=25	N=24	N=13	N=20 or 21*	N=17
Tot	67.5 (18.5)	77.8 (16.9)	81.0 (17.1)	73.0 (16.8)	71.5 (19.7)
PH	69.5 (25.3)	80.0 (20.7)	80.2 (27.3)	72.0 (29.1)	71.8 (28.3)
PS	66.8 (18.1)	77.0 (17.2)	81.7 (14.3)	74.6 (14.1)	71.8 (18.5)

PH = physical health; PS = psychosocial health; Tot = total score; PO = postoperatively
*N=20 (PS) 21(PH, Tot)
For the HRQoL total score, no statistically significant interactions were found between time and NI (p = 0.7721), time and preoperative delayed gastric emptying (p = 0.6478), and time and reflux symptoms (p = 0.2418). This indicates that these characteristics were not significantly associated with a different pattern of HRQoL over time.

DISCUSSION

The 5-year failure rate of 30% reported in this study falls within the range of 10-43% reported in previous studies.[5,6,9,10,27] The differences in failure rate across studies are likely caused by the lack of a uniform definition of success or failure of ARS. Some studies define failure by the presence of reflux symptoms, with different definitions of severity between studies, while others define failure only by redo fundoplication. The recurrence of reflux symptoms is assessed in some studies using objective measures, such as pH evaluation,[5,9] and sometimes with questionnaires or interviews.[19] Moreover, the questionnaires used in most studies have not been validated for children. This study is one of few studies that use validated questionnaires for assessment of both symptoms and HRQoL in a pediatric population.

None of the four variables examined was found to be a significant predictor of failure in this cohort. Two previous studies had found potentially elevated risk of GERD recurrence in NI children;[9,20] we saw virtually no difference in the proportions for NI and NN children. However, confidence intervals were wide and compatible with clinically meaningful differences in either direction.

Three months after LARS, only 12% of the patients reported persistent moderate or severe reflux symptoms. After 2 years that percentage increased to nearly 30%, and remained stable 5 years after LARS. An initial good short-term success rate of ARS, with a decline in effect in the long term, is in line with previous studies in children.[7,10,12]

The presence of moderate or severe dysphagia decreases postoperatively. Five years after LARS, 13% of the patients reported moderate or severe dysphagia, and all patients with reported dysphagia 5 years postoperatively also reported having reflux symptoms. Since reflux esophagitis is associated with dysmotility of the esophagus,[28] ongoing and late-onset dysphagia is likely to be a manifestation of recurrent reflux.

There was a clinically meaningful increase in HRQoL 3 months and 1 year postoperatively. Two and 5 years after LARS the HRQoL is decreased and, while still higher than preoperative levels, does not differ significantly from either pre- or postoperative levels. This is in contrast with previous studies, which reported a significant improvement in the quality of life both 6 months and 4 years after ARS in children[5] or high levels of patient "well-being" several years after surgery.[19,20] However, in those studies caregivers were asked only one question on the child's "overall quality of life" or "well-being," and in two[5,19] selective drop-out may also have biased results.

Three variables were examined for differing patterns of HRQoL over time. While NI patients reported a lower HRQoL at all times, they followed a similar pattern over time compared with NN patients; this is similar to a previous finding.[9] Children with

delayed preoperative GE did not have a different pattern in HRQoL over time than those with normal GE. And although children with and without reflux symptoms appeared to have diverging patterns of HRQoL over time, there was neither a significant interaction between reflux and time nor a significant difference in mean HRQoL averaged over all time points.

The current study has several limitations. As with many studies of pediatric ARS, the sample size is small. To compensate for the resulting limited power, we included one preoperative and four postoperative measurements per child on both symptoms and HRQoL, and used statistical methods that make efficient use of all available data and allow for valid estimation in the presence of missing data. Results of mixed effects models are less likely to be affected by selective dropout than analyses on complete cases (such as paired t-tests or repeated measures ANOVA).[25] The small sample may nevertheless have limited our ability to detect clinically meaningful differences, as reflected in some confidence intervals.

A second limitation is the use of two different surgical techniques for fundoplication. Consistent with the recommendations of Esposito et al.,[8] the type of procedure depended on the expertise of the participating center. Consequently the majority of children in this study were operated using a partial (Thal) fundoplication, while less than a third had a complete (Nissen) fundoplication. This is not expected to affect the results; previous studies have examined differences between partial and complete fundoplication and found similar success or reflux control rates between the two procedures.[8,29,30] However, we did see more, and more severe, dysphagia among children who underwent a Nissen procedure; this is consistent with a previous finding in children,[7] and with results in adults undergoing LARS.[31]

Finally, following the preoperative and 3 months' postoperative clinical assessments, the longer term follow-up in this study included only questionnaires to assess reflux symptoms. The use of questionnaires is less invasive than clinical examination, and the questionnaires were validated and easy to use. However, the questions may be interpreted differently by patients than by clinicians, resulting in over- or underestimation of both reflux and dysphagia. The presence of the assessed symptoms depends on multiple factors, such as comorbidity (such as obstipation) and the use of medication (such as anti-epileptics). A clinical examination could better determine whether symptoms are truly due to GERD or have another root cause.

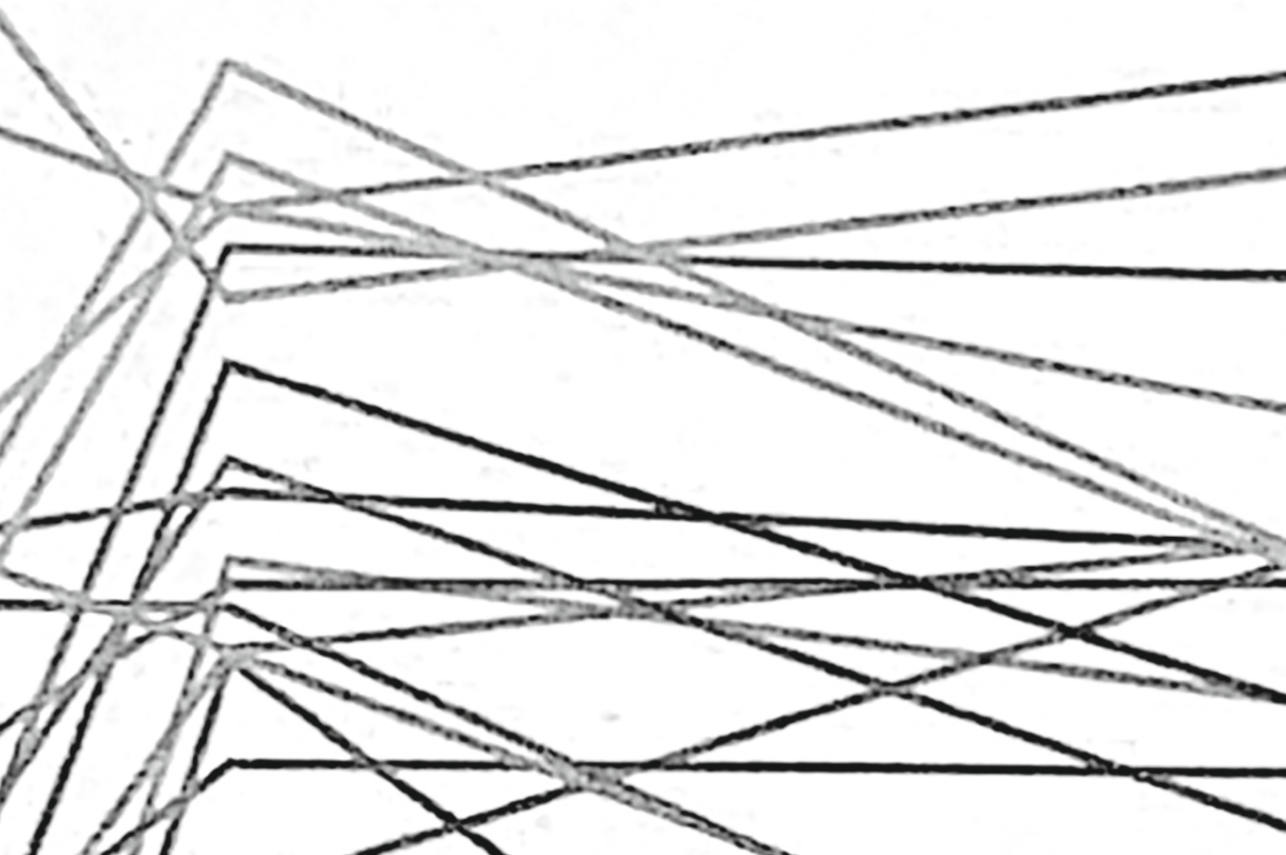
CONCLUSIONS

Guidelines for the treatment of pediatric GERD recommend conservative use of LARS, and clear communication to patients and/or caregivers regarding the potential benefits and risks associated with LARS.[32] Our results lead us to the same conclusion. After initial short-term improvement of reflux symptoms shortly after LARS, we found an increase in symptoms over time that appeared to stabilize after 2 years of follow-up; the 5-year prevalence of GERD symptoms was still significantly lower than before LARS. Similarly, following a short-term increase, HRQoL declines and then appears to remain stable 2 to 5 years after surgery.

REFERENCES

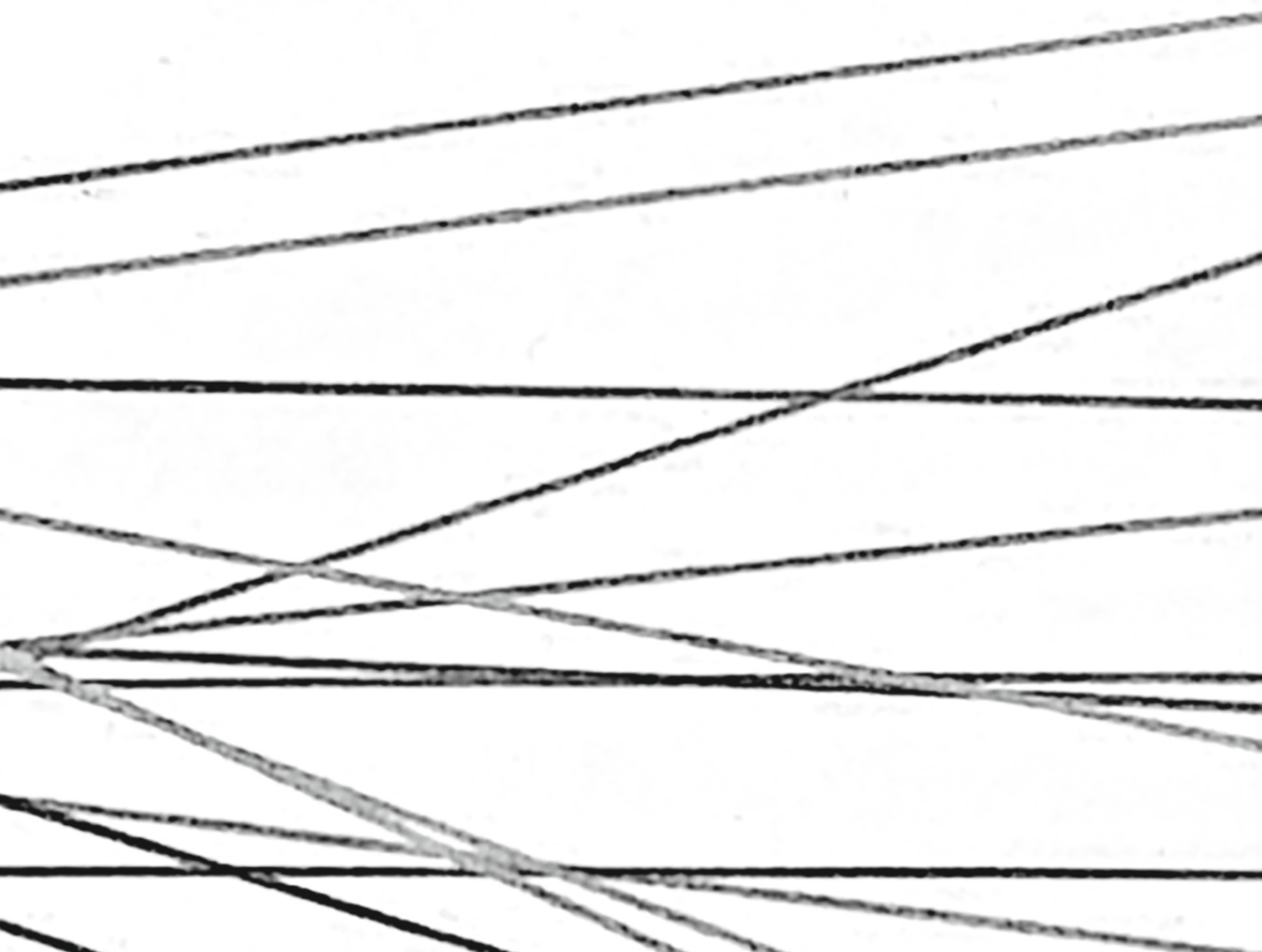
1. Mauritz FA, van Herwaarden-Lindeboom MY, Stomp W, et al. The effects and efficacy of antireflux surgery in children with gastroesophageal reflux disease: a systematic review. *J Gastrointest Surg*. Oct 2011;15(10):1872-8.
2. Srivastava R, Downey EC, Feola P, et al. Quality of life of children with neurological impairment who receive a fundoplication for gastroesophageal reflux disease. *J Hosp Med*. May 2007;2(3):165-73.
3. Engelmann C, Gritsa S, Ure BM. Impact of laparoscopic anterior 270 degrees fundoplication on the quality of life and symptoms profile of neurodevelopmentally delayed versus neurologically unimpaired children and their parents. *Surg Endosc*. Jun 2010;24(6):1287-95.
4. Mauritz FA, Conchillo JM, van Heurn LWE, et al. Effects and efficacy of laparoscopic fundoplication in children with GERD: a prospective, multicenter study. *Surg Endosc*. Mar 2017;31(3):1101-1110.
5. Pacilli M, Eaton S, McHoney M, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. *Arch Dis Child*. Jun 2014;99(6):516-21.
6. Koivusalo AI, Pakarinen MP. Outcome of Surgery for Pediatric Gastroesophageal Reflux: Clinical and Endoscopic Follow-up after 300 Fundoplications in 279 Consecutive Patients. *Scand J Surg*. Mar 2018;107(1):68-75.
7. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
8. Esposito C, Montupet P, van Der Zee D, et al. Long-term outcome of laparoscopic Nissen, Toupet, and Thal antireflux procedures for neurologically normal children with gastroesophageal reflux disease. *Surg Endosc*. Jun 2006;20(6):855-8.
9. Capito C, Leclair MD, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. *Surg Endosc*. Apr 2008;22(4):875-80.
10. Mauritz FA, van Herwaarden-Lindeboom MY, Zwaveling S, Houwen RH, Siersema PD, van der Zee DC. Laparoscopic Thal fundoplication in children: a prospective 10- to 15-year follow-up study. *Ann Surg*. Feb 2014;259(2):388-93.
11. Goessler A, Huber-Zeyringer A, Hoellwarth ME. Recurrent gastroesophageal reflux in neurologically impaired patients after fundoplication. *Acta Paediatr*. Jan 2007;96(1):87-93.
12. Fyhn T, Knatten C, Edwin B, et al. Randomized Controlled Trial of Laparoscopic and Open Nissen Fundoplication in Children. *Annals of surgery*. 2015;261(6):1061-7.
13. Rice S, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG. Laparoscopic anterior 180 degrees partial fundoplication: five-year results and beyond. *Arch Surg*. Mar 2006;141(3):271-5.
14. Mauritz FA, Stellato RK, van Heurn LWE, et al. Laparoscopic antireflux surgery increases health-related quality of life in children with GERD. *Surg Endosc*. Aug 2017;31(8):3122-3129.
15. Kamolz T, Granderath F, Schweiger U, Pointner R. Laparoscopic Nissen fundoplication in patients with nonerosive reflux disease. *Surg Endosc Interv Tech*. 2005;19(4):494–500.
16. Dallemagne B, Weerts J, Markiewicz S, et al. Clinical results of laparoscopic fundoplication at ten years after surgery. *Surg Endosc*. Jan 2006;20(1):159-65.
17. Zéman Z, Tihanyi T. Quality of life and patient satisfaction after laparoscopic antireflux surgery using the QOLARS questionnaire. *Surg Endosc*. 2007;21(8):1418-22.

18. O'Neill JK, O'Neill PJ, Goth-Owens T, Horn B, Cobb LM. Care-giver evaluation of anti-gastroesophageal reflux procedures in neurologically impaired children: what is the real-life outcome? *J Pediatr Surg*. Mar 1996;31(3):375-80.
19. Kristensen C, Avitsland T, Emblem R, Refsum S, Bjornland K. Satisfactory long-term results after Nissen fundoplication. *Acta Paediatr.* May 2007;96(5):702-5.
20. Knatten CK, Kvello M, Fyhn TJ, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. *J Pediatr Surg*. Jul 2016;51(7):1115-21.
21. Stellato RK, Mulder FVM, Tytgat SHA, et al. Two-year outcome after laparoscopic fundoplication in pediatric patients with gastroesophageal reflux disease. *J Laparoendosc Adv Surg Tech A*. 2020;in press
22. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *Journal of pediatric gastroenterology and nutrition*. Aug 2005;41(2):178-85.
23. Shaw MJ, Talley NJ, Beebe TJ, et al. Initial Validation of a Diagnostic Questionnaire for Gastroesophageal Reflux Disease. *Am J Gastroenterol*. 2001;96(1):52-57.
24. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(6):705–19.
25. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.
26. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing; 2018. https://www.R-project.org/
27. Lopez-Fernandez S, Hernandez F, Hernandez-Martin S, et al. Failed Nissen fundoplication in children: causes and management. *Eur J Pediatr Surg*. Feb 2014;24(1):79-82.
28. Diener U, Patti M, Molena D, Fisichella P, Way L. Esophageal dysmotility and gastroesophageal reflux disease. *J Gastrointest Surg*. 2001;5(3):260–5.
29. Mauritz FA, Blomberg BA, Stellato RK, van der Zee DC, Siersema PD, van Herwaarden-Lindeboom MY. Complete versus partial fundoplication in children with gastroesophageal reflux disease: results of a systematic review and meta-analysis. *J Gastrointest Surg*. Oct 2013;17(10):1883-92.
30. Wafa TA, El-Saied A, Dessouky NM, El-Ghazaly M, Barakat T. Laparoscopic Thal versus laparoscopic Nissen fundoplication in children. *Ann Pediatr Surg*. 2017;13(2):74-77.
31. Tan G, Yang Z, Wang Z. Meta-analysis of laparoscopic total (Nissen) versus posterior (Toupet) fundoplication for gastro-oesophageal reflux disease based on randomized clinical trials. *ANZ J Surg*. Apr 2011;81(4):246-52.
32. Vandenplas Y, Rudolph CD, Di Lorenzo C, et al. Pediatric gastroesophageal reflux clinical practice guidelines: joint recommendations of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN). *J Pediatr Gastroenterol Nutr.* Oct 2009;49(4):498-547.



