UNIVERSITÀ DEGLI STUDI DI PADOVA

Università degli Studi di Padova

Department of Cardiac, Thoracic, Vascular Sciences and Public Health

Ph.D. COURSE IN: Traslational Specialistic Medicine 'G.B. Morgagni'

CURRICULUM: Biostatistics and Clinical Epidemiology

SERIES: XXXIV

Statistical methods for estimating personalized risk profiles based on precision medicine tools in oncological patients

Metodi statistici per la stima di profili di rischio personalizzati basati sulla medicina di precisione del cancro nei pazienti oncologici

Coordinator: Prof. Annalisa Angelini

Supervisor: Prof. Giulia Barbati

Ph.D. Student: Fabiola Giudici

ABSTRACT

Precision medicine is beginning to emerge as a well-defined discipline with specific goals, areas of focus and tailored methodology. Specifically, the primary goal is to discover treatment rules that leverage heterogeneity to improve clinical decision making in a manner that is reproducible, generalizable and adaptable as needed. This endeavor spans a broad range of scientific areas including drug discovery, genetics/genomics, health communication and causal inference, all in support of evidence-based, i.e., data-driven, decision making. Precision Medicine aims to improve outcomes in medical practice by creating therapeutic strategies adapted to the characteristic individual patients. In other words, it allows patients to be discriminated according to their level of risk (e.g. low or high) and identifies subgroups of patients according to their characteristics in order to assign the treatment to those who are likely to benefit.
Statistics research in precision medicine is broadly focused on methodological development for estimation of and inference for treatment regimens that maximize some cumulative clinical outcome. The process for using statistical inference to establish personalized treatment strategies requires specific techniques for data-analysis that optimize the combination of competing therapies with candidate genetic features and characteristics of the patient and disease.

The present dissertation focuses on the implementation and application of statistical methods for establishing optimal treatment rules for personalized medicine and discuss specific examples in various medical contexts with oncology as an emphasis. I have focused my research activity mainly in the study of the following topics.

1) Statistical methods to analyze continuous biomarkers.
Biomarkers play an increasingly important role in many aspects of personalized medicine and the assessment of safety data. In oncology research, using biomarkers to identify patients who can benefit from an investigational anti-cancer treatment is becoming increasingly important. For continuous biomarkers, to determine patients' subgroups, it is often necessary to determine cut-off points based on their relationship to clinical response of interests (e.g. survival outcomes). Currently, there is no standard method or standard software for biomarker cut-off determination. The problem of the choosing the optimal cut-off is difficult to answer generally. The best method for cut-off determination may depend on the biomarker, the assay and the clinical application under investigation. Several approaches were considered according to the design of study: from classical approach - median or mean value, percentiles, optimal cut-point identified by means standard receiver operating characteristic (ROC) analysis - to more complex analysis - time-dependent ROC, conditional inferential tree and Subpopulation Treatment Effect Pattern (STEPP) method.

2) Statistical methods for time-to-event endpoints.
In oncology, several endpoints are used to compare clinical effectiveness. However, the primary therapeutic goal is to extend survivorship or delay recurrence/progression. Thus, time-to-event endpoints are often considered the most representative of clinical effectiveness. Competing risks occur commonly in medical research. For example, both treatment-related mortality and disease recurrence are important outcomes of interest and well-known competing risks in cancer research. In the analysis of competing risks data, methods of standard survival analysis such as the Kaplan-Meier method for estimation of cumulative incidence, the log-rank test for comparison of cumulative incidence curves and the standard Cox model for the assessment of covariates lead to

incorrect and biased results. In the presence of competing risks, data analysis has to be performed including methods to calculate the cumulative incidence of an event of interest, to compare cumulative incidence curves in the presence of competing risks and to perform competing risks regression analysis.

3) In the field of precision medicine, systematic reviews and meta-analysis are essential tools for synthesizing evidence needed to inform clinical decision-making and policy. This statistical approach i.e. the meta-analysis, the top of the evidence-based medicine pyramid, was performed 1) to improve the management of Early and Advanced breast cancer, 2) to investigate for the first time the conflicting literature data regarding cardiovascular manifestations of mild primary hyperparathyroidism and regarding the effect of thyroid hormone deficiency and excess on arterial stiffness.

4) The fourth topic reviews to use of several statistical methods that handle the issue of treatment switching. In particular, naïve and complex methods were applied to a clinical trial that aimed to estimate long-term of Tamoxifen administration in the setting of breast cancer. Randomized clinical trials on tamoxifen have presented a certain proportion of early tamoxifen stops in patients randomized to experimental arm. However, following the standard methodology, the analyses usually have been performed according to intention to treat (ITT). This approach can have led to an underestimation of treatment effect. The contribution aims at assessing tamoxifen treatment effect taking into account treatment switches, in order to provide a robust assessment of treatment effect applying causal inference methods.

5) The last study deals with the use of population-based registry and administrative databases. Access to health data is the next scalability challenge for personalized medicine. New evidence derived from large populations of data is needed to direct the development of new-targeted drugs, to investigate the disease pathway and patient characteristics, to discover unmet needs with the aim to deliver more personalized packages of care. The objective of this project is to develop an acceptable claims-based algorithm to identify second breast cancer events (local, regional and distant metastases) during a 10-year follow-up through a record-linkage of two data sources- 1) the Friuli Venezia Giulia population based-cancer registry and 2) the administrative individual-record regional database. Such an algorithm has the potential to be implemented in future data repositories to facilitate studies of disease surveillance, monitoring and quality assessment.

In conclusion, the dissertation provides an exposition of several statistical methods for identifying and evaluating appropriately the performance of personalized treatment using data acquired from clinical studies. The choice of different approaches is related to the study design and to the outcomes.

LIST OF PUBLICATIONS

WITHIN THE FIRST THREE AUTHORS' NAMES OR LAST-SECOND/LAST AMONG AUTHORS

1. **The role of NBI with flexible video-endoscope in the follow-up of head and neck cancer patients: a prospective study.** Francesca Boscolo Nata, Nicoletta Gardenal, *Fabiola Giudici*, Giancarlo Tirelli, Eur Arch Otorhinolaryngol. 2021 Jul 25. doi: 10.1007/s00405-021-07016-9. Epub ahead of print. PMID: 34304298.
2. **Clinical, Radiometabolic and Immunologic Effects of Olaparib in Locally Advanced Triple Negative Breast Cancer: The OLTRE Window of Opportunity Trial.** Schettini F, Corona SP, *Giudici F*, Strina C, Sirico M, Bernocchi O, Milani M, Ziglioli N, Aguggini S, Azzini C, Barbieri G, Cervoni V, Cappelletti MR, Molteni A, Lazzari MC, Ferrero G, Ungari M, Marasco E, Bruson A, Xumerle L, Zago E, Cerra D, Loddo M, Williams GH, Paris I, Scambia G and Generali D (2021) Front. Oncol. 11:686776. doi: 10.3389/fonc.2021.686776
3. **The added value of intraoperative near-infrared fluorescence imaging in elective laparoscopic cholecystectomy** Iacuzzo, Cristiana; Bressan, Livia; Troian, Marina; Germani, Paola; *Giudici, Fabiola*; Bortul, Marina, accepted on September 2021 in Surgical Innovation**.**
4. **Hemithyroidectomy versus total thyroidectomy in the intermediate-risk differentiated thyroid cancer: The Italian Societies of Endocrine Surgeons and Surgical Oncology Multicentric Study** Dobrinja C, Samardzic N, Giudici F, Raffaelli M, De Crea et al. Updates Surg. 2021 Aug 25. doi: 10.1007/s13304-021-01140-1. Epub ahead of print. PMID: 34435312**.**
5. **Prognostic value of H-index in patients surgically treated for squamous cell carcinoma of the larynx.** Boscolo-Rizzo P, Zanelli E, *Giudici F*, et al. Laryngoscope Investigative Otolaryngology. 2021;1–9. https://doi.org/10. 1002/lio2.603
6. **Investigation on the Loss of Taste and Smell and Consequent Psychological Effects: A Cross-Sectional Study on Healthcare Workers Who Contracted the COVID-19 Infection** Dudine L, Canaletti C, *Giudici F*, Lunardelli A, Abram G, Santini I, Baroni V, Paris M, Pesavento V, Manganotti P, Ronchese F, Gregoretti B, Negro C.. Front Public Health. 2021 May 28;9:666442. doi: 10.3389/fpubh.2021.666442. PMID: 34123991
7. **Comparison between male and female breast cancer survival using propensity score matching analysis.** Scomersi S, *Giudici F*, Cacciatore G, Losurdo P, Fracon S, Cortinovis S, Ceccherini R, Zanconati F, Tonutti M, Bortul M. Sci Rep. 2021 Jun 2;11(1):11639. doi: 10.1038/s41598-021-91131-4. PMID: 34079019
8. **Poly (ADP-ribose) polymerase inhibitors in solid tumours: Systematic review and meta-analysis.** Schettini F, *Giudici F*, Bernocchi O, Sirico M, Corona SP, Giuliano M, Locci M, Paris I, Scambia G, De Placido S, Rescigno P, Prat A, Curigliano G, Generali D. Eur J Cancer. 2021 Apr 13;149:134-152. doi: 10.1016/j.ejca.2021.02.035. Online ahead of print.PMID: 33862496 Review.
9. **Endocrine-Based Treatments in Clinically-Relevant Subgroups of Hormone Receptor-Positive/HER2-Negative Metastatic Breast Cancer: Systematic Review and Meta-Analysis.** Schettini F, Giuliano M, *Giudici F*, Conte B, De Placido P, Venturini S, Rognoni C, Di Leo A, Locci M, Jerusalem G, Del Mastro L, Puglisi F, Conte P, De Laurentiis M, Pusztai L, Rimawi MF, Schiff R, Arpino G, De Placido S, Prat A, Generali D. Cancers (Basel). 2021 Mar 22;13(6):1458. doi: 10.3390/cancers13061458. PMID: 33810205 Free PMC article. Review.
10. **Cryoablation of renal tumors: long-term follow-up from a multicenter experience**. Stacul F, Sachs C, *Giudici F*, Bertolotto M, Rizzo M, Pavan N, Balestreri L, Lenardon O, Pinzani A, Pola L, Cicero C, Celia A, Cova MA. Abdom Radiol (NY). 2021 Apr 29. doi: 10.1007/s00261-021-03082-z. Online ahead of print. PMID: 33912986
11. **Evaluation of Correlations between Genetic Variants and High-Resolution Computed Tomography Patterns in Idiopathic Pulmonary Fibrosis.** Baratella E, Ruaro B, *Giudici F*, Wade B,

Santagiuliana M, Salton F, Confalonieri P, SimboloM, Scarpa A, Tollot S, Marrocchio C, Cova MA, Confalonieri M. Diagnostics (Basel). 2021 Apr 23;11(5):762. doi: 10.3390/diagnostics11050762. PMID: 33922858

12. **Ileocecal valve syndrome and vitamin B12 deficiency after surgery: a multicentric prospective study.** Germani P, Zucca A, *Giudici F*, Terranova S, Troian M, Samardzic N, Greco M, Janez J, Gasparini C, Cagnazzo E, Vignali A, Giannone Codiglione F, Armellini A, Fumagalli UR, Rosati R, Piccinni G, Megevand J, Tomazic A, Corcione F, Palmisano S, de Manzini N. Updates Surg. 2021 Apr;73(2):569-580. doi: 10.1007/s13304-020-00845-z. Epub 2020 Jul 9. PMID: 32648110 Free PMC article.
13. **Meta-analysis on the effect of mild primary hyperparathyroidism and parathyroidectomy upon arterial stiffness.** Bernardi S, *Giudici F*, Barbato V, Zanatta L, Grillo A, Fabris B. J Clin Endocrinol Metab. 2021 Mar 8: dgab157. doi: 10.1210/clinem/dgab157. Online ahead of print. PMID: 33693666
14. **Direct comparison of three different mathematical models in two independent datasets of EUSOMA certified centers to predict recurrence and survival in estrogen receptor-positive breast cancer: impact on clinical practice.** Iacuzzo C., *Giudici F*., Scomersi S., Zanconati F., Ceccherini R., Bortul M., Generali D. accepted for publication in February 2021 Breast Cancer Research and Treatment
15. **HPV Infection in Middle Ear Squamous Cell Carcinoma: Prevalence, Genotyping and Prognostic Impact.** Cristalli G., Venuti A., *Giudici F*., Paolini F., Ferreli F., Mercante G., Spriano G., Boscolo Nata F., J Clin. Med. Feb 2021, 10, 738.
16. **Residual vital ratio predicts 5-year volume reduction and retreatment after radiofrequency ablation of benign thyroid nodules but not regrowth.** Bernardi S, *Giudici F*, Colombin G, Cavallaro M, Stacul F, Fabris B. Int J Hyperthermia. 2021;38(1):111-113. doi: 10.1080/02656736.2021.1876931.PMID: 33530757
17. **Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America.** Pachetti M, Marini B, *Giudici F,* Benedetti F, Angeletti S, Ciccozzi M, Masciovecchio C, Ippodrino R, Zella D. J Transl Med. 2020 Sep 2;18(1):338. doi: 10.1186/s12967-020-02501-x. PMID: 32878627; PMCID: PMC7463225.
18. **Surgical resection of oral cancer: en-bloc versus discontinuous approach".** Tirelli G, Marcuzzo AV, *Giudici F*, Tofanelli M. Eur Arch Otorhinolaryngol. 2020 Sep;277(9):2657-2658. doi: 10.1007/s00405-020-06135-z. Epub 2020 Jun 24. PMID: 32583179**.**
19. **Five-year results of radiofrequency and laser ablation of benign thyroid nodules: a multicenter study from the Italian minimally-invasive treatments of the thyroid group.** Bernardi S, *Giudici F,* Cesareo R, Antonelli G, Cavallaro M, Deandrea M, Giusti M, Mormile A, Negro R, Palermo A, Papini E, Pasqualini V, Raggiunti B, Rossi D, Sconfienza LM, Solbiati L, Spiezia S, Tina D, Vera L, Stacul F, Mauri G. Thyroid. 2020 Jun 24. doi: 10.1089/thy.2020.0202. Online ahead of print. PMID: 32578498
20. **Impact of BMI on the outcome of metastatic breast cancer patients treated with everolimus: a retrospective exploratory analysis of the BALLET study.** Corona S. P., *Giudici F*., Jerusalem G., Ciruelos E., Strina C., Sirico M., Bernocchi O., Milani M., Dester M., Ziglioli N., Barbieri G., Cervoni V., Montemurro F., et al. Oncotarget. 2020; 11: 2172-2181
21. **Early effects of attendance to the Friuli Venezia Giulia (Northern Italy) mammography screening programme on the incidence of advanced-stage breast cancer: a cohort study** *Giudici F*, Bortul M, Clagnan E, Del Zotto S, Franzo A, Giordano L, Gobbato M, Puliti D, Serraino D, Zucchetto A, Zanier L, Zanconati F, Bucchi L. Epidemiol Prev. 2020 Mar-Jun;44(2-3):145-153. doi: 10.19191/EP20.2-3.P145.037. PMID: 32631014 Italian
22. **ΔKi67 proliferation index as independent predictive and prognostic factor of outcome in luminal breast cancer: data from neoadjuvant letrozole-based treatment.** Ianza A, *Giudici F*, Pinello C, Corona SP, Strina C, Bernocchi O, Bortul M, Milani M, Sirico M, Allevi G, Aguggini S, Cocconi A, Azzini C, Dester M, Cervoni V, Bottini A, Cappelletti M, Generali D..Tumour Biol. 2020 Jun;42(6):1010428320925301. doi: 10.1177/1010428320925301. PMID: 32489146.

23. **Overall survival of CDK4/6-inhibitors-based treatments in clinically relevant subgroups of metastatic breast cancer: systematic review and meta-analysis.** Schettini F, *Giudici F*, Giuliano M, Cristofanilli M, Arpino G, Del Mastro L, Puglisi F, De Placido S, Paris I, De Placido P, Venturini S, De Laurentis M, Conte P, Juric D, Llombart-Cussac A, Pusztai L, Prat A, Jerusalem G, Di Leo A, Generali D. J Natl Cancer Inst. 2020 May 14:djaa071. doi: 10.1093/jnci/djaa071. Epub ahead of print. PMID: 32407488.
24. **Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. Version 2** Pachetti M, Marini B, Benedetti F, *Giudici F,* Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R.. J Transl Med. 2020 Apr 22;18(1):179. doi: 10.1186/s12967-020-02344-6. PMID: 32321524; PMCID: PMC7174922.
25. **Management of the axilla in breast cancer: outcome analysis in a series of ductal versus lobular invasive cancers.** Corona SP, Bortul M, Scomersi S, Bigal C, Bottin C, Zanconati F, Fox SB, *Giudici F*, Generali D. Breast Cancer Res Treat. 2020 Apr; 180(3):735-745. doi: 10.1007/s10549-020-05565-x. Epub 2020 Feb 14. PMID: 32060782.
26. **Comfortable walking speed and energy cost of locomotion in patients with multiple sclerosis.** Buoite Stella A, Morelli ME, *Giudici F*, Sartori A, Manganotti P, di Prampero PE.. Eur J Appl Physiol. 2020 Mar; 120(3):551-566. doi: 10.1007/s00421-019-04295-3. Epub 2020 Jan 9. PMID: 31919582
27. **The current role of touch imprint cytology in sentinel lymph node intra-operatory examination.** Iacuzzo C, Scomersi S, *Giudici F*, Bonadio L, Troian M, Bellio G, Rizzardi C, Bonazza D, Zanconati F, Bortul MBreast J. 2020 Mar;26(3):576-577. doi: 10.1111/tbj.13619. Epub 2019 Sep 4. PMID: 31486216.
28. **Frozen sections and complete resection in oral cancer surgery** Tirelli G., Comhinni M., Bussani R., Gatto A., Bonin P., *Giudici F*., Boscolo Nata F. Oral Dis. 2019 Jul;25(5):1309-1317. doi: 10.1111/odi.13101. Epub 2019 Apr 21.
29. **Elevated levels of eEF1A2 protein expression in triple negative breast cancer relate with poor prognosis.** *Giudici F***,** Petracci E, Nanni O, Bottin C, Pinamonti M, Zanconati F, Scaggiante B.PLoS One. 2019 Jun 20;14(6):e0218030. doi: 10.1371/journal.pone.0218030. eCollection 2019.
30. **A prospective study on the efficacy of patient simulation in heart and lung auscultation.** Bernardi S, *Giudici F*, Leone MF, Zuolo G, Furlotti S, Carretta R, Fabris B. BMC Med Educ. 2019 Jul 23;19(1):275. doi: 10.1186/s12909-019-1708-6. PMID: 31337361 Free PMC Article
31. **Predicting laryngeal exposure in microlaryngoscopy: External validation of the laryngoscore.** Tirelli G, Gatto A, Fortunati A, Marzolino R, *Giudici F*, Boscolo Nata F. Laryngoscope. 2018 Dec 3. doi: 10.1002/lary.27683.
32. **Association between benign thyroid disease and breast cancer: a single center experience.** Dobrinja C, Scomersi S, *Giudici F*, Vallon G, Lanzaro A, Troian M, Bonazza D, Romano A, Zanconati F, de Manzini N, Bortul M. BMC Endocr Disord. 2019 Oct 17;19(1):104. doi: 10.1186/s12902-019-0426-8. PMID: 31623603

UNDER REVIEW

1. **Meta-analysis on the effect of thyroid hormone deficiency and excess upon arterial stiffness** Bernardi S., Grillo A., Antonello RM., Cola M., Fabris B., *Giudici F*. under review- Journal of Clinical Endocrinology & Metabolism
2. **The real-word impact of breast and colorectal cancer surgery during the SARS-CoV-2 pandemic.** Pasquale Losurdo; Natasa Samardzic; Francesca Di Lenarda; Nicolò de Manzini; *Fabiola Giudici*; Marina Bortul under review- Updates in Surgery
3. **Hypovitaminosis D and "metabolic" inflammatory status in patients with obesity**. Bressan L., *Giudici F*., Zucchini N., Giuricin M., Casagranda B., de Manzini N., Palmisano S. under review Obesity Surgery

DRAFTS IN PREPARATION

1. **Breast Cancer Recurrence estimation from population-based cancer registry linked with administrative data**. *Giudici F*., Toffolutti F., Schettini F., Massarut S., Barbati G., Serraino D., Dal Maso L. This project is *in collaboration with* ***Unit of Cancer Epidemiology, Centro di Riferimento Oncologico (CRO) of Aviano, National Cancer Institute*****:** Working Group: Dal Maso Luigino, Massarut Samuele, Serraino Diego, Toffolutti Federica
2. **Long-term effect of adjuvant tamoxifen: adherence-based analysis.** *Giudici F.,* Bardet A, Pistilli B. Vaz-Duarte-Luis IM., Micheils S. This project is *in collaboration with* ***the Department of Biostatistics and Epidemiology, Gustave Roussy, Oncostat U1018, CESP, Inserm, Université Paris-Saclay, Villejuif, France***

PUBLISCHED AS CO-AUTHOR

1. **Interobserver agreement of PD-L1/SP142 immunohistochemis-try and tumor-infiltrating lymphocytes (TILs) in distant metas-tases of triple-negative breast cancer: a proof-of-concept study. A report on behalf of the International Immuno-Oncology Bi-omarker Working Group.** Mieke R. Van Bockstal , Maxine Cooks , Iris Nederlof , Mariël Brinkhuis , Annemiek Dutman , Monique Koopmans , Loes Kooreman , Bert van der Vegt , Leon Verhoog , Celine Vreuls , Pieter Westenend, Marleen Kok , Paul J. van Diest , Inne Nauwelaers, Nele Laudus, Carsten Denkert, David Rimm , Kalliopi P. Sizio-pikou, Scott Ely, Dimitrios Zardavas , Mustimbo Roberts, Giuseppe Floris , Johan Hartman, Balazs Acs, Dieter Peeters, John M.S. Bartlett, Els Dequeker, Roberto Salgado, *Fabiola Giudici*, Stefan Michiels, Hugo Horlings, Carolien H. M. van Deurzen -accepted on 23th September on Cancer
2. **Radiological-pathological signatures of patients with COVID-19-related pneumomediastinum: is there a role for sonichedgehog and Wnt5a pathways?** Baratella E., Bussani R., Zanconati F., Marrocchio C., Giudici F. Braga L., Maiocchi S., Berlot G., Volpe MC., Moro E., Confalonieri P., Cova MA., Confalonieri M., Salton F., Ruaro B.
   ERJ Open Res 2021; in press https://doi.org/10.1183/23120541.00346-2021
3. **Interstitial Lung Disease at High Resolution CT after SARS-CoV-2-Related Acute Respiratory Distress Syndrome According to Pulmonary Segmental Anatomy.** Baratella E., Ruaro B., Marrocchio C., Starvaggi N., Salton F., *Giudici F.,* Quaia E., Confalonieri M., Cova MA. Anatomy. J Clin Med. 2021 Sep 2;10(17):3985. doi: 10.3390/jcm10173985. PMID: 34501430
4. **20-Year Follow-up Study of Celiac Patients Identified in a Mass School Screening: Compliance to Gluten-Free Diet and Autoimmunity.** Cozzi G, Gabbana E, Zanchi C, *Giudici F*, De Leo L, Ziberna F, Bramuzzo M, Di Leo G, Not T. J Pediatr Gastroenterol Nutr. 2021 Sep 9. doi: 10.1097/MPG.0000000000003295. Epub ahead of print. PMID: 34508048.
5. **EGFR mutation analysis on circulating free DNA in NSCLC: a single-center experience**. Ianza A, Di Chicco A, Biagi C, *Giudici F,* Dicorato A, Guglielmi A, Variola F, Tomasi S, Roviello G, Generali D, Zanconati F. J Cancer Res Clin Oncol. 2021 May 18. doi: 10.1007/s00432-021-03658-8. Online ahead of print. PMID: 34003366
6. **State of the art in breast intraoperative electron radiation therapy after intraoperative ultrasound introduction**. Vidali C, Severgnini M, Bellio G, *Giudici F*, Milan V, Pellin Z, Savatovic S, Scomersi S, Fastner G, Ciabattoni A, Bortul M. Radiol Oncol. 2021 May 16. doi: 10.2478/raon-2021-0023. Online ahead of print. PMID: 33991470
7. **Serum carcinoembryonic antigen pre-operative level in colorectal cancer: revisiting risk stratification.** Iacuzzo C, Germani P, Troian M, Cipolat Mis T*, Giudici F*, Osenda E, Bortul M, de Manzini N. ANZ J Surg. 2021 Jun;91(6): E367-E374. doi: 10.1111/ans.16861. Epub 2021 Apr 18. PMID: 33870621.
8. **Contemporary snapshot of tumor regression grade (TRG) distribution in locally advanced rectal cancer: a cross sectional multicentric experience.** Germani P, Di Candido F, Léonard D,

Cuicchi D, Elmore U, Allaix ME, Barbieri VP, D'Allens L, Faes S, Milani M, Caputo D, Martinez C, Grosek J, Caracino V, Christou N, Roodbeen SX, Bracale U, Wildeboer A, Usai A, Benedetti M, Balani A, Piccinni G, Catarci M, Millo P, Bouvy N, Corcione F, Hompes R, Ris F, Basti M, Tomazic A, Targarona E, Coppola A, Pietrabissa A, Hahnloser D, Adamina M, Viola M, Morino M, Rosati R, Poggioli G, Kartheuser A, Spinelli A, de Manzini N; TRG Snapshot Study Group. Updates Surg. 2021 Apr 5. doi: 10.1007/s13304-021-01044-0. Epub ahead of print. PMID: 33818750.