PART III

SHORT- AND LONG-TERM EFFECTS OF LAPAROSCOPIC GASTROSTOMY PLACEMENT



CHAPTER 7

The Effect of Gastrostomy Placement on Health-Related Quality of Life in Children

Josephine Franken[1],
Rebecca K. Stellato[2]
Stefaan H.A.J. Tytgat[1]
David C. van der Zee[1]
Femke A. Mauritz[1¶]
Maud Y.A. Lindeboom[1¶]

[1]Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherlands

[2]Department of Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

[¶]F.A. Mauritz and M.Y.A. Lindeboom have shared last authorship

ABSTRACT

Background and purpose

A gastrostomy placement (GP) aims to improve nutritional status and health-related quality of life (HRQoL) in children who require long-term enteral tube feeding. We evaluated the effect of GP on HRQoL.

Methods

A prospective, longitudinal cohort study was performed including patients referred for laparoscopic GP. Children and/or caregivers were asked to fill out the validated PedsQL™ questionnaire before and 3 months after surgery. The aim was to compare preoperative with postoperative HRQoL and to identify predictors of HRQoL.

Results

Fifty patients were included with a median age of 3.4 years (interquartile range 1.4–5.6). After GP, total HRQoL did not significantly increase ($p = 0.30$). However, psychosocial health significantly increased: 55.8 (standard deviation±20.8) to 61.2(±19.6; $p = 0.03$) on a 100-point scale. This was mainly owing to an increase in social HRQoL: 58.2 (±32.3) to 68.3 (±27.9; $p = 0.04$). HRQoL both before and after GP was significantly lower in children with neurologic impairment ($p < 0.0005$). However, neurologic impairment did not influence the effect of surgery on HRQoL ($p = 0.66$). Low preoperative body mass index was a predictor for improvement in HRQoL after GP.

Conclusions

After GP in children, psychosocial HRQoL improved significantly. This was mainly owing to an improvement in social HRQoL.

INTRODUCTION

A gastrostomy placement (GP) is an effective treatment that provides long-term enteral tube feeding in children with feeding difficulties.[1,2] The main indications for GP are neurologic impairment (NI), cystic fibrosis and congenital cardiac disease.[3,4] The pediatric patients with these aforementioned conditions can suffer from poor nutritional status,[5] which may lead to increased morbidity. Also, feeding difficulties in these patients (e.g. refusal of food or prolonged feeding time) can have a negative impact on the lives of both patients and their caregivers.[6] GP, as a guaranteed route for enteral tube feeding, may not only lead to an improvement in nutritional status, but possibly lead to an improvement in other aspects in the lives of these patients as well, thereby increasing their health-related quality of life (HRQoL).

HRQoL is increasingly recognized as an essential part of patient care outcome. It aims to assess the impact of an illness and its treatment on the dimensions of physical and psychosocial health.[7-10] To our knowledge, no study has ever prospectively evaluated the effect of GP on HRQoL. One study reported on HRQoL before and after image-guided gastrostomy or gastrojejunostomy placement in neurologically impaired children. In this study no significant changes were reported, however, this study did not use validated HRQoL questionnaires.[11]

The lack of well-designed studies on GP and the effects of GP on its primary goal, improvement in HRQoL, led to the design of this study. Whereas the aforementioned study did not use validated questionnaires for HRQoL assessment, we used the Pediatric Quality of Life (PedsQL™) 4.0 generic core scales. The PedsQL™ is a validated diagnostic tool in healthy children as well as in children with numerous acute and chronic medical conditions. It has been proven to be reliable for both proxy-report by caregivers and parallel self-report for children.[8-10]

The primary aim of the current study is to evaluate the effect of GP on HRQoL in children prospectively. Although children in most cases have few alternatives for GP, it is important to understand the consequences of the operation on the lives of the children referred for GP, especially when providing information to caregivers. Our hypothesis is that GP in children leads to an improvement in HRQoL. We also aim to identify predictors of HRQoL and predictors of postoperative changes in HRQoL, thereby enabling us to identify the children who will gain the most benefit from GP. Finally, we investigated differences in HRQoL between self-report by patients and proxy-report by caregivers. We considered differences between proxy and self-reported HRQoL an interesting additional outcome, because various studies on HRQoL in children indicate that information provided by caregivers does not always correspond to what children report themselves.[12]

MATERIALS AND METHODS

Study design

Between May 2012 and April 2014, a prospective, longitudinal cohort study was performed including 50 pediatric patients that underwent laparoscopic GP at the Wilhelmina Children's Hospital, University Medical Center Utrecht (UMCU). Clinical assessment was performed before GP and 3 months after operation.

Surgical procedure

GP was performed laparoscopically under general anesthesia in all pediatric patients. All procedures were performed or supervised by an experienced pediatric surgeon. Operations were performed by 6 different pediatric surgeons.

Ethical approval and trial registration

This study was part of a larger trial on GP in children, registered under the name of 'The effect of laparoscopic gastrostomy on gastric emptying: A prospective observational study in children' at the Dutch trial register (NTR3314, 29-02-2012).

Ethical approval for the study was obtained from the UMCU Ethics Committee. Prior to initiating any study procedure, informed consent was obtained from the patients' parents or caregivers and the patients themselves (when 12 years or older and without NI).

Clinical assessment

Clinical assessment included the completion of the PedsQL™.[8-10] Questionnaires were completed in proxy-report by caregivers for all children. Additionally, children without NI completed a version of the questionnaires in self-report. Questionnaires were sent to the patients' house addresses, completed in private and sent back to the investigative team. The PedsQL™ is subdivided into four age-adjusted questionnaires (ages: 2–4; 5–7; 8–12 and 13–18 years) and a parallel self-report for children (ages: 5–7; 8–12 and 13–18 years). The inventory comprises 23 items. The total HRQOL score is divided into two main health scores: physical health summary score (8 items) and psychosocial health summary score (15 items), which in turn comprises the domains emotional scale score (5 items), social scale score (5 items), and functioning scale score (5 items). Scale scores per domain were computed as the sum of the items divided by the number of items answered. Items were then reverse-scored and transformed into a scale from 0 to 100, where higher scores indicate better HRQoL. The PedsQL™ for the age category 2–4 years is shown in **Table 1** as an illustration of HRQoL assessment.

Table 1. PedsQLTM questionnaire on health-related quality of life for age category 13-18 years.

Could you tell us to what extent your teenager had trouble with each of these things in the last month? There are no right or wrong answers. Please ask for help if you have any questions. **0** if it was **never** a problem **1** if it was **almost never** a problem **2** if it was **sometimes** a problem **3** If it was **often** a problem **4** if it was **almost always** a problem
Physical functioning (having trouble with…)
Walking more than 100 metres Running Doing sports or other physical exercise Heavy lifting Taking a bath or shower independently Having pain Feeling tired
Emotional functioning (having trouble with…)
Feeling afraid or scared Feelig sad Feeling angry Having trouble sleeping Being worried about what might happen to him/her
Social functioning (having trouble with…)
Getting along with other teenagers Other kids not wanting to be friends with her/him Begin bullied by other teenagers Not being able to do things other teenagers of his/her age can do Being able to keep up with other teenagers
Functioning at school (having trouble with…)
Paying attention in class Forgetting things Keeping up with work in class and doing his/her homework Not being able to go to school because he/she is not feeling well Not being able to go to school because he/she had to go to the doctor or hospital

Secondary outcomes

Additional data regarding complications and reinterventions were derived from the patient records. All encountered complications were registered. Complications were defined as such when adhering to fixed criteria: hypergranulation at the gastrostomy insertion requiring treatment with silver nitrate or surgical excision, infection at the gastrostomy insertion requiring treatment with antibiotics or antifungal medication and

dislodgement of the catheter requiring replacement. Leakage at the gastrostomy site was determined by the indication for (re)admission or gastrojejunostomy placement.

Feeding intolerance was determined with a questionnaire that was filled out by parents scoring the vomiting symptoms of their child on a frequency scale (0–7 days a week) and a severity scale (0–7). Patients with at least daily and moderately severe vomiting or at least weakly and severe vomiting were considered feeding intolerant.

Data concerning feeding regimen and weight and height values were collected with a gastrostomy-specific questionnaire. Weight and height measurements were converted to weight-for-height and height-for-age z-scores based on the Netherlands Organization for Applied Scientific Research (TNO) growth standards.[13] Z-scores allow comparison of an individual's weight or height, adjusting for age and sex relative to a reference population, expressed in standard deviations from the reference mean.

Statistical analysis

Continuous variables were expressed as mean ± standard deviations for symmetric variables or as median with interquartile ranges (IQR) for skewed variables. Correlations of continuous data were investigated with the Spearman's correlation coefficient.

A linear mixed model was used to compare pre- and postoperative HRQoL and to identify predictors of HRQoL and predictors of postoperative increase or decrease in HRQoL. Mixed models are appropriate for the analysis of repeated measurements, especially in the presence of missing data on the outcome variable.[14] Fixed effects were timing of the measurement (postoperative versus preoperative), age, neurologic impairment, cardiac disease, weight-for-length z-scores and postoperative complications of GP, and a random intercept per child was included. The variables included in the mixed model were chosen based on univariate analysis. Coefficients from the mixed model represent the predictive value of the variables on the outcome variable.

To examine the effects of the examined predictors on changes in HRQoL, interactions of all variables (except for the variable 'complications', because no preoperative values of this variable were available) with timing of the measurement (preoperative versus postoperative) were added to the mixed model analysis. A significant interaction indicates that the variable is associated with postoperative change in HRQoL.

A small subsample of children was also asked to complete HRQoL questionnaires. For this subsample, the responses of children and caregivers were compared using a linear mixed model. Fixed effects were timing (postoperative versus preoperative) and children versus caregivers; a random intercept per child was included to account for clustering of measurements within children.

Statistical significance was, when possible, expressed by 95% confidence intervals (CI). Where p-values were used, statistical significance was defined by *p*-values of less than 0.05. All analyses were performed using SPSS 24.0 statistical package (IBM, USA).

RESULTS

Patient inclusion

A total of 50 patients were included with a median age of 3.4 years (1.4 – 5.6). Patient characteristics are described in **Table 2**. The main underlying disease as a cause of feeding difficulty was NI (75.0%), which was clinically manifested as psychomotor retardation, epilepsy, spasticity, visual impairment and/or hypotonia. An overview of patient inclusion is depicted in **Fig 1**. In 28 out of 31 patients that were excluded from the study the reason was the refusal of parents to participate in the clinical tests that this study was combined with, namely 24-hour pH-Impedance monitoring studies and gastric emptying studies. Median follow-up time after GP was 4.6 months (3.7 – 5.6). Out of 50 included patients, 10 caregivers of patients (20.0%) did not fill out the postoperative PedsQL™ questionnaire resulting in missing data on HRQoL. There were no differences in preoperative HRQoL values between patients who completed both questionnaires and those who only filled out the preoperative questionnaires (**Table 3**).

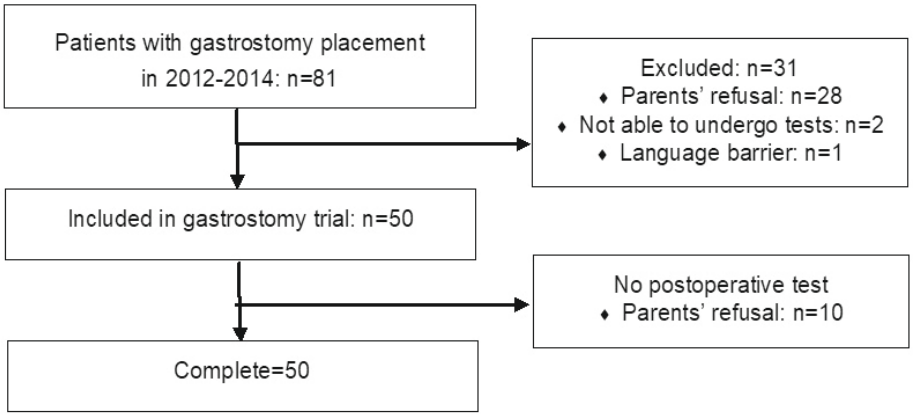


Figure 1. Flowchart of patient inclusion.

Table 2. Patient characteristics (n=50)

Demographics	
	n (%)
Male gender	29 (58%)
	Median (interquartile range)
Age at operation	3.4 (1.4 – 5.6)
Main underlying morbidity	
Neurologic impairment	39 (78%)
Cystic fibrosis	4 (8%)
Congenital cardiac disease	2 (4%)
Undiagnosed growth retardation	2 (4%)
Pulmonary disease	2 (4%)
Short bowel disease	1 (2%)

Health-related quality of life after gastrostomy placement

HRQoL results before and after operation are shown in **Table 4.** Differences between pre- and postoperative values are shown. Although the total score and all of its subdomains increased after GP, not all changes were statistically significant. The first row of **Table 5a** presents the estimated change in subcategories of HRQoL, adjusting for all other variables in the model. After GP, there was a nonsignificant increase in total HRQoL score of 2.8 points on a 100-point scale (CI -2.6 to 8.3; $p = 0.30$). However, analysis of subdomains of HRQoL found that, while physical health scores remained similar after operation (2.4 points; CI -1.7 to 6.3; $p = 0.24$), psychosocial health scores increased significantly (5.4 points; CI 0.5 – 10.3; $p = 0.03$). Further analysis of the subdomains of the psychosocial health scores found that this increase was mainly based on an increase in social scale score (10.1 points; $p = 0.04$). Patients improved in the items of 1. Getting along with other children of the same age; 2. Other kids not wanting to be friends with him/her; 3. Being able to keep up with other children. The emotional scale score (2.8 points; $p = 0.33$) and functional scale score (3.4 points; $p = 0.36$) did not increase significantly compared to the preoperative values.

Predictors of health-related quality of life

Results of the mixed model of HRQoL are shown in **Table 5a** (in which pre- and postoperative HRQoL is analyzed in one measure). Children with NI had significantly lower total HRQOL scores compared to children without NI (coefficient -30.5, CI -19.4 to -25.3; $p < 0.0005$). In analysis of subdomains of HRQoL, NI was predictive of both lower physical health scores (coefficient -43.8; CI -57.7 to -29.8; $p < 0.0005$) and lower psychosocial health scores (coefficient -22.6; CI -35.4 to -9.8; $p = 0.001$). Physical health scores increased with higher age at the time of operation (coefficient 1.8; CI 0.5 – 3.2;

p = 0.008). The other possible parameters, weight-for-length z-score (p = 0.73), cardiac disease (p = 0.09) and complications of GP (p = 0.43), did not predict HRQoL.

Predictors of changes in HRQoL after GP.

Results of the mixed model of changes in HRQoL after GP are shown in **Table 5b**. Analysis showed that preoperative weight-for-length z-score was negatively associated with a postoperative increase in total HRQoL score (coefficient -2.5 points per kg/m^2; CI -4.2 to -0.7; p = 0.01). Children with lower z-scores before operation showed a higher increase in postoperative HRQoL. In analysis of subdomains of HRQoL, the largest effect of weight-for-length z-scores was found in the domain of psychosocial health (coefficient -3.8 points per kg/m^2; CI -5.9 to -1.7; p = 0.001). For change in physical health scores, preoperative z-score was not a significant predictor. Age (p = 0.46), NI (p = 0.66) and cardiac disease (p = 0.79) were not predictive of postoperative change in HRQoL.

Differences in self-report by patients versus proxy-report by caregivers

Eleven patients were able to self-report on their HRQoL (27.5%). There was no statistically significant difference between patients' self-report and caregivers' proxy-report of total HRQoL scores, although children scored their own HRQoL on average higher than their caregivers with a difference of 4.27 points (p = 0.26).

Complications

In one patient the procedure could not be completed laparoscopically. In this patient, the sutures fixating the stomach to the abdominal fascia ruptured during insertion of the needle into the stomach. To complete the operation successfully, the surgeon converted to minilaparotomy. No other major events occurred during surgery. The majority of patients experienced postoperative complications as listed in **Table 6**. The occurrence of complications was not predictive of HRQoL (**Table 5b**).

In four patients a temporary nasoduodenal catheter was placed because of feeding intolerance with excessive vomiting or persistent leakage at the gastrostomy site. In these cases the gastrostomy was temporarily used for gastric drainage while administering tube feeding through the nasoduodenal catheter. Two of these patients ultimately underwent a gastrojejunostomy placement because of persisting symptoms.

No patients showed a deterioration of gastroesophageal reflux that required antireflux surgery during follow-up.

Table 3. Preoperative HRQoL in patients with only preoperative measures (n=10) and patients with both pre- and postoperative measures (n=40).

	Both measures (n=40)	Only preoperative measures (n=10)	Difference (CI)
Total Scale Score	53.9 ± 19.1	49.9 ± 15.8	-4.0 (-22.1 – 17.9)
Physical Health Summary Score	45.2 ± 24.8	39.0 ± 21.1	-6.2 (-26.3 – 14.3)
Psychosocial Health Summary Score	55.6 ± 21.4	54.8 ± 20.6	-0.8 (-17.2 – 15.2)
Emotional Scale Score	66.1 ± 18.1	63.3 ± 14.6	-2.8 (-16.0 – 11.5)
Social Scale Score	58.3 ± 32.7	59.4 ± 36.4	1.1 (-24.3 – 26.6)
Functioning Scale Score	45.8 ± 28.0	43.7 ± 29.5	-2.1 (-14.8 – 12.6)

Abbreviations: CI confidence interval; HRQoL health-related quality of life; NI neurologic impairment.

Table 4. Health-related quality of life in children before and after gastrostomy placement. Data are presented as mean **± standard deviation (n=40).**

	Before operation (n = 40)	After operation (n = 40)	Difference (n = 40)
Total Scale Score	53.9 ± 19.1	56.3 ± 20.5	2.1 ± 12.7
Physical Health Summary Score	45.2 ± 24.8	47.7 ± 28.0	2.1 ± 17.4
Psychosocial Health Summary Score	55.6 ± 21.4	61.2 ± 19.6	5.5 ± 15.7
Emotional Scale Score	66.1 ± 18.1	68.2 ± 19.2	2.8 ± 18.0
Social Scale Score	58.3 ± 32.7	68.3 ± 27.9	10.1 ± 24.9
Functioning Scale Score	45.8 ± 28.0	47.6 ± 26.9	3.4 ± 21.1

Abbreviations: CI confidence interval; HRQoL health-related quality of life; NI neurologic impairment.

Table 5a. Mixed model analysis of health-related quality of life.

	Total HRQoL score		Physical Health Summary Score		Psychosocial Health Summary Score	
	Coefficient (CI)	*p*-value	Coefficient (CI)	*p*-value	Coefficient (CI)	*p*-value
Postoperative (vs. preoperative)	+2.8 (-2.6 to 8.3)	0.30	+2.4 (-1.7 to 6.3)	0.24	+5.4 (0.5–10.3)	0.03
Age (years)	+0.9 (0.2–2.0)	0.09	+1.8 (0.5–3.2)	0.008	+0.2 (-1.0 to 1.4)	0.70
Preoperative weight-for-length z-score	-0.4 (-2.4 to 1.7)	0.73	+1.0 (-1.5 to 3.5)	0.42	-1.1 (-3.4 to 1.2)	0.34
NI (yes/no)	-30.5 (-41.6 to -19.4)	<0.0005	-43.8 (-57.7 to -29.8)	<0.001	-22.6 (-35.4 to -9.8)	0.001
Cardiac (yes/no)	-19.9 (-43.2 to 3.4)	0.09	-25.9 (-55.0 to 3.3)	0.08	-15.2 (-41.8 to 11.5)	0.26
Complication(s)	+1.3 (-2.0 to 4.7)	0.43	+1.4 (-3.0 to 5.9)	0.52	+0.7 (-3.4 to 4.8)	0.75

Abbreviations: CI confidence interval; HRQoL health-related quality of life; NI neurologic impairment.

Table 5b. Mixed model analysis of postoperative increase in health-related quality of life.

	Increase in Total HRQoL Score		Increase in Physical Health Summary Score		Increase in Psychosocial Health Summary Score	
	Coefficient (CI)	*p*-value	Coefficient (CI)	*p*-value	Coefficient (CI)	*p*-value
Age (years)	+0.4 (-0.6 – 1.4)	0.46	-0.2 (-1.7 – 1.2)	0.77	-0.05 (-1.2 – 1.1)	0.93
Weight-for-length z-score	-2.5 (-4.2 – -0.7)	0.01	-2.1 (-4.7 – 0.6)	0.12	-3.8 (-5.9 – -1.7)	0.001
NI (yes/no)	+2.1 (-7.5 – 11.7)	0.66	+4.7 (-9.5 – 18.8)	0.51	+2.3 (-9.1 – 13.6)	0.69
Cardiac (yes/no)	+2.6 (-17.2 – 22.4)	0.79	+6.0 (-23.2 – 35.2)	0.68	+3.9 (-19.5 – 27.2)	0.74

Abbreviations: CI confidence interval; HRQoL health-related quality of life; NI neurologic impairment.

Table 6. Complications and reinterventions (n=50)

Complications	**n (%)**
Perioperative events	
Rupture of fascial sutures managed by conversion to minilaparotomy	1 (2%)
Postoperative complications Total number of events*	75
Hypergranulation	39 (78%)
Infection at gastrostomy site	15 (30%)
Leakage at gastrostomy site	10 (20%)
Dislodgement of the catheter	11 (22%)
Reinterventions in operating theatre	
Excision of hypergranulation tissue	2 (4%)
Nasoduodenal catheter (temporary)	4 (8%)
Indication: feeding intolerance	1
Indication: persistent leakage	3
Gastrojejunostomy placement	2 (4%)
Indication: feeding intolerance	1
Indication: persistent leakage	1
Repositioning of gastrostomy balloon migrated into the subcutis (at radiology department)	1 (2%)

* Some patients had multiple minor complications

Nutritional status

Weight-for-height z-scores did not significantly change during the follow-up period of 4.6 months: from the 34th percentile (0.6 – 84.0) to the 38th percentile (1.4 – 71.3) with a mean difference of 4.6 (-5.1 to 14.4). Similarly, height-for-age scores did not significantly change: from the 6th percentile (0.2 – 18.5) before operation to the 18th percentile (0.3 – 26.3) after operation with a mean difference of 2.2 (-2.8 to 7.3). Changes in social HRQoL were not correlated to changes in weight-for-height values (p = 0.27) or height-for-age values (p = 0.31).

DISCUSSION

To our knowledge, this is the first study to prospectively investigate HRQoL after GP using validated questionnaires. Consequently, comparison to other published studies is limited.

We found that children undergoing GP significantly improved in the subdomain of psychosocial HRQoL. This was mainly based on an improvement in social HRQoL. Social HRQoL comprises the ability to function as other children of the same age. Presumably, GP helps children participate in normal daily life. This is an important finding for patients, caregivers and treating physicians when children are referred for GP.

Overall HRQoL however, remained unchanged 4.6 months after GP. It is important to consider the possibility that statistically significant or clinically relevant improvement of HRQoL may require a longer follow-up time after surgery. Similarly, another study investigating parental psychological distress after GP found that the beneficial effects of surgery were not seen after three months but first detected after six months.[15]

The fact that physical HRQoL remained unchanged after GP may be explained by the fact that physical HRQoL is heavily affected by the child's primary health condition.[16] The benefits of GP are therefore not sufficient to improve overall HRQoL in these patients. In our prediction model of HRQoL we found that NI was the main predictor of lower overall HRQoL, with the largest effect size for physical HRQoL. Even though NI was predictive of lower HRQoL, NI by itself did not influence the effect of GP on HRQoL. The predictive value of cardiac morbidity on HRQoL did not reach statistical significance ($p = 0.09$).

Physical health summary scores increased with higher age at the time of operation, indicating that children over time gain more physical well being, possibly owing to natural growth or medical assistance.

In our prediction model of postoperative changes in HRQoL, we found that the only predictor of change in HRQoL was preoperative weight-for-length z-score: children with lower preoperative weight-for-length z-score showed the largest improvement in HRQoL. This is in line with our hypothesis that children with the worst feeding difficulties gained the most benefit from a gastrostomy tube. There is a possibility that the large number of patients fed through nasogastric tubes in these patients prior to surgery influenced the parents' perception of their child's health after surgery when the nasogastric tubes were replaced for gastrostomy tubes.

Pediatric self-report is the standard for HRQoL measurement. However, in young children or in children with NI it can be difficult to obtain self-reports from children. Various studies on HRQoL in children indicate that caregivers' proxy-report does not always correspond to what children report themselves.[17] We found that patients consistently reported higher levels of HRQoL in comparison to their caregivers, although this was not statistically significantly different. This may be attributed to a small effect size or to the small number of children who were able to self-report on their HRQoL (27.5%), which was because of the large proportion of NI children in our study population. Similar effects were found in another study in pediatric patients undergoing laparoscopic

antireflux surgery, where patient's self-report of total HRQoL scores was significantly higher than parental proxy-report with small differences between both groups.[18,19]

Evaluation of adverse events showed that laparoscopic GP is a relatively safe procedure with no procedure related mortality and one case of a perioperative complication that could be solved during the operation. However, minor complications occurred frequently. These results are in line with previous studies on laparoscopic GP in children.[4,20] In the current study, complications were not predictive of HRQoL. The high incidence of these 'minor' complications (75 complications in 50 children) may have made it difficult to show a correlation between the occurrence of these complications and HRQoL.

In analysis of nutritional status, the increase in weight-for-height percentiles did not reach statistical significance. This could have been caused by our follow-up time of 3 months, which may have been too short to demonstrate significant weight gain. A retrospective survey of 300 children undergoing GP in our institute with a follow-up time of 2.63 years demonstrated a significant increase in weight-for-height percentile ($p < 0.0005$).[4]

Because of the heterogeneity of our included patients, the results of this study could theoretically be applied to all children undergoing GP. However, the relatively small sample size limits the power of the study. Therefore, the inclusion of a larger patient group would have been beneficial and would have provided the possibility to separately analyze the different morbidity groups.

CONCLUSIONS

In conclusion, after GP in children, psychosocial HRQoL improved significantly. This was mainly because of an improvement in social HRQoL. Presumably, GP helps children participate in normal daily life. Although children with NI had lower HRQoL, NI by itself did not predict improvement or deterioration in HRQoL after GP. Children with low preoperative BMI gained the most benefit from GP in terms of HRQoL.

The current study adds insight into the population of pediatric patients undergoing GP and the influence of the operation on the quality of life of these patients. This knowledge can help treating physicians provide better counseling to caregivers before and after GP.

REFERENCES

1. Ferluga ED, Sathe NA, Krishnaswami S, McPheeters ML. Surgical intervention for feeding and nutrition difficulties in cerebral palsy: a systematic review. *Dev Med Child Neurol*. Jan 2014;56(1):31-43.
2. Craig GM, Carr LJ, Cass H, et al. Medical, surgical, and health outcomes of gastrostomy feeding. *Dev Med Child Neurol*. May 2006;48(5):353-60.
3. Srinivasan R, Irvine T, Dalzell M. Indications for percutaneous endoscopic gastrostomy and procedure-related outcome. *J Pediatr Gastroenterol Nutr.* 2009;49(5):584–588.
4. Franken J, Mauritz FA, Suksamanapun N, Hulsker CC, van der Zee DC, van Herwaarden-Lindeboom MY. Efficacy and adverse events of laparoscopic gastrostomy placement in children: results of a large cohort study. *Surg Endosc*. Jun 2015;29(6):1545-52.
5. Sullivan PB, Lambert B, Rose M, Ford-Adams M, Johnson A, Griffiths P. Prevalence and severity of feeding and nutritional problems in children with neurological impairment: Oxford Feeding Study. *Dev Med Child Neurol*. Oct 2000;42(10):674-80.
6. Sullivan PB, Juszczak E, Bachlet AM, et al. Impact of gastrostomy tube feeding on the quality of life of carers of children with cerebral palsy. *Developmental Medicine & Child Neurology*. 2004;46:796–800.
7. Guyatt GH, Feeny DH, Patrick DL. Measuring Health-Related Quality of Life. *Ann Intern Med*. 1993;118(8):622-9.
8. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales in Healthy and Patient Populations. *Med Care*. 2001;39(8):800-812.
9. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales. *Health Qual Life Outcomes*. 2007;51.
10. Varni JW, Seid M, Rode CA. The PedsQL: Measurement model for the pediatric quality of life inventory. *Med Care*. 1999;37(2):126-139.
11. Mahant S, Friedman JN, Connolly B, Goia C, Macarthur C. Tube feeding and quality of life in children with severe neurological impairment. *Archives of Disease in Childhood* Sep 2009;94(9):668-73.
12. Eiser C, Morse R. Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res*. 2001;10(4):347-357.
13. Groeicalculator voor de vijfde landelijke Groeistudie TNO. Internet communication.
14. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.
15. Avitsland TL, Faugli A, Pripp AH, Malt UF, Bjørnland K, Emblem R. Maternal psychological distress and parenting stress after gastrostomy placement in children. *J Pediatr Gastroenterol Nutr.* Nov 2012;55(5):562-6.
16. Varni JW, Limbers CA, Burwinkle TM. Impaired health-related quality of life in children and adolescents with chronic conditions: a comparative analysis of 10 disease clusters and 33 disease categories/severities utilizing the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*. Jul 16 2007;5:43.
17. Rajmil L, Rodríguez López A, López-Aguilà S, Alonso J. Parent–child agreement on health-related quality of life (HRQOL): a longitudinal study. *Health Qual Life Outcomes*. 2013;11
18. Mauritz FA, Stellato RK, van Heurn LWE, et al. Laparoscopic antireflux surgery increases health-related quality of life in children with GERD. *Surg Endosc*. Aug 2017;31(8):3122-3129.
19. Hanneman MJ, Sprangers MA, De Mik EL, et al. Quality of life in patients with anorectal malformation or Hirschsprung's disease: development of a disease-specific questionnaire. *Dis Colon Rectum*. Nov 2001;44(11):1650-60.
20. Crosby J, Duerksen D. A retrospective survey of tube-related complications in patients receiving long-term home enteral nutrition. *Dig Dis Sci*. Sep 2005;50(9):1712-7.

CHAPTER 8

The Long-Term Effect of a Laparoscopic Gastrostomy: a Prospective, Observational Study in Children

Rebecca K. Stellato[1], MS
Josephine Franken, MD, PhD[2]
Stefaan H.A. Tytgat[3], MD, PhD
Femke A. van de Peppel – Mauritz[3,4], MD, PhD[¶]
Maud Y.A. Lindeboom[3], MD, PhD[¶]

[1] Department of Data Science and Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2] Department of Surgery, Spaarne Gasthuis, Haarlem, The Netherlands

[3] Department of Pediatric Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, The Netherlands

[4] Department of Family Medicine, Gezondheidcentrum Oog in Al, Utrecht

[¶]shared last author

ABSTRACT

Background
Laparoscopic-assisted gastrostomy placement (GP) is a frequently performed procedure to benefit pediatric patients with severe feeding difficulties. However, long-term health-related quality of life (HRQoL) and gastroesophageal reflux (GER) symptoms data in this population are lacking. This information is crucial in the decision-making process for both the children's physicians and their caregivers.

Methods
A longitudinal, observational study in children was performed. Between May 2012 and April 2014, 50 patients in the Wilhelmina Children's Hospital were included. Caregivers and children were asked about HRQoL and GER symptoms before and 4-6 months after GP; this was repeated six years later. The longitudinal patterns of HRQoL and GER symptoms were examined.

Results
In 20 children long-term questionnaires were successfully completed for evaluation. The majority (81%) still had a well-functioning gastrostomy in place. Psychosocial HRQoL was significantly increased at short-term. Long-term psychosocial HRQOL was lower than at short-term, but not significantly. Compared to preoperative levels, long-term HRQoL was higher, but not significantly. Children with neurologic impairment had significantly lower HRQoL than neurologically normal children at all time points, and followed a similar trend in HRQoL over time. At long-term follow-up, 20% of respondents reported moderate to severe GER compared to 39% at short term follow-up.

Conclusions
Psychosocial HRQoL did not decrease significantly at long-term follow-up. However, the initial increase in psychosocial HRQoL shortly after GP was no longer present at long-term follow-up. GER symptoms did not change significantly in the long term.

INTRODUCTION

Gastrostomy placement (GP) is a frequently performed procedure that provides long-term enteral tube feeding in children with swallowing or other feeding difficulties.[1,2] The majority of these patients have significant neurologic impairment or congenital heart disease. Other indications for GP include inadequate caloric intake in children with chronic medical diseases e.g. cystic fibrosis, and chronic lung, renal or metabolic disease.[3,4]

Most children referred for GP have few alternatives for achieving sufficient nutritional intake. Nonetheless, it is important to understand the impact of the operation on the quality of life in these children. Most studies in children after GP have focused on the quality of life of the caregivers.[5-7] One previous study examined health-related quality of life (HRQoL) in neurologically impaired children after GP, and found a small, non-significant increase in HRQoL at 6 and 12 months after GP.[8] However, that study used self-created questionnaires to examine HRQoL.

In a previous study, using a validated and age-appropriate questionnaire, Franken et al. found a significant increase in psychosocial HRQoL in a group of 50 pediatric patients at 4.5 months after GP. The total and physical HRQoL, however, did not increase. Some studies have found an increase in gastroesophageal reflux (GER) symptoms after GP. Franken et al, however, observed an equal number of children before (44%) and after (39%) GP with GER.[9]

Little is known about long-term effects of GP on the HRQoL of children and gastrointestinal symptoms in this population. The primary objective of the current study was to examine the long-term consequences of GP on HRQoL and GER symptoms. A secondary objective was to identify predictors of HRQoL after GP.

METHODS

Study design

A prospective, longitudinal study was previously conducted at the Wilhelmina Children's Hospital, University Medical Center Utrecht (UMCU) between May 2012 and April 2014 and included 50 children undergoing GP. Questionnaires regarding HRQoL and GER symptoms were completed before GP and 4-5 months postoperatively. Surgical and HRQoL outcomes 4.5 months after the procedure have been published previously.[9-11] All children who participated in the original study, and were still alive at the time of follow-up, were eligible for the current study. Caregivers and (where possible) the children were again asked to fill in questionnaires on HRQoL and GER symptoms between five and eight years after GP. Data collection for the current study occurred between March

2019 and October 2022 (with a suspension in data collection of approximately two years due to the Covid-19 pandemic).

Surgical procedure

In all patients, GP was performed laparoscopically under general anesthesia. All procedures were performed or supervised by an experienced pediatric surgeon. A full description of the procedure has been published previously.[9]

Study measures

HRQoL and GER symptoms were assessed using validated questionnaires. Both questionnaires were used in earlier measurements (preoperatively and four months postoperatively) in this group of patients.[10,11]

The primary outcome of the current study was HRQoL, as measured by the Pediatric Quality of Life (PedsQL™)[12] 4.0 Generic Core Scales,[13] completed 5-8 years after surgery. Values were compared to preoperative and short-term postoperative levels of HRQoL. The PedsQL™ is a reliable HRQoL assessment tool that has been frequently used in pediatric HRQoL publications and is validated for specific age groups and languages, including the Dutch language (8-10). The PedsQL™ consists of three summary scales: the psychosocial health summary score (which comprises 15 items on emotional, social and school functioning scales); the physical health summary score (8 items); and the total scale score.[12] Parents/caregivers filled in a parental questionnaire, and children over the age of 5 without neurological impairment (NI) were also asked to complete a questionnaire. When both the child and parent/caregiver filled in the questionnaire, the average of the two scores was used.

The secondary study parameter was the proportion of children with self-reported reflux. Reflux symptoms were assessed using the validated, age-adjusted Gastroesophageal Reflux Symptom Questionnaire (GSQ).[14] In the GSQ, parents/caregivers are asked about the frequency of the patient's symptoms in the last seven days and about the severity of their symptoms, ranging from not at all (1) to most severe (7). Heartburn, regurgitation, or vomiting were combined and coded as "none" or "mild" (mild symptoms weekly) vs. "moderate" (mild symptoms daily or severe symptoms weekly) vs. "'severe" (severe symptoms daily).

Patients were also asked at follow-up if the gastrostomy tube was still in situ and if it was still functioning properly. The baseline variables age at operation, neurologic impairment (NI), weight-for-length (expressed as a percentage of expected), preoperative pH, and cardiac disease were collected in the original study[11] and were examined as potential predictors of HRQoL.

Ethical approval and trial registration

The current study is a continuation of the original trial registered under the name of "Gastric Emptying in children with a gastrostomy" in the Dutch trial registry, now registered in the International Clinical Trial Registry Platform (NTR3314, 29-02-2012). Ethical approval was obtained from the UMCU Medical Research Ethics Committee. Prior to sending follow-up questionnaires, informed consent was again obtained from the patients' caregivers, and from the patients themselves (if patients were 12 years or older and without NI).