9. **Baseline Characteristics and Outcomes of Cancer Patients Infected with SARS-CoV-2 in the Lombardy Region, Italy (AIOM-L CORONA): A Multicenter, Observational, Ambispective, Cohort Study.** Di Cosimo S, Tagliaferri B, Generali D, *Giudici F*, Agustoni F, Bernardo A, Borgonovo K, Farina G, Luchena G, Luciani A, Nolè F, Palmeri L, Pietrantonio F, Poggi G, Zucali PA, Balletti E, Catania G, Bernocchi O, D'Antonio F, Giordano M, Grossi F, Lasagna A, La Verde N, Manzoni M, Montagna B, Olgiati A, Raimondi A, Rampinelli I, Verri E, Zaniboni A, Di Maio M, Beretta G, Danova M. Cancers (Basel). 2021 Mar 16;13(6):1324. doi: 10.3390/cancers13061324. PMID: 3380946
10. **Percutaneous CT-Guided Renal Cryoablation: Technical Aspects, Safety, and Long-Term Oncological Outcomes in a Single Center.**Cernic S, Marrocchio C, Ciabattoni R, Fiorese I, Stacul , Rizzo M, Cova MA. Medicina (Kaunas). 2021 Mar 20;57(3):291. doi: 10.3390/medicina57030291. PMID: 33804740; PMCID: PMC8003936.
11. **Initial ablation ratio predicts volume reduction and retreatment after 5 years from radiofrequency ablation of benign thyroid nodule**. Bernardi S, Cavallaro M, Colombin G, *Giudici F*., Zuolo G., Zdjelar A. Dobrinja C., De Manzini N., Zanconati F., Cova MA., Stacul F., Fabris B. Front. Endocrinol., Feb 2021
12. **Phenotype and Natural History of Children With Coexistent Inflammatory Bowel Disease and Celiac Disease**. Bramuzzo M, Lionetti P, Miele E, Romano C, Arrigo S, Cardile S, Di Nardo G, Illiceto MT, Pastore M, Felici E, Fuoti M, Banzato C, Citrano M, Congia M, Norsa L, Pozzi E, Zuin G, Agrusti A, Bianconi M, Grieco C, *Giudici F*, Aloi M, Alvisi P.
13. **Canakinumab as treatment for COVID-19-related pneumonia: a prospective case-control study.** Generali D, Bosio G, Malberti F, Cuzzoli A, Testa S, Romanini L, Fioravanti A, Morandini A, Pianta L, Giannotti G, Viola EM, Pier-Franceschi MG, Foramitti M, Tira RA, Zangrandi I, Chiodelli G, Machiavelli A, Cappelletti MR, Giossi A, De Giuli V, Costanzi C, Campana C, Bernocchi O, Sirico M, Zoncada A, Molteni A, Venturini S, *Giudici F*, Scaltriti M, Pan A. Int J Infect Dis. 2020 Dec 29:S1201-9712(20)32597-2. doi: 10.1016/j.ijid.2020.12.073. Epub ahead of print. PMID: 33385581; PMCID: PMC7771302.
14. **Anastomosis Leak: Is There Still a Place for Indocyanine Green Fluorescence Imaging in Colon-Rectal Surgery? A Retrospective, Propensity Score-Matched Cohort Study**. Losurdo P, Mis TC, Cosola D, Bonadio L, *Giudici F,* Casagranda B, Bortul M, de Manzini N.. Surg Innov. 2020 Nov 25:1553350620975258. doi: 10.1177/1553350620975258. Epub ahead of print. PMID: 33236661.
15. **Circulating PV-1 as a marker of celiac disease-associated liver injury** Leo L, Naviglio S, Vatta S, Benelli E, Stera G, Santon D, Ziberna F, Taddio A, Martelossi S, *Giudici F,* Giuffrida P, Sabatino AD, Corazza GR, Ventura A, Not T. Biomark Med. 2020 Dec;14(18):1675-1681. doi: 10.2217/bmm-2020-0281. PMID: 33346700.
16. **Early Changes of the Standardized Uptake Values (SUVmax) Predict the Efficacy of Everolimus-Exemestane in Patients with Hormone Receptor-Positive Metastatic Breast Cancer**. Sirico M, Bernocchi O, Sobhani N, *Giudici F*, Corona SP, Vernieri C, Nichetti F, Cappelletti MR, Milani M, Strina C, Cervoni V, Barbieri G, Ziglioli N, Dester M, Bianchi GV, De Braud F, Generali D. Cancers (Basel). 2020 Nov 10;12(11):3314. doi: 10.3390/cancers12113314. PMID: 33182575; PMCID: PMC7697290.
17. **Immune system and angiogenesis-related potential surrogate biomarkers of response to everolimus-based treatment in hormone receptor-positive breast cancer: an exploratory study** Schettini F, Sobhani N, Ianza A, Triulzi T, Molteni A, Lazzari MC, Strina C, Milani M, Corona SP,

Sirico M, Bernocchi O, *Giudici F*, Cappelletti MR, Ciruelos E, Jerusalem G, Loi S, Fox SB, Generali D.. Breast Cancer Res Treat. 2020 Aug 7. doi: 10.1007/s10549-020-05856-3. Epub ahead of print. PMID: 32770287.

18. **Sex Differences in Proatherogenic Cytokine Levels**. Bernardi S, Toffoli B, Tonon F, Francica M, Campagnolo E, Ferretti T, Comar S, *Giudici F*, Stenner E, Fabris B.. Int J Mol Sci. 2020 May 29;21(11):E3861. doi: 10.3390/ijms21113861. PMID: 32485823.
19. **Indocyanine green-enhanced fluorangiography (ICGf) in laparoscopic extraperitoneal rectal cancer resection**. Bonadio L, Iacuzzo C, Cosola D, Cipolat Mis T, *Giudici F*, Casagranda B, Biloslavo A, de Manzini N. Updates Surg. 2020 Feb 18. doi: 10.1007/s13304-020-00725-6. Epub ahead of print. PMID: 32072407.
20. **Diagnostic accuracy and applicability of intestinal auto-antibodies in the wide clinical spectrum of coeliac disease.** De Leo L, Bramuzzo M, Ziberna F, Villanacci V, Martelossi S, Leo GD, Zanchi C, *Giudici F*, Pandullo M, Riznik P, Mascio AD, Ventura A, Not T.. EBioMedicine. 2020 Jan;51:102567. doi: 10.1016/j.ebiom.2019.11.028. Epub 2020 Jan 2. PMID: 31901853; PMCID: PMC6940709
21. **Colorectal cancer-screening program improves both short- and long- term outcomes: a single-center experience in Trieste.** Losurdo P, Giacca M, Biloslavo A, Fracon S, Sereni E, *Giudici F,* Generali D, De Manzini N.. Updates Surg. 2020 Mar;72(1):89-96. doi: 10.1007/s13304-020-00703-y. Epub 2020 Jan 8. PMID: 31965546.
22. **Tumour infiltrating lymphocytes and PD-L1 expression as potential predictors of outcome in patients with malignant pleural mesothelioma.** Sobhani N, Roviello G, Pivetta T, Ianza A, Bonazza D, Zanconati F, *Giudici F,* Bottin C, Corona SP, Guglielmi A, Rizzardi C, Milione M, Cortale M, Confalonieri M, Generali D.Mol Biol Rep. 2019 Mar 6.
23. **Machine learning in clinical and epidemiological research: isn't it time for biostatisticians to work on it?** Azzolina, Danila; Baldi, Ileana; Barbati, Giulia; et al. Group Author(s): Machine Learning Clinical Res Grp Epidemiology biostatistics and public health, 2019 Volume: 16 Issue: 4
24. **Efficacy of extended aromatase inhibitors for hormone-receptor-positive breast cancer: A literature-based meta-analysis of randomized trials.** Corona SP, Roviello G, Strina C, Milani M, Madaro S, Zanoni D, Allevi G, Aguggini S, Cappelletti MR, Francaviglia M, Azzini C, Cocconi A, Sirico M, Bortul M, Zanconati F, *Giudici F*, Rosellini P, Meani F, Pagani O, Generali D. Breast. 2019 Apr 25;46:19-24. doi: 10.1016/j.breast.2019.04.004.
25. **Tumour infiltrating lymphocytes and immune-related genes as predictors of outcome in pancreatic adenocarcinoma**. D'Angelo A, Sobhani N, Roviello G, Bagby S, Bonazza D, Bottin C, *Giudici F*, Zanconati F, De Manzini N, Guglielmi A, Generali D. PLoS One. 2019 Aug 5;14(8):e0219566. doi: 10.1371/journal.pone.0219566. eCollection 2019. PMID: 31381571
26. **The role of distress and pain catastrophizing on the health-related quality of life of children with inflammatory bowel disease.** De Carlo C, Bramuzzo M, Canaletti C, Udina C, Cozzi G, Pavanello PM, Rampado S, Martelossi S*, Giudici F*, Di Leo G, Barbi E.. J Pediatr Gastroenterol Nutr. 2019 Jul 22. doi: 10.1097/MPG.0000000000002447.PubMed PMID: 31335840.
27. **Parental Psychological Factors and Quality of Life of Children with Inflammatory Bowel Disease.** Bramuzzo M, De Carlo C, Arrigo S, Pavanello PM, Canaletti C, *Giudici F*, Agrusti A, Martelossi S, Di Leo G, Barbi E. J Pediatr Gastroenterol Nutr. 2019 Nov 6. doi: 10.1097/MPG.0000000000002548. PMID: 31703041

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

2. Five-year results of radiofrequency and laser ablation of benign thyroid nodules: a multicenter study from the Italian minimally invasive treatments of the thyroid group

3. Poly (adp-ribose) polymerase inhibitors in solid tumors: systematic review and meta-analysis

4. Adjusting survival outcomes for treatment switches: statistical methods and a practical application in breast cancer

5. Breast cancer recurrence estimation from population-based cancer registry linked with administrative data

ACKWNOLEDGEMENTS

- **For chapter 2**

Source of data: Italian minimally-invasive treatments of the thyroid (MITT) group (Centres: Milano, Genova, Napoli, Lecce, Latina, Teramo, Torino, Trieste)

Major collaborators: Principal Investigator**:** Stella Bernardi (Dipartimento di Scienze Mediche, Università degli Studi di Trieste, Trieste-UO Medicina Clinica, Ospedale di Cattinara, ASUGI (Azienda Sanitaria Universitaria Giuliano Isontina), Trieste

- **For Chapter 4**

Source of Data: The analysis was based on the derived data from trial TAM-01 (Delozier T., Spielmann M., Macé-Lesec'h J. et al.,Tamoxifen Adjuvant Treatment Duration in Early Breast Cancer: Initial Results of a Randomized Study Comparing Short-Term Treatment With Long-Term Treatment. J Clin Oncol., 2000, 8:3507-3512)

This analysis was conducted with the supervision of Aurelie Bardet at Department of Biostatistics and Epidemiology, Gustave Roussy, Oncostat U1018, CESP, Inserm, Université Paris-Saclay, Villejuif, in the framework of the project "Tamoxifen & compliance"

Major collaborators: Aurelie Bardet and Stefan Micheils (Department of Biostatistics and Epidemiology, Gustave Roussy, Paris-Saclay University, Villejuif, France; Oncostat U1018, Inserm, Paris-Saclay University, Labeled Ligue Contre le Cancer, Villejuif, France), Barbara Pistilli (Department of Medical Oncology, Gustave Roussy, Villejuif, France)

- **For Chapter 5**

Source of data: Friuli Venezia Giulia Cancer registry, Director Diego Serraino

This analysis was conducted with the supervision of Luigino Dal Maso at Centro di Riferimento Oncologico di Aviano in the framework of the project AIRC: "Long-term survival and cancer cure: estimates of population –based indicators in Italy and Europe"

Major Collaborators: Federica Toffolutti and Luigino Dal Maso (Cancer Epidemiology Unit, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS), Francesco Schiettini (Translational Genomics and Targeted Therapies in Solid Tumours, IDIBAPS, Barcelona, Spain; SOLTI Breast Cancer Research Group, Barcelona), Marina Bortul (Azienda Sanitaria Universitaria Giuliano Isontina, ASUGI, SSD Breast Unit, Cattinara Hospital, Trieste) ,Samuele Massarut (Breast Surgery Unit, Centro di Riferimento Oncologico di Aviano (CRO Aviano), IRCCS, National Cancer Institute, Aviano), Giulia Barbati (Biostatistics Unit, Department of Medical Surgical and Health Sciences-University of Trieste)

1 INTRODUCTION

1.1 PRECISION MEDICINE: OVERVIEW

The idea of improving health outcomes by tailoring treatment to individual patient characteristics is centuries old and remains a core component of medical practice. The scientific method began to affect medical treatment with statistical inference by the late 1700s, but advances began to dramatically increase after the success of the first randomized controlled clinical trial, conducted by Austin Bradford Hill in 1946, which demonstrated the efficacy of streptomycin for treating tuberculosis (1). Following Hill's trial was a period of rapid methodological progress in the design and analysis of clinical trials as well as observational studies. Systematic study of the integration of data, experience, and clinical judgment into the clinical decision process led to the concept of evidence-based medicine, wherein clinical decision-making is based on empirical evidence with randomized controlled trials being a gold standard for generating such evidence (2). However, the primary scientific aim in most clinical trials is the identification of the best treatment for a given disease area, with any heterogeneity in patient characteristics or outcomes being viewed as a nuisance to the research process. Awareness that patient heterogeneity was important in evaluating treatments began to emerge late in the twentieth century among both clinicians (3) and biostatisticians (4). That patient heterogeneity implied the need to individualize therapy in the context of evidence-based medicine was nicely articulated in Kravitz et al. (5). These constituent concepts, combined together, yield the modern concept of precision medicine, the paradigm wherein patient heterogeneity is leveraged through data-driven approaches to improve treatment decisions so that the right treatment is given to the right patient at the right time. We note that precision medicine is conceptually the same as stratified medicine (6) and personalized medicine (7). The chief priority of statistical research in precision medicine is to use data to inform decision making in health care; thus, precision medicine encompasses a wide range of tasks including drug discovery, biomarker identification, estimation and inference for causal treatment effects, modeling health communication and shared decision-making and study design. An estimated optimal treatment regime might be used as part of a decision support system within a health care organization or to generate new clinical hypotheses for future study. Thus, it is critical that statistical methodology for precision medicine be rigorous, transparent, reproducible and generalizable. Precision medicine fits within the broader concepts of precision public health and data-driven decision science. However, the focus on data-driven, patient-centered care with its inherent challenges (e.g., patient heterogeneity, implementation cost, and causal confounding) distinguishes precision medicine as its own field of study. To this point, precision medicine has led to new methodologies and insights in semiparametric modeling, causal inference, clinical trial design, and machine learning (7-10). There have also been major advancements in genetics driven by the vision for precision medicine (11-12); however, the focus is broader, in that while the biomarkers used to inform treatment selection could be genetic or genomic factors, the precision medicine allows that treatment selections could be based on demographic and physiological measurements, comorbid conditions, individual patient preferences, lifestyle, and so on.

1.2 PRECISION MEDICINE AND STATISTICAL INFERENCE

The goal in precision medicine is to use data to improve decision making in health care. Dynamic treatment regimens formalize decision making as a sequence of decision rules, one per decision point, that map available information to a recommended intervention. The decision points may be

either fixed in calendar time or driven by patient outcomes. Thus, the timing and number of decision points may be random and can vary considerably across patients in some application domains. From a statistical perspective, personalized medicine is a process involving six fundamental steps as showed in Figure 1 (13-14). Intrinsic to any statistical inference, initially one must select an appropriate method of inference based on the available source of training data and clinical endpoints (e.g., steps (1) and (2)). Step (3) is the fundamental component of personalized treatment selection, deriving the individualized treatment rule (ITR) for the chosen method of inference. An ITR is a decision rule that identifies the optimal treatment given patient/disease characteristics (15-16).

Individualized treatment rules are functions of model parameters (usually treatment contrasts reflecting differences in treatment effects) which must be estimated from the assumed statistical model and training data. Statistical estimation takes place in step (4). After estimating the optimal treatment rule in step (4), the resulting estimated ITR's performance and reliability must be evaluated before the model can be used to guide treatment selection (17). The manner in which one assesses the performance of the derived ITR depends on the appropriate clinical utility (i.e., increased response rate or prolonged survival duration). Evaluation of model goodness-of-fit and appropriate summary statistics that use the available information to measure the extent to which future patients would benefit from application of the ITR is conducted in step (5). The ITR is applied to guide treatment selection for a future patient based on his/her baseline clinical and genetic characteristics as the final step.

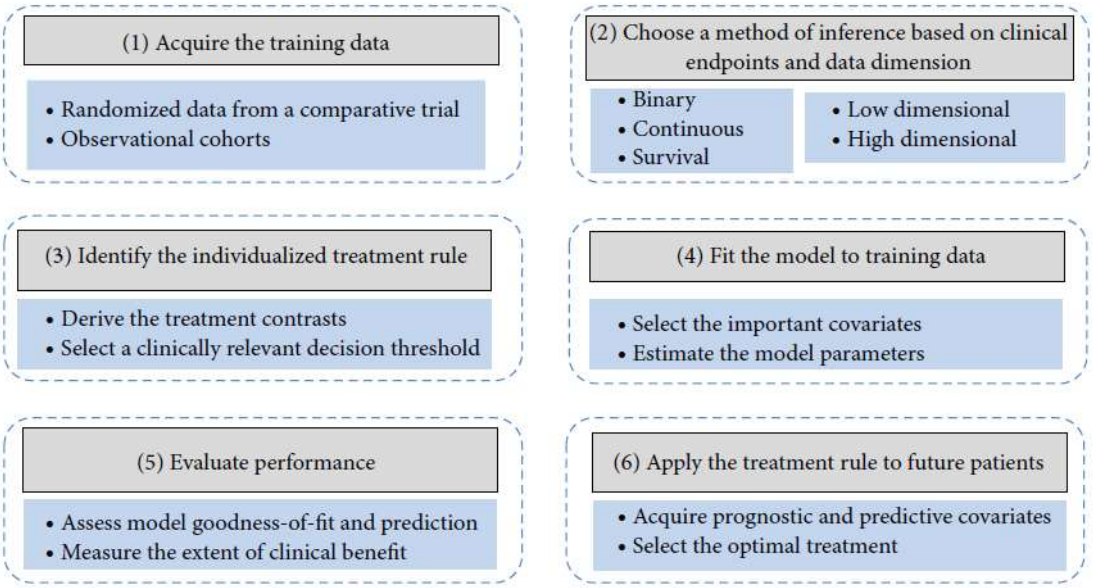


Figure 1*:* The process of using statistical inference to establish personalized treatment rules (14)

1.3 STATISTICAL METHODS TO ANALYZE CONTINUOUS BIOMARKERS

A biological marker (biomarker) is "a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions." (18). Biomarkers have various applications, such as risk estimation, disease screening and detection, diagnosis, estimation of prognosis, prediction of benefit from therapy and disease monitoring (Figure 2).

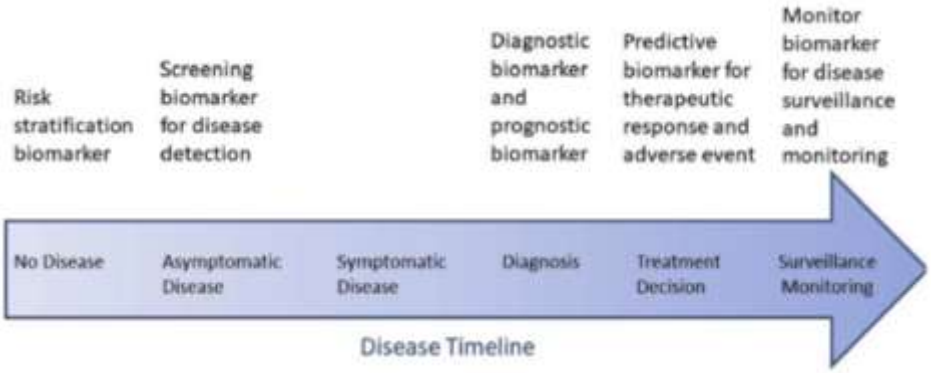


Figure 2: Use of biomarkers in relation to the course of disease (19)

In this era of precision medicine, having validated biomarkers to inform clinical decision-making is more important than ever. Biomarkers can provide a basis for the selection of lead candidates for clinical trials, for contribution to the understanding of the pharmacology of candidates and for characterization of the subtypes of disease for which a therapeutic intervention is most appropriate. In oncology, biomarkers are typically classified as either prognostic or predictive (see Figure 3). Prognostic biomarkers are correlates for the extent of disease or extent to which the disease is curable. Therefore, prognostic biomarkers influence the likelihood of achieving a therapeutic response regardless of the type of treatment. By way of contrast, predictive biomarkers select patients who are likely or unlikely to benefit from a particular class of therapies (20). Thus, predictive biomarkers are used to guide treatment selection for individualized therapy based on the specific attributes of a patient's disease. A prognostic biomarker can be identified in properly conducted retrospective studies that do not rely solely on convenience samples but use biospecimens prospectively collected from a cohort that represents the target screening population, case-control studies, and single arm trials. A prognostic biomarker is identified through a main effect test of association between the biomarker and the outcome in a statistical model. Predictive biomarker needs to be identified in secondary analyses using data from a randomized clinical trial, through an interaction test between the treatment and the biomarker in a statistical model. Secondary analyses refer to subsequent correlative studies that may or may not be predefined as a protocol objective.

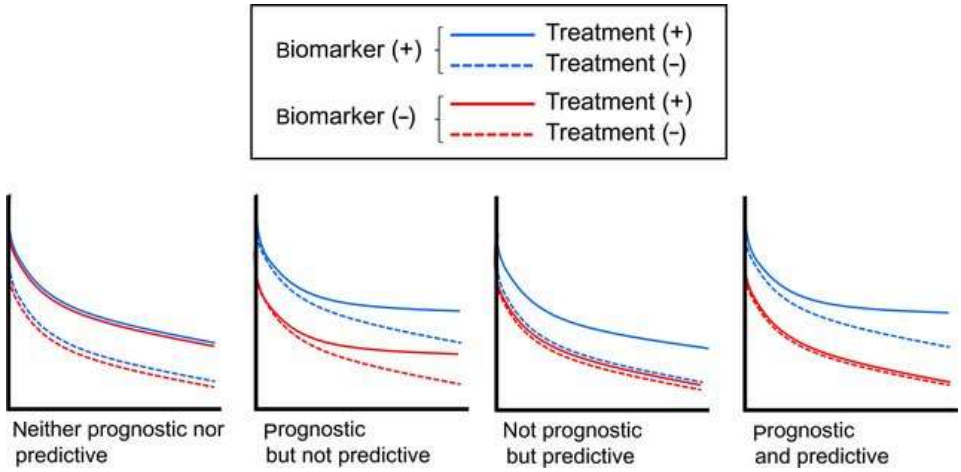


Figure 3: Difference between prognostic and predictive biomarkers (21)

For continuous biomarkers, to determine patients' subgroups, it is often necessary to determine cut-off points based on their relationship to clinical response of interests (e.g. survival outcomes). Currently, there is no standard method or standard software for biomarker cut-off determination. The problem of the choosing the optimal cut-off is difficult to answer generally. The best method for cut-off determination may depend on the biomarker, the assay and the clinical application under investigation.
To find cut-offs values of continuous biomarkers several approaches can be applied:

1) Diagnostic study: the marker and disease are measured at the same time
-*classical approaches*: median or mean value, percentiles according to biomarker values distribution, optimal cut-point identified by means standard receiver operating characteristic (ROC) analysis

2) Prognostic studies: the marker is measured at a given time while the disease may occur at any time thereafter
- *time-dependent ROC:* extension of ROC-based cut-point finding methods to the case of censored failure time outcomes. This analysis aims to assess the discrimination ability of a binary marker measured at baseline to identify patients who will relapse in time

-*conditional inferential tree:* the conditional inference tree method applies binary recursive partition sequentially on each independent predictor that is associated with the given response variable. The order of the predictors to be partitioned depends on the significance of the association between the predictor and the response variable. Partition steps will be repeated until a pre-define level of statistical significance is reached

-*Subpopulation Treatment Effect Pattern (STEPP) method*: the approach of categorizing biomarker expression may fail to fully identify the worth of the biomarker as a predictor of treatment efficacy because categorization results in a loss of information. Alternatives to such dichotomized analyses should be applied for evaluating treatment-effect heterogeneity when the biomarker is measured on a continuous scale.
The STEPP methodology examines treatment-effect heterogeneity by estimating treatment effect within overlapping subpopulations of patients, where the subpopulations are defined with respect to values of the variable of interest along its range (22).

1.4 STATISTICAL METHODS FOR TIME-TO-EVENT ENDPOINTS

Competing risk methods are time-to-event analyses that account for fatal and/or nonfatal events that may potentially alter or prevent a subject from experiencing the primary endpoint. Competing risk methods may provide a more accurate and less biased estimate of the incidence of an outcome, but are rarely applied. Kaplan–Meier (KM) estimates of survival curves and Cox proportional hazard models are widely used to describe survival trends and identify significant prognostic factors. All these statistical analyses deal with only one type of event, for example death, independently of its cause. A particular situation arises when interest is focused on a specific cause of failure in the presence of other different causes, which alter the probability of experiencing the event of interest. This is the case of competing risk events, which refers to a situation where an individual is exposed to two or more causes of failure, and its eventual failure can be attributed exactly to only one. In this case, the occurrence of one type of event hinders the occurrence of any other event. In the analysis of competing risks data, the Kaplan-Meier method (1-KM) for estimation of cumulative incidence function (CIF) lead to incorrect and biased results because treat competing events as censored at the time they occurred, but this censoring is inappropriate because after a competing event has occurred, failure from the cause of interest is no longer possible. 1-KM correctly estimates the probability of failure independently of any specific cause, while the probability of one type of competing event is correctly estimated using the CIF. CIF partitions the probability of failure into the probability corresponding to each competing event: at any point in time, the overall 1-KM is equal to the sum of the CIFs for each type of event. (23). Moreover, to assess the statistical significance of a prognostic factor in a cumulative incidence analysis, Gray's test (24) is one of the appropriate tests to perform. As regard as regression analysis, the most commonly used regression model for analyzing event-time data is the Cox proportional hazards model. In the presence of competing risks, the standard Cox proportional hazards model is not adequate because the cause-specific Cox model treats competing risks of the event of interest as censored observations. In addition, the cause-specific hazard function does not have a direct interpretation in terms of survival probability. Direct regression modeling of the effect of covariates on the cumulative incidence function (CIF) for competing risks data has been proposed, among others, by Fine and Gray (25). Fine and Gray proposed a model for the subdistribution hazard of the CIF. The subdistribution hazard is a key concept in this approach, and it is defined as the hazard of failing from a given cause in the presence of competing events, given that a subject has survived or has already failed due to different causes. In summary, the important first step for the analysis of competing risks data is the recognition that competing risks are present. Following this, the analysis should include a calculation of cumulative incidence of an event of interest in the presence of competing risks, a proper test for cumulative incidence curves of an event, and competing risk regression analyses.

1.5 META-ANALYSIS AS STATISTICAL APPROACH IN PRECISION MEDICINE

Systematic reviews and meta-analyses are essential tools for synthesizing evidence needed to inform clinical decision making and policy. Systematic reviews summarize available literature using specific search parameters followed by critical appraisal and logical synthesis of multiple primary studies. Meta-analysis refers to the statistical analysis of the data from independent primary studies focused on the same question, which aims to generate a quantitative estimate of the studied phenomenon, for example, the effectiveness of the intervention (26). In clinical research, systematic reviews and meta-analyses are a critical part of evidence-based medicine. (Figure 4)

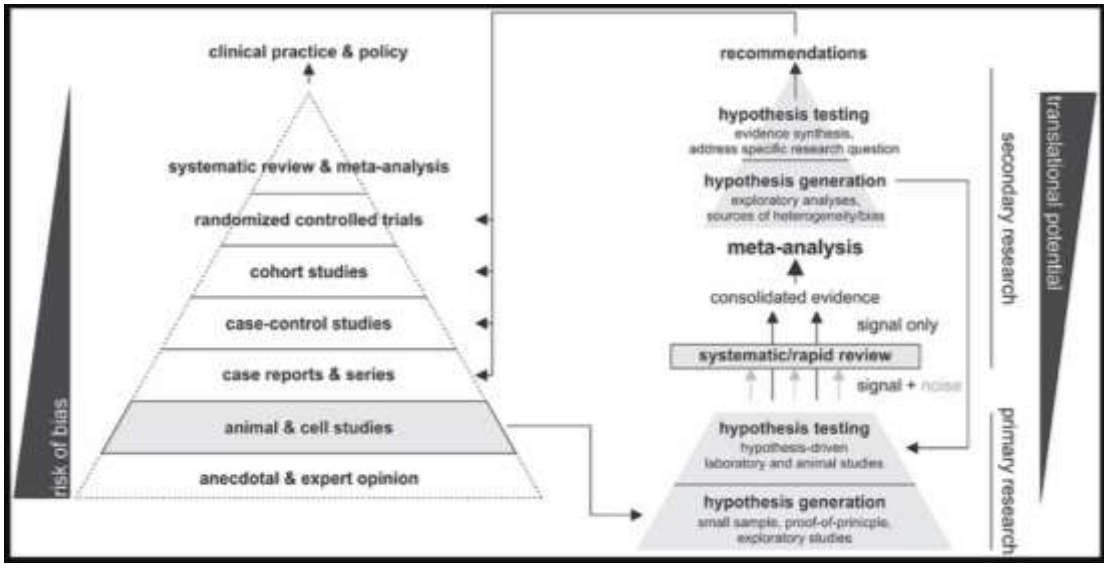


Figure 4. Schematic of proposed hierarchy of translational potential in basic research (27)

Meta-analysis, a set of statistical techniques for synthesizing the results of multiple studies, is used when the guiding research question focuses on a quantitative summary of study results

Traditional methods of meta-analysis attempt to combine results in order to obtain a single summarized 'effect size'. Transparency and reproducibility are key quality indicators of a meta-analysis. A high-quality meta-analysis, like any high-quality primary study, must provide a strong argument that the methods and analytic strategy can support claims about the distribution of effect sizes across studies and thus about the quantitative results in a given literature base. Many meta-analyses focus on questions that have direct policy implications**.** For this reason, it is fundamental to apply statistical method adequate to investigate the heterogeneity among studied, the presence of outliers and the publication bias. According to the number of studies included, advanced methods, such as metaregression models, are suggested to investigate whether particular covariates (potential effect modifiers) explain any of the heterogeneity of treatment effects between studies.