Statistical Analysis

Continuous variables were expressed as mean and standard deviation or median and interquartile range (IQR), and categorical variables as number and percentage.

To examine changes in HRQoL and GER symptoms over time and to identify predictors of HRQoL, mixed models were used. Mixed models adjust estimates and standard errors for the clustering of repeated measures within patients and allow for the estimation of effects in the presence of missing outcomes.[15] Estimated marginal means were used to compare levels of the outcome at the different time points.

For HRQoL linear mixed models were applied. Separate models were estimated using each of the three PedsQL scale scores (total, psychosocial and physical) as the outcome variable. Fixed effects were time (categorical), NI, age at operation, and weight-for-length. An interaction between NI and time was added to examine whether the patterns of HRQoL differed between NI and neurologically normal (NN) children; if not significant according to a likelihood ratio test, it was removed from the model. A random intercept per patient was used to account for clustering of repeated measurements within individuals, and a continuous first-order autoregressive correlation matrix for the residuals was added to account for unevenly spaced time points.

To examine change in GER symptoms over time, a logistic mixed model was estimated. Presence of moderate or severe GER symptoms (vs. none or mild) was the outcome, and time in three categories was the fixed predictor. A random intercept per patient was added to account for repeated measures within patients.

All analyses were performed in R Statistical Software (version 4.2.2).[16] Linear mixed models were performed using the R package nlme (version 3.1.160),[17] the logistic mixed model in the lme4 package (version 1.1.31),[18] and estimated marginal means using the package emmeans (version 1.8.3).[19] Graphs were made using the ggplot2 package (version 3.4.0).[20]

Sample size calculation

The sample size estimation was performed on the basis of the primary outcome of the initial study, assuming a moderate Cohen's effect size (0.5 standard deviation) for the paired difference in gastric emptying before and after surgery (which was the primary outcome in that study). With a two-sided α of 0.05, 44 children were required to obtain 90% power. To compensate potential loss to follow-up, 50 patients were included in the original study.

RESULTS

Forty-five of the original 50 children (or their parents/caretakers) who participated in the original study gave consent to be contacted again for future research. Five children had died in the years since GP, two due to complications from underlying pathology; for the remaining three the cause of death could not be obtained but was likely also due to underlying pathology. Of the remaining 40 children, long-term questionnaires were returned by 20 (Figure 1). Median long-term follow-up time was 70 months, IQR [66; 76]. At least one PedsQL questionnaire (at 0, 4.5 or 70 months) was obtained from all but one child and at least one GSQ from all children; therefore 49 children are included in the longitudinal HRQoL analyses, and 50 in the GER analyses.

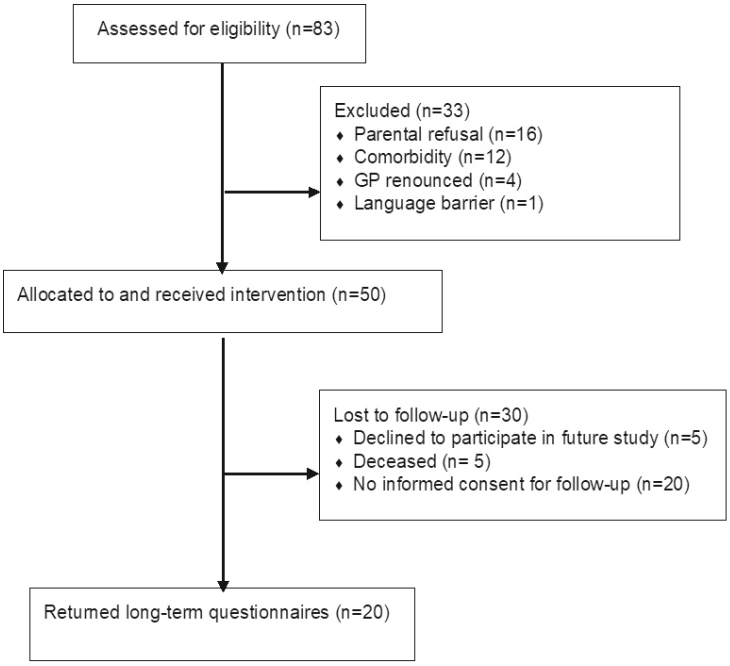


Figure 1. Flow chart of patient inclusion for long-term follow-up

Patient characteristics

The baseline characteristics of the 20 responders and the remaining 30 patients from the original study are described in Table 1. Responders were 1.5 years younger at operation, slightly more likely to have NI and less likely to have cystic fibrosis. Responders also had a lower preoperative pH level on 24-hour pH-monitoring than non-responders.

Among the 16 children for whom data was available, the gastrostomy tube was still in situ and functioning properly in 13 (81.2%).

Table 1. Patient characteristics at baseline (n = 50)

Characteristic	Long-term follow-up (n = 20)	No long-term follow-up (n = 30)
Age at time of operation in years, mean (SD)	3.9 (3.9)	5.5 (4.1)
Male sex, N (%)	11 (55%)	18 (60%)
Neurologically impaired, N (%)	17 (85%)	22 (73%)
Cardiac disease	1 (5%)	1 (3%)
Weight for length, percentage of expected	30.2 (38.2)	31.8 (33.6)
Cystic fibrosis	0 (0%)	4 (13%)
Total preoperative pH, mean (SD)	7.0 (8.4)	10.5 (7.8)

Health-related quality of life over time

Table 2 and Figure 2 display the observed mean total, physical and psychosocial health summary scores for the three time points, separately for NI and NN children. All three scores increased by a few points in the months after GP, followed by a decline to preoperative levels in the NN group; the mean HRQoL in the NI group remained stable or increased slightly. Only the increase of 5 points [95% CI: 0.3; 9.7] in psychosocial health shortly after GP was statistically significant (Table 3). The interaction between NI and time was not significant for any of the three outcomes (total p = 0.1435, physical p = 0.3989, and psychosocial p = 0.2909). This indicates that the patterns of HRQoL did not differ significantly between NI and NN children.

Predictors of health-related quality of life

NI was a statistically significant predictor of HRQoL, with NI children scoring, on average, 17.1 points [95% CI: 6.1; 28.1] lower on the psychosocial scale, 37.6 points [95% CI: 23.2; 52.1] lower on the physical scale, and 24.6 points [95% CI: 14.2; 35.0] lower on the total scale compared to NN children. Higher age at operation was associated with lower physical health scores, with a mean decrease of 1.7 [95% CI 0.1; 3.2] points per year. There was no evidence for an effect of weight-for-length values on HRQoL (Table 3).

Table 2. Mean caregiver proxy and child self-reported PedsQLTM scores and reflux symptoms over time

	Preoperative (N = 48)	± 4.5 months postoperative (N = 41 or 44)[†]	± 70 months postoperative (N = 20)
Total scale score, mean (SD)	52.7 (18.7)	56.8 (20.5)	53.7 (16.6)
Physical health score, mean (SD)	42.8 (24.3)	48.8 (28.6)	46.1 (32.2)
Psychosocial health score, mean (SD)	55.6 (20.6)	61.5 (19.4)	56.9 (14.1)

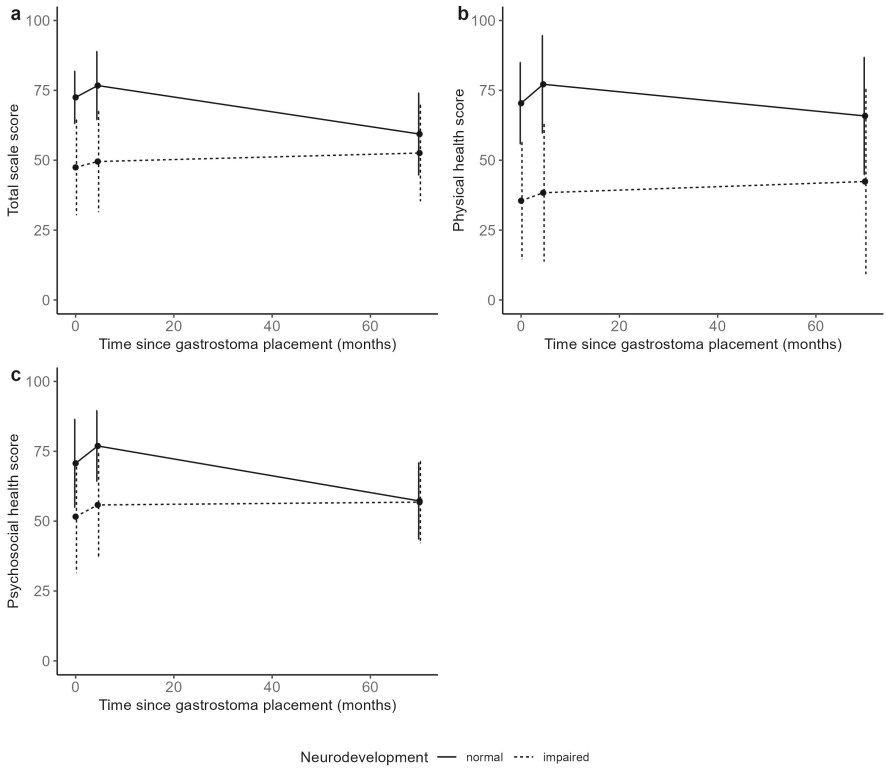


Figure 2. Mean caregiver proxy and child self-reported PedsQLTM scores over time for the (a) total, (b) physical, and (c) psychosocial scale scores

Reflux symptoms

At long-term follow-up, 20% of respondents reported moderate to severe GER symptoms. This was lower than, but not statistically significantly different from, levels before and shortly after GP (44% and 39%; p-values 0.0641 and 0.1468, respectively).

Table 3. Predictors of caregiver proxy and child self-reported PedsQLTM scores from longitudinal analysis: the total, physical, and psychosocial scale scores

	Total Scale Score		Physical Health Scale Score		Psychosocial Health Scale Score	
Variable	Estimate [95% CI†]	p-value	Estimate [95% CI†]	p-value	Estimate [95% CI†]	p-value
4.5 months postoperatively vs. preoperative values	1.9 [-1.8; 5.7]	0.3022	2.3 [-2.3; 7.4]	0.3761	5.0 [0.3; 9.7]	0.0384
6 years postoperatively vs. preoperative values	0.02 [-7.2; 7.3]	0.9951	1.2 [-9.9; 12.3]	0.8294	1.6 [-6.7; 10.0]	0.7014
Age at operation (years)	-0.7 [-1.8; 0.4]	0.1803	-1.7 [-3.2; -0.1]	0.0342	-0.1 [-1.3; 1.0]	0.8242
Neurological impairment	-24.6 [-35.0; -14.2]	<0.0001	-37.6 [-52.1; -23.2]	<0.0001	-17.1 [-28.1; -6.1]	0.0030
Weight-for-length (%)	0.01 [-0.1; 0.1]	0.8637	-0.03 [-0.2; 0.1]	0.7362	0.01 [-0.1; 0.1]	0.8175

†CI = confidence interval

DISCUSSION

To our knowledge, this is the first study to examine long-term HRQoL and GER symptoms prospectively after GP in a pediatric population using validated questionnaires.

In an earlier report on this group of patients, a significant increase in psychosocial HRQoL was seen 4.5 months after GP, primarily driven by improvement in social HRQoL.[10] Five to eight years after GP, physical, psychosocial and total HRQoL were close to baseline levels and confidence intervals for long-term mean change in HRQoL were wide. One other study prospectively examined short-term (6 and 12 month) follow-up after GP and similarly found no statistically significant changes in HRQoL.[8]

Long-term studies for this intervention are not available, so comparison is limited to a few long-term studies that could be found for other pediatric surgical interventions. These studies described different interventions for varying underlying illnesses, used different instruments to examine HRQoL, and follow-up times differed considerably (2 to 5 years). Nevertheless, most studies had a similar pattern of improving HRQoL up to 1.5 or 2 years after intervention, followed by waning to baseline levels.[21-25] Only one study reported a post-intervention increase followed by stabilization up to 48-months post-adenotonsillectomy.[26] The effect of pediatric surgical interventions on HRQoL may only be noticeable for about two years after the intervention, at which point other factors (e.g. adverse events due to the underlying illnesses or from the GP placement) may become more important to perceived quality of life. Furthermore, parents who are uncertain about GP ahead of time may be relieved by the success of the procedure in the months following surgery, causing a temporary improvement in proxy-reported HRQoL that wanes over time.

Children with NI scored consistently lower on all three HRQoL scale scores. Although trends in HRQoL over time appeared to differ slightly for children with and without NI, those differences were not statistically significant. This is likely due to the large variation in HRQoL scores and small numbers of children in both groups, especially at the long-term follow-up.

Higher age at operation was associated with lower physical HRQoL. One previous study also used the PedsQL to examine HRQoL approximately four years after GP. That study likewise found lower HRQoL among children who were older at the time of operation.[27] This is consistent with earlier studies that reported decreasing HRQoL in healthy children after around age 6 through the mid-teens,[28,29] suggesting that entry into adolescence may be associated with lower perceived quality of life.

Long-term GER symptoms were lower than before GP, although not significantly so. The lower levels of self-reported GER may be due to chance, to the maturing of the child,

or to the increased use in blended diet food in these children during the last years. Parents and children may also grow used to certain symptoms over time and report lower severity for the same level of symptoms.

Several limitations impair the generalizability of the current findings. The sample size was relatively small, especially at long-term follow-up. The original design only included short-term follow-up, and patients or their caregivers had to provide informed consent for participation in the long-term study. A design in which more regular follow-up questionnaires were completed might have kept participants more involved in the study and improved long-term response. There were also incomplete questionnaires, especially among caregivers of NI children (questions regarding functioning at school were especially difficult, as were questions about GER symptoms). The use of mixed models should help correct for bias due to selective dropout, and the models make use of all available data; however, the low number of respondents resulted in less certainty, especially concerning the long-term results. The small sample also limited the ability to examine subgroups.

Five children died before the final measurement. Ideally, a joint model[30] would have been used to correct HRQoL levels for missing due to death; however, dates of death were not available for three of the five deceased children.

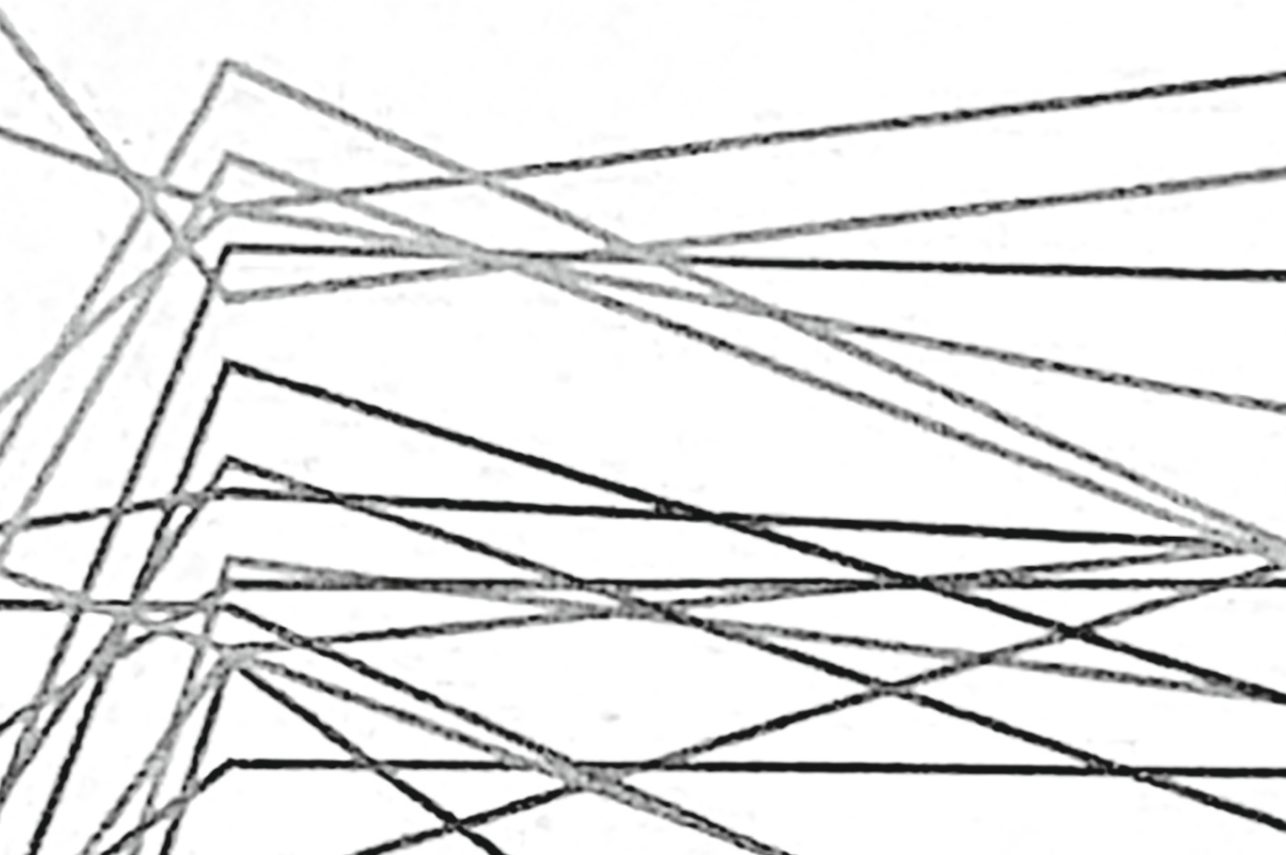
The results are based on self- or proxy-report of both HRQoL and GER symptoms. Proxy report may not be a reliable estimate of the HRQoL perceived by the child,[31] but was necessary in the case of very young children and children with NI. GER symptoms are more reliably obtained by physical examination than by questionnaire. However, since all three assessments were done using the same validated questionnaire, results from the three time points can be directly compared.

In conclusion, while the modest increase in psychosocial HRQoL short-term was not sustained 6 years after GP placement, long-term HRQoL among GP patients remained largely unchanged from preoperative and short-term postoperative levels. GER symptoms did not change significantly in the long term.

REFERENCES

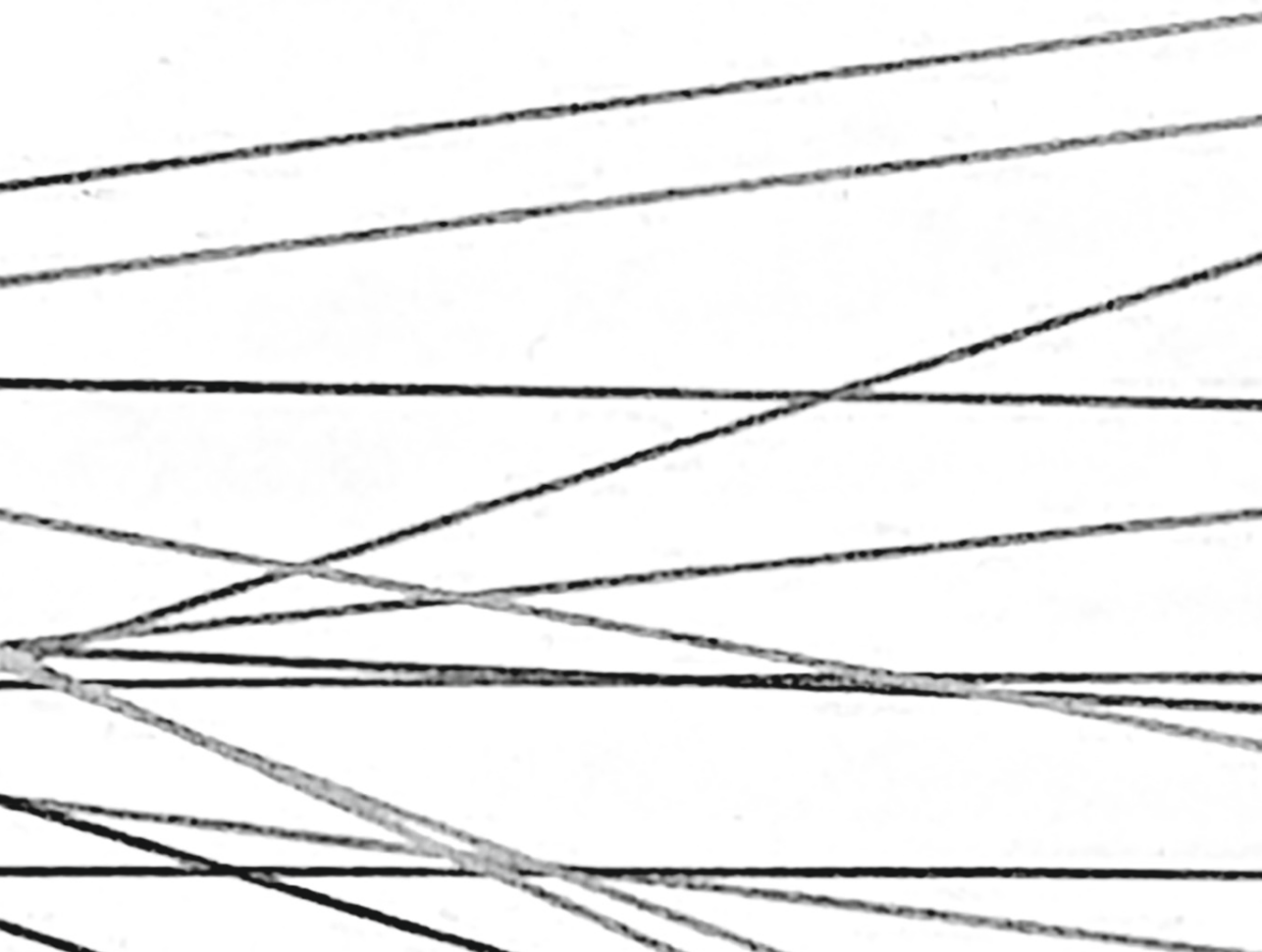
1. Ferluga ED, Sathe NA, Krishnaswami S, McPheeters ML. Surgical intervention for feeding and nutrition difficulties in cerebral palsy: a systematic review. *Dev Med Child Neurol*. Jan 2014;56(1):31-43.
2. Craig GM, Carr LJ, Cass H, et al. Medical, surgical, and health outcomes of gastrostomy feeding. *Dev Med Child Neurol*. May 2006;48(5):353-60.
3. Khattak IU, Kimber C, Kiely EM, Spitz L. Percutaneous endoscopic gastrostomy in paediatric practice: Complications and outcome. *Journal of Pediatric Surgery.* 1998;33(1):67-72.
4. Srinivasan R, Irvine T, Dalzell M. Indications for percutaneous endoscopic gastrostomy and procedure-related outcome. *J Pediatr Gastroenterol Nutr.* 2009;49(5):584–588.
5. Heyman MB, Harmatz P, Acree M, et al. Economic and psychologic costs for maternal caregivers of gastrostomy-dependent children. *The Journal of Pediatrics*. Oct 2004;145(4):511-6.
6. Pemberton J, Frankfurter C, Bailey K, Jones L, Walton JM. Gastrostomy matters--the impact of pediatric surgery on caregiver quality of life. *Journal of Pediatric Surgery.* May 2013;48(5):963-70.
7. Sullivan PB, Juszczak E, Bachlet AM, et al. Impact of gastrostomy tube feeding on the quality of life of carers of children with cerebral palsy. *Developmental Medicine & Child Neurology.* 2004;46:796–800.
8. Mahant S, Friedman JN, Connolly B, Goia C, Macarthur C. Tube feeding and quality of life in children with severe neurological impairment. *Archives of Disease in Childhood* Sep 2009;94(9):668-73.
9. Franken J, Mauritz FA, Stellato RK, Van der Zee DC, Van Herwaarden-Lindeboom MYA. The Effect of Gastrostomy Placement on Gastric Function in Children: a Prospective Cohort Study. *Journal of Gastrointestinal Surgery.* Jul 2017;21(7):1105-1111.
10. Franken J, Stellato RK, Tytgat S, van der Zee DC, Mauritz FA, Lindeboom MYA. The Effect of Gastrostomy Placement on Health-Related Quality of Life in Children. *J Pediatr Surg*. Nov 2019;54(11):2268-2273.
11. Franken J, Stellato RK, Tytgat SHAJ, Van der Zee DC, Mauritz FA, Lindeboom MYA. Gastro-esophageal Reflux After Laparoscopic Gastrostomy Placement in Children. *J Pediatr Gastroenterol Nutr.* Feb 2020;70(2):e41-e47.
12. Varni JW, Bendo CB, Denham J, et al. PedsQL gastrointestinal symptoms module: feasibility, reliability, and validity. *Journal of Pediatric Gastroenterology and Nutrition*. Sep 2014;59(3):347-55.
13. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(6):705–19.
14. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr.* Aug 2005;41(2):178-85.
15. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley-Interscience; 2006.
16. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2022. https://www.R-project.org/
17. Pinheiro J, Bates D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-160. 2022. https://CRAN.R-project.org/package=nlme
18. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1-48.

19. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.3. 2022. https://CRAN.R-project.org/package=emmeans
20. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag 2016.
21. DiFazio RL, Vessey JA, Miller PE, Snyder BD, Shore BJ. Health-related quality of life and caregiver burden after hip reconstruction and spinal fusion in children with spastic cerebral palsy. *Dev Med Child Neurol*. Jan 2022;64(1):80-87.
22. Bekkering WP, Vliet Vlieland TP, Koopman HM, et al. A prospective study on quality of life and functional outcome in children and adolescents after malignant bone tumor surgery. *Pediatr Blood Cancer*. Jun 2012;58(6):978-85.
23. Dewberry L, Khoury J, Schmiege S, Jenkins T, Boles R, Inge T. Gastrointestinal symptoms in relation to quality of life after metabolic surgery in adolescents. *Surg Obes Relat Dis*. Apr 2020;16(4):554-561.
24. Stellato RK, Colmer N, Tytgat SHA, van der Zee DC, van de Peppel-Mauritz FA, Lindeboom MYA. Five-Year Outcome of Laparoscopic Fundoplication in Pediatric GERD Patients: a Multicenter, Prospective Cohort Study. *J Gastrointest Surg*. Jun 2021;25(6):1412-1418.
25. Widjaja E, Puka K, Speechley KN, et al. Trajectory of Health-Related Quality of Life After Pediatric Epilepsy Surgery. *JAMA Netw Open*. Mar 1 2023;6(3):e234858.
26. Lushington K, Kennedy D, Martin J, Kohler M. Quality-of-life but not behavior improves 48-months post-adenotonsillectomy in children with SDB. *Sleep Med*. May 2021;81:418-429.
27. Franken J, Stellato RK, Tytgat SHAJ, van der Zee DC, Mauritz FA, Lindeboom MYA. Health-related quality of life in children after laparoscopic gastrostomy placement. *Qual Life Res*. Jan 2020;29(1):171-178.
28. Simon AE, Chan KS, Forrest CB. Assessment of children's health-related quality of life in the United States with a multidimensional index. *Pediatrics*. Jan 2008;121(1):e118-26.
29. Wang J, Jin W, Shi L, Geng Y, Zhu X, Hu W. Health-Related Quality of Life in Children: The Roles of Age, Gender and Interpersonal Trust. *Int J Environ Res Public Health*. Nov 21 2022;19(22)
30. Rizopolous D. *Joint Models for Longitudinal and Survival Data With Applications in R*. CRC Press; 2012.
31. Rajmil L, Rodríguez López A, López-Aguilà S, Alonso J. Parent–child agreement on health-related quality of life (HRQOL): a longitudinal study. *Health Qual Life Outcomes*. 2013;11



PART IV

APPENDICES



CHAPTER 9

Summary and General Discussion

SUMMARY

This thesis described the current state of longitudinal data analysis in the pediatric gastrointestinal (GI) surgical literature. Studies on two procedures were performed to gather the best available evidence for minimally invasive pediatric upper GI surgery. The studies were performed using the best possible design and analysis given the constraints.

Chapter one provided an introduction to the current state of methodology and statistics, especially regarding longitudinal data analysis, in the pediatric GI surgical literature. Several challenges in the reporting and analysis of the effects of pediatric surgical studies were summarized, such as the preponderance of case studies, the small number of patients in studies, ethical issues regarding a vulnerable population, poor reporting of methodology, and the suspicion that the statistical analysis of repeated measures is not always appropriate. In addition, the effects of two surgical procedures were described: anti-reflux surgery (LARS) for children with severe gastrointestinal reflux disease that does not respond to medication; and gastrostomy placement (GP) for children with severe feeding problems.

I: Longitudinal data analysis in pediatric gastrointestinal surgical literature

To identify potential methodological problems in the pediatric surgical literature, a literature review was performed on all scientific articles that reported on repeated measures in a pediatric GI surgical study in the years 2010-2019. The results were presented in **chapter two**. A large majority of the studies examined reported on the number of included patients, which was generally low (median sample size 40). Most also clearly stated the objective of the study and interpreted the results in that context. Poor adherence to reporting was found on several items related to data collection and analysis: sample size justification; reliability and validity of methods; statistical methods used; numbers of participants at each wave; and generalizability of results were not, or not clearly, reported in a majority of the articles examined. This information is crucial to the assessment of the quality of the study.

Furthermore, more than half the pediatric GI surgery studies classified as “longitudinal” or “cohort” in PubMed did not make efficient use of that design by collecting either time-to-event data or repeated measures on one or more outcomes. In the majority of the articles that did collect repeated measures, we observed inefficient, inappropriate or incorrect analysis of longitudinal data. The inefficient use of the cohort/longitudinal design identified in many publications on pediatric GI surgery has likely led to reduced power of the studies. In addition, inappropriate or incorrect analysis of repeated measures data at best make inefficient use of the available information, and at worst may have led to biased estimates of treatment effects. While these methodological

issues are important in any medical study, they are especially so in studies on vulnerable (pediatric) populations.

In many of the studies examined by the literature review, *t*-tests and repeated measures ANOVA were often used. To visually demonstrate the bias and relative inefficiency of these methods for analyzing repeated measures in the presence of missing data (especially in the case of selective dropout), **chapter three** reported the results of a simulation study. All methods examined gave fairly comparable results for low levels of missingness when data was missing completely at random (MCAR). Paired *t*-tests and repeated measures ANOVA produced biased results and poor coverage and precision in the presence of selective dropout ("missing at random" or MAR). Power was also lower for *t*-tests and repeated measures ANOVA in the presence of higher levels of missing data. Generalized estimating equations models produced unbiased results for both MCAR and MAR data in this simulation and performed well in terms of power and coverage. As expected, linear mixed effects models and covariance pattern models performed best in terms of bias and coverage, even in the case of higher levels of MAR data. These results were not unexpected, as the simulation was primarily intended to support the recommendation in **chapter two** that appropriate models be used when analyzing repeated measurements.

II: Short- and long-term effects of laparoscopic anti-reflux surgery

Short-term follow-up studies have shown that laparoscopic antireflux surgery (LARS) in children was effective in 57%–100% of children with proton pump inhibitor-resistan gastroesophageal reflux disease (GERD).[1] However, prospective studies on the long-term efficacy of LARS were scarce. This was the first study on health-related quality of life (HRQoL) in children undergoing LARS that used a validated pediatric HRQoL questionnaire,[2-4] and examined both children's self-reported and parental proxy-reported HRQoL. Reflux symptoms were also reported via validated, age-appropriate questionnaires.[5,6]

In **chapter four**, the short-term effects of LARS were reported. Three months postoperatively HRQoL increased significantly, and after LARS HRQoL scores were comparable to the normal HRQoL scores measured in a healthy population.[7] The increase was seen in the total scale score and in the psychosocial and physical scale scores. Reflux symptoms significantly decreased from 64% with severe reflux symptoms before LARS to 4% after LARS. Children with reflux symptoms scored 4 points lower on the total scale score. Before and after LARS, HRQoL was significantly lower in neurologically impaired (NI) than in neurologically normal (NN) children, though there was no evidence that the change in HRQoL was different for NI and NN children.

The higher levels of HRQoL were sustained one and two years after LARS for the total and psychosocial scale scores. The increase in physical scale score was observed up

to one year after LARS; two years postoperatively, the mean physical scale score was slightly lower (though not significantly) than the preoperative level. **Chapter five** also examined potential predictors of HRQoL, such as type of fundoplication (Thal vs. Nissen), age at operation, and delayed gastric emptying (GE) before LARS, though little evidence was found for associations. Preoperative delayed GE was significantly associated with lower physical scale score and marginally associated with lower total HRQoL. The only consistent predictor of HRQoL across the three scale scores was NI, with NI children scoring 21-48 points lower than NN children. Further, no evidence was found for a difference between parent proxy reports of HRQoL and child self-reports.

Five years after LARS, the HRQoL and reflux symptom questionnaires were repeated and the five-year failure rate of LARS was examined. In **chapter six** we found no sustained increase in HRQoL. The total, physical and psychosocial HRQoL scores were decreased compared to short-term scores. And while long-term scores were higher than preoperative levels, they did not differ significantly from either pre- or early postoperative levels. Again, while NI patients reported a lower HRQoL at all times, they followed a similar pattern over time compared with NN patients. Reflux symptom severity remained significantly improved 5 years after LARS compared to preoperative levels, although more children (26%) reported moderate or severe reflux symptoms compared to 12% at 3 months. Type of fundoplication and neurologic impairment were not significantly associated with reflux severity. The rate of failure (redo fundoplication and/or recurring or persistent moderate to severe reflux symptoms) within five years was 30%. None of the variables examined (age, GE, NI, type of fundoplication) was found to be a significant predictor of failure in this cohort.

III: Short- and long-term effects of laparoscopic gastrostomy placement

Gastrostomy placement (GP) is a frequently performed procedure that provides long-term enteral tube feeding in children with swallowing or other feeding difficulties.[8,9] The primary reason for GP is to improve the nutritional status of the child; this will presumably improve the HRQoL of the child.[10] There is conflicting evidence on whether GP causes or helps alleviate reflux.[11]

The short-term effects (approximately 4.5 months postoperatively) of GP on HRQoL and reflux symptoms is reported in **chapter seven**. Total and physical HRQoL increased by 2-3 points, however, these increases were not significant. Only the 5-point increase in psychosocial HRQoL was statistically significant. This increase was mainly due to an improvement in social HRQoL, which captures functioning socially as other children of the same age. There was very little change in reflux symptoms before and after GP, as reported elsewhere.[12] Preoperative weight-for-length was negatively associated with a postoperative increase in total HRQoL score; none of the other variables examined (age, NI, cardiac disease) was found to be associated with change in HRQoL.

Five to eight years after GP, physical, psychosocial and total HRQoL were close to baseline levels, as reported in **chapter eight**. Confidence intervals for long-term mean change in HRQoL were wide, and compatible with meaningful changes in HRQoL in either direction. At long-term follow-up, 20% of respondents reported moderate to severe GER symptoms. This was lower than, but not statistically significantly different from, levels before and shortly after GP (44% and 39%, respectively). Again, NI children scored lower on all three scale scores, but there was insufficient evidence for differing trends over time for NI and NN children.

GENERAL DISCUSSION

In both the LARS and GP studies long-term HRQoL was lower than short-term, though usually still slightly higher than baseline. However, neither difference (long-term vs. short-term or long-term vs. baseline) was statistically significant, leaving both studies with inconclusive results. In the case of LARS, the long-term waning of HRQoL was in contrast to previous studies, which reported a significant improvement in the quality of life both 6 months and 4 years after ARS in children[13] or high levels of patient "well-being" several years after surgery.[14,15] However, in those studies, caregivers were asked only one question on the child's "overall quality of life" or "well-being," and in two studies[13,14] selective drop-out may also have biased results. Systematic reviews indicate that most LARS studies have reported only shorter term (1-2 years at most) follow-up.[16,17] No long-term studies of HRQoL for GP were available, but one previous study prospectively examined short-term (6 and 12 month) follow-up after GP and similarly found no statistically significant changes in HRQoL.[18]

The initial good short-term success rate of LARS, with a decline in effect in the long term, is in line with previous studies in children.[19-21] The 5-year failure rate of 30% reported in this study falls within the range of 10–43% reported in previous studies.[13,20,22-24] That no variables were found to be a significant predictor of failure is not surprising, given the small number of patients (n=25). In the GP study, the gastrostomy tube was still in situ and functioning properly in 81% of responders. However, one or two non-responders gave as their reason for not participating in long-term follow-up that the gastrostomy was no longer in situ, so the true proportion of patients with functioning tubes may have been lower.

Despite the careful planning of the studies, both the LARS and GP studies were limited by the problems common to many pediatric surgical studies. Both studies were developed as single-arm trials with a protocol that stipulated the collection of important baseline data on all participants. Sample size estimation was performed in advance, and an attempt was made (ARS) and achieved (GP) to reach the required sample size. Nevertheless, the samples were ultimately small. Neither study provided enough power to detect potentially relevant changes in HRQOL from short- to long-term, even when appropriate statistical methods were used to account for the longitudinal nature of the

data, and for missing data. This is compatible with the results of the simulation study in **chapter three**: using 25 patients per group (NI and NN), CPM & LME had slightly higher power than RMA or *t*-tests, but generally not enough to detect an increase (or decrease) of 5-10 points within a group.