1.6 CAUSAL INFERENCE METHODS TO ADJUST FOR TREATMENT SWITCHING IN RANDOMIZED TRIALS

Although randomized clinical trials (RCTs) often suffer from non-adherence or noncompliance of trial participants to the intervention(s) protocol to which they are randomized, the analyses are usually performed according to intention-to-treat (ITT) principle. These analyses do not provide the answer to the key question of the true benefit of the experimental drug when appropriately undertaken. Several commonly used statistical methods are available to estimate survival benefit while adjusting for treatment switching, ranging from naive exclusion or censoring approaches, to time-dependent Cox regression models. However, all these methods are prone to serious selection bias and cannot claim causal treatment effect. More complex methods are required to improve upon the ITT analysis and account for treatment switching. Rank-preserving structural failure time models (RPSFTM) and inverse probability of censoring weights (IPCW) are well-established methods that may be used for this purpose (28-29). The first methods rely on g-computation whereas the last one on the counterfactual framework relies. In the presence of treatment switching, several sensitivity analyses should be performed to evaluate robustness of the complex models: if a

range of methods are shown to be potentially appropriate for a particular case, and each provides similar estimates of the treatment effect, decision makers may have more confidence in the results. This project is in collaboration with the Oncostat Team, Department of Biostatistics and Epidemiology, Gustave Roussy, Paris-Saclay University, Villejuif, France, where I have spent a period as visiting researcher.

1.7 POPULATION-BASED REGISTRY AND ADMINISTRATIVE DATABASES

Access to health data is the next scalability challenge for personalized medicine. New evidence derived from large populations of data is needed to direct the development of new-targeted drugs, to investigate the disease pathway and patient characteristics, to discover unmet needs with the aim to deliver more personalized packages of care. In the era of precision medicine, overall survival alone is not an adequate endpoint for assessing healthcare quality, comparing treatment efficacy, or informing decision making for patients with cancer, especially for cancers with long survival times such as breast cancer. Knowing the risks of second breast cancer events is important for improving quality of life (30), for patients making decisions about their treatment as well as for cancer control experts identifying research priorities and health services planning (31). Consequently, the importance of studying long-term outcomes in breast cancer patients is growing: second breast cancer events (i.e. loco-regional recurrences, metastases and second primary breast cancers) are of interest in these studies, and efficient methods of identifying and collecting data on the occurrence of second breast cancer events are needed. Although population-based cancer registries data are useful in tracking and reporting the evolving burden of cancer in the population, the information they recorded reflects the outcomes of diagnosis and death and in particular, registries do not routinely collect information on cancer progression or recurrence. A method that in recent years is spreading is the use of the growing bulk of population-based administrative data from hospitals and other health care-related institutions as proxy for follow-up of patient. Such data offer new possibilities for the generations of disease models for health evaluation: in particular, records of breast cancer patients from administrative data can be potentially used for identifying recurrences. This contribution to my Ph.D. Thesis is in collaboration with the Cancer Epidemiology Unit, CRO (Centro di Riferimento Oncologico) Aviano National Cancer Institute. This is the main project of my last year of PhD course: the main goal of this study is to develop an acceptable claims-based algorithm to identify breast cancer recurrences (local (LR), regional (RR) and distant metastases (DM)) during a 10 years' follow-up through a record-linkage of two data sources- 1) Friuli Venezia Giulia population based-cancer registry and 2) administrative individual-record regional database. Such an algorithm has the potential to be implemented in future data repositories to facilitate studies of disease surveillance, monitoring, and quality assessment.

The main aim of this Ph.D. thesis is to provide adequate statistical methods in the field of statistical research for precision medicine according to the study design (observational studies or randomized clinical trials) and outcomes of interest, with particular interest in the oncological setting. In the following chapters, I report the most important contributes of Ph.D. research period regarding the mentioned topics.

1.8 REFERENCES

1. Stusser RJ. 2006. Reflections on the scientific method in medicine. In Medical Sciences, Vol. 1, ed. YLG Verhasselt, AVJablensky, A Pellegrini, CSayegh, A Kazanjian, et al., pp. 23–47. Paris: UNESCO-EOLSS Joint Comm. Secr
2. Eddy DM. 1990. Practice policies: guidelines for methods. J. Am. Med. Assoc. 263:1939–41
3. Sorensen TIA. 1996. Which patients may be harmed by good treatments? Lancet 348:351–2
4. Longford NT, Nelder JA. 1999. Statistics versus statistical science in the regulatory process. Stat. Med.18:2311–20
5. Kravitz RL, Duan N, Braslow J. 2004. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q. 82:661–87
6. Lonergan M, Senn SJ, McNamee C, Daly AK, Sutton R, et al. 2017. Defining drug response for stratified medicine. Drug Discov. Today 22:173–79
7. KosorokMR, Moodie EEM. 2016. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. Philadelphia: SIAM
8. Chakraborty, B., & Moodie, E. (2013). Statistical methods for dynamic treatment regimes: Reinforcement learning, causal inference, and personalized medicine. New York: Springer
9. Laber EB, Lizotte DJ, Qian M, Pelham WE, Murphy SA. 2014. Dynamic treatment regimes: technical challenges and applications. Electron. J. Stat. 8:1225–72
10. Zhao YQ, Zeng D, Laber EB, Kosorok MR. 2015a. New statistical learning methods for estimating optimal dynamic treatment regimes. J. Am. Stat. Assoc. 110:583–98
11. Liang F, LiQ, Zhou L. 2018. Bayesian neural networks for selection of drug sensitive genes. J. Am. Stat. Assoc. 113:955–72
12. Torkamani A, Wineinger NE, Topol EJ. 2018. The personal and clinical utility of polygenic risk scores. Nat.Rev. Genet. 19:581–90
13. F. E. Harrell, K. L. Lee, and D. B. Mark, "Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," Statistics in Medicine, vol. 15, no. 4, pp. 361– 387, 1996
14. Ma, J., Hobbs, B. P., & Stingo, F. C. (2015). Statistical Methods for Establishing Personalized Treatment Rules in Oncology. BioMed research international, 2015, 670691. https://doi.org/10.1155/2015/670691
15. M. Qian and S. A. Murphy, "Performance guarantees for individualized treatment rules," The Annals of Statistics, vol. 39, no. 2, pp. 1180–1210, 2011
16. W. Lu, H. H. Zhang, and D. Zeng, "Variable selection for optimal treatment decision," Statistical Methods in Medical Research, vol. 22, no. 5, pp. 493–504, 2013)
17. E.W. Steyerberg, A. J. Vickers, N. R. Cook et al., "Assessing the performance of prediction models: a framework for traditional and novel measures," Epidemiology, vol. 21, no. 1, pp. 128–138, 2010
18. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and Other Tools. Resource. Silver Spring,MD: Food and Drug Administration; 2016.
19. Biomarker Discovery and Validation: Statistical Considerations Ou, Fang-Shu et al. Journal of Thoracic Oncology, Volume 16, Issue 4, 537 – 545)

20. R. Simon, "Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology," *Personalized Medicine*, vol. 7, no. 1, pp. 33–47, 2010
21. Mukohara T. PI3K mutations in breast cancer: prognostic and therapeutic implications. Breast Cancer (Dove Med Press). 2015 May 15;7:111-23. doi: 10.2147/BCTT.S60696. PMID: 26028978; PMCID: PMC4440424.)
22. Lazar AA, Cole BF, Bonetti M, Gelber RD. Evaluation of treatment-effect heterogeneity using biomarkers measured on a continuous scale: subpopulation treatment effect pattern plot. J Clin Oncol. 2010 Oct 10;28(29):4539-44).
23. Klein JP, Rizzo JD, Zhang MJ, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: unadjusted analysis. Bone Marrow Transplant 2001; 28: 909–915
24. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 1988; 16: 1141–115)
25. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999; 94: 496–50)
26. Gopalakrishnan, S., and Ganeshkumar, P. (2013). Systematic reviews and meta-analysis: understanding the best evidence in primary healthcare. J Fam. Med. Prim. Care 2, 9–14. doi: 10.4103/2249-4863.109934
27. Mikolajewicz N and Komarova SV (2019) Meta-Analytic Methodology for Basic Research: A Practical Guide. Front. Physiol. 10:203. doi: 10.3389/fphys.2019.00203)
28. Robins, J. M., and A. A. Tsiatis. 1991. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun. Stat. Theory Methods 20:2609–26317, 14–18
29. Latimer NR, Abrams KR, Lambert PC, et al., Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. Stat Methods Med Res. 2018 Mar;27(3):765-784
30. Ferreira AR, Di Meglio A, Pistilli B, et al. Differential impact of endocrine therapy and chemotherapy on quality of life of breast cancer survivors: a prospective patient-reported outcomes analysis. Ann Oncol. 2019;30(11):1784–95. https://doi.org/10.1093/annonc/mdz298
31. Shimozuma K. Quality of life assessment. Breast Cancer. 2002;9(2):100–6. https://doi.org/10.1007/bf02967573

2 FIVE-YEAR RESULTS OF RADIOFREQUENCY AND LASER ABLATION OF BENIGN THYROID NODULES: A MULTICENTER STUDY FROM THE ITALIAN MINIMALLY INVASIVE TREATMENTS OF THE THYROID GROUP

ABSTRACT

Background: Radiofrequency ablation (RFA) and laser ablation (LA) are effective treatments for benign thyroid nodules. Due to their relatively recent introduction into clinical practice, there are limited long-term follow-up studies. This study aimed to evaluate technique efficacy, rate of regrowth, and retreatment over 5 years after RFA or LA and to identify predictive factors of outcome.

Methods: In this multicenter retrospective study, the rates of technique efficacy, regrowth, and retreatment were evaluated in 406 patients treated with either RFA or LA, and followed for 5 years after initial treatment. Propensity score matching was used to compare treatments. Cumulative incidence studies with hazard models were used to describe regrowth and retreatment trends, and to identify prognostic factors. Logistic regression models and receiver operating characteristic analyses were used for risk factors and their cutoffs.

Results: RFA and LA significantly reduced benign thyroid nodule volume, and this reduction was generally maintained for 5 years. Technique efficacy (defined as a reduction ‡50% after 1 year from the treatment) was achieved in 74% of patients (85% in the RFA and 63% in the LA group). Regrowth occurred in 28% of patients (20% in the RFA and 38% in the LA group). In the majority of cases, further treatment was not required as only 18% of patients were retreated (12% in the RFA and 24% in the LA group). These data were confirmed by propensity score matching. Cumulative incidence studies showed that RFA was associated with a lower risk of regrowth and a lower risk of requiring retreatment over time. Overall, technique inefficacy and regrowth were associated with low-energy delivery. Retreatments were more frequent in young patients, in large nodules, in patients with lower volume reduction at 1 year, and in cases of low-energy delivery (optimal cutoff was 918 J/mL for RFA).

Conclusions: Both thermal ablation techniques result in a clinically significant and long-lasting volume reduction of benign thyroid nodules. The risk of regrowth and needing retreatment was lower after RFA. The need for retreatment was associated with young age, large baseline volume, and treatment with low-energy delivery

Summary of Statistical Methods applied: propensity score methods to attenuate the effects of the important confounding variables due to selection bias; competing risk analysis: estimate of cumulative incidence function, Gray's test and Fine-Gray regression analysis; Receiver operating characteristic (ROC) analyses to identify optimal cut-point for continuous variables

This chapter was published as:

Stella Bernardi***, Fabiola Giudici***, Roberto Cesareo, Giovanni Antonelli, Marco Cavallaro, Maurilio Deandrea, Massimo Giusti, Alberto Mormile, Roberto Negro, Andrea Palermo, Enrico Papini, Valerio Pasqualini, Bruno Raggiunti, Duccio Rossi, Luca Maria Sconfienza, Luigi Solbiati, Stefano Spiezia, Doris Tina, Lara Vera, Fulvio Stacul, and Giovanni Mauri. ***Five-Year Results of Radiofrequency and Laser Ablation of Benign Thyroid Nodules: A Multicenter Study from the Italian Minimally Invasive Treatments of the Thyroid Group***. Thyroid. Dec 2020.1759-1770.http://doi.org/10.1089/thy.2020.0202

2.1 INTRODUCTION

Benign thyroid nodules are a common clinical finding. Although most of them are asymptomatic, in a small percentage of patients (10-15%) they increase over time, causing local symptoms and cosmetic concerns (1). In these cases, the conventional remedy is thyroid surgery. Recently, however, minimally invasive non-surgical treatments, mostly image-guided thermal ablations, such as laser and radiofrequency ablation (LA and RFA), have emerged as an alternative approach to treat symptomatic benign thyroid nodules (2-4).

Both RFA and LA are outpatient procedures, which are generally performed under local anesthesia. Technically, either an electrode-needle generating an alternating electric field (in case of RFA) or one or more optical fibers conveying laser light (in case of LA), are inserted into the nodule under ultrasound (US) guidance, to induce rapid heating of the target zone. Treatment is accompanied by the formation of coagulative necrosis, and, over time, by fibrotic changes and progressive nodule shrinkage. RFA is generally performed with the moving shot technique, whereby the tip of the electrode is sequentially moved from the medial and deepest part of the nodule to its most superficial and lateral parts. By contrast, LA requires the positioning of one or more optical fibers into the target nodule, which might be eventually pulled back in case of larger nodules (1). The mean costs of RFA and LA are substantially similar with a fixed charge for RFA (i.e., one device per nodule, whose cost ranges from $500-1000) and a variable one for LA depending on the number of fibers (i.e., the larger the nodule volume to be treated the higher the cost, as one fiber costs $300-500) (5).

The use of RFA and LA to treat symptomatic benign thyroid nodules is supported by robust evidence of efficacy and tolerability. Both treatments demonstrated a significant reduction of thyroid nodule volume (6-8), paralleled by significant improvement of local symptoms. Procedures are reported as well tolerated in large retrospective series, with a risk of major complications (recurrent laryngeal nerve injury or damage to cervical structures) lower than 1% (9, 10). Unfortunately, long-term follow-up studies evaluating not only volume reduction and technique efficacy, but also regrowth and retreatment rates are currently limited. Based on this background, this multicenter retrospective study aimed (*i*) to evaluate the rate of technique efficacy, regrowth, and retreatment following thyroid thermal ablations, as well as (*ii*) to use logistic regression models and receiver operating characteristics (ROC) analyses in order to identify potential risk factors and cut-off values predicting efficacy, regrowth, and retreatment. Cumulative incidence studies and hazard models were used to describe regrowth and retreatment over time and to identify prognostic factors

2.2 MATERIALS AND METHODS

Study design

This is a retrospective multicenter study, whose primary outcome was to describe the rate of technique efficacy, regrowth, and retreatment during the 5 years after thermoablation of a benign thyroid nodule. Secondary outcomes were the identification of predictive variables of efficacy, regrowth, and retreatment. Inclusion criteria were: (*i*) benign cytology prior to ablation (diagnostic category Thy2/Tir2 or Bethesda II, (11, 12), as assessed by FNAB; (*ii*) no prior thyroid treatment (radioiodine, ethanol injection); (*iii*) yearly follow-up of at least 5 consecutive years after the first ablation; and (*iv*) patient consent to use their data for this study. Patients fitting the criteria (*i*) and

(*ii*), who had been treated before 2015, whose follow-up was interrupted because they underwent surgery were also included. By contrast, all the patients treated from 2015 onwards were excluded (as they could not

complete a 5-year follow-up by the time of data analysis). This study was conducted in accordance with the declaration of Helsinki, and the protocol of this retrospective analysis was approved by the Institutional Review Board (268_2019 FYTNAB).

The study protocol was presented during the 2nd meeting of the Minimally-invasive treatments thyroid (MITT) Group (4, 13), held in Milan in February 2019. The Italian centers belonging to the MITT group were invited with an open call to contribute with patient data. Centres were invited to contribute with data of entire annual cohorts of patients treated before 2015 (**Table 1**). The following parameters were collected: age, sex, year of treatment, type of procedure, energy delivered (J), baseline nodule volume (mL), nodule structure, nodule function, nodule volume after 1, 2, 3, 4, 5 years from the procedure (mL), symptom relapse, type of retreatment, final pathology (in case of surgery), nodule volume after a second thermal ablation (mL). In case of LA, the number of fibers was recorded, while in case of RFA, the type of electrode was specified. Nodule volume was measured by ultrasound examination. Ultrasound scans were generally performed with linear transducers except for very large nodules, whose volume was quantified with convex transducers. To measure nodule volume (V) the following formula was used: $V = \pi abc/6$ (where V is the volume, a is the maximum diameter, and b and c are the other two perpendicular diameters). Energy delivered was expressed as J/mL. Joules (or kilocalories) were either provided by the machine or calculated as Watt * s. Nodule function was assessed with laboratory examinations as well as thyroid scintigraphy, which was performed only in case TSH was < 0.4 microU/mL (2, 14).

Definitions

Nodule structure was classified as solid if the fluid component was ≤10%, predominantly solid if the fluid component was between 11-50%, predominantly cystic if the fluid component was between 51-90%, and cystic if the fluid component was >90% (15). ***Volume reduction ratio*** (VRR) is the percentage reduction in volume and it is calculated as follows: VRR = ((initial volume – final volume)/initial volume) x 100. Given that our cohort included some patients who were retreated, in order to analyze nodule volume reduction induced by the first procedure/ablation, data after retreatments were censored. ***Technique efficacy*** was defined as a volume reduction ≥50% after 1 year from the treatment (15, 16). As opposed to technique efficacy, technique inefficacy was defined as any volume reduction <50% after 1 year from the treatment. ***Regrowth*** was defined as a ≥50% increase compared to the previous smallest volume at US examination (15, 17).

Statistical analyses.

All statistical analyses were carried out in R system for statistical computing (Ver. 5.0; R Development Core Team, 2018). Statistical significance was set at $p<0.05$.

Shapiro-Wilk test was applied to quantitative (continuous) variables to check for distribution normality. Continuous variables were reported as median with range (minimum-maximum). Qualitative (categorical) variables were reported as absolute frequencies and/or percentages (rates of technique efficacy, regrowth and retreatment). Continuous variables were compared by student's

t test (and ANOVA) or by Mann–Whitney test (and Kruskall Wallis test), depending on data distribution and number of groups. Categorical variables were compared by Chi-square test or Fischer's exact test whenever appropriate. Variations over time of nodules' volume were evaluated with linear mixed-effects models (LME) for repeated measures. Multiple comparisons of nodules' volume respect to different follow-up periods (baseline vs 1, 2, 3, 4, and 5 years) were performed with Friedman test for repeated measures and p-values adjusted with Bonferroni post-hoc test.

To compare the patients treated with RFA to those treated with LA, in order to control potential confounders and selection bias, we performed a sensitivity analysis using propensity score matching with the R package 'MatchIt' (method nearest neighbor). The patients were matched 1:1 by age, sex, nodule volume, nodule structure (solidity), and function.

To describe regrowth trends we used the cumulative incidence function (CIF), which takes into account the presence of competing risks (such as retreatment in our case). Then, cumulative incidence of regrowth in RFA and LA groups was compared with the Gray test. To identify significant predictors of regrowth over time (hazard ratio with 95% confidence interval) we used the Fine and Gray competing risk regression model (18). CIF and CRR analyses were performed with the R package *cmprsk* (19). To describe the likelihood of not being retreated we used the standard Kaplan-Meier method. Cox proportional hazard regression model was implemented to identify predictors of retreatment and to estimate HR with 95% CI.

To identify potential risk factors of technique inefficacy, regrowth, and retreatment, we conducted a univariate logistic regression analysis and calculated the odds ratios of age, sex, baseline volume, nodule structure and function, 1-year nodule reduction, technique efficacy and regrowth for the outcome technique inefficacy, regrowth, and retreatment. Statistically significant variables at 10% level at univariate analysis were selected as candidate prognostic factors for multivariate logistic regression analyses. It has to be noted that energy/volume and technique could not be tested simultaneously for collinearity. So, we decided to evaluate/prioritize the association between energy delivered and outcome of thermal ablations.

Receiver operating characteristic (ROC) analyses were used to calculate the accuracy of volume, 1-year volume reduction, and energy, as predictors of technique efficacy, regrowth, and retreatment. Area under the (ROC) curves with 95% confidence interval, were interpreted according to Sweets criteria, and were used to identify a cut-off value of baseline volume, 1-year volume nodule and energy delivered that best predicted technique efficacy, regrowth, and retreatment. Specificity and sensitivity were also calculated (95% confidence interval, CI). The best possible cut-off point was defined as the highest Youden Index ((specificity + sensitivity) – 1 (R package 'OptimalCutPoints')). DeLong method was used to test the statistical significance of the difference between the areas under the curve.

2.3 RESULTS

Study population and general characteristics.

Eight centers participated in the trial (Genova, Latina, Lecce, Milano, Napoli, Teramo, Torino, Trieste). Each center provided data of all consecutive patients (entire cohorts of patients) treated in the years reported in **Table 1.** Data from 477 patients with benign thyroid nodules were collected. Among these patients, 59 patients were lost during the follow-up and 12 patients had undergone

other treatments before the procedure, such that they were excluded. Inclusion criteria were met by 406/477 patients (85%), who were selected for this study (**Table 1**).

Median age was 57 years (17-87); there were 304/406 women (75%) and 102/406 men (25%). Among the 406 patients selected for this study, 216 patients (53%) were treated with radiofrequency ablation (RFA), while 190 patients (47%) were treated with percutaneous laser ablation (LA). Treatments with LA were performed between 2009 and 2014, consistent with the fact that LA is the first thermal ablation technique that was introduced in clinical practice to treat thyroid nodules (20), while RFA is more recent (21). LA was performed with 1-3 optical fibers and a 1064 nm diode laser source (20, 23). The number of fibers depended on nodule volume and morphology. Treatments with RFA were performed between 2012 and 2014. RFA was performed with the moving shot technique and the monopolar 18-G needle (21, 22).

Nodule volume reduction and technique efficacy

A total of 75% of patients had a solid nodule, 19% had a predominantly solid nodule, 5% had a predominantly cystic nodule, and 1% had a cystic nodule. Nodules were non-functioning in 91% of patients.

Overall, median baseline nodule volume was 14.26 mL (0.44-179.0). Specifically, it was 17.2 mL (0.44-179) in the RFA group and 12.2 mL (1.7-86) in the LA group (**Table 2**). Nodule volume was significantly reduced by the first ablation (**Table 2** and **Figure 1**), $p<0.001$ for repeated measures. Overall, median thyroid nodule volume decreased by 63%, 67%, 68%, 68%, and 70% at 1, 2, 3, 4, and 5 years after the first ablation. In all the patients treated with RFA (n=216), median thyroid nodule volume decreased by 72%, 75%, 76%, 76%, and 77% at 1, 2, 3, 4, and 5 years after the ablation. In all the patients treated with LA (n=190), median thyroid nodule volume decreased by 55%, 58%, 59%, 57%, and 57%, at 1, 2, 3, 4, and 5 years after the ablation (**Table 2**).

Overall, technique efficacy was achieved in 74.4% of the patients (302/406), specifically in 84.7% of patients treated with RFA (183/216) and 62.6% of patients treated with LA (119/190), $p<0.001$.

Regrowth and retreatment rates

A total of 28% of patients (115/406) had a regrowth. Among the 115 patients with a regrowth, 69% of patients (79/115) lost technique efficacy, 26% of patients (30/115) had symptom relapse, and 28% of patients (32/115) were retreated. When looking at RFA and LA groups, regrowth was observed in 19.9% of patients treated with RFA (43/216) and in 37.9% of patients treated with LA (72/190), $p<0.001$. Consistent with the efficacy of both procedures and the lower tendency to regrow after RFA, we found a good correlation between 1-year and 5-year volume reduction after both treatments, even if it was more pronounced after RFA (**Figure 2**). **Figure 3A-B** shows the non-cumulative and cumulative regrowth rates over the 5 years of follow-up.

The vast majority of patients (82%) did not receive any further treatment after the first thyroid ablation, while 18% (72/406) underwent a second procedure. In particular, in the RFA group 12% of patients (26/216) were retreated, while in the LA group 24.2% of patients (46/190) were retreated ($p<0.001$). In terms of type of retreatment, 43/406 patients were operated on (11%), 13 patients (3%) underwent a second LA, 10 patients (2%) underwent a second RFA, 2 patients (0.5%) were

treated with radioiodine, 2 patients underwent a second RFA and surgery, while 1 patient (0.25%) underwent ethanol injection, and 1 patient underwent a second LA and surgery (**Figure 3C)**.

Patients who underwent a second ablation exhibited a median nodule volume of 12.50 (3.00-114.0) mL before the retreatment, which was reduced to 6.80 (1.49-40.8) mL after 1 year from the retreatment, with a median volume reduction of 44%. **Figure 3D** shows nodule volume reductions of every single patient at further time points after retreatment.

Comparison between RFA and LA with propensity score matching analysis

After propensity score matching analysis we selected 76 patients treated with RFA and 76 patients treated with LA, who did not differ in terms of age, sex, baseline volume, nodule structure, nodule function (**Table 2**). It was impossible to match the two groups in terms of delivered energy, because LA is associated with a significantly lower amount of energy delivery (due to a more rapid energy/heat decay around the thermal source). **Figure 4** reports the volume reduction ratios after RFA and LA (p<0.001 for technique), before and after propensity score matching. Both procedures significantly reduced nodule volume (p<0.001 vs baseline), but nodule volume reduction after RFA was greater than after LA (p=0.019) (**Figure 5A**). Specifically, after propensity score matching, in the patients treated with RFA (=76), thyroid nodule volume decreased by 72%, 74%, 75%, 75%, and 75% at 1, 2, 3, 4, and 5 years after the first ablation. In the patients treated with LA (=76), thyroid nodule volume decreased by 54%, 57%, 55%, 55%, and 56%, at 1, 2, 3, 4, and 5 years after the first ablation. RFA was associated with a greater rate of technique efficacy (p=0.001) (**Figure 5B**), with a significantly lower percentage of regrowth (p=0.016) (**Figure 5C**), and a significantly lower percentage of retreatments (p=0.01) as compared to LA (**Figure 5D**). Also after propensity score matching there was a good correlation between 1-year and –year volume reduction, which was more pronounced after RFA, being ρ=0.79, p<0.001 in the RFA group and ρ=0.69 p<0.001 in the LA group.