In both the LARS and GP studies, the decrease in long-term HRQoL compared to short-term was not statistically significant; nor was long-term HRQoL significantly increased compared to baseline. It is therefore difficult to know if the initial increase shortly after surgery truly wanes over time, and if so, why. Since either no (GP) or only a small number of studies (ARS) are available for comparison, we additionally examined results of the few long-term studies that could be found for other pediatric surgical interventions. These studies described different interventions for varying underlying illnesses, used different instruments to examine HRQoL, and follow-up times differed considerably (2 to 5 years). Nevertheless, most studies had a similar pattern of improving HRQoL up to 1.5 or 2 years after intervention, followed by waning to baseline levels.[25-29] Only one study reported a post-intervention increase followed by stabilization up to 48-months post-adenotonsillectomy.[30] The effect of pediatric surgical interventions on HRQoL may only be noticeable for about two years after the intervention, at which point other factors (e.g. adverse events due either the surgical intervention or to the underlying illnesses) may become more important to perceived quality of life.

It is also conceivable that HRQoL, and in particular generic HRQoL, is not the most sensitive measure for determining the effects of a surgical procedure in the long term. To better evaluate the long-term effects of GP, relatively non-invasive markers of nutritional status such as height, weight, and weight-for-height could be measured several times postoperatively. For LARS, it is not desirable or feasible to repeat invasive testing such as pH monitoring, but the use of a reflux symptoms questionnaire and a disease-specific HRQoL questionnaire could be more sensitive to long-term improvement in symptoms and HRQoL. In future research among pediatric LARS patients, the GI supplement to the PedsQL questionnaire[31] or another reliable, validated questionnaire, could be used for long-term follow-up. A recent LARS study used a different GI-specific HRQoL questionnaire at 3 and 12 months and found an improvement in the first three months after surgery, followed by a further improvement 12 months postoperatively.[32]

The time between some questionnaires was quite long. In the LARS study, no measurements were taken between two and five years after surgery; in the GP study there was nearly a six-year gap after six months. Because of these gaps, it is difficult to estimate the pattern of HRQoL over time, and to know when (if at all) the levels decrease again over time. In order to study the patterns over time, the questionnaires could be repeated more frequently.

Both studies also had to contend with the heterogeneity of the populations (a mix of children with various underlying morbidities), which resulted in large variation in HRQoL scores. This large variation accurately reflects the populations of children undergoing ARS and GP and increases the generalizability of the results, but has a negative effect on statistical power.

While the retention in the LARS study was excellent, the total number of participants was quite low due to slow recruitment. In the GP study, there were only 20 children with long-term follow-up, which is very small. It is questionable whether the results of the long-term responders are generalizable to the population of children needing GP.

In **chapter two** we recommended the use of multi-center studies to increase numbers of patients. The GP study was a single-center study, and might indeed have benefitted from the addition of other centers in order to achieve a larger sample size. Also, long-term follow-up was not planned, and consent had to be obtained again, which proved difficult 5+ years later. While the LARS findings are based on a multi-center study, there were only three centers and the majority of the children were operated in one center. The protocol was also not standardized; according to European guidelines at the time, the surgeons involved used the technique in which they had the most experience and the best outcome.[33] This led to seven children operated with Nissen and 18 with Thal.

Both studies also tested a relatively large number of potential explanatory variables and potential effect modifiers, while the sample sizes were small. No adjustment was made on the p-values or alpha levels to account for the multiple testing. Nevertheless, none of the potential explanatory variables for HRQoL was consistently significant except for NI vs NN, and no effect modifiers over time were identified. This could well be due to the small samples and low power of both studies.

The LARS study was performed in three centers in the Netherlands, though most of the patients came from one of those centers. The GP study was performed entirely in one center. The results of these two studies may not be generalizable to other centers performing the same operations either within or outside the Netherlands. Both studies could – and should – be repeated in larger samples, and in other countries, preferably using more frequent (and potentially more sensitive) questionnaires.

CONCLUSION

When medical studies report on small numbers of patients, it is generally not possible to draw conclusions based on one sample. It is crucial that the reports of those studies provide enough information to be included later in systematic reviews. It is vital that researchers provide enough information in their manuscript so that the evidence can be assessed, and where possible, included in systematic reviews. Guidelines exist

for proper reporting on trials (the CONSORT statement[34]) and observational studies (STROBE[35]), among others.

The systematic review in **chapter two** with its recommendations was performed after both the LARS and GP studies had been designed (and partially reported). However, both the CONSORT and STROBE statements predate the reporting of the studies. Several items on the STROBE checklist[35] were consistently not reported in the chapters four through eight: insufficient attention was paid to addressing potential sources of bias or quantifying their influence (items 9 and 19), and to the generalizability of the results (item 21). Furthermore, item 10 "explain how the study size was arrived at" is only explicitly mentioned in chapter eight; other chapters refer to the original publication in which sample size was justified or explained, but do not give any pertinent information on sample size. Not all chapters explicitly summarize follow-up time (item 14c), though in most chapters the children had fairly equal lengths of follow-up at each moment in time. Source of funding (item 22) was always reported to, but not always by, the journal in which the publications appeared. There was a noticeable improvement over time in the compliance to STROBE for the chapters in this book.

This thesis was born of a longstanding cooperation between a surgeon and a biostatistician, and grew into a cooperation between a surgical department and a biostatistics department. Both bring their own expertise. To quote Pirosca, et al., "we would not tolerate statisticians doing surgery so why do we tolerate the reverse?"[36] **Chapter two** ended with a number of recommendations. Surgical researchers were encouraged to: use prospective cohort studies instead of case series; make use of the prospective design by collecting repeated measures or a time-to-event outcome; to use an appropriate analysis method for the repeated measures; and report on the studies using to the appropriate guideline; and to involve statisticians in their research, from the design through the analysis and reporting phases.

REFERENCES

1. Mauritz FA, Conchillo JM, van Heurn LWE, et al. Effects and efficacy of laparoscopic fundoplication in children with GERD: a prospective, multicenter study. *Surg Endosc*. Mar 2017;31(3):1101-1110.
2. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(6):705–19.
3. Varni JW, Bendo CB, Denham J, et al. PedsQL gastrointestinal symptoms module: feasibility, reliability, and validity. *Journal of Pediatric Gastroenterology and Nutrition*. Sep 2014;59(3):347-55.
4. Varni JW, Bendo CB, Denham J, et al. PedsQL Gastrointestinal Symptoms Scales and Gastrointestinal Worry Scales in pediatric patients with functional and organic gastrointestinal diseases in comparison to healthy controls. *Quality of Life Research*. Feb 2015;24(2):363-78.
5. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr.* Aug 2005;41(2):178-85.
6. Shaw M, Talley N, Beebe T, et al. Initial Validation of a Diagnostic Questionnaire for Gastroesophageal Reflux Disease. *Am J Gastroenterol*. 2001;96(1):52-7.
7. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. Dec 2005;5(6):705-19.
8. Ferluga ED, Sathe NA, Krishnaswami S, McPheeters ML. Surgical intervention for feeding and nutrition difficulties in cerebral palsy: a systematic review. *Dev Med Child Neurol*. Jan 2014;56(1):31-43.
9. Craig GM, Carr LJ, Cass H, et al. Medical, surgical, and health outcomes of gastrostomy feeding. *Dev Med Child Neurol*. May 2006;48(5):353-60.
10. Samson-Fang L, Butler C, O'Donnell M. Effects of gastrostomy feeding in children with cerebral palsy: an AACPDM evidence report*. *Developmental Medicine & Child Neurology*. 2007;45(6):415-426.
11. Noble LJ, Dalzell AM, El-Matary W. The relationship between percutaneous endoscopic gastrostomy and gastro-oesophageal reflux disease in children: a systematic review. *Surg Endosc*. 2012;26(9):2504-12.
12. Franken J, Stellato RK, Tytgat SHAJ, Van der Zee DC, Mauritz FA, Lindeboom MYA. Gastro-esophageal Reflux After Laparoscopic Gastrostomy Placement in Children. *J Pediatr Gastroenterol Nutr.* Feb 2020;70(2):e41-e47.
13. Pacilli M, Eaton S, McHoney M, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. *Arch Dis Child*. Jun 2014;99(6):516-21.
14. Kristensen C, Avitsland T, Emblem R, Refsum S, Bjornland K. Satisfactory long-term results after Nissen fundoplication. *Acta Paediatr.* May 2007;96(5):702-5.
15. Knatten CK, Kvello M, Fyhn TJ, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. *J Pediatr Surg*. Jul 2016;51(7):1115-21.
16. Cullis PS, Siminas S, Losty PD. Efficacy of antireflux surgery in children with or without neurological impairment: a systematic review. *Br J Surg*. May 2020;107(6):636-646.
17. Mauritz FA, Blomberg BA, Stellato RK, van der Zee DC, Siersema PD, van Herwaarden-Lindeboom MY. Complete versus partial fundoplication in children with gastroesophageal reflux disease: results of a systematic review and meta-analysis. *J Gastrointest Surg*. Oct 2013;17(10):1883-92.

18. Mahant S, Friedman JN, Connolly B, Goia C, Macarthur C. Tube feeding and quality of life in children with severe neurological impairment. *Archives of Disease in Childhood* Sep 2009;94(9):668-73.
19. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
20. Mauritz FA, van Herwaarden-Lindeboom MY, Zwaveling S, Houwen RH, Siersema PD, van der Zee DC. Laparoscopic Thal fundoplication in children: a prospective 10- to 15-year follow-up study. *Ann Surg*. Feb 2014;259(2):388-93.
21. Fyhn T, Knatten C, Edwin B, et al. Randomized Controlled Trial of Laparoscopic and Open Nissen Fundoplication in Children. *Ann Surg*. 2015;261(6):1061-7.
22. Capito C, Leclair MD, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. *Surg Endosc*. Apr 2008;22(4):875-80.
23. Lopez-Fernandez S, Hernandez F, Hernandez-Martin S, et al. Failed Nissen fundoplication in children: causes and management. *Eur J Pediatr Surg*. Feb 2014;24(1):79-82.
24. Koivusalo AI, Pakarinen MP. Outcome of Surgery for Pediatric Gastroesophageal Reflux: Clinical and Endoscopic Follow-up after 300 Fundoplications in 279 Consecutive Patients. *Scand J Surg*. Mar 2018;107(1):68-75.
25. DiFazio RL, Vessey JA, Miller PE, Snyder BD, Shore BJ. Health-related quality of life and caregiver burden after hip reconstruction and spinal fusion in children with spastic cerebral palsy. *Dev Med Child Neurol*. Jan 2022;64(1):80-87.
26. Bekkering WP, Vliet Vlieland TP, Koopman HM, et al. A prospective study on quality of life and functional outcome in children and adolescents after malignant bone tumor surgery. *Pediatr Blood Cancer*. Jun 2012;58(6):978-85.
27. Dewberry L, Khoury J, Schmiege S, Jenkins T, Boles R, Inge T. Gastrointestinal symptoms in relation to quality of life after metabolic surgery in adolescents. *Surg Obes Relat Dis*. Apr 2020;16(4):554-561.
28. Stellato RK, Colmer N, Tytgat SHA, van der Zee DC, van de Peppel-Mauritz FA, Lindeboom MYA. Five-Year Outcome of Laparoscopic Fundoplication in Pediatric GERD Patients: a Multicenter, Prospective Cohort Study. *J Gastrointest Surg*. Jun 2021;25(6):1412-1418.
29. Widjaja E, Puka K, Speechley KN, et al. Trajectory of Health-Related Quality of Life After Pediatric Epilepsy Surgery. *JAMA Netw Open*. Mar 1 2023;6(3):e234858.
30. Lushington K, Kennedy D, Martin J, Kohler M. Quality-of-life but not behavior improves 48-months post-adenotonsillectomy in children with SDB. *Sleep Med*. May 2021;81:418-429.
31. Varni JW, Kay MT, Limbers CA, Franciosi JP, Pohl JF. PedsQL gastrointestinal symptoms module item development: qualitative methods. *Journal of Pediatric Gastroenterology and Nutrition*. May 2012;54(5):664-71.
32. Haffreingue A, Dupont-Lucas C, Dolet N, Marret JB, Petit T, Rod J. Assessment of quality of life after laparoscopic GERD surgery in children: a prospective study. *Eur J Pediatr*. May 2023;182(5):2005-2012.
33. Fuchs KH, Babic B, Breithaupt W, et al. EAES recommendations for the management of gastroesophageal reflux disease. *Surg Endosc*. Jun 2014;28(6):1753-73.
34. Altman D, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-94.
35. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. Dec 2014;12(12):1500-24.
36. Pirosca S, Shiely F, Clarke M, Treweek S. Tolerating bad health research: the continuing scandal. *Trials*. Jun 2 2022;23(1):458.

CHAPTER 10

Nederlandse Samenvatting en Discussie

SAMENVATTING

Dit proefschrift beschreef de huidige stand van zaken van longitudinale gegevensanalyse in de chirurgische literatuur over pediatrische gastro-intestinale (GI) ingrepen. Er werden onderzoeken uitgevoerd naar twee procedures om het best beschikbare bewijs te verzamelen voor minimaal invasieve kinderchirurgie van het bovenste deel van de GI. De onderzoeken werden uitgevoerd met de best mogelijke opzet en analyse gezien de beperkingen.

Hoofdstuk één gaf een inleiding tot de huidige stand van methodologie en statistiek, vooral met betrekking tot longitudinale gegevensanalyse, in de pediatrische GI-chirurgische literatuur. Verschillende uitdagingen in de rapportage en analyse van de effecten van pediatrische chirurgische onderzoeken werden samengevat, zoals het overwicht van casestudies, het kleine aantal patiënten in onderzoeken, ethische kwesties met betrekking tot een kwetsbare populatie, slechte rapportage van methodologie en het vermoeden dat de statistische analyse van herhaalde metingen niet altijd geschikt is. Daarnaast werden de effecten van twee chirurgische ingrepen beschreven: anti-reflux chirurgie (ARC) voor kinderen met ernstige gastro-intestinale refluxziekte (GOR) die niet reageert op medicatie; en gastrostomie voor kinderen met ernstige voedingsproblemen.

I: Longitudinale data analyse in de pediatrische gastro-intestinale chirurgische literatuur

Om potentiële methodologische problemen in de pediatrische chirurgische literatuur te identificeren, werd een literatuuronderzoek uitgevoerd op alle wetenschappelijke artikelen die rapporteerden over herhaalde metingen in een pediatrische GI-chirurgische studie in de jaren 2010-2019. De resultaten werden gepresenteerd in **hoofdstuk twee**. Een grote meerderheid van de onderzochte studies rapporteerde over het aantal geïncludeerde patiënten, dat over het algemeen laag was (mediane steekproefgrootte 40). De meeste vermeldden ook duidelijk het doel van de studie en interpreteerden de resultaten in die context. Er werd een slechte naleving van de regels voor rapportage gevonden op verschillende punten met betrekking tot gegevensverzameling en analyse: rechtvaardiging van de steekproefgrootte; betrouwbaarheid en validiteit van methoden; gebruikte statistische methoden; aantallen deelnemers bij elke ronde; en generaliseerbaarheid van resultaten werden niet of niet duidelijk gerapporteerd in een meerderheid van de onderzochte artikelen. Deze informatie is cruciaal voor de beoordeling van de kwaliteit van het onderzoek.

Bovendien maakte meer dan de helft van de onderzoeken naar kinderheelkunde die in PubMed werden geclassificeerd als “longitudinaal” of “cohort” geen efficiënt gebruik van dat ontwerp door ofwel tijd-tot-event gegevens of herhaalde metingen op een of meer uitkomsten te verzamelen. In de meerderheid van de artikelen die wel herhaalde

metingen verzamelden, zagen we inefficiënte, ongeschikte of onjuiste analyse van longitudinale gegevens. Het inefficiënte gebruik van het cohort/longitudinale ontwerp in veel publicaties over kinderheelkunde heeft waarschijnlijk geleid tot een verminderde onderscheidend vermogen van de onderzoeken. Daarnaast maakt inadequate of onjuiste analyse van gegevens met herhaalde metingen in het beste geval inefficiënt gebruik van de beschikbare informatie en in het slechtste geval kan dit hebben geleid tot vertekende schattingen van behandelingseffecten. Hoewel deze methodologische kwesties belangrijk zijn in elk medisch onderzoek, zijn ze dat vooral in onderzoeken naar kwetsbare (pediatrische) populaties.

In veel van de studies die in de literatuurstudie werden onderzocht, werd vaak gebruik gemaakt van t-toetsen en ANOVA met herhaalde metingen. Om de vertekening en relatieve inefficiëntie van deze methoden voor het analyseren van herhaalde metingen in aanwezigheid van ontbrekende gegevens (vooral in het geval van selectieve uitval) visueel aan te tonen, werden in **hoofdstuk drie** de resultaten van een simulatiestudie gerapporteerd. Alle onderzochte methoden gaven redelijk vergelijkbare resultaten voor lage niveaus van ontbrekende gegevens bij volledig willekeurig ontbrekende gegevens (MCAR). Gekoppelde t-tests en herhaalde metingen ANOVA gaven vertekende resultaten en een slechte dekking en precisie bij selectieve uitval ("missing at random" of MAR). Het onderscheidend vermogen was ook kleiner voor t-tests en ANOVA met herhaalde metingen in de aanwezigheid van grotere aantallen ontbrekende gegevens. Modellen met gegeneraliseerde schattingsvergelijkingen leverden ongebiaste resultaten op voor zowel MCAR- als MAR-gegevens in deze simulatie en presteerden goed in termen van onderscheidend vermogen en dekking. Zoals verwacht presteerden lineaire modellen met gemengde effecten en modellen met covariantiepatronen het best in termen van vertekening en dekking, zelfs in het geval van hogere niveaus van MAR-gegevens. Deze resultaten waren niet onverwacht, aangezien de simulatie in de eerste plaats bedoeld was ter ondersteuning van de aanbeveling in **hoofdstuk twee** om geschikte modellen te gebruiken bij het analyseren van herhaalde metingen.