Cumulative incidence of regrowth and retreatment

Given that regrowth and retreatment are time-dependent events, we assessed their cumulative incidence over time. When looking at regrowth, we calculated the cumulative incidence of regrowth in the presence of retreatment as a competing risk (i.e. an event precluding the occurrence of regrowth). The estimated cumulative incidence rates of regrowth in the entire patient cohort are reported in **Figure 6A-B**. The Fine and Gray competing risk regression model showed that energy delivered was the only parameter that was independently associated with the risk of regrowth (**Table 4**). When looking at the cumulative incidence of regrowth (and retreatment as competing event) for type of treatment, we found that RFA was associated with a significantly lower risk of regrowth as compared to LA (p<0.001 Gray Test), while there were no differences in terms of retreatment (p=0.08), these differences remained significant also after propensity score matching (**Figure 6C**). When looking at retreatment, we used the Kaplan-Meier estimates of "not being retreated" intervals and Cox proportional hazard models, to describe retreatment trends and identify significant prognostic factor. There was a significant difference in the risk of being retreated between RFA and LA (after propensity score matching), as in the RFA group more patients did not need retreatment (p<0.01) (**Figure 6D-E**). Multivariate Cox model showed that young age, greater

baseline volume, lower energy delivery, lower technique efficacy, and regrowth were all significantly associated with the risk of being retreated (**Table 4**).

Risk factors of technique inefficacy, regrowth, and retreatment and their cut-offs

Consistent with HR (**Table 4**), after logistic regression models, a lower amount of energy delivered per mL of tissue was the only parameter that was significantly associated with technique inefficacy and regrowth (**Supplementary Table 1**). By contrast, younger age, greater baseline volume, lower amount of energy, technique inefficacy, and regrowth were all significantly and independently associated with the likelihood of being retreated (**Supplementary Table 1**).

Taking into account logistic regression model results, ROC curves were designed to evaluate the accuracy of baseline parameters independently associated with technique efficacy, regrowth, and retreatment. Unfortunately, energy delivered had a poor accuracy as a predictor of regrowth. When looking at technique efficacy, we found that energy delivered had an AUC of 0.65 (0.59, 0.72) and the cut-off value best predicting technique efficacy was 566.06 J/mL (sensitivity =0.72; specificity =0.56). After technique stratification, only the energy delivered by RFA had a moderate accuracy to predict technique efficacy with an AUC of 0.72 (0.60, 0.83) and a cut-off value of 1360.45 J/mL (p=0.01). When looking at retreatments, baseline volume had an AUC of 0.63 (0.56, 0.70), which increased to 0.68 (0.57, 0.79) in the RFA group and to 0.67 (0.58, 0.76) in the LA group. Baseline volume cut-offs best predicting retreatment were 22.1 mL for RFA and 14.5 mL for LA. On the other hand, the 1-year volume reduction resulted moderately accurate to predict retreatment with an AUC of 0.79 (0.74, 0.85) and a cut-off corresponding to a 58% reduction. After technique stratification, the 1-year volume reduction after RFA had an AUC of 0.82 (0.73, 0.91) and a cut-off best predicting retreatment of 66%. Likewise, the 1-year volume reduction after LA had an AUC of 0.74 (0.66, 0.88) and a cut-off best predicting retreatment of 54%. Last, delivered energy had an AUC of 0.70 and its cut-off value best predicting retreatment was 556.5 J/mL (sensitivity =0.82; specificity =0.55). After technique stratification, only the energy delivered by RFA had a good accuracy to predict retreatment, with an AUC of 0.83 (0.75, 0.92) and a cut-off value of 918.37 J/mL ($p<0.001$).

Risk of overlooking non-benign pathology

A total of 46/406 patients (11%) were operated on during follow-up. Final histologic diagnosis showed benign nodules in 27/46 patients (59%), non-benign pathology in 16/46 patients (35%), while in 3 patients final pathology results went missed (6%). Non-benign pathology included: 6 incidental microcarcinomas outside the ablated nodule, 4 follicular carcinomas, 3 papillary carcinomas, 3 follicular tumors of uncertain malignant potential. When looking at the entire patient cohort, non-benign pathology was found in 16/406 patients (3.9%), and excluding microcarcinomas it was found in 10/406 (2.4%). In all centres except one, patients underwent 2 FNAB for cytology assessment. Be it one or two assessments, one FNAB cytology was always assessed in the year before the procedure. Of note, there were no differences in the rate of non-benign pathology among the patients who underwent one FNAB (4/103) and two FNAB (12/303) (p=0.99).

The odds ratios of malignancy for age, sex, baseline volume, 1-year volume reduction, nodule structure, success, regrowth, and energy delivered, showed that only male sex was associated with a greater risk of malignancy, as shown in **Supplementary Table 2**.

Looking at the volume reduction of the 16 cases with non-benign pathology, we noticed that most patients had been retreated after 1 year, and the only aspect that could be compared to the other nodules was the 1-year volume reduction. We analyzed the median 1-year volume reduction of patients who did not require further treatments (n=334), the median 1-year volume reduction of patients who were operated on and were found having a benign nodule (n=27), and of patients who were found having a non-benign pathology (n=16). The 1-year volume reduction resulted 67%, 46%, and 27%, respectively ($p<0.001$ for all groups). A ROC analysis was performed to evaluate if the 1-year volume reduction could be a predictive marker of non-benign pathology in patients treated with thermal ablations. The ROC curve showed that the 1-year volume reduction had an AUC of 0.823 (95% CI) and its cut-off value best predicting non-benign pathology was 20% (sensitivity = 50%; specificity = 98%), according to the maximum of the Youden Index. Consistent with these results, when the ROC analysis was repeated excluding microcarcinomas, the ROC curve showed that the 1-year volume reduction had a AUC of 0.853 (95% CI) and the cut-off value best predicting malignancy was still 21% (sensitivity =50%; specificity = 98%).

2.4 DISCUSSION

Several short-term studies have demonstrated that US-guided thermal ablation is a safe and clinically effective procedure for the treatment of benign thyroid nodules that become symptomatic. Only few studies with extended follow-up (i.e. 5 years), however, addressed the issue of nodule regrowth and need of retreatment (17, 24-26). So, this is the first multicenter study enrolling patients who were followed for five consecutive years after a single session of RFA and/or LA.

Nodule volume decrease and technique efficacy. Consistent with previous reports, a single session of RFA or LA significantly reduced thyroid nodule volume and this result was substantially maintained during a five-year period (27). In our study nodule volume reduction after RFA was lower than in a few former trials (25) that reported a 89% and 90% nodule volume decrease at 1 and 3 years, respectively. Notably, in these previous studies, only part of the patients completed the 5-year follow-up (follow-up range was 36-81 months) and, most importantly, they were treated on average twice (mean number of session was 2.2 ± 1.4) (25). Conversely, our volume reductions after RFA are similar to those reported by Sim (17), who found a volume reduction of 77%, and by Deandrea (26), who found a volume reduction of 70% after the first RFA session. As for LA, in our study, thyroid nodule volume decreased by 55%, 58%, 59%, 57%, and 57%, at 1, 2, 3, 4, and 5 years after a single ablation, which is line with the percentages reported by Papini (8) and Dossing (24).

Treatment efficacy vs delivered energy. Technique efficacy was achieved in 74% of patients and was significantly associated with the delivered energy. The energy cut-off best predicting technique efficacy was 566 J/mL. Although it had a poor accuracy, this cut-off is consistent with previous data by Gambelunghe (28) and De Freitas. Of note, the accuracy of the energy cut-off increased after technique stratification only for RFA, where energy cut-off was 1360 J/mL. Propensity score matching showed that technique efficacy was achieved more frequently in patients treated with RFA (82%) than in those treated with LA (66%) possibly because RFA was associated with a greater amount of energy delivered. This variability could be due to the different modalities of action of RFA and LA, which are not only two operator-dependent techniques, but they have also specific modalities of production and distribution of thermal energy (22, 29). For instance, laser technology directs high-level energy to a well-delimited area of tissue, heat deposition is greatest near the thermal source with a rapid energy/heat decay in the surrounding tissue (Ritz J Lasers in

Surgery and Medicine 2009 479). When performing RFA with monopolar electrodes, which are the ones that we have used, the patient is part of a closed-loop circuit that includes the radiofrequency generator, the electrode needle, and a large dispersive electrode (ground pads), such that heat is distributed in a larger area of surrounding tissue (Goldberg 1995 pp399).

RFA vs LTA outcomes. The direct comparison of the two techniques was assessed in recent studies reaching differing conclusions (27, 29-32). Our results are consistent with the conclusions of two metanalyses and the only randomized controlled trial comparing these techniques. Ha et al. reported that RFA was more effective than LA in terms of volume reduction after 6 months from the procedure (77.8% vs 49.5% after RFA and LA, respectively) (30). Trimboli et al. similarly reported that volume reduction after 1, 2, and 3 years was 68%, 75%, and 87% with RFA as compared to 48%, 52%, and 45% with LA (27). Finally, in the only randomized controlled trial comparing these two treatment modalities, technique efficacy was achieved in 86.7% of patients treated with RFA as compared to 66.7% of patients treated with LA (32) and RFA was associated with a significantly greater nodule volume reduction after 6 months (64.3% vs 53.2% with LA) (32). These data appear consistent with our results.

Long-term nodule regrowth. Nodule regrowth occurred in 28% of patients. Nodules regrowth rate increased progressively over time. Importantly, in our study nodule regrowth did not always represent a problem on clinical grounds, as symptom recurrence occurred in 26% of cases of regrowth, and a second treatment was requested in 28% of patients whose nodules regrew. Our results are similar to those of Sim and colleagues who reported a regrowth in 24% of the nodules, mostly after 2-4 years of follow-up. Nevertheless, it is this difficult to compare our data to those of other Authors, due to the different definitions used (8, 33) and the significant patient loss at follow-up (17, 25, 26) of their studies. Although we found a good correlation between 1-year and 5-year volume reduction, odds ratio assessment demonstrated that the only variable significantly associated with nodule regrowth after thermoablation was the quantity of delivered energy. However, given that energy was a poorly accurate predictor of regrowth, our findings suggest that nodule regrowth may be associated, not only to energy delivery (35), but also to the type of technique (34), as RFA resulted in a significantly lower regrowth rate (17%) as compared to LA (34%). Consistent with these rates, cumulative risk curves showed that RFA had a significantly lower risk of regrowth over time. One of the reasons accounting for this difference could be that RFA is performed by sequentially moving the tip of the electrode across the entire nodule area, which allows the tailoring of the procedure to the variable features of the nodules, maximizing the ablation of the marginal areas of the lesion. The undertreatment of nodule margins (34) and the nodule structure (specifically, solid vs spongiform structure) (36), together with other minor biological characteristics (24) are additional factors that could account for nodule regrowth.

Risk factors of nodule retreatment. In our study, the vast majority of patients did not require multiple treatments, as only 18% of them underwent a second procedure over the 5 years of follow-up. LA was associated with a significantly higher rate of retreatments as compared to RFA (32% vs 14%, respectively). The rate of retreatments after LA is consistent with the rate reported by Dossing, which was 35% (24). Consistent with this finding, Kaplan-Meier curves showed that patients with RFA were more likely not to be retreated over time. Retreatments were more likely to happen in young patients, in larger nodules, in patients with lower 1-year volume reduction, and when delivered energy was low (37).

Importantly, this real world study provided a few relevant cut-offs for the prediction of retreatment. Specifically, the baseline volume cut-off that best predicted the need of retreatment was 22 mL after

RFA and 14.5 mL after LA. This is consistent with data from a few previous trials, stating that nodules larger than 20 mL generally require more than one session (6, 25) and that in nodules larger than 20 mL the results might not be as satisfactory as surgery (38). The 1-year volume reduction cut-off that best predicted retreatment was a reduction <66% after RFA and a reduction <54% after LA. As for energy delivered, the cut-off best predicting retreatment was 556 J/mL, and it improved in accuracy after technique stratification, changing to 918 J/mL after RFA.

Risk of overlooking malignancy. Thyroid surgery represented 60% of the retreatments (46/406 patients) and 16/46 (35%) of these patients resulted to have non-benign pathology at histologic examination (3.9% of all the treated patients and 2.4% if we excluded microcarcinomas). It has to be taken into account that although thermal ablations should not affect substantially pathology results in case of a thyroid carcinoma (39), spots of invasions as well as microcarcinomas within the ablated area might be no longer found (40). Male sex was significantly associated with the risk of non-benign pathology and, importantly, most patients with non-benign pathology did not achieve technique efficacy and were retreated after 1 year from the first ablation. Due to the timing of surgery, we could not observe an association between regrowth and non-benign pathology. ROC analysis showed that a nodule volume decrease less than 20% after 1 year was a predictive factor of the risk of non-benign pathology. So, for patients whose nodule decrease is less than 20% after thermal ablation, a repeat cytological assessment and, possibly, surgery appear more appropriate than a repeat thermal ablation procedure.

Main limitations of the present study are its retrospective design, and the collection of data from different centers with possible selection bias. In addition, the procedures were performed by different operators, which has to be taken into account as thermal ablation is an operator-dependent technique. Despite these limitations, due to its multicenter design, our data provide a real world assessment of thermal ablation outcomes. In particular, this is the first follow-up study where all the patients were followed entirely for 5 years, allowing us to report cumulative risk of regrowth and retreatment over time, as well as hazard ratios, not only for RFA but also for LA.

In conclusion, both RFA and LA induce a clinically relevant volume reduction of benign thyroid nodules that persists several years after the procedure. Technique efficacy is achieved in the vast majority of patients, if energy delivery is above the observed cut-offs. Regrowth occurs in one third of patients but in the majority of cases does not require a further treatment. Retreatments are more likely in young patients, in larger nodules, and in patients with a low 1-year volume reduction. RFA is associated with a lower risk of regrowth and retreatment as compared to LA, which may be due to the different amount of energy released inherent to the technique. Finally, a VRR ≤20% after one year should raise suspicion of an underlying malignancy and prompt for FNAB.

2.5 REFERENCES

1. Papini E, Gugliemi R, Pacella CM 2016 Laser, radiofrequency, and ethanol ablation for the management of thyroid nodules. Curr Opin Endocrinol Diabetes Obes 23:400-406.
2. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, Paschke R, Valcavi R, Vitti P, Nodules AAATFoT 2016 American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules--2016 Update. Endocr Pract 22:622-639.
3. Kim JH, Baek JH, Lim HK, Ahn HS, Baek SM, Choi YJ, Choi YJ, Chung SR, Ha EJ, Hahn SY, Jung SL, Kim DS, Kim SJ, Kim YK, Lee CY, Lee JH, Lee KH, Lee YH, Park JS, Park H, Shin JH, Suh CH, Sung JY, Sim JS, Youn I, Choi M, Na DG, Guideline Committee for the Korean Society of Thyroid R, Korean Society of R 2018 2017 Thyroid Radiofrequency Ablation Guideline: Korean Society of Thyroid Radiology. Korean J Radiol 19:632-655.
4. Papini E, Pacella CM, Solbiati LA, Achille G, Barbaro D, Bernardi S, Cantisani V, Cesareo R, Chiti A, Cozzaglio L, Crescenzi A, De Cobelli F, Deandrea M, Fugazzola L, Gambelunghe G, Garberoglio R, Giugliano G, Luzi L, Negro R, Persani L, Raggiunti B, Sardanelli F, Seregni E, Sollini M, Spiezia S, Stacul F, Van Doorne D, Sconfienza LM, Mauri G 2019 Minimally-invasive treatments for benign thyroid nodules: a Delphi-based consensus statement from the Italian minimally-invasive treatments of the thyroid (MITT) group. Int J Hyperthermia 36:376-382.
5. Gharib H, Hegedus L, Pacella CM, Baek JH, Papini E 2013 Clinical review: Nonsurgical, image-guided, minimally invasive therapy for thyroid nodules. J Clin Endocrinol Metab 98:3949-3957.
6. Cesareo R, Pasqualini V, Simeoni C, Sacchi M, Saralli E, Campagna G, Cianni R 2015 Prospective study of effectiveness of ultrasound-guided radiofrequency ablation versus control group in patients affected by benign thyroid nodules. J Clin Endocrinol Metab 100:460-466.
7. Deandrea M, Sung JY, Limone P, Mormile A, Garino F, Ragazzoni F, Kim KS, Lee D, Baek JH 2015 Efficacy and Safety of Radiofrequency Ablation Versus Observation for Nonfunctioning Benign Thyroid Nodules: A Randomized Controlled International Collaborative Trial. Thyroid 25:890-896.
8. Papini E, Rago T, Gambelunghe G, Valcavi R, Bizzarri G, Vitti P, De Feo P, Riganti F, Misischi I, Di Stasio E, Pacella CM 2014 Long-term efficacy of ultrasound-guided laser ablation for benign solid thyroid nodules. Results of a three-year multicenter prospective randomized trial. J Clin Endocrinol Metab 99:3653-3659.
9. Kim C, Lee JH, Choi YJ, Kim WB, Sung TY, Baek JH 2017 Complications encountered in ultrasonography-guided radiofrequency ablation of benign thyroid nodules and recurrent thyroid cancers. Eur Radiol 27:3128-3137.
10. Pacella CM, Mauri G, Achille G, Barbaro D, Bizzarri G, De Feo P, Di Stasio E, Esposito R, Gambelunghe G, Misischi I, Raggiunti B, Rago T, Patelli GL, D'Este S, Vitti P, Papini E 2015 Outcomes and Risk Factors for Complications of Laser Ablation for Thyroid Nodules: A Multicenter Study on 1531 Patients. J Clin Endocrinol Metab 100:3903-3910.

11. Nardi F, Basolo F, Crescenzi A, Fadda G, Frasoldati A, Orlandi F, Palombini L, Papini E, Zini M, Pontecorvi A, Vitti P 2014 Italian consensus for the classification and reporting of thyroid cytology. Journal of endocrinological investigation 37:593-599.
12. Cibas ES, Ali SZ, Conference NCITFSotS 2009 The Bethesda System For Reporting Thyroid Cytopathology. American journal of clinical pathology 132:658-665.
13. Mauri G, Pacella CM, Papini E, Sconfienza LM, Solbiati L 2018 Proceedings of the first Italian conference on thyroid minimally invasive treatments and foundation of the Italian research group for thyroid minimally invasive procedures. Int J Hyperthermia 34:603-605.
14. American Thyroid Association Guidelines Taskforce on Thyroid N, Differentiated Thyroid C, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM 2009 Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 19:1167-1214.
15. Mauri G, Pacella CM, Papini E, Solbiati L, Goldberg SN, Ahmed M, Sconfienza LM 2019 Image-Guided Thyroid Ablation: Proposal for Standardization of Terminology and Reporting Criteria. Thyroid : official journal of the American Thyroid Association 29:611-618.
16. Ahmed M, Solbiati L, Brace CL, Breen DJ, Callstrom MR, Charboneau JW, Chen MH, Choi BI, de Baere T, Dodd GD, 3rd, Dupuy DE, Gervais DA, Gianfelice D, Gillams AR, Lee FT, Jr., Leen E, Lencioni R, Littrup PJ, Livraghi T, Lu DS, McGahan JP, Meloni MF, Nikolic B, Pereira PL, Liang P, Rhim H, Rose SC, Salem R, Sofocleous CT, Solomon SB, Soulen MC, Tanaka M, Vogl TJ, Wood BJ, Goldberg SN, International Working Group on Image-guided Tumor A, Interventional Oncology Sans Frontieres Expert P, Technology Assessment Committee of the Society of Interventional R, Standard of Practice Committee of the C, Interventional Radiological Society of E 2014 Image-guided tumor ablation: standardization of terminology and reporting criteria--a 10-year update. Radiology 273:241-260.
17. Sim JS, Baek JH, Lee J, Cho W, Jung SI 2017 Radiofrequency ablation of benign thyroid nodules: depicting early sign of regrowth by calculating vital volume. International journal of hyperthermia : the official journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group 33:905-910.
18. Kim HT 2007 Cumulative incidence in competing risks data and competing risks regression analysis. Clin Cancer Res 13:559-565.
19. Scrucca L, Santucci A, Aversa F 2007 Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant 40:381-387.
20. Pacella CM, Bizzarri G, Guglielmi R, Anelli V, Bianchini A, Crescenzi A, Pacella S, Papini E 2000 Thyroid tissue: US-guided percutaneous interstitial laser ablation-a feasibility study. Radiology 217:673-677.
21. Kim YS, Rhim H, Tae K, Park DW, Kim ST 2006 Radiofrequency ablation of benign cold thyroid nodules: initial clinical experience. Thyroid 16:361-367.
22. Baek JH, Lee JH, Valcavi R, Pacella CM, Rhim H, Na DG 2011 Thermal ablation for benign thyroid nodules: radiofrequency and laser. Korean journal of radiology 12:525-540.

23. Pacella CM, Bizzarri G, Spiezia S, Bianchini A, Guglielmi R, Crescenzi A, Pacella S, Toscano V, Papini E 2004 Thyroid tissue: US-guided percutaneous laser thermal ablation. Radiology 232:272-280.
24. Dossing H, Bennedbaek FN, Hegedus L 2011 Long-term outcome following interstitial laser photocoagulation of benign cold thyroid nodules. Eur J Endocrinol 165:123-128.
25. Lim HK, Lee JH, Ha EJ, Sung JY, Kim JK, Baek JH 2013 Radiofrequency ablation of benign non-functioning thyroid nodules: 4-year follow-up results for 111 patients. European radiology 23:1044-1049.
26. Deandrea M, Trimboli P, Garino F, Mormile A, Magliona G, Ramunni MJ, Giovanella L, Limone PP 2019 Long-Term Efficacy of a Single Session of RFA for Benign Thyroid Nodules: A Longitudinal 5-Year Observational Study. The Journal of clinical endocrinology and metabolism 104:3751-3756.
27. Trimboli P, Castellana M, Sconfienza LM, Virili C, Pescatori LC, Cesareo R, Giorgino F, Negro R, Giovanella L, Mauri G 2020 Efficacy of thermal ablation in benign non-functioning solid thyroid nodule: A systematic review and meta-analysis. Endocrine 67:35-43.
28. Gambelunghe G, Fede R, Bini V, Monacelli M, Avenia N, D'Ajello M, Colella R, Nasini G, De Feo P 2013 Ultrasound-guided interstitial laser ablation for thyroid nodules is effective only at high total amounts of energy: results from a three-year pilot study. Surg Innov 20:345-350.
29. Pacella CM, Mauri G, Cesareo R, Paqualini V, Cianni R, De Feo P, Gambelunghe G, Raggiunti B, Tina D, Deandrea M, Limone PP, Mormile A, Giusti M, Oddo S, Achille G, Di Stasio E, Misischi I, Papini E 2017 A comparison of laser with radiofrequency ablation for the treatment of benign thyroid nodules: a propensity score matching analysis. Int J Hyperthermia 33:911-919.
30. Ha EJ, Baek JH, Kim KW, Pyo J, Lee JH, Baek SH, Dossing H, Hegedus L 2015 Comparative efficacy of radiofrequency and laser ablation for the treatment of benign thyroid nodules: systematic review including traditional pooling and bayesian network meta-analysis. J Clin Endocrinol Metab 100:1903-1911.
31. Mauri G, Cova L, Monaco CG, Sconfienza LM, Corbetta S, Benedini S, Ambrogi F, Milani V, Baroli A, Ierace T, Solbiati L 2017 Benign thyroid nodules treatment using percutaneous laser ablation (PLA) and radiofrequency ablation (RFA). Int J Hyperthermia 33:295-299.
32. Cesareo R, Pacella CM, Pasqualini V, Campagna G, Iozzino M, Gallo A, Lauria Pantano A, Cianni R, Pedone C, Pozzilli P, Taffon C, Crescenzi A, Manfrini S, Palermo A 2020 Laser Ablation versus Radiofrequency Ablation for benign non-functioning thyroid nodules: Six-month results of a randomised, parallel, open-label, trial (LARA trial). Thyroid.
33. Valcavi R, Riganti F, Bertani A, Formisano D, Pacella CM 2010 Percutaneous laser ablation of cold benign thyroid nodules: a 3-year follow-up study in 122 patients. Thyroid 20:1253-1261.
34. Sim JS, Baek JH 2019 Long-Term Outcomes Following Thermal Ablation of Benign Thyroid Nodules as an Alternative to Surgery: The Importance of Controlling Regrowth. Endocrinol Metab (Seoul) 34:117-123.
35. de Freitas RMC, Miazaki AP, Tsunemi MH, de Araujo Filho VJF, Marui S, Danilovic DLS, Buchpiguel CA, Chammas MC 2019 Laser Ablation of Benign Thyroid Nodules: A

Prospective Pilot Study With a Preliminary Analysis of the Employed Energy. Lasers Surg Med.
36. Negro R, Greco G 2019 Unfavorable Outcomes in Solid and Spongiform Thyroid Nodules Treated with Laser Ablation. A 5-Year Follow-up Retrospective Study. Endocr Metab Immune Disord Drug Targets 19:1041-1045.
37. Sim JS, Baek JH, Cho W 2018 Initial Ablation Ratio: Quantitative Value Predicting theTherapeutic Success of Thyroid Radiofrequency Ablation. Thyroid 28:1443-1449.
38. Bernardi S, Dobrinja C, Fabris B, Bazzocchi G, Sabato N, Ulcigrai V, Giacca M, Barro E, De Manzini N, Stacul F 2014 Radiofrequency ablation compared to surgery for the treatment of benign thyroid nodules. Int J Endocrinol 2014:934595.
39. Dobrinja C, Bernardi S, Fabris B, Eramo R, Makovac P, Bazzocchi G, Piscopello L, Barro E, de Manzini N, Bonazza D, Pinamonti M, Zanconati F, Stacul F 2015 Surgical and Pathological Changes after Radiofrequency Ablation of Thyroid Nodules. Int J Endocrinol 2015:576576.
40. Valcavi R, Piana S, Bortolan GS, Lai R, Barbieri V, Negro R 2013 Ultrasound-guided percutaneous laser ablation of papillary thyroid microcarcinoma: a feasibility study on three cases with pathological and immunohistochemical evaluation. Thyroid 23:1578-1582.

2.6 TABLES

Table 1. Centres, techniques, year of treatment, and number of patients included and excluded.

Centre	Technique	Patient cohorts (years of treatment)	End of 5-year follow-up	Patients included (n=406)	Patients excluded (n=59)
Milano	RFA	2014	2019	17	1
Genova	RFA	2012-2014	2017-2019	19	12
Napoli	RFA	2013	2018	24	0
Lecce	LA	2009-2014	2014-2019	87	26
Latina	RFA	2014	2019	37	1
Teramo	LA	2009-2014	2014-2019	103	2
Torino	RFA	2014	2019	40	4
Trieste	RFA	2012-2014	2017-2019	78	13

Table 2. Nodule volumes and nodule volume reduction

	Baseline	1 Year	2 Years	3 Years	4 Years	5 Years
ALL PATIENTS						
Nodule Volume (mL)[§]	14.3	5.2*	4.8*	4.3*	4.2*	4.0*
	(0.4-179.0)	(0.0-242.0)	(0.0-214.0)	(0.0-96.0)	(0.0-88.7)	(0.0-62.0)
Volume Reduction Ratio (%)[§]	-	63.3	67.5	68.3	68.7	70.4
		(-50.0;99.7)	(-80.4;99.9)	(-63.5;1.0)	(-54.9;1.0)	(-50.0;1.0)
Number (cumulative) of patients not retreated	406	406	387	363	352	334
Number (cumulative) of patients retreated (%)	0	0	19 (4.7%)	43 (10.3%)	53 (13.1%)	72 (17.7%)
Surgery			*12*	*13*	*4*	*14*
MITT			*5*	*9*	*6*	*4*
MITT + surgery			*2*	*1*		
I-131				*1*		*1*
RFA GROUP						
Nodule Volume (mL)[§]	17.2	4.9*	4.7*	4.4*	4.0*	3.9*
	(0.4-179.0)	(0.0-242.0)	(0.0-214.0)	(0.0-96.0)	(0.0-89.0)	(0.0-62.0)
Volume Reduction Ratio (%)[§]	-	72.4	74.6	75.9	76.3	77.1
		(-35.2-99.7)	(-24.9-99.9)	(-48.2; 1.0)	(-34.5; 1.0)	(-34.5; 1.0)
Number (cumulative) of patients not retreated	216	216	214	203	197	192
Number (cumulative) of patients retreated (%)	0	0	8 (3.7%)	13 (6.0%)	19 (8.8%)	26 (11.1%)
Surgery			*4*	*2*	*2*	*5*
MITT			*2*	*3*	*4*	*2*

MITT + surgery			*2*			
LA GROUP						
Nodule Volume (mL)[§]	12.2	5.5*	4.8*	4.3*	4.2*	4.1*
	(1.7-86.0)	(0.3-52.0)	(0.2-39.0)	(0.2-46.8)	(0.2-39.7)	(0.1-35.0)
Volume Reduction Ratio (%)[§]	-	54.9	58.3	58.8	57.5	56.7
		(-50.0-95.7)	(-80.0-97.0)	(-63.5; 93.8)	(-54.9; 1.0)	(-50.0; 97.8)
Number (cumulative) of patients not retreated	190	190	179	160	155	144
Number (cumulative) of patients retreated (%)	0	0	11 (5.8%)	30 (15.8%)	34 (18.4%)	46 (24.2%)
Surgery			*8*	*11*	*2*	*9*
MITT			*3*	*6*	*2*	*2*
MITT + surgery				*1*		
I-131				*1*		*1*

Nodule volume and volume reduction are presented as Median (Min-Max). § Nodule volume and volume reduction do not include data after retreatments. *p<0.001, Friedman test for repeated measures. MITT is for minimally invasive treatments of the thyroid and include radiofrequency ablation, laser ablation, and ethanol injection.

Table 3. Characteristics of RFA and LA groups after propensity score matching analysis

	RFA (n=76)	*LA (n=76)*
Age (years)	58.5 (33-85)	63.5 (29-78)
F (%)	57 (75.0%)	55 (72.4%)
Baseline volume (mL)	15.9 (1.2-67.0)	17.5 (2.5-86.0)
Solid nodules (%)	76 (100.0%)	76 (100.0%)
Non-functioning nodules (%)	63 (82.9%)	73 (96.1%)
Energy/volume (J/mL)	1397.9 (175.6-2409.8	348.1 (61.0-1100.4) *

*p<0.05

Table 4. Fine-Gray competing risk and Cox proportional hazard regression models

		FINE-GRAY COMPETING RISK REGRESSION MODEL REGROWTH			
		Univariate CRR model		**Multivariate CRR model**	
		HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Age (years)		0.98 (0.97-0.99)	0.008*	0.99 (0.98-1.01)	0.60
Sex	M	1.00 (ref)		1.00 (ref)	
	F	1.53 (0.96-2.43)	0.09	1.41(0.84-2.36)	0.19
Baseline volume (mL)		0.99 (0.97-1.00)	0.09	0.99 (0.98-1.01)	0.36
Nodule structure	S	1.00 (ref)			
	PS	0.70 (0.40-1.19)	0.18	//	//
	PC/C	0.94 (0.45-1.93)	0.86		
Nodule function	AFTN	1.00 (ref)		1.00 (ref)	
	Non-AFTN	2.72 (1.02-7.26)	0.04*	2.39 (0.30-18.93)	0.41
1-year reduction (%)		0.76 (0.36-1.60)	0.47	//	//
Energy/volume (J/mL)		0.99 (0.99-1.00)	<0.001*	0.99 (0.99-1.00)	**0.001***

		COX PROPORTIONAL HAZARD REGRESSION MODEL RETREATMENT			
		Univariate Cox model		**Multivariate Cox model**	
		HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Age (years)		0.98 (0.96-1.00)	0.01*	0.98 (0.96-0.99)	**0.01***
Sex	M	1.00 (ref)			
	F	1.01 (0.59-1.71)	0.99	//	//
Baseline volume (mL)		1.017 (1.001-1.002)	<0.001*	1.03 (1.02-1.04)	**<0.001***
Nodule structure	S	1.00 (ref)		1.00 (ref)	
	PS	0.45 (0.20-0.98)	0.04*	0.51 (0.11-2.39)	0.39
	PC/C	0.33 (0.88-1.35)	0.12	0.91 (0.13-6.71)	0.93
Nodule function	AFTN	1.00 (ref)			
	Non-AFTN	1.59 (0.58-4.36)	0.38	//	//
1-year reduction		0.03 (0.02-0.06)	<0.001*	0.04 (0.02-0.09)	**<0.001***
Regrowth	No	1.00 (ref)		1.00 (ref)	
	Yes	2.00 (1.16-3.19)	0.003*	1.68 (0.99-2.87)	**<0.001***
Energy/volume (J/mL)		0.99 (0.99-1.00)	<0.001*	0.99 (0.99-1.00)	**0.04***

Multivariable model was performed inclusind parameters assessed in the univariable analysis with a p-value of less than the prespecified cut-off of 0.10. *p<0.05 AFTN is for autonomously functioning thyroid nodules, S is for solid, C is for cystic, PC is for predominantly cystic, and PS is for predominantly solid.

2.7 FIGURES

Figure 1. Nodule volume reduction

Box plots of nodule volume at baseline and after 1, 2, 3, 4, 5 years from the thermal ablation. In case of a retreatment, volumes after retreatment were not included. Volume reduction over time is statistically significant, $p<0.001$ Friedman Test for repeated measures.

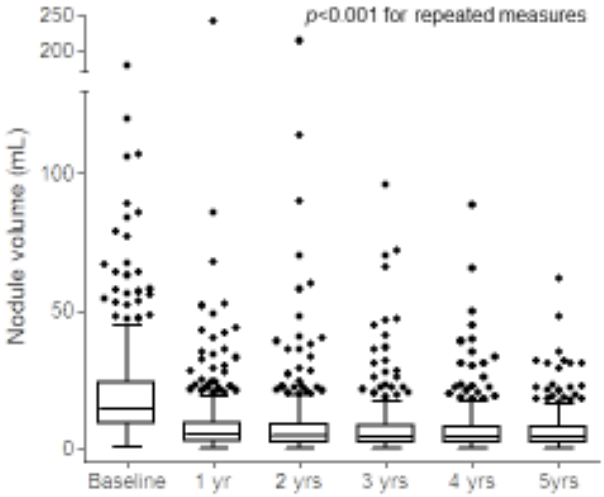


Figure 2. Correlation between 1-year and 5-year volume reduction

Scatter plot for entire patient cohort (A), LA group (B), and RFA group (C).

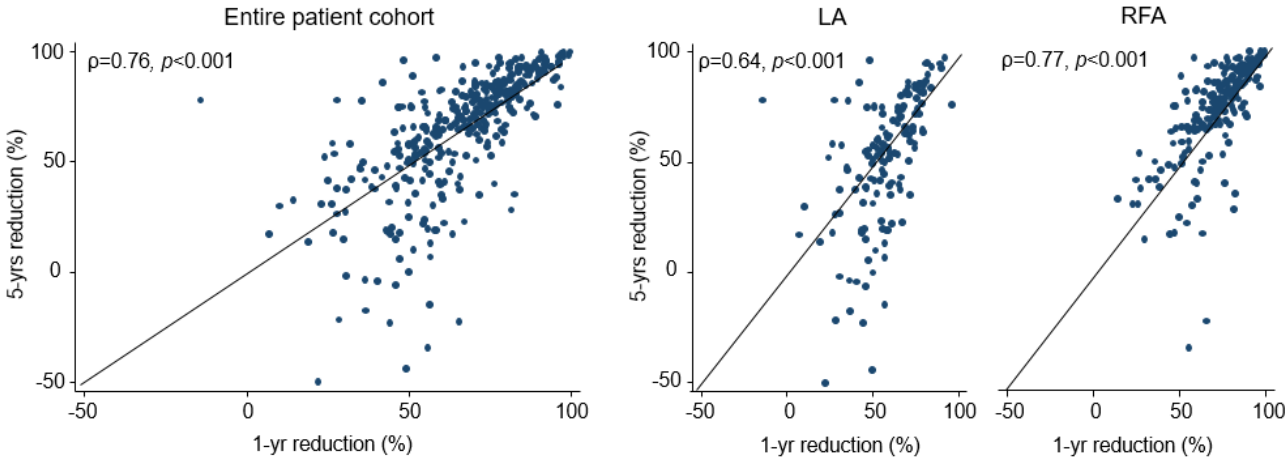


Figure 3. Regrowth rates, type of retreatments, and volume reductions after a second procedure. **A.** Non-cumulative regrowth rate. This figure describes the distribution of the first regrowth over time. In the RFA group regrowth was observed in 19.9% of nodules (43/216). In the LA group regrowth was observed in 37.9% of nodules (72/190). RFA group: 0% nodules (0/216) at 1 year; 6.1% (13/214) at 2 years; 2.5% (5/203) at 3 years; 8.1% (16/197) at 4 years; 4.7% (9/192) at 5 years. LA group: 0.5% nodules (1/190) at 1 year; 8.4% (179) at 2 years; 13.1% (21/160) at 3 years; 11% (17/155) at 4 years; 12.5% (18/144) at 5 years. **B.** Cumulative regrowth rate. This figure describes the distribution of nodule regrowth over time, taking into account that some nodules regrew more than once. **C**. Type of retreatment distribution (%) without the patients non-retreated. **D**. Spaghetti plot showing single patient nodule volume reduction after a second thermal ablation (one patient with an outlayer volume has been excluded). Dotted line is for LA and continuous line is for RFA

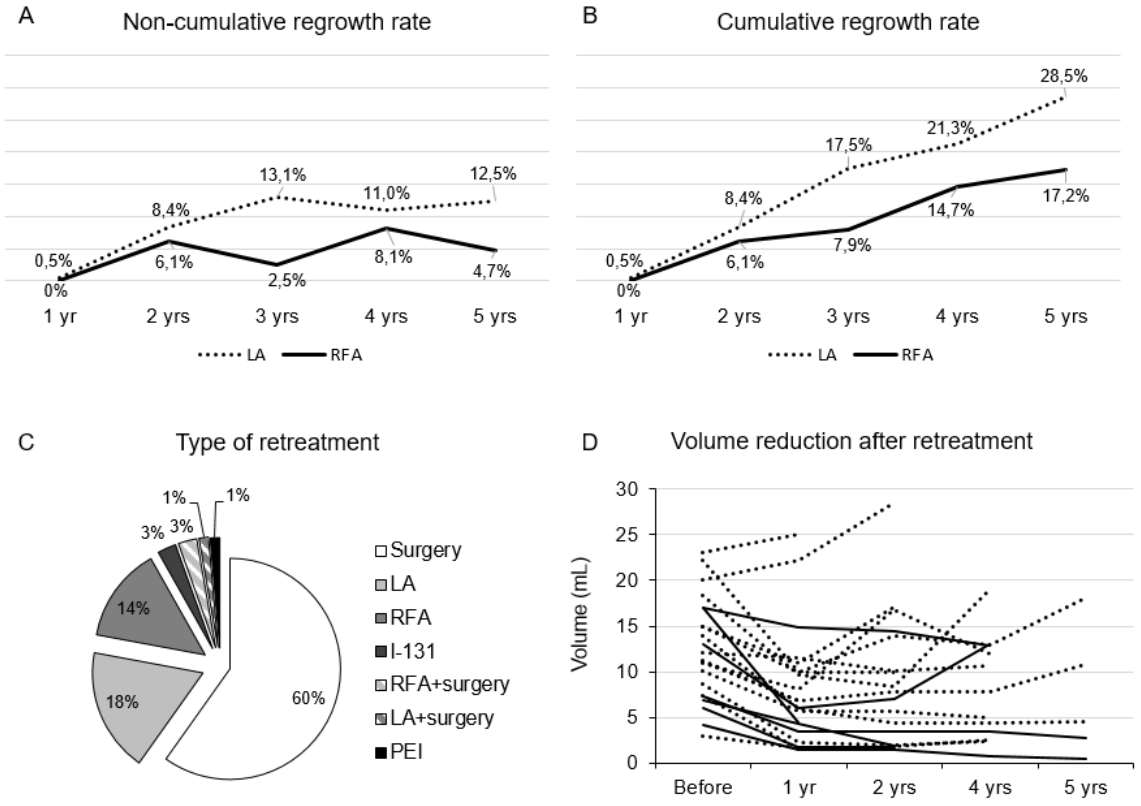


Figure 4. Distribution of volume reduction ratio (%) 1 year after the procedure.

Distribution of volume reduction ratio (%) in RFA and LA groups 1 year after the procedure in the entire patient cohort (A, C), and after propensity score matching (B, D). In both cases, VRR was significantly different in RFA and LA groups, $p<0.001$. PSM is for propensity score matching VRR is for volume reduction ratio.

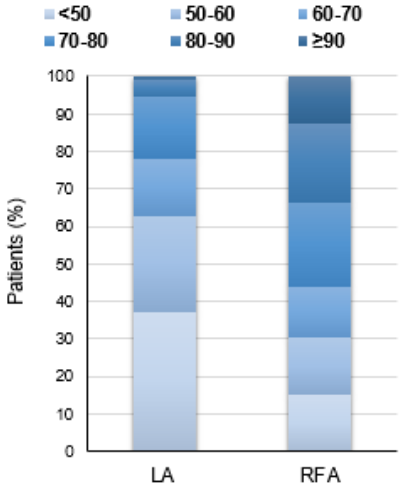
A

1-yr VRR (%)	LA (N=190)	RFA (N=216)	*p*-value
<50 (no treatment efficacy)	71 (37.4%)	33 (15.3%)	<0.001
50-60	48 (25.3%)	33 (15.3%)	
60-70	29 (15.3%)	29 (13.4%)	
70-80	32 (16.8%)	48 (22.2%)	
80-90	8 (4.2%)	46 (21.3%)	
≥90	2 (1.1%)	27 (12.5%)	

B

1-yr VRR (%) after PSM	LA (N=76)	RFA (N=76)	*p*-value
<50 (no treatment efficacy)	32 (42.1%)	14 (18.4%)	<0.001
50-60	19 (25.0%)	13 (17.1 %)	
60-70	9 (11.8%)	7 (9.2%)	
70-80	14 (18.4%)	15 (19.7%)	
80-90	2 (2.6%)	18 (23.7%)	
≥90	0 (0.0%)	9 (11.8%)	

C



D

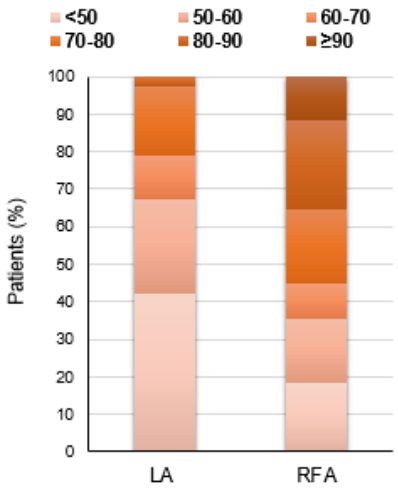


Figure 5. Comparison between RFA and LA after propensity score matching analysis.

Trends of volume reduction after RFA and LA (baseline and 1, 2, 3, 4, and 5 years from the thermal ablation). Both procedures significantly reduced nodule volume (p<0.001 vs baseline), but RFA reduced nodule volume more than LA (p=0.019, Linear Mixed Effect model). **B.** Rate of technique efficacy and inefficacy. **C.** Rate of regrowth **D.** Rate of retreatments.

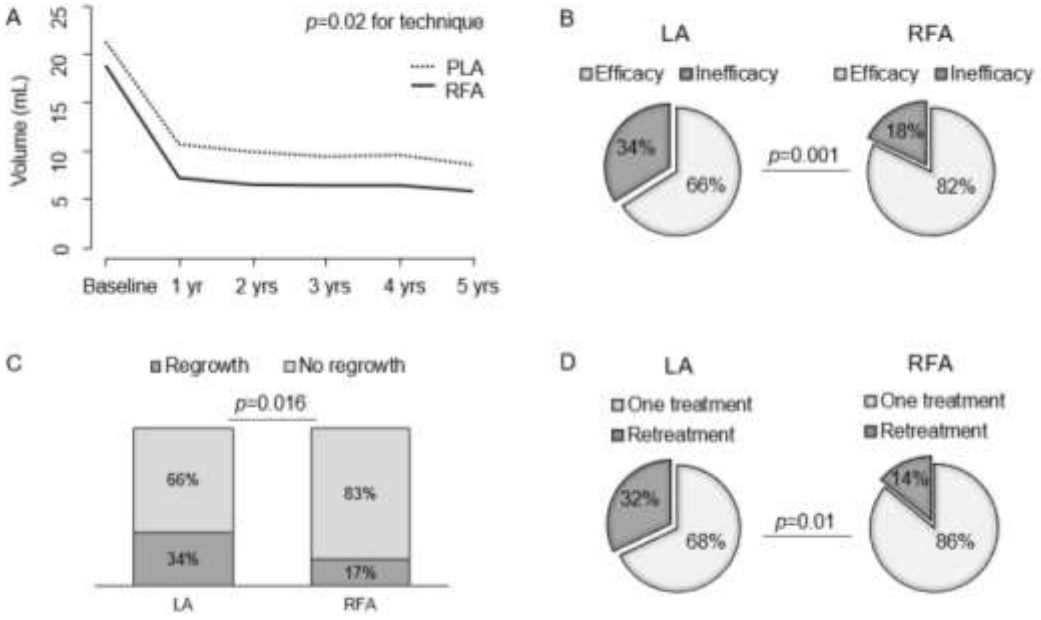
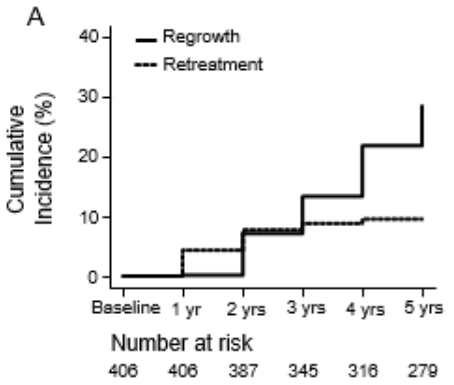


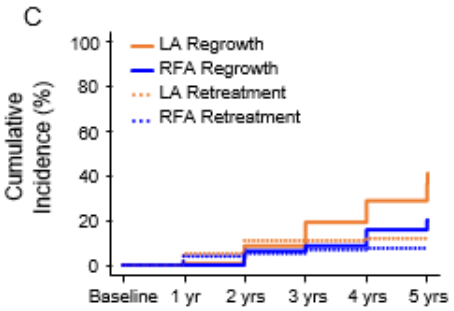
Figure 6. Cumulative incidence of regrowth and retreatment

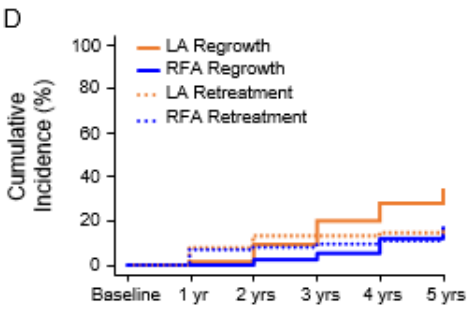
A-B. Cumulative incidence of regrowth and retreatment (as the competing event) with the Competing Risk method. Numbers at risk are the patients who were not retreated. **C.** Cumulative incidence of regrowth and retreatment for RFA and LA in the entire patient cohort. RFA and LA significantly differed in terms of regrowth (p<0.001, Gray Test) but not in terms of retreatment (p=0.08). **D.** Cumulative incidence of regrowth and retreatment for RFA and LA after propensity score matching. RFA and LA significantly differed in terms of regrowth (p<0.01, Gray Test) but not in terms of retreatment (p=0.34). **E-F**. Curves describing no retreatment probability for RFA and LA with the Kaplan Meier method (p=0.01).

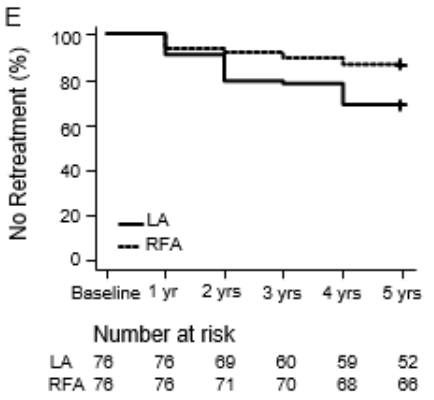


B

Year	Cumulative incidence of Regrowth (95% CI)	Cumulative incidence of Retreatment (95% CI)
1	0.2% (0%-1%)	4% (3%-7%)
2	7% (5%-10%)	8% (6%-11%)
3	13% (10%-17%)	9% (6%-12%)
4	22% (18%-26%)	10% (7%-12%)
5	28% (24%-33%)	10% (7%-13%)







F

Year	RFA (n=76) No Retreatment (95%CI)	LA (n=76) No Retreatment (95% CI)
1	93% (85%-97%)	91% (82%-96%)
2	92% (83%-96%)	79% (68%-87%)
3	90% (80%-95%)	78% (67%-86%)
4	87% (77%-93%)	68% (57%-78%)
5	86%(75%-92%)	68% (57%-78%)

2.8 SUPPLEMENTARY TABLES

SUPPLEMENTARY TABLE S1. CENTERS, TECHNIQUES, CONSECUTIVE ANNUAL COHORTS, AND NUMBER OF PATIENTS INCLUDED AND EXCLUDED

					Patients excluded (n=71)	
Center	*Technique*	*Patient cohorts*	*End of 5-year follow-up*	*Patients included (n=406)*	*Not meeting criteria (n=12)*	*Lost (n=59)*
Genova	RFA	2012–2014	2017–2019	19	0	12
Latina	RFA	2014	2019	37	0	1
Lecce	LA	2009–2014	2014–2019	87	0	26
Milano	RFA	2014	2019	17	0	1
Napoli	RFA	2013	2018	24	0	0
Teramo	LA	2009–2014	2014–2019	103	0	2
Torino	RFA	2014	2019	40	5	4
Trieste	RFA	2012–2014	2017–2019	79	7	13

LA, laser ablation; RFA, radiofrequency ablation.

SUPPLEMENTARY TABLE S2. CHARACTERISTICS OF RADIOFREQUENCY ABLATION AND LASER ABLATION GROUPS AFTER PROPENSITY SCORE MATCHING ANALYSIS

	RFA (n=76)	*LA (n=76)*
Age (range), years	58.5 (33–85)	63.5 (29–78)
Female (%)	57 (75.0)	55 (72.4)
Baseline volume (mL)	15.9 (1.2–67.0)	17.5 (2.5–86.0)
Solid nodules (%)	76 (100.0)	76 (100.0)
Nonfunctioning nodules (%)	63 (82.9)	73 (96.1)
Energy/volume (J/mL)	1397.9 (175.6–2409.8)	348.1 (61.0–1100.4)*

*$p<0.001$ with the Mann–Whitney test.

3. POLY (ADP-RIBOSE) POLYMERASE INHIBITORS IN SOLID TUMOURS: SYSTEMATIC REVIEW AND META-ANALYSIS

ABSTRACT

Background: PARP-inhibitors (PARPi) showed antitumor activity in *BRCA*1/2 mutated cancers, with more heterogeneous outcomes in tumors harboring mutations impairing other genes involved in the DNA homologous recombination repair (HRR) or wild-type (wt).

Methods: We conducted a systematic review and meta-analysis to better assess the role of PARPi for the treatment of metastatic solid tumors, with and without *BRCA*1/2 mutations. The primary endpoint was progression-free survival (PFS). Secondary endpoints were overall response rate (ORR) and overall survival (OS). A random-effect model was applied.

Results: 29 studies (8,839 patients) were included. PFS was significantly improved (HR:0.59, 95%CI:0.51-0.68, p<0.001), without being affected by *BRCA* mutational status (p=0.65). Significant subgroup differences were observed according to tumor site (p=0.001), line of therapy (p=0.002), control arm (p<0.001), type of PARPi (p<0.001) and trials' phase (p=0.006). PARPi were associated with ORR (RR:1.35, 95%CI:1.16–1.56, p<0.001), with significant subgroup differences observed according to treatment line (p=0.03), control arm (p=0.04) and PARPi (p<0.001), and independently from mutational status (p=0.44), tumor site (p=0.86) and trials' phase (p=0.09). OS was significantly improved by PARPi (HR:0.86, 95%CI:0.80–0.92, p<0.001), regardless to mutational status (p=0.57), tumor site (p=0.82), treatment line (p=0.22), control arm (p=0.21), PARPi (p=0.30) and trials' phase (p=0.26). Finally, an exploratory subgroup analysis showed a significant PFS improvement (HR:0.51, 95%CI:0.43–0.60, p<0.001) with PARPi in *BRCA*-wt/HRR deficient (HRD) tumors.

Conclusion: Our results confirm the efficacy of already approved PARPi-based treatments in *BRCA*1/2-mutant solid tumors, support their role also in tumors with *BRCA*-independent HRD and suggest a potentially broader efficacy in some wt tumors, perhaps with appropriate therapeutic partners. Prospective studies are warranted.

Summary of Statistical Methods applied: Exploring heterogeneity in a meta-analysis with quantitative and graphical methods; detailed outlier and influence diagnostics analysis; subgroups analysis and metaregression; publication bias; risk bias assessment

This chapter was published as:

Poly (ADP-ribose) polymerase inhibitors in solid tumours: Systematic review and meta-analysis. Schettini F, ***Giudici F***, Bernocchi O, Sirico M, Corona SP, Giuliano M, Locci M, Paris I, Scambia G, De Placido S, Rescigno P, Prat A, Curigliano G, Generali D. Eur J Cancer. 2021 Apr 13;149:134-152. doi: 10.1016/j.ejca.2021.02.035. Online ahead of print.PMID: 33862496 Review.