II: Korte- en langetermijneffecten van laparoscopische antirefluxchirurgie

Kortetermijn follow-up onderzoeken hebben aangetoond dat laparoscopische antirefluxchirurgie (LARC) bij kinderen effectief was bij 57%-100% van de kinderen met protonpompremmer-resistente gastro-oesofageale refluxziekte (GOR).[1] Prospectieve onderzoeken naar de werkzaamheid van LARC op de lange termijn waren echter schaars. Dit was het eerste onderzoek naar de gezondheidsgerelateerde kwaliteit van leven (GKvL) bij kinderen die LARC ondergingen, waarbij gebruik werd gemaakt van een gevalideerde pediatrische GKvL-vragenlijst[2-4], en waarbij zowel de zelfgerapporteerde GKvL van kinderen als de proxy-gerapporteerde GKvL van ouders werd onderzocht. Refluxsymptomen werden ook gerapporteerd via gevalideerde, leeftijdsgeschikte vragenlijsten.[5,6]

In **hoofdstuk vier** werden de kortetermijneffecten van LARC gerapporteerd. Drie maanden postoperatief nam de GKvL significant toe en na LARC waren de GKvL -scores vergelijkbaar met de normale GKvL-scores gemeten in een gezonde populatie.[7] De toename werd gezien in de totale schaalscore en in de psychosociale en fysieke schaalscores. Refluxsymptomen namen significant af van 64% met ernstige refluxsymptomen vóór LARC tot 4% na LARC. Kinderen met refluxklachten scoorden 4 punten lager op de totale schaalscore. Voor en na LARC was GKvL significant lager bij kinderen met psychomotore retardatie (PR) dan bij kinderen met normale neurologische ontwikkeling (NN), hoewel er geen bewijs was dat de verandering in GKvL verschillend was voor PR en NN kinderen.

De hogere niveaus van GKvL bleven één en twee jaar na LARC aanhouden voor de totale en psychosociale schaalscores. De toename in de fysieke schaalscore werd waargenomen tot één jaar na LARC; twee jaar postoperatief was de gemiddelde fysieke schaalscore iets lager (maar niet significant) dan het preoperatieve niveau. Hoofdstuk vijf onderzocht ook potentiële voorspellers van GKvL, zoals het type fundoplicatie (Thal vs. Nissen), leeftijd bij operatie en vertraagde maaglediging (GE) vóór LARC, hoewel er weinig bewijs werd gevonden voor associaties. Preoperatieve vertraagde GE was significant geassocieerd met een lagere fysieke schaalscore en marginaal geassocieerd met een lagere totale GKvL. De enige consistente voorspeller van GKvL over de drie schaalscores was kinderen met PR, waarbij kinderen met PR 21-48 punten lager scoorden dan NN kinderen. Verder werd er geen bewijs gevonden voor een verschil tussen proxy-rapportages van ouders over GKvL en zelfrapportages van kinderen.

Vijf jaar na LARC werden de GKvL- en refluxsymptoomvragenlijsten herhaald en werd het mislukkingspercentage onderzocht. In **hoofdstuk zes** vonden we geen blijvende toename in GKvL. De totale, fysieke en psychosociale GKvL-scores waren gedaald ten opzichte van de scores op de korte termijn. En hoewel de scores op lange termijn hoger waren dan preoperatieve niveaus, verschilden ze niet significant van pre- of vroeg-postoperatieve niveaus. Nogmaals, terwijl PR-patiënten op elk moment een lagere GKvL rapporteerden, volgden ze in de loop van de tijd een vergelijkbaar patroon als NN-patiënten. De ernst van de refluxsymptomen bleef 5 jaar na LARC significant verbeterd in vergelijking met preoperatieve niveaus, hoewel meer kinderen (26%) matige of ernstige refluxsymptomen rapporteerden in vergelijking met 12% op 3 maanden. Type fundoplicatie en psychomotorische retardatie waren niet significant geassocieerd met ernst van de reflux. Het percentage mislukkingen (heroperatie van de fundoplicatie en/of terugkerende of aanhoudende matige tot ernstige refluxklachten) binnen vijf jaar was 30%. Geen van de onderzochte variabelen (leeftijd, GE, PR, type fundoplicatie) bleek een significante voorspeller van falen in dit cohort.

III: Korte- en langetermijneffecten van laparoscopische gastrostomie
Gastrostomie is een vaak uitgevoerde procedure die zorgt voor langdurige enterale sondevoeding bij kinderen met slik- of andere voedingsproblemen.[8,9] De primaire reden voor gastrostomie is het verbeteren van de voedingsstatus van het kind; dit zal vermoedelijk de GKvL van het kind verbeteren[10]. Er is tegenstrijdig bewijs over de vraag of gastrostomie reflux veroorzaakt of helpt verlichten[11].

De kortetermijneffecten (ongeveer 4,5 maanden postoperatief) van laparoscopische gastrostomie (LG) op GKvL en refluxsymptomen worden gerapporteerd in **hoofdstuk zeven**. De totale en fysieke GKvL stegen met 2-3 punten, maar deze stijgingen waren niet significant. Alleen de toename van 5 punten in psychosociale GKvL was statistisch significant. Deze toename was voornamelijk te danken aan een verbetering in de sociale GKvL, die het sociaal functioneren van kinderen van dezelfde leeftijd weergeeft. Er was weinig verandering in refluxsymptomen voor en na de LG, zoals elders gerapporteerd[12]. Preoperatief gewicht-voor-lengte was negatief geassocieerd met een postoperatieve toename van de totale GKvL-score; geen van de andere onderzochte variabelen (leeftijd, PR, hartziekte) bleek geassocieerd met verandering in GKvL.

Vijf tot acht jaar na de LG lagen de fysieke, psychosociale en totale GKvL dicht bij de uitgangsniveaus, zoals gerapporteerd in **hoofdstuk acht**. De betrouwbaarheidsintervallen voor de gemiddelde verandering in GKvL op lange termijn waren breed en compatibel met betekenisvolle veranderingen in GKvL in beide richtingen. Bij de follow-up op lange termijn rapporteerde 20% van de respondenten matige tot ernstige GOR-symptomen. Dit was lager dan, maar niet statistisch significant verschillend van, de niveaus voor en kort na de LG (respectievelijk 44% en 39%). Opnieuw scoorden PR-kinderen lager op alle drie de schaalscores, maar er was onvoldoende bewijs voor verschillende trends in de tijd voor PR- en NN-kinderen.

DISCUSSIE

In zowel de LARC- als de gastrostomie-studies was de GKvL op lange termijn lager dan op korte termijn, maar meestal toch nog iets hoger dan op baseline. Geen van beide verschillen (lange termijn versus korte termijn of lange termijn versus baseline) was echter statistisch significant, waardoor de resultaten van beide onderzoeken niet overtuigend waren. In het geval van LARC stond de afname van de GKvL op lange termijn in contrast met eerdere onderzoeken, die een significante verbetering van de kwaliteit van leven rapporteerden, zowel 6 maanden als 4 jaar na LARC bij kinderen,[13] of hoge niveaus van “welzijn” bij patiënten enkele jaren na de operatie.[14,15] In die onderzoeken waren de verzorgers echter niet altijd in staat om de kwaliteit van leven te verbeteren.

In die onderzoeken werd aan verzorgers echter slechts één vraag gesteld over de “algehele kwaliteit van leven” of het “welzijn” van het kind, en in twee onderzoeken13,

14 kan selectieve uitval de resultaten ook vertekend hebben. Systematische reviews geven aan dat de meeste LARC-onderzoeken alleen kortdurende (maximaal 1-2 jaar) follow-up hebben gerapporteerd. [16,17] Er zijn geen langetermijnonderzoeken naar GKvL gerapporteerd. Er waren geen langetermijnstudies van GKvL voor LG beschikbaar, maar één eerdere studie onderzocht prospectief de kortetermijnfollow-up (6 en 12 maanden) na LG en vond eveneens geen statistisch significante veranderingen in GKvL.[18]

Het aanvankelijke goede succes van LARC op korte termijn, met een afnemend effect op de lange termijn, is in overeenstemming met eerdere onderzoeken bij kinderen.[19-21] Het 5-jaars mislukkingspercentage van 30% dat in dit onderzoek werd gerapporteerd, valt binnen het bereik van 10-43% dat in eerdere onderzoeken werd gerapporteerd.[13,20,22-24] Dat er geen variabelen werden gevonden die een significante voorspeller van falen waren, is niet verrassend gezien het kleine aantal patiënten (n=25). In de LG-studie was de catheter bij 81% van de responders nog in situ en functioneerde deze naar behoren. Een of twee non-responders gaven echter als reden voor het niet deelnemen aan langetermijnfollow-up op dat de gastrostoma niet meer in situ was, dus het werkelijke aandeel patiënten met een goed functionerend buisje kan lager zijn geweest.

Ondanks de zorgvuldige planning van de onderzoeken werden zowel het LARC- als het LG-onderzoek beperkt door de problemen die veel kinderchirurgisch onderzoek met zich meebrengt. Beide onderzoeken werden ontwikkeld als enkelvoudige onderzoeken met een protocol dat het verzamelen van belangrijke basislijngegevens van alle deelnemers voorschreef. De steekproefgrootte werd van tevoren geschat en er werd geprobeerd (LARC) en het lukte (LG) om de vereiste steekproefgrootte te bereiken. Desondanks waren de steekproeven uiteindelijk klein. Geen van beide onderzoeken bood voldoende onderscheidend vermogen om potentieel relevante veranderingen in HRQOL van korte naar lange termijn te detecteren, zelfs niet wanneer de juiste statistische methoden werden gebruikt om rekening te houden met de longitudinale aard van de gegevens en met ontbrekende gegevens. Dit komt overeen met de resultaten van de simulatiestudie in **hoofdstuk drie**: met 25 patiënten per groep (PR en NN) hadden CPM & LME een iets hoger onderscheidend vermogen dan RMA of t-tests, maar over het algemeen niet genoeg om een toename (of afname) van 5-10 punten binnen een groep te detecteren.

In zowel het LARC- als het LG-onderzoek was de afname van de GKvL op lange termijn ten opzichte van de korte termijn niet statistisch significant; ook was de GKvL op lange termijn niet significant toegenomen ten opzichte van de uitgangswaarde. Het is daarom moeilijk om te weten of de aanvankelijke toename kort na de operatie echt afneemt na verloop van tijd, en zo ja, waarom. Aangezien er geen (LG) of slechts een klein aantal onderzoeken (LARC) beschikbaar zijn voor vergelijking, hebben we aanvullend de resultaten onderzocht van de weinige langetermijnonderzoeken die konden worden gevonden voor andere pediatrische chirurgische ingrepen. Deze onderzoeken beschreven verschillende interventies voor verschillende onderliggende ziekten,

gebruikten verschillende instrumenten om GKvL te onderzoeken en de follow-up tijden verschilden aanzienlijk (2 tot 5 jaar).

Niettemin vertoonden de meeste onderzoeken een vergelijkbaar patroon van verbetering van de GKvL tot 1,5 of 2 jaar na de interventie, gevolgd door terugval naar het beginniveau.[25-29] Slechts één onderzoek rapporteerde een toename na de interventie gevolgd door stabilisatie tot 48 maanden na deadenotonsillectomie.[30] Het effect van kinderchirurgische ingrepen op GKvL is mogelijk pas ongeveer twee jaar na de ingreep merkbaar, waarna andere factoren (bijv. bijwerkingen door de chirurgische ingreep of door de onderliggende ziekten) belangrijker worden voor de ervaren kwaliteit van leven.

Het is ook denkbaar dat GKvL, en in het bijzonder generieke GKvL, niet de meest gevoelige maat is om de effecten van een chirurgische ingreep op lange termijn te bepalen. Om de langetermijneffecten van LG beter te evalueren, zouden relatief niet-invasieve markers van de voedingstoestand zoals lengte, gewicht en gewicht-voor-lengte verschillende keren postoperatief kunnen worden gemeten. Voor LARC is het niet wenselijk of haalbaar om invasieve testen zoals pH-monitoring te herhalen, maar het gebruik van een vragenlijst over refluxsymptomen en een ziektespecifieke GKvL-vragenlijst zou gevoeliger kunnen zijn voor verbetering van symptomen en GKvL op de lange termijn. In toekomstig onderzoek onder pediatrische ARC-patiënten zou het GI-supplement van de PedsQL-vragenlijst[31] of een andere betrouwbare, gevalideerde vragenlijst gebruikt kunnen worden voor langetermijnfollow-up. Een recent LARC-onderzoek gebruikte een andere GI-specifieke GKvL-vragenlijst op 3 en 12 maanden en vond een verbetering in de eerste drie maanden na de operatie, gevolgd door een verdere verbetering 12 maanden postoperatief.[32]

De tijd tussen sommige vragenlijsten was vrij lang. In de ARC-studie werden geen metingen gedaan tussen twee en vijf jaar na de operatie; in de LG-studie was er bijna een gat van zes jaar na zes maanden. Door deze hiaten is het moeilijk om het patroon van GKvL in de loop van de tijd in te schatten en om te weten wanneer (als dat al gebeurt) de niveaus in de loop van de tijd weer afnemen. Om de patronen in de tijd te bestuderen, zouden de vragenlijsten vaker herhaald kunnen worden.

Beide onderzoeken hadden ook te maken met de heterogeniteit van de populaties (een mix van kinderen met verschillende onderliggende ziekten), wat resulteerde in grote variatie in GKvL-scores. Deze grote variatie weerspiegelt nauwkeurig de populaties van kinderen die ARC en gastrostomie ondergaan en vergroot de generaliseerbaarheid van de resultaten, maar heeft een negatief effect op het statistisch onderscheidend vermogen.

Hoewel de retentie in de LARC-studie uitstekend was, was het totale aantal deelnemers vrij laag door de trage werving. In de LG-studie waren er slechts 20 kinderen met langdurige follow-up, wat erg klein is. Het is twijfelachtig of de resultaten van de responders op de langetermijn generaliseerbaar zijn naar de populatie van kinderen die een gastrostomie nodig hebben.

In **hoofdstuk twee** hebben we het gebruik van multicenterstudies aanbevolen om het aantal patiënten te vergroten. De LG-studie was een ééncenteronderzoek en had inderdaad baat kunnen hebben bij de toevoeging van andere centra om een grotere steekproef te bereiken. Er was ook geen langetermijnfollow-up gepland en er moest opnieuw toestemming worden verkregen, wat 5+ jaar later moeilijk bleek te zijn. Hoewel de LARC-bevindingen gebaseerd zijn op een multicenterstudie, waren er slechts drie centra en werd het merendeel van de kinderen in één centrum geopereerd. Het protocol was ook niet gestandaardiseerd; volgens de Europese richtlijnen van die tijd gebruikten de betrokken chirurgen de techniek waarmee ze de meeste ervaring hadden en het beste resultaat.[33] Dit leidde tot zeven kinderen geopereerd met Nissen en 18 met Thal.

Beide onderzoeken testten ook een relatief groot aantal potentiële verklarende variabelen en potentiële effectmodificatoren, terwijl de steekproefomvang klein was. Er werden geen aanpassingen gedaan aan de p-waarden of significantieniveaus om rekening te houden met meervoudige vergelijkingen. Desondanks was geen van de potentiële verklarende variabelen voor GKvL consistent significant, behalve PR vs NN, en er werden geen effectmodificatoren in de tijd geïdentificeerd. Dit zou te wijten kunnen zijn aan de kleine steekproeven en het lage onderscheidend vermogen van beide onderzoeken.

Het LARC-onderzoek werd uitgevoerd in drie centra in Nederland, hoewel de meeste patiënten uit één van die centra kwamen. Het LG-onderzoek werd volledig in één centrum uitgevoerd. De resultaten van deze twee onderzoeken zijn mogelijk niet generaliseerbaar naar andere centra die dezelfde operaties binnen of buiten Nederland uitvoeren. Beide onderzoeken zouden herhaald kunnen - en moeten - worden in grotere steekproeven en in andere landen, bij voorkeur met frequentere (en mogelijk gevoeligere) vragenlijsten.

CONCLUSIE

Wanneer medische onderzoeken rapporteren over kleine aantallen patiënten, is het meestal niet mogelijk om conclusies te trekken op basis van één steekproef. Het is van cruciaal belang dat de verslagen van deze onderzoeken voldoende informatie bevatten om later te kunnen worden opgenomen in systematische reviews. Het is van vitaal belang dat onderzoekers in hun manuscript voldoende informatie geven, zodat het bewijs kan worden beoordeeld en waar mogelijk kan worden opgenomen

in systematische reviews. Er bestaan richtlijnen voor een goede rapportage van onder andere trials (het CONSORT statement[34]) en observationele studies (STROBE[35]).

De systematische review in **hoofdstuk twee** met zijn aanbevelingen werd uitgevoerd nadat zowel de LARC- als de LG-studies waren ontworpen (en gedeeltelijk gerapporteerd). Zowel de CONSORT als STROBE statements dateren echter van voor de rapportage van de studies. Verschillende items van de STROBE-checklist[35] werden consequent niet gerapporteerd in de: er werd onvoldoende aandacht besteed aan het aanpakken van potentiële bronnen van vertekening of het kwantificeren van hun invloed (item 9 en 19), en aan de generaliseerbaarheid van de resultaten (item 21). Verder wordt item 10 "leg uit hoe de onderzoeksgrootte tot stand is gekomen" alleen expliciet genoemd in **hoofdstuk acht**; andere hoofdstukken verwijzen naar de oorspronkelijke publicatie waarin de steekproefgrootte werd verantwoord of uitgelegd, maar geven geen relevante informatie over de steekproefgrootte. Niet alle hoofdstukken geven een expliciet overzicht van de follow-up tijd (item 14c), hoewel in de meeste hoofdstukken de follow-up tijd van de kinderen op elk moment redelijk gelijk was. Financieringsbron (item 22) werd altijd gerapporteerd aan, maar niet altijd door het tijdschrift waarin de publicaties verschenen. Er was een merkbare verbetering over de tijd in de naleving van STROBE voor de hoofdstukken in dit boek.

Dit proefschrift is ontstaan uit een jarenlange samenwerking tussen een chirurg en een biostatisticus, en groeide uit tot een samenwerking tussen een chirurgische afdeling en een biostatistische afdeling. Beiden brengen hun eigen expertise in. Om Pirosca et al. te citeren: "we zouden niet toestaan dat statistici chirurgie bedrijven, dus waarom zouden we het omgekeerde wel toestaan?"[36] **Hoofdstuk twee** eindigde met een aantal aanbevelingen. Chirurgische onderzoekers werden aangemoedigd om: prospectieve cohortstudies te gebruiken in plaats van case series; gebruik te maken van het prospectieve ontwerp door herhaalde metingen of een tijd-tot-event uitkomst te verzamelen; een geschikte analysemethode te gebruiken voor de herhaalde metingen; en te rapporteren over de studies met behulp van de juiste richtlijn; en statistici te betrekken bij hun onderzoek, vanaf het ontwerp tot en met de analyse- en rapportagefasen.

REFERENTIES

1. Mauritz FA, Conchillo JM, van Heurn LWE, et al. Effects and efficacy of laparoscopic fundoplication in children with GERD: a prospective, multicenter study. *Surg Endosc*. Mar 2017;31(3):1101-1110.
2. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(6):705–19.
3. Varni JW, Bendo CB, Denham J, et al. PedsQL gastrointestinal symptoms module: feasibility, reliability, and validity. *Journal of Pediatric Gastroenterology and Nutrition*. Sep 2014;59(3):347-55.
4. Varni JW, Bendo CB, Denham J, et al. PedsQL Gastrointestinal Symptoms Scales and Gastrointestinal Worry Scales in pediatric patients with functional and organic gastrointestinal diseases in comparison to healthy controls. *Quality of Life Research*. Feb 2015;24(2):363-78.
5. Deal L, Gold BD, Gremse DA, et al. Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr.* Aug 2005;41(2):178-85.
6. Shaw M, Talley N, Beebe T, et al. Initial Validation of a Diagnostic Questionnaire for Gastroesophageal Reflux Disease. *Am J Gastroenterol*. 2001;96(1):52-7.
7. Varni JW, Burwinkle TM, Seid M. The PedsQL as a pediatric patient-reported outcome: reliability and validity of the PedsQL Measurement Model in 25,000 children. *Expert Rev Pharmacoecon Outcomes Res*. Dec 2005;5(6):705-19.
8. Ferluga ED, Sathe NA, Krishnaswami S, McPheeters ML. Surgical intervention for feeding and nutrition difficulties in cerebral palsy: a systematic review. *Dev Med Child Neurol*. Jan 2014;56(1):31-43.
9. Craig GM, Carr LJ, Cass H, et al. Medical, surgical, and health outcomes of gastrostomy feeding. *Dev Med Child Neurol*. May 2006;48(5):353-60.
10. Samson-Fang L, Butler C, O'Donnell M. Effects of gastrostomy feeding in children with cerebral palsy: an AACPDM evidence report*. *Developmental Medicine & Child Neurology*. 2007;45(6):415-426.
11. Noble LJ, Dalzell AM, El-Matary W. The relationship between percutaneous endoscopic gastrostomy and gastro-oesophageal reflux disease in children: a systematic review. *Surg Endosc*. 2012;26(9):2504-12.
12. Franken J, Stellato RK, Tytgat SHAJ, Van der Zee DC, Mauritz FA, Lindeboom MYA. Gastro-esophageal Reflux After Laparoscopic Gastrostomy Placement in Children. *J Pediatr Gastroenterol Nutr.* Feb 2020;70(2):e41-e47.
13. Pacilli M, Eaton S, McHoney M, et al. Four year follow-up of a randomised controlled trial comparing open and laparoscopic Nissen fundoplication in children. *Arch Dis Child*. Jun 2014;99(6):516-21.
14. Kristensen C, Avitsland T, Emblem R, Refsum S, Bjornland K. Satisfactory long-term results after Nissen fundoplication. *Acta Paediatr.* May 2007;96(5):702-5.
15. Knatten CK, Kvello M, Fyhn TJ, et al. Nissen fundoplication in children with and without neurological impairment: A prospective cohort study. *J Pediatr Surg*. Jul 2016;51(7):1115-21.
16. Cullis PS, Siminas S, Losty PD. Efficacy of antireflux surgery in children with or without neurological impairment: a systematic review. *Br J Surg*. May 2020;107(6):636-646.
17. Mauritz FA, Blomberg BA, Stellato RK, van der Zee DC, Siersema PD, van Herwaarden-Lindeboom MY. Complete versus partial fundoplication in children with gastroesophageal reflux disease: results of a systematic review and meta-analysis. *J Gastrointest Surg*. Oct 2013;17(10):1883-92.

18. Mahant S, Friedman JN, Connolly B, Goia C, Macarthur C. Tube feeding and quality of life in children with severe neurological impairment. *Archives of Disease in Childhood* Sep 2009;94(9):668-73.
19. Kubiak R, Andrews J, Grant HW. Long-term outcome of laparoscopic nissen fundoplication compared with laparoscopic thal fundoplication in children: a prospective, randomized study. *Ann Surg*. Jan 2011;253(1):44-9.
20. Mauritz FA, van Herwaarden-Lindeboom MY, Zwaveling S, Houwen RH, Siersema PD, van der Zee DC. Laparoscopic Thal fundoplication in children: a prospective 10- to 15-year follow-up study. *Ann Surg*. Feb 2014;259(2):388-93.
21. Fyhn T, Knatten C, Edwin B, et al. Randomized Controlled Trial of Laparoscopic and Open Nissen Fundoplication in Children. *Ann Surg*. 2015;261(6):1061-7.
22. Capito C, Leclair MD, Piloquet H, Plattner V, Heloury Y, Podevin G. Long-term outcome of laparoscopic Nissen-Rossetti fundoplication for neurologically impaired and normal children. *Surg Endosc*. Apr 2008;22(4):875-80.
23. Lopez-Fernandez S, Hernandez F, Hernandez-Martin S, et al. Failed Nissen fundoplication in children: causes and management. *Eur J Pediatr Surg*. Feb 2014;24(1):79-82.
24. Koivusalo AI, Pakarinen MP. Outcome of Surgery for Pediatric Gastroesophageal Reflux: Clinical and Endoscopic Follow-up after 300 Fundoplications in 279 Consecutive Patients. *Scand J Surg*. Mar 2018;107(1):68-75.
25. DiFazio RL, Vessey JA, Miller PE, Snyder BD, Shore BJ. Health-related quality of life and caregiver burden after hip reconstruction and spinal fusion in children with spastic cerebral palsy. *Dev Med Child Neurol*. Jan 2022;64(1):80-87.
26. Bekkering WP, Vliet Vlieland TP, Koopman HM, et al. A prospective study on quality of life and functional outcome in children and adolescents after malignant bone tumor surgery. *Pediatr Blood Cancer*. Jun 2012;58(6):978-85.
27. Dewberry L, Khoury J, Schmiege S, Jenkins T, Boles R, Inge T. Gastrointestinal symptoms in relation to quality of life after metabolic surgery in adolescents. *Surg Obes Relat Dis*. Apr 2020;16(4):554-561.
28. Stellato RK, Colmer N, Tytgat SHA, van der Zee DC, van de Peppel-Mauritz FA, Lindeboom MYA. Five-Year Outcome of Laparoscopic Fundoplication in Pediatric GERD Patients: a Multicenter, Prospective Cohort Study. *J Gastrointest Surg*. Jun 2021;25(6):1412-1418.
29. Widjaja E, Puka K, Speechley KN, et al. Trajectory of Health-Related Quality of Life After Pediatric Epilepsy Surgery. *JAMA Netw Open*. Mar 1 2023;6(3):e234858.
30. Lushington K, Kennedy D, Martin J, Kohler M. Quality-of-life but not behavior improves 48-months post-adenotonsillectomy in children with SDB. *Sleep Med*. May 2021;81:418-429.
31. Varni JW, Kay MT, Limbers CA, Franciosi JP, Pohl JF. PedsQL gastrointestinal symptoms module item development: qualitative methods. *Journal of Pediatric Gastroenterology and Nutrition*. May 2012;54(5):664-71.
32. Haffreingue A, Dupont-Lucas C, Dolet N, Marret JB, Petit T, Rod J. Assessment of quality of life after laparoscopic GERD surgery in children: a prospective study. *Eur J Pediatr*. May 2023;182(5):2005-2012.
33. Fuchs KH, Babic B, Breithaupt W, et al. EAES recommendations for the management of gastroesophageal reflux disease. *Surg Endosc*. Jun 2014;28(6):1753-73.
34. Altman D, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-94.
35. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. Dec 2014;12(12):1500-24.
36. Pirosca S, Shiely F, Clarke M, Treweek S. Tolerating bad health research: the continuing scandal. *Trials*. Jun 2 2022;23(1):458.

CHAPTER 11

Dankwoord

DANKWOORD

Dr. M.Y.A. Lindeboom, beste Maud, dank voor de mogelijkheid die je me gaf om in een min of meer rijdende trein te stappen. Je zag dat het een promotietraject kon worden, en je vertrouwde mij dat toe. Helaas kwam de vertraging vanwege COVID, en mijn aandacht werd opgeslokt door mijn onderwijstaken. Je hebt geduldig gewacht tot mijn aandacht weer bij het traject kon zijn, en je hebt begeleiding gegeven waar dat nodig was. In deze periode is mijn respect voor jou toegenomen vanwege jouw doorzettingsvermogen en harde werk ook in de tijden waarin het verder niet meezat.

Prof. dr. M.J.C Eijkemans, beste René: je bent een mentor en rolmodel voor mij geweest, niet alleen in de afgelopen vijf jaar maar ook in de tien jaar daarvoor. Hartelijk dank voor alles wat je me geleerd hebt - niet alleen over de statistiek, maar ook over statisticus, docent en wetenschapper zijn, over persoonlijke en wetenschappelijke integriteit. Het is een grote eer dat je mij deze kans hebt gegeven, en het was een enorm plezier om met jou op deze manier samen te werken. Ik mis nog dagelijks je relativeringsvermogen, je warmte, je gulle lach en je optimisme op de werkvloer.

Dr. M. Schipper, beste Maria: waar moet ik beginnen? Bedankt voor de leiding die je geeft aan ons team, en de begeleiding die je mij gaf bij dit proefschrift en dit traject. Je stapte er wat later (en een beetje noodgedwongen) in, maar hebt enorm veel voor mij betekend. Ook bedankt voor alles wat je mij hebt geleerd over onderwijs geven en docent zijn. Maar vooral: bedankt voor je vriendschap, voor de lange gesprekken die meestal over werk beginnen, maar bijna altijd ontsporen in beschouwingen over familie, psychologie, politiek, filosofie – kortom, de belangrijke thema's van het leven.

Dr. S.H.A.J. Tytgat, beste Stefaan, bedankt voor je vriendelijkheid, de bemoedigende woorden en voor de plezierige samenwerking.

Prof. dr. D.C. van der Zee, het is langer geleden dan de bedoeling was, maar u hebt een belangrijke rol gespeeld aan de start van mijn promotietraject, mijn dank daarvoor.

Leden van de leescommissie, **Prof. dr. M.C.J.M. Sturkenboom**, **Prof. dr. M.A. Boermeester**, **Prof. dr. H.C. Boshuizen**, **Prof. dr. M.M. Rovers**, en **Dr. J. Ruiterkamp**, dank voor het plaatsnemen in de commissie en voor het kritisch lezen en beoordelen van dit proefschrift.

Femke Mauritz en **Josephine Franken**, bedankt voor de fijne samenwerking. Jullie hebben twee mooie studies opgezet, dank voor alle hulp bij het voortzetten daarvan.

Femke Mulder, **Nadia Colmer**, en **Auke van de Kuil**, dank voor jullie hulp met data verzameling, invoer en checken. Zonder jullie was ik nu nog mailings aan het sturen.

Mijn collega's bij biostatistiek. **Caroline**, dank voor de aanmoediging om dit te gaan doen, en voor de samenwerking aan één van de artikelen. Ook voor je vriendschap en alle fijne gesprekken in de afgelopen jaren. **Cas**, je ben een voorbeeld voor mij met je eindeloze nieuwsgierigheid en je passie voor lesgeven. Aan jou mag ik altijd alle (ook domme) vragen stellen. Dank voor je fijne manier van meedenken. **Rutger**, je hebt me veel geleerd over programmeren en simuleren, en het is een eer dat je op mijn favoriete artikel staat. Ik kijk uit naar verdere samenwerking. **Daniela**, je hebt naar me geluisterd, me gesteund en aangemoedigd. En me lekker afgeleid met broodbak successen (en frisbees) en schattige foto's van je kleine meid. **Paul**, je hebt me veel geleerd over consultaties voeren, en me later gesteund (en geplaagd) tijdens dit traject. **Peter Z**, fijn dat we altijd kunnen sparren, en samen lachen. Nu mag je me eindelijk vragen hoe het gaat met mijn promotietraject. **Peter v.d. V**, dank voor de vroege ochtend gesprekken en dank dat ik jou zo vaak als vraagbak mag gebruiken. **Hae Won**, van echt samenwerken komt het niet, maar ik geniet van onze uitstapjes! Wanneer kom je lunchen? **Said**, leuk dat we wat vaker samenwerken aan onderwijs de laatste tijd, en fijn dat ik bij jou terecht kan voor moeilijke R en Github vragen. **Marga** en **Jessica**, mijn nieuwste onderwijscollega's, bedankt dat jullie ons team zijn komen versterken en voor alle ondersteuning de laatste tijd. **Maarten**, wat fijn om je weer bij biostatistiek te hebben, en leuk om weer samen te werken op onderwijsvlak. **Jinke,** dank voor je ondersteuning aan ons team en **Cindy**, wat fijn dat jij het nu overneemt! Ik kijk uit naar een fijne samenwerking. En **Bert**, die op de achtergrond alles regelt. Stiekem hou jij de afdeling draaiende, hoewel je dat nooit zelf zou zeggen. Bedankt voor alle gesprekken, lunchwandelingen, wandeltips, steun, en vooral bedankt dat ik altijd mijn hart bij jou mag luchten.

Sofie van Tuyll, dank voor de gezelligheid op mijn WKZ-dagen, en voor al je praktische hulp. Fijn dat we elkaar konden helpen in het proces van promoveren!

Miranda van den Oetelaar, je bent me voor gegaan in het promoveren-naast-je-werk. Je hebt veel geluisterd de laatste jaren en bemoedigende woorden gesproken. En je hebt laten zien dat het kon. **Everlien de Graaf**, ook jij hebt me laten zien hoe je proefschrift (en nog een masterthesis) kunt schrijven naast je meer-dan-fulltime baan. Je hebt ook fijn meegedacht en goede vragen gesteld, en mij over de streep gehaald om met de DRE te werken. En ondertussen heb je me veel geleerd over imputeren van longitudinale data!

Mijn andere **Julius collega's**, vooral van de kankerlijn, dank voor de fijne samenwerking. Jullie vragen sporen mij altijd aan om nieuwe dingen te leren en uit te proberen.

I also want to thank all **my students** over the years, for teaching me so much. Because of your questions and curiosity I've learned more than I ever would have about statistics, teaching, and clearer communication.

Mijn paranimfen. **Petra**, zonder jou was ik gillend gek geworden in de COVID-tijd. Dank voor alle wandelingen, de eindeloze gesprekken die altijd alle kanten opschieten, het luisterend oor bij al mijn moppers over werk en promoveren, en dat je mij telkens aan het lachen maakt. Dat we nog jarenlang het Edese Bos onveilig mogen blijven maken! Met de 5 kilometer (+/- een heel kleine standaarddeviatie) waar we altijd weer op uit komen. **Dorieke**, feestcommissie, geboren coördinator, fellow *would-be cat lady* en Jane Austen fan. Dank voor je collegialiteit, je humor, het altijd in-zijn voor iets gezelligs met collega's. En veel dank voor het geregel rondom de uitnodigingen.

My friends in the US (especially **Kay**), thank you for the decades of friendship, for being our homes away from home, and for your support in this project - and in all other aspects of life - even if (unfortunately) usually from a great distance.

Wilmy, je bent altijd betrokken bij grote en kleine gebeurtenissen in mijn leven. Je bent een maatje voor altijd. Ook superfijn dat je foto's wilt maken op deze dag.

Naast alle mensen die ik hierboven heb genoemd, wil ik alle familie en vrienden bedanken voor hun aandacht, altijd al, maar speciaal in deze tijd. Stoom afblazen tijdens wandelingen, etentjes, spelletjes; jullie belangstelling en betrokkenheid doen mij altijd goed.

Candice and **Mike**, thank you for the loving care you gave to my parents, and for taking me (and us) into your families.

Mijn lieve schoonfamilie, bedankt dat jullie mij hebben opgenomen in jullie warme gezin. **Dini**, het betekent veel voor me dat je hier bij kunt zijn, en hoe mooi is het dat jij alweer mijn kleding hebt gemaakt. **Johan**, wat had jij deze promotie prachtig gevonden, ik zie je in gedachten vrienden maken van alle onbekenden in de zaal. **Bertine**, ik heb vroeger nooit een zus gehad, en nu staat mijn lieve zus naast me op mijn promotie. Fijn dat je er bent!

Patrick and **Mary Nell**, my dear Dad and Mom, everything you taught me, everything you did for me over the years brought me to this moment. This very late moment, unfortunately too late for you. I miss you both, and know you would be proud of me.

En lieve **Marieke**, natuurlijk. Bedankt voor je geduld, je heerlijke humor, je warmte en betrokkenheid. Dank voor je steun in alles, al 25 jaar lang, maar vooral in de laatste (pittige) jaren. Zonder jou was dit project om vele redenen niet gelukt. Je hebt maaltijden gekookt en heel wat geregeld voor ons terwijl ik alweer met mijn neus tegen het scherm zat. Je hebt me af en toe weggetrokken bij de computer (of ik het wilde of niet) en gezorgd voor de nodige afleiding. Je hebt geluisterd naar al mijn verhalen. Je hebt ook teksten kritisch gelezen, vertaald en geholpen met de vormgeving. Wat bof ik toch met jou! Ik kijk uit naar nog een lange, fijne, gezellige, grappige, proefschrift-vrije toekomst samen.

CHAPTER 12

Curriculum Vitae

CURRICULUM VITAE

Rebecca Stellato was born on December 24, 1968 in New Orleans, Louisiana (United States). After graduating from high school at the Bethlehem Central High School in Delmar, New York in 1986 she attended Bryn Mawr College in Bryn Mawr, Pennsylvania. There she earned her Bachelor's degree (cum laude) in mathematics with a minor in biology in 1990. Two years later she combined her interest in the two fields when she enrolled in a program in biostatistics at the Harvard School of Public Health, where she obtained her Master's degree in 1994. After working several years at the New England Research Institutes in Watertown, Massachusetts, she moved to the Netherlands in 1999. There she worked first at the National AIDS Therapy Evaluation Center at the Amsterdam Medical Center and at the National Institute for Public Health and the Environment in Bilthoven. In 2006 she joined the Center for Biostatics at the University of Utrecht and two years later that group moved to the Julius Center for Health Sciences and Primary Care at the University Medical Center Utrecht. There she teaches and coordinates biostatistics courses and collaborates with epidemiologists and clinical researchers. In 2015 the mixed models course she coordinates won the Graduate School of Life Sciences "best Master course of the year" prize and in 2019 she received the "best teacher" award from Elevate Health for the online mixed models and survival analysis courses she coordinates. Since 2018, she has combined her teaching and consultation work with a PhD project under the supervision of M.Y.A. Lindeboom at the department of Pediatric Surgery at the Wilhelmina Children's Hospital and M.J.C. Eijkemans at the department of Data Science and Biostatistics at the Julius Center, resulting in this thesis.