3.1 INTRODUCTION

Around 1 in 400 to 800 people harbors a germline pathogenetic variant of *BRCA*1 and/or *BRCA*2 genes (1). These genes are involved in the homologous recombination mechanism of repair (HRR) of DNA double-strand breaks (DSBs), a substantially error-free procedure (2). Inactivation of *BRCA*1/2 due to pathogenetic mutations impairs HRR, thus indirectly inducing an incorrect processing of DSBs through inappropriate and error-prone alternative mechanisms of repair (i.e. nonhomologous end joining (NHEJ) and single-strand annealing (SSA)). This can lead to either a progressive accumulation of DNA lesions, that ultimately induce cell death *via* apoptosis (3, 4), or to an increasing chromosomal instability, cell mutability and subsequent neoplastic transformation(2, 5). Indeed, inactivating germline mutations in *BRCA*1 and 2 significantly increase the risk for early-onset breast cancer (45-65% lifetime risk) in both women and men, and ovarian cancer (15-40%)(1, 6). Proportionally to its prevalence, *BRCA*1 mutations also increase the risk for fallopian tube, peritoneal, testicular, prostate and pancreatic cancer (1, 6, 7). Pathogenic variants in *BRCA*2 are also associated with prostate and pancreatic cancer, as well as melanoma(1, 6, 7). The risk for lung and gastric cancer is also slightly increased in *BRCA*-mutant (mut) patients (7). Overall, the tumors with the highest prevalence of *BRCA*1/2 mutations are epithelial ovarian (10-15%), breast (2-10%), prostate (5-13%) and pancreatic cancer (4-7%)(8–11).

Differently from *BRCA1* and 2, poly (ADP-ribose) polymerases (PARP) is a family of nuclear enzymes involved in the recognition and repair of DNA single-strand breaks (SSBs). When the PARP-dependent mechanism of repair is impaired, DSBs develop. As previously mentioned, in normal cells, DSBs are primarily repaired through HRR. However, when HRR is constitutionally dysfunctional (as in *BRCA*-mutant tumors), if other events that impair DNA damage repair occur, the damage is likely to become permanent, with progressive accumulation of DNA lesions that ultimately lead cells to apoptosis(3, 4). This mechanism is on the basis of the theory of synthetic lethality, which justified the development of PARP inhibitors (PARPi) for the treatment of *BRCA*-deficient tumors(12). Two proof-of-concept phase II studies demonstrated the significant activity and good safety of the oral PARPi olaparib in metastatic breast cancer (MBC) and metastatic ovarian cancer (MOC), paving the way for its further development in those and other solid tumors, along with other PARPi, including talazoparib, niraparib, rucaparib and veliparib (7). This drug class was originally considered to act through the mere inhibition of PARP1/2 by competing with NAD+ for the enzymes' catalytic site (catalytic inhibition)(13, 14). On the contrary, PARPi have been recently demonstrated to elicit synthetic lethality in HRR-deficient cancers mostly by inhibiting PARylation (PAR-mediated dissociation of PARP enzymes from DNA), hence trapping PARP molecules on DNA, especially PARP1 (PARP trapping). In this way, aberrant PARP1-DNA complexes impair DNA replication fork and elicit a cytotoxic effect (15). Actually, this mechanism seems to be the most relevant contribution to synthetic lethality provided by PARPi, with different impact on both efficacy and toxicity(12, 15). Importantly, this mechanism of action might be also on the basis of a PARPi's synergistic effect with other cytotoxic drugs, such as alkylating agents(15).

At present, there are more than 150 trials for multiple PARP inhibitors (e.g. niraparib, olaparib, rucaparib, talazoparib and veliparib) in different stage of development, combined with other drug classes or as single agent (16). Of those, at least 59 in several metastatic cancers have been published so far, including 29 phase II/III randomized controlled trials providing compelling

evidence of efficacy in *BRCA*-mutant (mut) tumors(7, 8, 17–48). Nevertheless, there are still conflicting results with regard to PARPi efficacy in *BRCA*-wild-type (wt) tumors, with or without deleterious mutations occurring in other HRR genes and depending on the PARPi administered. We thus performed a systematic literature review and meta-analysis to more precisely assess the role of PARPi for the treatment of metastatic solid tumors, with or without *BRCA* mutations.

3.2 METHODS

Study Objectives

The objective of our study was to comprehensively evaluate the activity and efficacy of PARPi in metastatic solid tumors, with or without *BRCA*1/2 mutations. Primary endpoint was progression-free survival (PFS), while secondary endpoints were overall response rate (ORR) and overall survival (OS), as *per* US Food and Drug Administration (FDA) Guidance Document(49).**Search Strategy and Selection Criteria**

After a systematic review of the literature conducted in August 2020 on Pubmed®, Cochraine Library and Embase®, we selected all phase II/III randomized clinical trials (RCT) published until 31th July 2020 that studied the activity and/or efficacy of PARPi, combined or not with chemotherapy (CT) or other therapies, in metastatic solid tumors, independently from *BRCA* mutational status(8, 17–47). All other types of studies were excluded, including early-stage trials, because different clinical settings imply different prognosis, therapeutic approaches (e.g. curative surgery, radiotherapy etc.) and endpoints.

We used a query based on the words “parp inhibitors”, “niraparib”, “olaparib”, “talazoparib”, “veliparib”, “rucaparib” and “solid tumors”. The search was conducted by three independent reviewers (SPC, MS, OB) and a fourth reviewer was consulted in case of controversy (FS). No language restrictions were adopted. Some novel or updated results were published between August and December 2020 and were also included (50–52).

Data Extraction

Details concerning study design, patient characteristics, current and previous treatment were extracted from each paper, together with hazard ratios (HR) and associated 95% confidence intervals (CI) for PFS and OS, when reported, and the proportion of patients responding to evaluated treatments in each trials’ arms. These data had to be publicly available or computable from published paper/abstracts, otherwise studies were excluded. Prespecified subgroup analyses for all end-points were performed, independently from the presence of heterogeneity, to highlight any differences between studies.

Some of the included studies provided only results for the intention-to-treat (ITT) population, either *BRCA*-mut, *BRCA*-wt or with unknown *BRCA* status; others, showed subgroup results for *BRCA*-mut patients and/or patients with HRR deficiency (HRD) due to other causes, as well. We collected and analyzed the results for the overall populations included in each study, so to avoid improper comparisons between overall and nested subpopulations in subgroup analyses.

We categorized the studies according to the following subgroups: 1) *BRCA*1/2 mutational status (Mutant vs Mixed/Wild-type); 2) tumor type (breast, ovarian, gastrointestinal, pancreatic, prostate cancer, NSCLC, SCLC and melanoma); 3) control protocol (CT +/- placebo, placebo, enzalutamide or abiraterone, a PARPi without antiangiogenic agent, olaparib inferior dose, bevacizumab); 4) treatment line (1^{st}-line +/- maintenance, $\geq 2^{nd}$-line, Maintenance only); 5) different PARP inhibitor drug (olaparib, talazoparib, veliparib, rucaparib, niraparib); 6) Trials' phase (phase III vs II). Finally, an exploratory analysis on tumors with *BRCA*-independent HRD was carried out.

For the dichotomous variables (ORR), relative risks (RR) with 95% CI were calculated for each study. The time-to-event variables (PFS and OS) were analyzed with HR and 95%CI. The Mantel–Haenszel method and the generic inverse-variance method were used to estimate RR and HR with their 95% CI, respectively. Heterogeneity among the studies was assessed by the χ^2-based Cochran Q statistic and the inconsistency index (I^2 statistic)(53). We preplanned to conduct the analyses using the random-effects (RE) model of DerSimonian and Laird(54). In case of non-significant heterogeneity, a fixed-effects (FE) model was subsequently applied to confirm the result and perform the pre-specified subgroup analyses(54). To further investigate heterogeneity, we used the Baujat plot graphical method (55). To assess the stability of the pooled results, multiple sensitivity analysis (influence analysis) were performed(56). A more extensive explanation is reported in **Supplementary methods**.

Publication bias for each endpoint was explored by visual inspection of funnel plots, Egger's regression test, Begg's test and *trim-and-fill* analysis (57, 58).

Data were analyzed using the R statistical software (version 4.0.2-packages*: meta, metafor, dmetar*) and Revman 5.4 (59, 60). A two-tailed *P*-value$\leq$0.05 was considered statistically significant.

The risk of bias for each trial was assessed according to the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions*(61). Internal validity of eligible studies was assessed according to the Cochrane Collaboration's *Risk of Bias* tool in Review Manager(60).

The study was registered in the Open Science Framework online repository (www.osf.io), with doi:10.17605/OSF.IO/NGY6D.

3.3 RESULTS

Overall 264 records were screened and 29 studies (8,839 patients) met the inclusion criteria (**Figure 1**). Eight studies (27.6%) evaluated PARPi in the first-line setting, with 4 of them also including patients in more advanced lines (second-line and/or further) and 1 studying a subsequent maintenance strategy. Thirteen studies (44.8%) included patients in second-line and/or further and 8 (27.6%) more studies only focused on maintenance after first-line or more advanced. Fifteen (51.7%) studies were phase II RCT, while the other 14 (48.3%) were phase III RCT. Twelve (41.4%) studies included exclusively *BRCA*1/2-mutant patients, 1 (3.4%) study included a cohort *BRCA*-mut and a *BRCA*-wt cohort. All other 16 (55.2%) studies recruited patients independently from their *BRCA* mutational status. Four (13.8%) studies recruited breast cancer patients, 1 (3.4%) colon 1 (3.4%) gastric , 2 (6.9%) small cell lung cancer (SCLC), 1 (3.4%) non-small cell lung

cancer (NSCLC), 1 (3.4%) melanoma, 4 (13.8%) prostate cancer and 15 (51.9%) ovarian cancer (some also including fallopian tube and primary peritoneal cancers). Main study characteristics are reported in **Table 1**. Only 9 (31%) studies reported subgroup analysis according to HRD status. Of these, 2 (22.3%) studies did not provide separate information for *BRCA*-independent and *BRCA*-dependent HRD(29, 44), 1 (11.1%) study reported data on *ATM*-negative gastric tumors(22), 3 (33.3%) studies reported separate data for a subpopulation of HRR-mutant genes different from *BRCA1/2(37, 38, 45)* and 3 (33.3%) studies reported separate data for a subgroup affected by BRCA-wt tumors with HRD detected through the assessment of deleterious mutations in HRR genes, but also characteristic genomic scar signatures (28, 30, 43). All of these studies reported PFS data, while only 3 reported OS(22, 30, 45) and ORR(22, 28, 29) results (**Table 1**).

Primary endpoint: PFS

Overall 26 studies provided data for the PFS analysis, with 30 different comparisons. The pooled effect on PFS was statistically significant, with a considerable improvement provided by the experimental arms, although heterogeneity was high (HR: 0.59, 95%CI: 0.51 - 0.68 p<0.001, I^2=85%) (**Figure 2**).

No significant difference was observed according to patients' mutational status (p=0.65), with a significant pooled effect observed in both mixed/wt cases (HR: 0.61, 95%CI: 0.52 - 0.71) and in *BRCA*-mut patients (HR: 0.56, 95%CI: 0.42 - 0.75). Significant subgroup differences were observed according to tumor site (p=0.001), line of therapy (p=0.002), control arm (p<0.001), type of PARPi (p<0.001) and trials' phase (p=0.006). More specifically, PFS was significantly improved in melanoma, SCLC, ovarian, prostate and pancreatic cancer, while results were non-significant for NSCLC, breast and gastrointestinal cancers (**Table 2**). PARPi improved PFS in all treatment lines, with a more pronounced effect observed in the maintenance setting, followed by second-line or further and first-line (**Table 2**). With respect to treatment comparison, PARPi were also superior to all control arms, with different degree of benefit. The effect was more pronounced over placebo, followed by enzalutamide or abiraterone acetate (Enza/Abi) and CT +/- placebo. Moreover, olaparib + bevacizumab (beva) was superior to beva and PARPi + an antiangiogenic drug (bevacizumab or cediranib) were superior to a PARPi alone (**Table 2**). All PARPi were effective with a more pronounced benefit obtained with rucaparib, followed by niraparib, olaparib, talazoparib and veliparib (**Table 2**). Finally, a significant result was observed in both phase II and III trials, with a more pronounced benefit observed within the latter (**Table 2**).
The main result, as well as numerous subgroup pooled estimates, were affected by significant heterogeneity. According to the Baujat plot, the second comparison of the study from Han et al. 2018 (i.e. "Han (2) 2018"), confronting veliparib + temozolamide (VT) to carboplatin + paclitaxel + placebo (CPP) in breast cancer(17), and the study by Coleman et al. 2017, comparing rucaparib to placebo in ovarian cancer (28), were the most relevant contributors to heterogeneity (**Supplementary figure 1**). A sensitivity analysis was performed to examine the stability and reliability of the pooled HR results. In the leave-one-out sensitivity analyses, the pooled overall effect estimate remained similar (data not shown). Considering the influence diagnostics plot (**Supplementary figure 1**), the study Han (2) 2018 was the most influential case. In fact, its omission from each subgroup improved the pooled effect estimates, as well as the main pooled PFS result (data not shown). Importantly, in the subgroup of breast cancer, the pooled HR showed a

clinically meaningful and statistically significant result, when omitting this study (HR: 0.62, 95%CI: 0.51 - 0.76, p<0.001).

Secondary endpoint: ORR

Overall 24 studies provided data for the ORR analysis, for a total of 27 comparisons. The pooled result showed a significant correlation between ORR and the experimental arms (RR: 1.35, 95%CI: 1.16 – 1.56, p<0.001, I^2=74%), with high heterogeneity (**Figure 3**). The test for subgroup differences was non-significant according to tumor mutational status, with comparable effect observed for *BRCA*-mut patients (RR: 1.44, 95%CI: 1.09 – 1.91), and mixed or wt population (RR: 1.27, 95%CI: 1.08 – 1.49). No difference was observed according to trials' phase (p=0.09), with both phase II and III trials showing significant association between the experimental arm and ORR (**Table 3**). No significant difference was also observed according to tumor site (p=0.86), although the only study group with an individual statistically significant result was the one concerning ovarian cancer (p<0.001).

Significant subgroup differences were observed according to treatment line (p=0.03), control arm (p=0.04) and type of PARPi (p<0.001). With respect to the first subgroup, a significant better association with ORR for PARPi over the control was observed in ≥2nd-line and maintenance (**Table 3**). PARPi showed a stronger association with ORR when compared to placebo and CT. Moreover, the combination with an antiangiogenic drug showed a significantly superior activity when compared to the same PARPi as single agent. Conversely, there was no significant difference when PARPi were compared to Enza/Abi or when olaparib at higher dose was compared to an inferior dose (**Table 3**).

The effect was also significant with olaparib, niraparib and talazoparib but not with rucaparib or veliparib (**Table 3**).

Also in this case, the main result, as well as several subgroup pooled estimates, were affected by significant heterogeneity. According to the Baujat plot, the study by Litton et al. ("Litton 2018") comparing talazoparib to CT (18) and, again, Han (2) 2018 (17) were the most relevant contributors to the observed heterogeneity (**Supplementary figure 2**). Both trials were focused on breast cancer.

In the leave-one-out sensitivity analyses, the pooled overall effect estimate remained similar also when removing the above mentioned studies (data not shown). Based on the influence diagnostics plot, Han (2) 2018 and Litton 2018 considered as a potential influential cases (**Supplementary figure 2**). Although the main pooled effect remained significant when omitting the studies, subgroup results were affected within the breast cancer subset, where the omission of Han (2) 2018 led to a statistically significant result (RR: 1.54, 95%CI: 1.01 – 2.36, p=0.05). On the contrary, by removing Litton 2018 the result remained non-significant (p=0.74). A similar influence was observed in the subgroup of first-line trials, where the removal of Han (2) 2018 led to a statistically significant result (RR: 1.26, 95%CI: 1.03 – 1.53, p=0.02), while the removal of Litton 2018 did not impact the non-significance of the result (p=0.71). In the subgroup of control arms, when removing Litton 2018 the comparison with CT shifted to a non-significant result (p=0.08), while the removal

of Han (2) 2018, strengthened the pooled result in favor of PARPi-based treatments, which however remained significant (RR: 1.27, 95%CI: 1.09 – 1.47, p=0.002). With respect to the PARPi subgroup, Litton 2018 was the only contributor to the talazoparib subset, however when removing Han (2) 2018 from the veliparib subgroup, the effect become only marginally non-significant (RR: 1.10, 95%CI: 1.00 - 1.22, p=0.06).

Secondary endpoint: OS

Overall 19 studies provided data for the OS analysis, for a total of 22 comparisons. Pooled OS was significantly improved by the experimental arm (HR: 0.86, 95%CI: 0.80 – 0.93, p<0.001, I^2=7%), with no significant heterogeneity (**Figure 4**). Consequently, we performed again the analysis under the fixed effect model, obtaining a comparable result (HR: 0.86, 95%CI: 0.80 – 0.92, p<0.001, I^2=7%).

Due to the absence of substantial heterogeneity, we performed prespecified subgroup analyses using the same FE model. No significant difference was observed according to *BRCA* mutational status (p=0.57), tumor site (0.82), treatment line (p=0.22), control arm (p=0.21), PARPi (p=0.30) and trials' phase (p=0.26).

Within subgroups, however, the subsets with an individually significant OS benefit associated with the experimental arms were both mixed/wt (p<0.001) and *BRCA*-mut tumors (p=0.02), ovarian (p=0.004) and prostate cancer (p=0.04), ≥2nd-line (p=0.005) and maintenance (p=0.003), CT (p=0.02), Enza/Abi (p=0.04), placebo (p=0.003), olaparib (p<0.001) and both phase III (p<0.001) and phase II RCT (p=0.05). Subgroup analyses results are detailed in **Table 4.**

When omitting each study in the leave-one-out sensitivity analysis, the overall result was never affected significantly (data not shown). The influence diagnostic plot identified Han (2) 2018 as a potential influential case (**Supplementary figure 3**). When re-perfoming subgroup analyses by omitting it, the most affected subgroups were the one of tumor site, treatment line, control arm, type of PARPi and RCT phase. More specifically, the pooled effect in breast cancer was improved and became only marginally non-significant (HR: 0.87, 95%CI: 0.75 - 1.01, p=0.07) while the pooled effect in first-line (HR: 0.87, 95%CI: 0.78 - 0.98, p=0.02), in the veliparib group (HR: 0.88, 95%CI: 0.78 – 1.00, p=0.05) and in phase II RCT (HR: 0.85, 95%CI: 0.76 – 0.96, p=0.006) became significant.

Subgroup analysis on HRD tumors

We performed an exploratory subgroup analysis by pooling the treatment effects observed in the patients subpopulations affected by HRD tumors not exclusively due to *BRCA1/2* mutations. PARPi-based treatments appeared to be significantly effective in prolonging PFS (HR: 0.51, 95%CI: 0.43 – 0.60, p<0.001, I^2=6%), with no significant heterogeneity observed (**Figure 5**). A numerical but non-significant correlation with higher response rates (RR: 1.57, 95%CI: 0.55 – 4.49, p=0.40, I^2=73%) and better OS (HR: 0.85, 95%CI: 0.65 – 1.10, p=0.21, I^2=0%) compared to the control arm was also observed, with significant heterogeneity for the former endpoint and no heterogeneity for the latter (**Figure 5**).

Given the substantial absence of heterogeneity, we performed again the analyses under a fixed effect model, obtaining comparable results in both PFS (HR: 0.51, 95%CI: 0.44 – 0.59, p<0.001, I^2=6%) and OS (HR: 0.85, 95%CI: 0.65 – 1.10, p=0.21, I^2=0%).

Risk of bias analysis and publication bias

The analysis of bias did not raise any specific concern. The only domain that showed higher risk, compared to the others, concerned the "performance bias", which takes into account the blinding of study participants and personnel. In detail, 12/29 (41.4%) of the included studies were open-label. However, there were no, or very few, risk for other biases suggesting an overall good internal validity of the studies included (**Figure 6 and Supplementary figure 4**).

With respect to publication bias, the funnel plots for PFS and OS did not show asymmetry (**Supplementary figure 1 and 3**), as also confirmed by non-significant Eggers' test (p=0.963 for PFS and p=0.599 for OS) and Begg's test (p=0.832 for PFS and p=0.402 for OS). In the case of ORR, Eggers' test indicated the presence of funnel plot asymmetry (p=0.025), while Begg's test was not significant (p=0.288). Therefore, we evaluated the effect of publication bias through a "*trim-and-fill"* analysis (**Supplementary figure 2**). By using the L0 estimator, we obtained a significant pooled result (RR: 1.26, 95%CI: 1.10 - 1.46, p=0.001). A confirmatory *trim-and-fill* with another estimator (R0) showed similar results (RR: 1.33, 95%CI: 1.16 - 1.54, p<0.001).

3.4 DISCUSSION

Main results

Our study included 29 published phase II/III RCT of metastatic solid tumors where PARPi-containing regimens were compared to a therapeutic standard, represented in the majority of studies by either CT, hormonal treatment (HT) or placebo. Only a minority of studies (3/29) compared a PARPi in different doses (1 study) or compared the combination of a PARPi with an antiangiogenic drug vs the same PARPi alone (2 studies). Hence, our results substantially reflected the effect of PARPi-containing regimens against a different therapeutic standard.

The pooled analyses showed that PARPi regimens are associated with a consistent and statistically significant benefit in all clinical endpoints, with an overall reduction in the instantaneous risk of progression of 41%, a strong association with ORR (1.35 more than the therapeutic standard) and ~14% reduction in the instantaneous risk of death. When observing prespecified subgroup analyses results, a differential PFS effect was observed according to tumor site, line of therapy, the type of control arm, the type of PARPi and the trial phase. With respect to ORR, the treatment line, type of control arm and PARPi were the subgroups showing statistically significant different within-subgroup results. Conversely, OS subgroup analyses did not identify subsets that might specifically benefit more than others. However, to note, significant individual subgroup results were observed for ovarian and prostate cancer, for olaparib (the most studied PARPi so far), in second-/further lines or maintenance, over CT, placebo and Enza/Abi as control and in both phase II/III RCT. To note, it is plausible that the lack of OS data in 10 out of 29 studies limited the possibility to observe significant differences within subgroups.

Efficacy and activity according to tumor type

When dissecting subgroup analyses, we observed that PARPi-based combinations seemed to be associated with prolonged PFS in ovarian, prostate, pancreatic cancer, melanoma and SCLC. Additionally, after sensitivity analyses and the following selective removal of the VT vs CPP comparison(17), a clinically meaningful and statistically significant PFS improvement in the breast cancer subgroup was also observed, consistently with results from olaparib and talazoparib pivotal trials(18, 20). Apparently, the choice of temozolomide as the CT companion for veliparib in one BC trial, turned out to produce such a profound significantly inferior performance for the experimental combination, that the whole breast cancer subgroup pooled result was affected, despite being still numerically in favor of PARPi. Importantly, a significant association with better ORR was only observed for ovarian cancer, though the overall subgroup result did not show a statistical significance. At least two explanations might be given for this result. Firstly, ovarian cancer regrouped the highest number of studies (11), while pooled results for other cancer types relied only on 2 or 1 trials, with the exception of breast (5) and prostate (3) cancers. Secondly, ovarian cancer has proven to be particularly sensitive to PARPi, due to a higher prevalence of both *BRCA*-dependent and -independent HRD, compared to other solid tumors(62).

Efficacy and activity according to treatment line and control arm

Experimental regimens improved PFS in all treatment lines, with a more profound effect in maintenance and pretreated patients. However, the comparisons in first-line trials were mostly over CT (e.g. platinum-based regimens in ovarian cancer), while maintenance trials and some advanced line trials were against placebo, which might explain the differential effect observed. At the same time, the association with ORR was significant in advanced lines and maintenance, but not in first-line trials. It is high likely that this is the result of the higher concentration of CT control arms in earlier-line RCT. This might appear contradictory with what observed in the control arm subgroup, where PARPi-regimens were superior to CT +/- placebo. Despite this result, it has to be considered that the most effective and active CT regimens are usually administered in upfront schedules. Hence this might lead to differential effects observed on tumor shrinkage capabilities according to treatment line. Similarly, the comparison with Enza/Abi in prostate cancer trials did not show a clear superiority for PARPi in terms of ORR. However, while several comparisons against CT involved a PARPi combined with a CT regimen, this was not the case for comparisons against antiandrogen therapy (**Table 1**). This might potentially explain the reason of the comparable activity observed, but could also mean that, due to the prominent growth-driver role played by the androgen receptor pathway in prostate cancers(63), novel HT might retain their activity independently from the presence of HRR-deficiency, despite PARPi being more effective in terms of PFS in this context. Importantly, compared to other solid tumors, metastatic castration resistant prostate cancer (mCRPC) patients present a higher prevalence of bone-only disease(64). Therefore, some of the trials in this setting have used a composite endpoint to evaluate response rates, by including also the percentage of prostate specific antigen reduction from baseline (PSA) and circulating tumor cells (CTCs) conversion (from more/equal than 5, to less than 5)(46), or by including progression on bone scan as *per* Prostate Cancer Working Group (PCWG) criteria(44).

These differences might affect pooled ORR result interpretation, as well. In any case, when considering the efficacy over different control arms, PARPi regimens were superior to all competing regimens in terms of PFS, with a more pronounced effect over placebo, Enza/Abi (in prostate cancer), and over PARPi monotherapy when a combo with antiangiogenic drugs represented the experimental comparator (cediranib or bevacizumab).

Results based on mutational status

Notably, for all endpoints (i.e. PFS, ORR and OS), no difference was observed between the subgroup of *BRCA*-mut and *BRCA*-wt/mixed population. This is a surprising, yet not completely unexpected finding. The first solid tumor where PARPi demonstrated a clear clinical benefit that translated into FDA approval was MOC, and half comparisons included in our analyses were conducted in this tumor type. Notably, individual pooled results for ovarian cancer, were all uniformly in favor of PARPi in terms of PFS, ORR and OS, despite including numerous studies with *BRCA*-wt or unknown/mixed mutational status(28–30, 33, 35, 37–39, 43). It is high likely that such benefit was driven by a subgroup of patients with HRD, a condition that can be caused by *BRCA*1/2 mutations, as well as by an impairment in other genes involved in the homologous recombination DNA repair mechanism, such as *ATM, CHECK1/2, RAD51* or *PALB2*, either due to somatic/germinal mutations or epigenetic mechanisms(15, 33, 65). Actually, roughly 50% of all high-grade serous ovarian cancers present some form of HRD, either because of germline/somatic mutations in *BRCA*1/2 (20%), epigenetic silencing of *BRCA*1 (11%), amplification/mutation of *EMSY* (8%), deletion of *PTEN* (7%), hypermethylation of *RAD51C* (3%), or mutations in *ATM*/*ATR* (2%) and Fanconi anemia genes (5%)(66). In this perspective, it is important to highlight that we performed an exploratory subgroup analysis on HRD tumors, which comprised 9 studies. Of these, 6 were on MOC(28–30, 37, 38, 43), 2 on mCRPC(44, 45) and 1 on gastric cancer(22). We observed a strikingly 49% significant reduction in the risk of progression or death with PARPi-based treatments, compared to control. This result was undoubtably driven by MOC studies, but also 1 out of 2 mCRPC studies showed a significant result in favor of PARPi. This result strengthen the arising theory that PARPi might be particularly effective not only in *BRCA*-mutant tumors, but also in tumors with other forms of defective HRR. Unfortunately, up to now, this condition has been assessed in methodologically heterogeneous ways, and with different definitions, depending on the clinical trial (46, 67, 68). Therefore, the implementation of a more homogeneous characterization of HRD status across different solid tumors is highly recommended.

Intriguingly, the PRIMA trial in ovarian cancer was able to identify a PFS improvement with niraparib monotherapy in HRR-proficient MOC (30). Similarly, a study of olaparib + abiraterone vs abiraterone in mCRPC showed improved radiologic PFS irrespective of HRR status, with exploratory analyses suggesting efficacy also in non-dysfunctional tumors(44). Additionally, the combination of veliparib, cisplatin and etoposide within the ECO-ACRIN 2511 study in SCLC also showed a significant PFS improvement in the absence of HRD, possibly due to a synergistic effect with CT agents capable of directly damaging DNA (23, 29, 40). Nevertheless, some individual trials still failed to demonstrate the efficacy of PARPi in non-mutant tumors like melanoma (26), NSCLC and SCLC (24, 25), colon and gastric cancer (21, 22), albeit PARPi had been combined with several effective CT partners. Given the low frequency of HRR genes such *BRCA1/2* in tumors like melanoma, colon, gastric cancer, SCLC and NSCLC(7, 65), it is not particularly surprising that some results observed in unselected populations have been disappointing. Still, there are preclinical

evidences for potential alternative biomarkers of response to PARPi in subgroups of those solid tumors, such as low ERCC1 expression in NSCLC and melanoma(26, 69), ARID1A deficiency in solid tumors, including gastric and colon cancer(70), detectable p16 expression in melanoma(26) or biomarkers of resistance, like TRIP12, which has been recently demonstrated to constrain the PARP1-trapping mechanism of PARPi(71). A more extensive evaluation of these biomarkers in future studies, so to better select target population for PARPi in *BRCA*-wt solid tumors, is highly recommended. Furthermore, recent findings have shown that the overall frequency of mutations affecting HRR genes is around 17%, with the maximum prevalence observed in endometrial cancer (34.4%) and the lowest in gastrointestinal stromal tumors (3.7%)(65).

In light of our results supporting PARPi efficacy also in *BRCA*-independent HRD-positive solid tumors, a potential way to assess the efficacy of PARPi in rarer *BRCA*-mut or -wt/HRD cancers might be the development of basket trials, with the objective to prove a class effect as tumor-agnostic therapeutic option. This has been already observed, for example, with NTRK fusion-positive or with high microsatellite instability tumors(72, 73). Few trials of this kind are already planned/ongoing (i.e. NCT03742895, NCT04123366, NCT04171700, NCT04503265, NCT04174716).

Another possibility to overcome patient recruitment issues in trials involving rare tumor types might be through the comparison of single arm trials involving the experimental drug with a synthetic control arm represented, for example, by historical observational data, already published results from RCT or other external control data(74). This strategy is not new and is gaining more attention in recent years, having led, among others, to the expanded indication of palbociclib + HT for men with hormone receptor-positive/HER2-negative MBC (74).

Results according to PARPi molecule

All PARPi prolonged PFS. The most pronounced effect was observed with rucaparib, while the less potent effect was observed with veliparib. Similarly for ORR, the most potent PARPi was rucaparib and the less was veliparib, although individual rucaparib and veliparib pooled results were non-significant. It is important to underline that PARPi's main therapeutic effect seem to be related to PARylation and subsequent PARP trapping,(12, 15). In this perspective, different PARPi have shown different trapping potency, with talazoparib>niraparib>olaparib=rucaparib>veliparib; the latter substantially lacking PARP trapping capability (15, 75, 76). This is well represented by the poor performance observed in our pooled analyses, despite the numerous veliparib-containing studies included (11 out of 29). In apparent contrast, rucaparib has provided the best individual result in PFS and ORR (OS data were unavailable), compared to the other PARPi. However, it is high likely that this result has to be attributable to the fact that only one trial contained rucaparib, and in this study, the PARPi was compared to placebo as maintenance treatment after response to platinum agents in ovarian cancer(28). Good responses to platinum agents, at least in MOC and mCRPC have been linked to the presence of HRD (77, 78) and the potential to be a predictor of response to PARP inhibition (77); thus a particularly sensitive population might have been tested. At the same time, no active comparison was administered. As a consequence, differently from rucaparib, the effect of other PARPi might have been diluted in pooled analyses regrouping several studies with different combinations, comparisons and tumor type. This is not true for talazoparib,

the most potent PARPi in terms of PARP trapping, which was also tested in only 1 trial included in our study and still did not outperformed the other inhibitors. However, it is necessary to underline that it was administered in a poor prognosis breast cancer subgroup (i.e. triple negative) and, differently from rucaparib, it was compared to potentially effective therapeutic alternatives, like eribulin and capecitabine(18). In any case, our results substantially confirm the superiority of PARPi with trapping capacity over veliparib.

Additionally, some other mechanism of action for PARPi have been proposed, such as the blocking of PARP-regulated gene transcription, interference with ribosome biogenesis, mitophagy and apoptosis, which might differ between different PARPi molecules and might show different impact in different types of cancers (37, 79–82). In addition to this, other molecular and cellular mechanisms, like immune pathway activation, PD-L1 expression modulation on cancer cells, and the genomic instability produced by PARP inhibition, might increase tumor immunogenicity and responsiveness to immune checkpoint inhibitors (ICI)(83–85). In this perspective, promising evidence of efficacy for PARPi+ICI combination has been recently observed with the TOPACIO and MEDIOLA single arm phase 2 trials, and RCT are already ongoing (7, 48, 86, 87).

All in all, a deeper characterization of all these mechanisms is warranted, so to identify the best combination strategies and the most adequate PARPi for the appropriate context.

Limitations and strengths

The major limitation of our study relies in the considerable heterogeneity observed for PFS and ORR pooled results. It is high likely that such heterogeneity was related to the design of the study itself, having included in our analysis trials of different phase, conducted in different lines, several solid tumors, in both mutant and wt populations and with different PARPi and control arms. In fact, subgroup analyses identified specific subsets where the efficacy and activity of PARPi-based regimens seem to be modest with respect to the therapeutic standard. Importantly, when performing leave-one-out sensitivity analyses, the main pooled results were not affected significantly by a single study/treatment comparison. Moreover, a RE model was applied to take into account such heterogeneity. We addressed it also through visual inspection of Baujat plots and influential analyses, that helped identifying the most problematic comparisons and assess their impact on each subgroup.

Another limitation is represented by the use of individual patients' data (IPD). We were not able to perform an IPD meta-analysis, due to the lack of the necessary resources. Although this kind of meta-analysis is usually considered the best option to summarize the results of multiple studies, scientific literature recognizes that such studies are not always feasible(88). Moreover, while some guidance is available to help understand when aggregate patient data (APD) meta-analyses, such ours, might suffice and when IPD might add value, this is not backed by empirical evidence(89) and is still not clear when the collection of more detailed IPD is truly needed(90). In fact, for meta-analyses of published time-to-event outcomes, individual case studies have shown that they can produce effects that are either larger than, smaller than, or similar to their IPD equivalents(89). Moreover, HR from published APD meta-analyses seem to most likely agree with those from IPD when the information size is large(90). Finally, considering the complexity of the topic and the high

number of studies and patients involved, is high likely that an IPD meta-analysis on the same topic won't be conducted.
Another limitation is represented by the publication bias observed regarding the ORR result. Importantly, though, according to the *tream-and-fill* analyses performed, ORR pooled result was confirmed to be statistically significant even when controlling for selective publication, thus suggesting that this bias had little effect and the results were relatively robust. Finally, we did not analyzed here the toxicity data emerging from those trials. In any case, PARPi are usually well-tolerated drugs, with nausea, vomiting, seizures, fatigue, leukopenia, anemia and thrombocytopenia being the most frequent, albeit manageable side effects. The incidence is different with respect to PARPi molecule, as also well described elsewhere(7, 76, 91).

The strength of our study relies in its comprehensive assessment of PARPi activity and efficacy in solid tumors, the most complete and up-to-date. The methodology was solid and reliable, with numerous sensitivity analyses conducted to overcome the main issues observed related to heterogeneity and robustness of results.

3.5 CONCLUSIONS

Although our study confirms and reinforce the role of already approved PARPi-based treatments, especially in *BRCA*1/2-mutant tumors, a more comprehensive effort is needed to identify other forms of HRD along with a better characterization of secondary mechanisms of action and further predictive biomarkers of response. We envision that this approach will better elucidate PARPi efficacy in a broader scenario, alone or in combination with other therapeutic agents.

Acknowledgements

PR is supported by a Prostate Cancer Foundation's Young Investigator award. FS is supported by a European Society for Medical Oncology (ESMO) Translational Research Fellowship.

3.6 REFERENCES

1. National Cancer Institute. Genetics of breast and gynecologic cancers (PDQ®) - health professional version. 2019.

2. Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell 2002; 108(2):171–182.

3. Featherstone C, Jackson SP. DNA double-strand break repair. Curr. Biol. 1999; 9(20):R759–R761.

4.Wang JYJ. DNA damage and apoptosis. Cell Death Differ. 2001; 8(11):1047–1048.

5. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. Nat. Rev. Cancer 2004; 4(9):665–676.

6. Susan G. Komen Foundation. Inherited Genetic Mutations. .

7. Bayraktar S, Glück S, Darling H. Update on PARP inhibitor therapy for solid tumors. J Cancer Prev Curr Res 2019; 10(4):98–107.

8. Golan T, Hammel P, Reni M et al. Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. N. Engl. J. Med. 2019; 381(4):317–327.

9. McClain MR, Palomaki GE, Nathanson KL, Haddow JE. Adjusting the estimated proportion of breast cancer cases associated with BRCA1 and BRCA2 mutations: public health implications. Genet. Med. Off. J. Am. Coll. Med. Genet. 2005; 7(1):28–33.

10. Messina C, Cattrini C, Soldato D et al. BRCA Mutations in Prostate Cancer: Prognostic and Predictive Implications. J. Oncol. 2020. doi:10.1155/2020/4986365.

11. Manchana T, Phoolcharoen N, Tantbirojn P. BRCA mutation in high grade epithelial ovarian cancers. Gynecol. Oncol. Rep. 2019; 29:102–105.

12.Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science 2017; 355(6330):1152–1158.

13. Farmer H, McCabe N, Lord CJ et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005; 434(7035):917–921.

14. McCabe N, Turner NC, Lord CJ et al. Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. Cancer Res. 2006; 66(16):8109–8115.

15. Murai J, Huang SN, Das BB et al. Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. Cancer Res. 2012; 72(21):5588–5599.

16.Hopkins TA, Ainsworth WB, Ellis PA et al. PARP1 Trapping by PARP Inhibitors Drives Cytotoxicity in Both Cancer Cells and Healthy Bone Marrow. Mol. Cancer Res. 2019; 17(2):409–419.

17.Han HS, Diéras V, Robson M et al. Veliparib with temozolomide or carboplatin/paclitaxel versus placebo with carboplatin/paclitaxel in patients with BRCA1/2 locally recurrent/metastatic

breast cancer: randomized phase II study. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2018; 29(1):154–161.

18.Litton JK, Rugo HS, Ettl J et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. N. Engl. J. Med. 2018; 379(8):753–763.

19.Robson ME, Tung N, Conte P et al. OlympiAD final overall survival and tolerability results: Olaparib versus chemotherapy treatment of physician's choice in patients with a germline BRCA mutation and HER2-negative metastatic breast cancer. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2019; 30(4):558–566.

20.Robson M, Im S-A, Senkus E et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N. Engl. J. Med. 2017; 377(6):523–533.

21.Gorbunova V, Beck JT, Hofheinz R-D et al. A phase 2 randomised study of veliparib plus FOLFIRI±bevacizumab versus placebo plus FOLFIRI±bevacizumab in metastatic colorectal cancer. Br. J. Cancer 2019; 120(2):183–189.

22.Bang Y-J, Xu R-H, Chin K et al. Olaparib in combination with paclitaxel in patients with advanced gastric cancer who have progressed following first-line therapy (GOLD): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2017; 18(12):1637–1651.

23.Owonikoko TK, Dahlberg SE, Sica GL et al. Randomized Phase II Trial of Cisplatin and Etoposide in Combination With Veliparib or Placebo for Extensive-Stage Small-Cell Lung Cancer: ECOG-ACRIN 2511 Study. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2019; 37(3):222–229.

24.Pietanza MC, Waqar SN, Krug LM et al. Randomized, Double-Blind, Phase II Study of Temozolomide in Combination With Either Veliparib or Placebo in Patients With Relapsed-Sensitive or Refractory Small-Cell Lung Cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2018; 36(23):2386–2394.

25.Ramalingam SS, Blais N, Mazieres J et al. Randomized, Placebo-Controlled, Phase II Study of Veliparib in Combination with Carboplatin and Paclitaxel for Advanced/Metastatic Non-Small Cell Lung Cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2017; 23(8):1937–1944.

26. Middleton MR, Friedlander P, Hamid O et al. Randomized phase II study evaluating veliparib (ABT-888) with temozolomide in patients with metastatic melanoma. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2015; 26(10):2173–2179.

27. Audeh MW, Carmichael J, Penson RT et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. Lancet Lond. Engl. 2010; 376(9737):245–251.

28.Coleman RL, Oza AM, Lorusso D et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Lond. Engl. 2017; 390(10106):1949–1961.

29.Coleman RL, Fleming GF, Brady MF et al. Veliparib with First-Line Chemotherapy and as Maintenance Therapy in Ovarian Cancer. N. Engl. J. Med. 2019; 381(25):2403–2415.

30.González-Martín A, Pothuri B, Vergote I et al. Niraparib in Patients with Newly Diagnosed

Advanced Ovarian Cancer. N. Engl. J. Med. 2019; 381(25):2391–2402.

31.Kaye SB, Lubinski J, Matulonis U et al. Phase II, open-label, randomized, multicenter study comparing the efficacy and safety of olaparib, a poly (ADP-ribose) polymerase inhibitor, and pegylated liposomal doxorubicin in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2012; 30(4):372–379.

32.Kummar S, Oza AM, Fleming GF et al. Randomized Trial of Oral Cyclophosphamide and Veliparib in High-Grade Serous Ovarian, Primary Peritoneal, or Fallopian Tube Cancers, or BRCA-Mutant Ovarian Cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2015; 21(7):1574–1582.

33.Ledermann J, Harter P, Gourley C et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. N. Engl. J. Med. 2012; 366(15):1382–1392.

34.Ledermann JA, Harter P, Gourley C et al. Overall survival in patients with platinum-sensitive recurrent serous ovarian cancer receiving olaparib maintenance monotherapy: an updated analysis from a randomised, placebo-controlled, double-blind, phase 2 trial. Lancet Oncol. 2016; 17(11):1579–1589.

35.Liu JF, Barry WT, Birrer M et al. Combination cediranib and olaparib versus olaparib alone for women with recurrent platinum-sensitive ovarian cancer: a randomised phase 2 study. Lancet Oncol. 2014; 15(11):1207–1214.

36.Liu JF, Barry WT, Birrer M et al. Overall survival and updated progression-free survival outcomes in a randomized phase II study of combination cediranib and olaparib versus olaparib in relapsed platinum-sensitive ovarian cancer. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2019; 30(4):551–557.

37Mirza MR, Monk BJ, Herrstedt J et al. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. N. Engl. J. Med. 2016; 375(22):2154–2164.

38. Mirza MR, Åvall Lundqvist E, Birrer MJ et al. Niraparib plus bevacizumab versus niraparib alone for platinum-sensitive recurrent ovarian cancer (NSGO-AVANOVA2/ENGOT-ov24): a randomised, phase 2, superiority trial. Lancet Oncol. 2019; 20(10):1409–1419.

39.Moore K, Colombo N, Scambia G et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N. Engl. J. Med. 2018; 379(26):2495–2505.

40.Oza AM, Cibula D, Benzaquen AO et al. Olaparib combined with chemotherapy for recurrent platinum-sensitive ovarian cancer: a randomised phase 2 trial. Lancet Oncol. 2015; 16(1):87–97.

41.Penson RT, Valencia RV, Cibula D et al. Olaparib Versus Nonplatinum Chemotherapy in Patients With Platinum-Sensitive Relapsed Ovarian Cancer and a Germline BRCA1/2 Mutation (SOLO3): A Randomized Phase III Trial. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2020; 38(11):1164–1174.

42.Pujade-Lauraine E, Ledermann JA, Selle F et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2017; 18(9):1274–1284.

43.Ray-Coquard I, Pautier P, Pignata S et al. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. N. Engl. J. Med. 2019; 381(25):2416–2428.

44.Clarke N, Wiechno P, Alekseev B et al. Olaparib combined with abiraterone in patients with metastatic castration-resistant prostate cancer: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Oncol. 2018; 19(7):975–986.

45.de Bono J, Mateo J, Fizazi K et al. Olaparib for Metastatic Castration-Resistant Prostate Cancer. N. Engl. J. Med. 2020; 382(22):2091–2102.

46.Mateo J, Porta N, Bianchini D et al. Olaparib in patients with metastatic castration-resistant prostate cancer with DNA repair gene aberrations (TOPARP-B): a multicentre, open-label, randomised, phase 2 trial. Lancet Oncol. 2020; 21(1):162–174.

47.Poveda A, Floquet A, Ledermann JA et al. Final overall survival (OS) results from SOLO2/ENGOT-ov21: A phase III trial assessing maintenance olaparib in patients (pts) with platinum-sensitive, relapsed ovarian cancer and a BRCA mutation. J. Clin. Oncol. 2020; 38(15_suppl):6002–6002.

48.Vikas P, Borcherding N, Chennamadhavuni A, Garje R. Therapeutic Potential of Combining PARP Inhibitor and Immunotherapy in Solid Tumors. Front. Oncol. 2020. doi:10.3389/fonc.2020.00570.

49.Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2018:19.

50.Hussain M, Mateo J, Fizazi K et al. Survival with Olaparib in Metastatic Castration-Resistant Prostate Cancer. N. Engl. J. Med. 2020. doi:10.1056/NEJMoa2022485.

51.Diéras V, Han HS, Kaufman B et al. Veliparib with carboplatin and paclitaxel in BRCA-mutated advanced breast cancer (BROCADE3): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2020; 21(10):1269–1282.

52.Litton JK, Hurvitz SA, Mina LA et al. Talazoparib versus chemotherapy in patients with germline BRCA1/2-mutated HER2-negative advanced breast cancer: final overall survival results from the EMBRACA trial. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 2020; 31(11):1526–1535.

53.Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat. Med. 2002; 21(11):1539–1558.

54.Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res. Synth. Methods 2010; 1(2):97–111.

55.Baujat B, Mahé C, Pignon J-P, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat. Med. 2002; 21(18):2641–2652.

56.Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 2010; 1(2):112–125.

57.Sterne JAC, Sutton AJ, Ioannidis JPA et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011; 343:d4002.

58.Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of

axis. J. Clin. Epidemiol. 2001; 54(10):1046–1055.

59.R Core Team. R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing, 2017.

60.The Cochrane Collaboration. Review Manager (RevMan), Copenhagen: The Nordic Cochrane Centre: 2014.

61.Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011), The Cochrane Collaboration, .

62.Pan-cancer landscape of homologous recombination deficiency | Nature Communications. (https://www.nature.com/articles/s41467-020-19406-4).

63.Lonergan PE, Tindall DJ. Androgen receptor signaling in prostate cancer development and progression. J. Carcinog. 2011. doi:10.4103/1477-3163.83937.

64.Hernandez RK, Wade SW, Reich A et al. Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. BMC Cancer 2018. doi:10.1186/s12885-017-3922-0.

65. Heeke AL, Pishvaian MJ, Lynce F et al. Prevalence of Homologous Recombination–Related Gene Mutations Across Multiple Cancer Types. JCO Precis. Oncol. 2018. doi:10.1200/PO.17.00286.

66.Integrated genomic analyses of ovarian carcinoma | Nature. (https://www.nature.com/articles/nature10166).

67.Myriad Genetics, Inc. Myriad myChoice® CDx description. Available at: https://myriad.com/products-services/precision-medicine/mychoice-cdx/. 2020.

68. Foundation Medicine, Inc. FoundationOne®CDx Technical Specifications, 2018. Available at: www.rochefoundationmedicine.com/f1cdxtech. 2020.

69.Postel-Vinay S, Bajrami I, Friboulet L et al. A high-throughput screen identifies PARP1/2 inhibitors as a potential therapy for ERCC1-deficient non-small cell lung cancer. Oncogene 2013; 32(47):5377–5387.

70.Shen J, Peng Y, Wei L et al. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov. 2015; 5(7):752–767.

71.Gatti M, Imhof R, Huang Q et al. The Ubiquitin Ligase TRIP12 Limits PARP1 Trapping and Constrains PARP Inhibitor Efficiency. Cell Rep. 2020. doi:10.1016/j.celrep.2020.107985.

72.FDA Approves Entrectinib for Tumors with NTRK Fusions - National Cancer Institute. (https://www.cancer.gov/news-events/cancer-currents-blog/2019/fda-entrectinib-ntrk-fusion).

73.FDA approves pembrolizumab for first-line treatment of MSI-H/dMMR colorectal cancer. (https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-first-line-treatment-msi-hdmmr-colorectal-cancer).

74.Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and External Controls in Clinical Trials – A

Primer for Researchers. Clin. Epidemiol. 2020; 12:457–467.

75.Murai J, Huang S-YN, Renaud A et al. Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. Mol. Cancer Ther. 2014; 13(2):433–443.

76.Pommier Y, O'Connor MJ, de Bono J. Laying a trap to kill cancer cells: PARP inhibitors and their mechanisms of action. Sci. Transl. Med. 2016; 8(362):362ps17.

77. Jiang X, Li X, Li W et al. PARP inhibitors in ovarian cancer: Sensitivity prediction and resistance mechanisms. J. Cell. Mol. Med. 2019; 23(4):2303–2313.

78.Zafeiriou Z, Bianchini D, Chandler R et al. Genomic Analysis of Three Metastatic Prostate Cancer Patients with Exceptional Responses to Carboplatin Indicating Different Types of DNA Repair Deficiency. Eur. Urol. 2019; 75(1):184–192.

79.Kim D-S, Camacho CV, Nagari A et al. Activation of PARP-1 by snoRNAs Controls Ribosome Biogenesis and Cell Growth via the RNA Helicase DDX21. Mol. Cell 2019; 75(6):1270-1285.e14.

80.Brenner JC, Ateeq B, Li Y et al. Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. Cancer Cell 2011; 19(5):664–678.

81.Arun B, Akar U, Gutierrez-Barrera AM et al. The PARP inhibitor AZD2281 (Olaparib) induces autophagy/mitophagy in BRCA1 and BRCA2 mutant breast cancer cells. Int. J. Oncol. 2015; 47(1):262–268.

82.Lee EK, Konstantinopoulos PA. PARP inhibition and immune modulation: scientific rationale and perspectives for the treatment of gynecologic cancers. Ther. Adv. Med. Oncol. 2020. doi:10.1177/1758835920944116.

83.Chabanon RM, Soria J-C, Lord CJ, Postel-Vinay S. Beyond DNA repair: the novel immunological potential of PARP inhibitors. Mol. Cell. Oncol. 2019. doi:10.1080/23723556.2019.1585170.

84.Hodgson DR, Dougherty BA, Lai Z et al. Candidate biomarkers of PARP inhibitor sensitivity in ovarian cancer beyond the BRCA genes. Br. J. Cancer 2018; 119(11):1401–1409.

85. Gay CM, Byers LA. PARP Inhibition Combined with Immune Checkpoint Blockade in SCLC: Oasis in an Immune Desert or Mirage? J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer 2019; 14(8):1323–1326.

86.Domchek SM, Postel-Vinay S, Im S-A et al. Olaparib and durvalumab in patients with germline BRCA-mutated metastatic breast cancer (MEDIOLA): an open-label, multicentre, phase 1/2, basket study. Lancet Oncol. 2020; 21(9):1155–1164.

87.Konstantinopoulos PA, Waggoner SE, Vidal GA et al. TOPACIO/Keynote-162 (NCT02657889): A phase 1/2 study of niraparib + pembrolizumab in patients (pts) with advanced triple-negative breast cancer or recurrent ovarian cancer (ROC)—Results from ROC cohort. J. Clin. Oncol. 2018; 36(15_suppl):106–106.

88.Berlin JA, Santanna J, Schmid CH et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head.

Stat. Med. 2002; 21(3):371–387.

89.Tudur Smith C, Clarke M, Marson T et al. A framework for deciding if individual participant data are likely to be worthwhile. Abstracts of the 23rd Cochrane Colloquium, Vienna, Austria. Cochrane Database Syst Rev 2015; 10((Suppl)):RO 6.1.

90.Tierney JF, Fisher DJ, Burdett S et al. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study. PLoS Med. 2020; 17(1):e1003019.

91.Poggio F, Bruzzone M, Ceppi M et al. Single-agent PARP inhibitors for the treatment of patients with BRCA-mutated HER2-negative metastatic breast cancer: a systematic review and meta-analysis. ESMO Open 2018; 3(4):e000361.

3.7 TABLES

Table 1. Main characteristics of the included studies

FIRST AUTHOR	YEAR	JOURNAL	PHASE	LINE	CANCER TYPE	*BRCA* STATUS	N. ARMS	N. PTS EXPERIMENTAL ARM	N. PTS CONTROL ARM	TREATMENTS	META-ANALYSIS ENDPOINTS	SEPARATE HRD DATA*
Mirza	2016	New Engl J Med	III	Maintenance after ≥2 line	OC	Mut	2	138	65	NIRAPARIB vs PLACEBO	PFS	Yes (PFS)
						Wt	2	234	116	NIRAPARIB vs PLACEBO	PFS	
Ledermann	2012/2016	New Engl J Med/Lancet Oncol	II	Maintenance after ≥2 line	OC	Mixed	2	68	21	OLAPARIB vs PLACEBO	ORR, PFS, OS	No
Gonzalez Martin	2019	New Engl J Med	III	Maintenance after 1st line	OC	Mixed	2	487	246	NIRAPARIB vs PLACEBO	PFS, OS	Yes (PFS)
Liu	2014/2019	Lancet Oncol/Ann Oncol	II	≥2 line	OC	Mixed	2	46	44	OLAPARIB+CEDIRANIB vs OLAPARIB	ORR, PFS, OS	No
Oza	2015	Lancet Oncol	II	≥2 line	OC	Mixed	2	81	81	OLAPARIB+PACLITAXEL+CARBOPLATIN --> OLAPARIB vs PACLITAXEL+CARBOPLATIN	ORR, PFS, OS	No
Robson	2017/2019	New Engl J Med/Ann Oncol	III	≥1st line	BC	Mut	2	205	97	OLAPARIB vs CHEMOTHERAPY	ORR, PFS, OS	No
Clarke	2018	Lancet Oncol	II	≥2 line	PC	Mixed	2	71	71	OLAPARIB+ABIRATERONE vs PLACEBO+ABIRATERONE	ORR, PFS, OS	Yes (PFS)
Moore	2018	New Engl J Med	III	Maintenance after 1st line	OC	Mut	2	260	131	OLAPARIB vs PLACEBO	PFS, OS	No
Golan	2019	New Engl J Med	III	Maintenance after 1st line	PC	Mut	2	92	62	OLAPARIB vs PLACEBO	ORR, PFS, OS	No
Ramalingam	2016	Clin Cancer Res	II	1st line	NSCLC	Mixed	2	105	53	VELIPARIB+ PACLITAXEL+CARBOPLATIN vs PLACEBO+PACLITAXEL+CARBOPLATIN	ORR, PFS, OS	No

Han	2018	Ann Oncol	II	≥1st line	BC	Mut	3	97	99	VELIPARIB+CARBOPLATINO+PACLITAXEL vs PACLITAXEL +CARBOPLATINO+PLACEBO	ORR, PFS, OS	No
								94	99	VELIPARIB+TEMOZOLAMIDE vs PACLITAXEL+ CARBOPLATINO+PLACEBO	ORR, PFS, OS	
Pietanza	2018	J Clin Oncol	II	≥2 line	SCLC	Mixed	2	49	55	TEMOZOLOMIDE+VELIPARIB vs TEMOZOLOMIDE+PLACEBO	ORR, PFS	No
Coleman	2019	New Engl J Med	III	1st line and Maintenance after 1st line	OC	Mixed	3	382	375	CARBOPLATINO+PACLITAXEL+VELIRPARIB --> VELIPARIB vs CARBOPLATINO+PACLITAXEL --> PLACEBO	ORR, PFS	Yes (PFS, ORR)
Gorbunova	2019	Br J Cancer	II	1st line	CRC	Mixed	2	65	65	FOLFIRI+VELIPARIB±BEVACIZUMAB vs PLACEBO+FOLFIRI±BEVACIZUMAB	ORR, PFS, OS	No
Litton	2018	New Engl J Med	III	≥1st line	BC	Mut	2	287	144	TALAZOPARIB vs CHEMOTHERAPY	ORR, PFS, OS	No
Owonikoko	2018	J Clin Oncol	II	1st line	SCLC	Mixed	2	64	64	CISPLATIN+ETOPOSIDE+VELIPARIB CISPLATIN+ETOPOSIDE+PLACEBO	ORR, PFS, OS	No
Middelton	2015	Ann Oncol	II	≥2nd line	ME	Mixed	3	116	115	VELIPARIB (20MG)+TEMOZOLAMIDE vs TEMOZOLAMIDE+PLACEBO	ORR, PFS, OS	No
								115	115	VELIPARIB (40MG)+TEMOZOLAMIDE vs TEMOZOLAMIDE+PLACEBO	ORR, PFS, OS	
Mirza	2019	Lancet Oncol	II	≥2nd line	OC	Mixed	2	48	49	NIRAPARIB+BEVACIZUMAB vs NIRAPARIB	ORR, PFS	Yes (PFS)
Bang	2017	Lancet Oncol	II	2nd line	GC	Mixed	2	263	262	OLAPARIB +PACLITAXEL vs PLACEBO+PACLITAXEL	ORR, PFS, OS	Yes (PFS, ORR, OS)

Kummar	2015	Clin Cancer Res	II	≥2nd line	OC	Mixed	2	37	38	VELIPARIB + CYCLOPHOSPHAMIDE vs CYCLOPHOSPHAMIDE	ORR	No
Kaye	2011	J Clin Oncol	II	≥2nd line	OC	Mut	3	32	33	OLAPARIB (200MG) vs PEGYLATED LIPOSOMAL DOXORUBICIN	ORR, PFS, OS	No
								32	33	OLAPARIB (400MG) vs PEGYLATED LIPOSOMAL DOXORUBICIN	ORR, PFS, OS	
Coleman	2017	Lancet	III	≥3rd line	OC	Mixed	2	375	189	RUCAPARIB vs PLACEBO	ORR, PFS	Yes (PFS, ORR)
De Bono/Hussain	2020	New Engl J Med	III	≥2nd line	PC	Mut	2	162	83	OLAPARIB vs ENZALUTAMIDE/ABIRATERONE	ORR, PFS, OS	Yes (PFS, OS)
Penson	2020	J Clin Oncol	III	≥3rd line	OC	Mut	2	178	88	OLAPARIB vs CHEMOTHERAPY	ORR, PFS	No
Ray-Coquard	2019	New Engl J Med	III	Maintenance after 1st line	OC	Mixed	2	535	269	OLAPARIB+BEVACIZUMAB vs BEVACIZUMAB+PLACEBO	PFS	Yes (PFS)
Mateo	2020	Lancet Oncol	II	≥2nd line	PC	Mut	2	49	49	OLAPARIB (400mg twice) vs OLAPARIB (300mg twice)	ORR	No
Audeh	2010	Lancet	II	≥2nd line	OC	Mut	2	33	24	OLAPARIB (400mg twice) vs OLAPARIB (100mg twice)	ORR	No
Dieras	2020	Lancet Oncol	III	1st/2nd line	BC	Mut	2	337	172	CARBOPLATIN+PACLITAXEL+VELIRPARIB vs PLACEBO+CARBOPLATIN+PACLITAXEL	ORR, PFS, OS	No
Pujade-Lauraine/Poveda	2017/2020	Lancet Oncol/J Clin Oncol	III	Maintenance after ≥2 line	OC	Mut	2	196	99	OLAPARIB vs PLACEBO	PFS, OS	No

Legend. OC: ovarian cancer; BC: breast cancer; PC: prostate cancer; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; ME: melanoma; CRC: colo-rectal cancer; GC: gastric cancer; OS: overall survival; PFS: progression-free survival; ORR: overall response rates; Mut: mutant; Wt: wild-type; N: number; PTS: patients; FOLFIRI: 5-fluorouracil + oxaliplatin + irinotecan; HRD: homologous recombination deficiency; *: homologous recombination deficiency not due to *BRCA1/2* mutation

Table 2. Progression-free survival results

PFS	No. of comparisons	Pooled HR (95% CI)	P pooled	I^2%	P subgroups
Overall	30	0.59 (0.51 - 0.68)	<0.001	85%	N/A
Mutation Status					
Mixed/Wild-type	17	0.61 (0.52 - 0.71)	<0.001	79%	0.65
Mutant	13	0.56 (0.42 - 0.75)	<0.001	89%	
Tumor Site					
Ovarian	15	0.48 (0.40 - 0.58)	<0.001	83%	0.001
Breast	5	0.77 (0.52 - 1.14)	0.19	88%	
Prostate	2	0.46 (0.25 - 0.88)	0.02	85%	
Melanoma	2	0.78 (0.62 - 0.98)	0.03	0%	
NSCLC	1	0.72 (0.45 - 1.15)	0.17	N/A	
SCLC	2	0.77 (0.63 - 0.95)	0.01	0%	
Pancreatic	1	0.53 (0.35 - 0.80)	0.003	N/A	
Gastrointestinal	2	0.86 (0.70 - 1.04)	0.12	0%	
Line of Therapy					
1st-line +/- maintenance	9	0.76 (0.61 - 0.93)	0.009	77%	0.002
≥2nd-line	13	0.61 (0.48 - 0.76)	<0.001	83%	
Maintenance only	8	0.42 (0.32 - 0.53)	<0.001	83%	
Control Arm					
CT +/- placebo	17	0.75 (0.66 - 0.86)	<0.001	63%	<0.001
Placebo	8	0.39 (0.31 - 0.48)	<0.001	77%	
PARPi w/o Antiangiogenic drug	2	0.42 (0.27 - 0.64)	<0.001	14%	
Bevacizumab	1	0.63 (0.51 - 0.78)	<0.001	N/A	
Enzalutamide/Abiraterone	2	0.46 (0.25 - 0.88)	0.02	85%	
Type of PARP inhibitor					
Olaparib	14	0.52 (0.42 - 0.64)	<0.001	83%	<0.001
Veliparib	10	0.82 (0.70 - 0.97)	0.02	63%	
Niraparib	4	0.42 (0.29 - 0.60)	<0.001	77%	
Rucaparib	1	0.36 (0.30 – 0.43)	<0.001	N/A	

Talazoparib	1	0.54 (0.41 - 0.71)	<0.001	N/A	
Trials' phase					
Phase II	15	0.72 (0.59 – 0.88)	0.001	76%	0.006
Phase III	15	0.49 (0.41 – 0.59)	<0.001	86%	

Legend. PFS: progression-free survival; HR: hazard ratio; CI: confidence interval; N/A: not applicable; CT: chemotherapy; P pooled: p value of the pooled results; P subgroups: p values for the subgroup analyses; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; PARPi: PARP inhibitor.

Table 3. Overall response rates results

ORR	**No. of comparisons**	**Pooled RR (95% CI)**	**P pooled**	**I^2%**	**P subgroups**
Overall	27	1.35 (1.16 - 1.56)	<0.001	74%	N/A
Mutational Status					
Mixed/Wild-type	15	1.27 (1.08 - 1.49)	0.003	50%	0.44
Mutant	12	1.44 (1.09 - 1.91)	0.01	85%	
Tumor Site					
Ovarian	11	1.42 (1.17 - 1.73)	<0.001	50%	0.86
Breast	5	1.24 (0.82 - 1.87)	0.30	93%	
Prostate	3	1.75 (0.62 - 4.94)	0.29	78%	
Melanoma	2	1.37 (0.73 - 2.54)	0.32	0%	
NSCLC	1	0.93 (0.63 - 1.38)	0.97	N/A	
SCLC	2	1.64 (0.59 - 4.53)	0.34	82%	
Pancreatic	1	2.00 (0.85 - 4.70)	0.11	N/A	
Gastrointestinal	1	1.17 (0.68 - 2.04)	0.57	78%	
Line of Therapy					
1st-line +/- maintenance	9	1.14 (0.90 - 1.41)	0.29	87%	0.03
≥2nd-line	15	1.50 (1.23 - 1.82)	<0.001	39%	
Maintenance only	3	2.29 (1.28 – 4.07)	0.005	0%	
Control arm					
CT +/- placebo	18	1.20 (1.02 – 1.41)	0.03	77%	0.04
Placebo	3	2.29 (1.28 – 4.07)	0.005	0%	

Olaparib inferior dose	2	1.55 (0.95 - 2.54)	0.08	11%	
PARPi w/o Antiangiogenic drug	2	1.83 (1.37 – 2.45)	<0.001	5%	
Enzalutamide/Abiraterone	2	3.17 (0.11 – 89.64)	0.50	90%	
Type of PARP inhibitor					
Olaparib	13	1.52 (1.26 - 1.84)	<0.001	41%	
Veliparib	11	1.04 (0.89 - 1.22)	0.62	64%	
Niraparib	1	2.28 (1.35 - 3.83)	0.002	N/A	<0.001
Rucaparib	1	2.43 (0.98 - 6.06)	0.06	N/A	
Talazoparib	1	2.30 (1.67 - 3.16)	<0.001	N/A	
Trials' phase					
Phase II	18	1.22 (1.01 – 1.47)	0.04	62%	0.09
Phase III	9	1.63 (1.22 – 2.17)	<0.001	87%	

Legend. ORR: overall response rates; RR: relative risk; CI: confidence interval; CT: chemotherapy; P pooled: p value of the pooled results; P subgroups: p values for the subgroup analyses; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; N/A: not applicable.

Table 4. Overall survival results

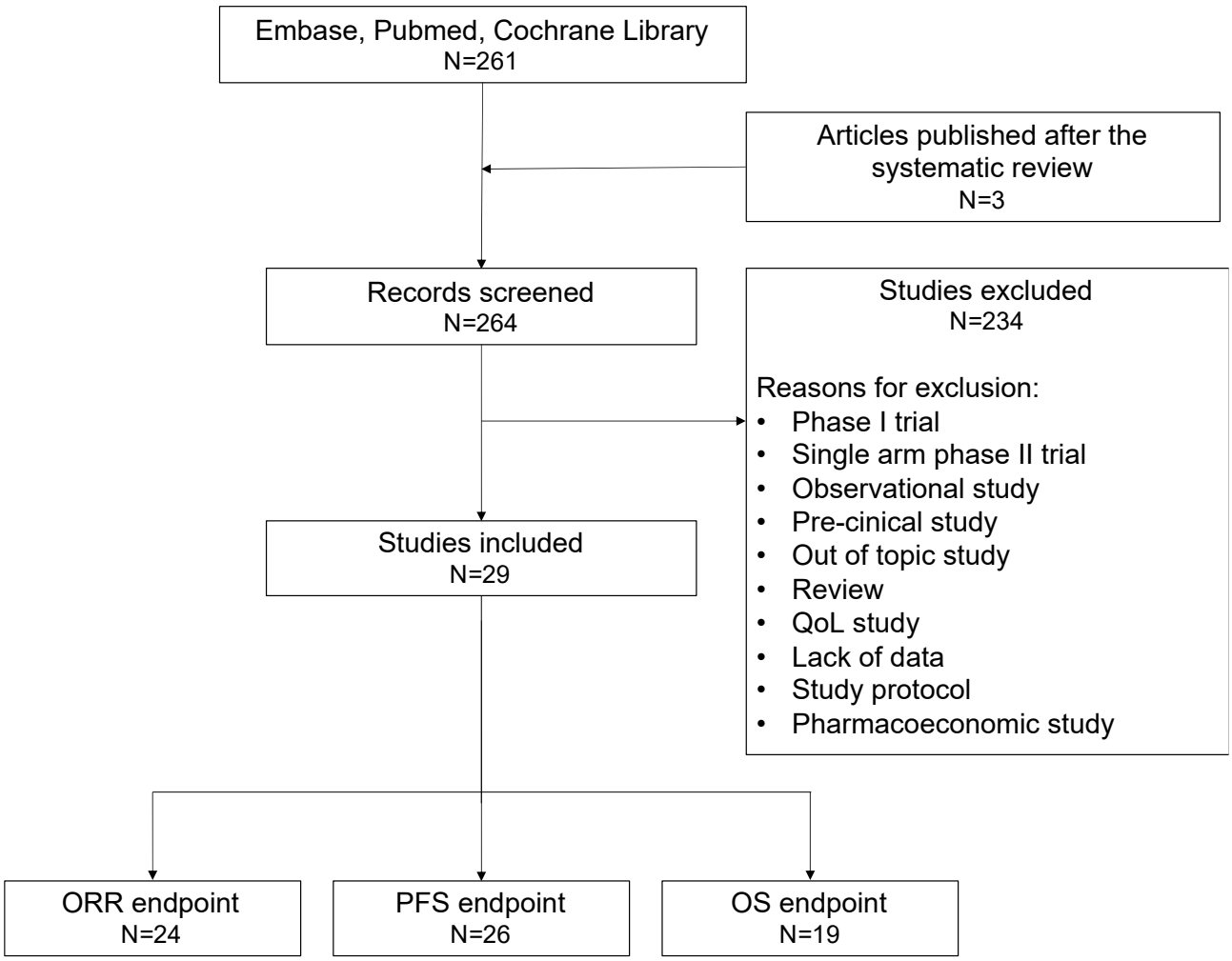
OS	**No. of comparisons**	**Pooled HR (95% CI)**	**P pooled**	**I^2%**	**P subgroups**
Overall RE	22	0.86 (0.80 - 0.93)	<0.001	7%	N/A
Overall FE	22	0.86 (0.80 – 0.92)	<0.001	7%	N/A
Mutation Status					
Mixed/Wild-type	11	0.84 (0.76 - 0.93)	<0.001	0%	0.57
Mutant	11	0.88 (0.79 - 0.98)	0.02	23%	
Tumor Site					
Ovarian	8	0.80 (0.69 - 0.93)	0.004	0%	0.82
Breast	5	0.94 (0.82 - 1.08)	0.37	51%	
Prostate	2	0.77 (0.59 - 0.99)	0.04	6%	
Melanoma	2	0.89 (0.71 - 1.13)	0.35	6%	
NSCLC	1	0.80 (0.54 – 1.19)	0.27	N/A	
SCLC	1	0.83 (0.64 - 1.08)	0.16	N/A	

Pancreatic	1	0.91 (0.56 - 1.48)	0.70	N/A	
Gastrointestinal	2	0.85 (0.69 - 1.05)	012	60%	
Line of Therapy					
1st-line +/- maintenance	8	0.92 (0.82 - 1.03)	0.14	34%	
≥2nd-line	9	0.84 (0.74 - 0.95)	0.005	0%	0.22
Maintenance only	5	0.77 (0.66 - 0.91)	0.003	0%	
Control arm					
CT +/- placebo	14	0.90 (0.83 - 0.99)	0.02	15%	
Placebo	5	0.77 (0.66 - 0.91)	0.003	0%	
PARPi w/o Antiangiogenic drug	1	0.64 (0.36 - 1.14)	0.13	N/A	0.21
Enzalutamide/Abiraterone	2	0.77 (0.59 - 0.99)	0.04	6%	
Type of PARP inhibitor					
Olaparib	12	0.81 (0.73 - 0.90)	<0.001	0%	
Veliparib	8	0.93 (0.83 - 1.05)	0.25	38%	0.30
Niraparib	1	0.70 (0.44 - 1.11)	0.13	N/A	
Talazoparib	1	0.85 (0.67 - 1.07)	0.17	N/A	
Trials' phase					
Phase II	13	0.90 (0.81 – 1.00)	0.05	30%	0.26
Phase III	9	0.82 (0.74 – 0.91)	<0.001	0%	

Legend. RE: random-effect model; FE: fixed-effect model; OS: overall survival; HR: hazard ratio; CI: confidence interval; CT: chemotherapy; P pooled: p value of the pooled results; P subgroups: p values for the subgroup analyses; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer. Subgroup analyses were conducted under a FE model.

3.8 FIGURES

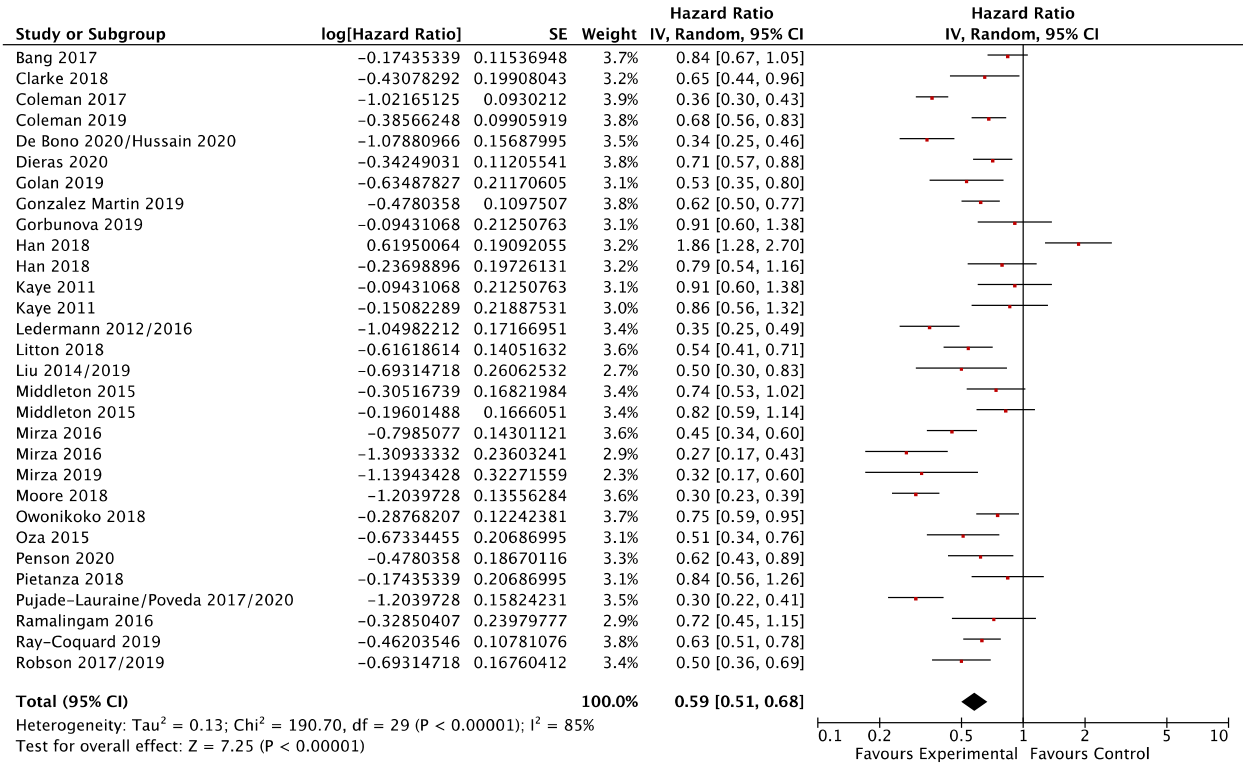
Figure 1. PRISMA flow-chart



Legend. ORR: overall response rates; PFS: progression-free survival; OS: overall survival

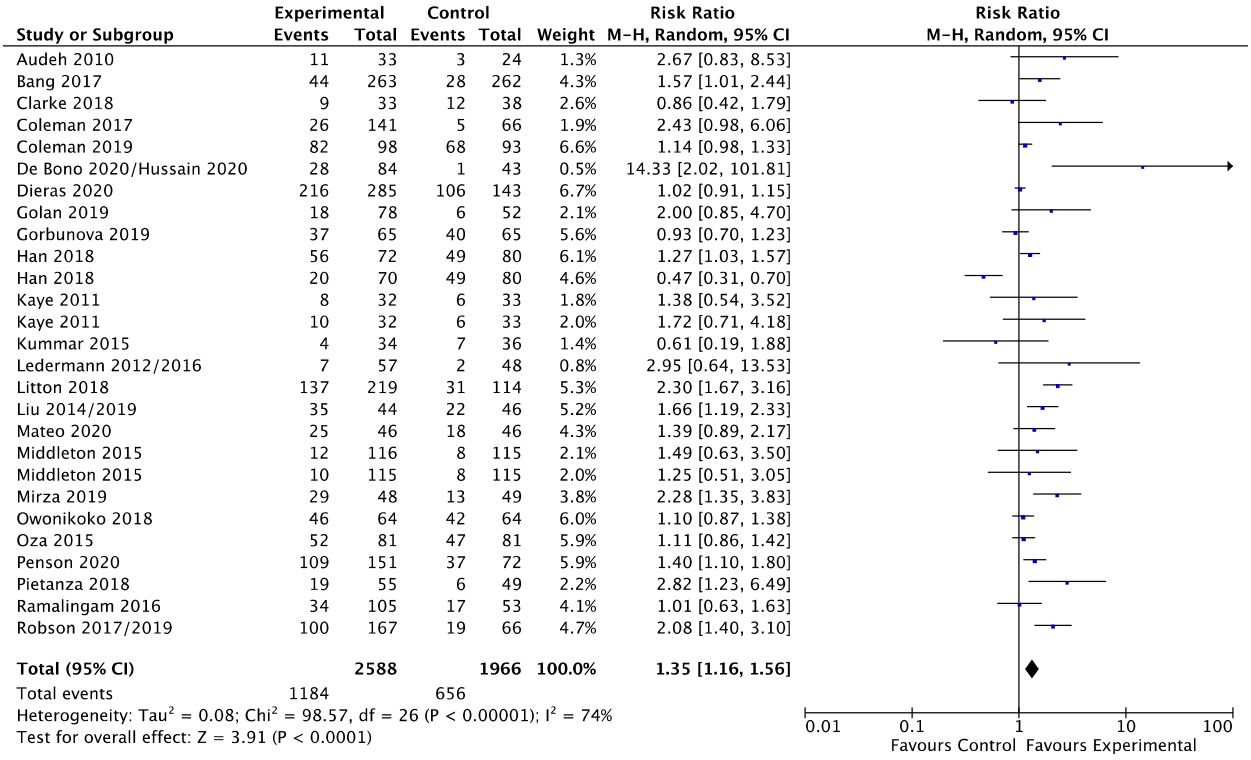
.

Figure 2. Forest plot for progression-free survival



Legend. SE: standard error; HR: hazard ratio; IV: inverse variance method; Random: random-effect model; CI: confidence interval.

Figure 3. Forest plot for overall response rates



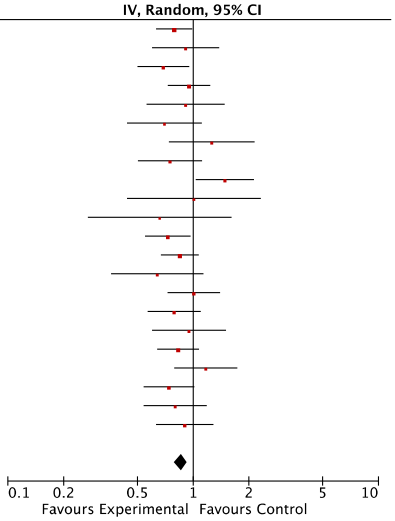
Legend. SE: standard error; RR: relative risk; IV: inverse variance method; Random: random-effect model; CI: confidence interval.

Figure 4. Forest plot for overall survival

A

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Bang 2017	−0.23572233	0.11546588	10.1%	0.79 [0.63, 0.99]
Clarke 2018	−0.09431068	0.21250763	3.3%	0.91 [0.60, 1.38]
De Bono 2020/Hussain 2020	−0.37106368	0.16432832	5.4%	0.69 [0.50, 0.95]
Dieras 2020	−0.05129329	0.13439666	7.8%	0.95 [0.73, 1.24]
Golan 2019	−0.09431068	0.24770807	2.5%	0.91 [0.56, 1.48]
Gonzalez Martin 2019	−0.35667494	0.23689062	2.7%	0.70 [0.44, 1.11]
Gorbunova 2019	0.23111172	0.27153919	2.1%	1.26 [0.74, 2.15]
Han 2018	−0.28768207	0.20381788	3.6%	0.75 [0.50, 1.12]
Han 2018	0.39406706	0.18498388	4.3%	1.48 [1.03, 2.13]
Kaye 2011	0.00995033	0.42394433	0.9%	1.01 [0.44, 2.32]
Kaye 2011	−0.41551544	0.45602953	0.8%	0.66 [0.27, 1.61]
Ledermann 2012/2016	−0.31471075	0.14445217	6.8%	0.73 [0.55, 0.97]
Litton 2020	−0.16487464	0.12020557	9.4%	0.85 [0.67, 1.07]
Liu 2014/2019	−0.4462871	0.29355314	1.8%	0.64 [0.36, 1.14]
Middleton 2015	0.00895974	0.16653774	5.3%	1.01 [0.73, 1.40]
Middleton 2015	−0.23572233	0.16922125	5.1%	0.79 [0.57, 1.10]
Moore 2018	−0.05129329	0.23445527	2.8%	0.95 [0.60, 1.50]
Owonikoko 2018	−0.18632958	0.13263139	7.9%	0.83 [0.64, 1.08]
Oza 2015	0.15700375	0.20037045	3.7%	1.17 [0.79, 1.73]
Pujade-Lauraine/Poveda 2017/2020	−0.30110509	0.16075564	5.6%	0.74 [0.54, 1.01]
Ramalingam 2016	−0.22314355	0.20053193	3.7%	0.80 [0.54, 1.19]
Robson 2017/2019	−0.10536052	0.18197701	4.5%	0.90 [0.63, 1.29]
Total (95% CI)			**100.0%**	**0.86 [0.80, 0.93]**

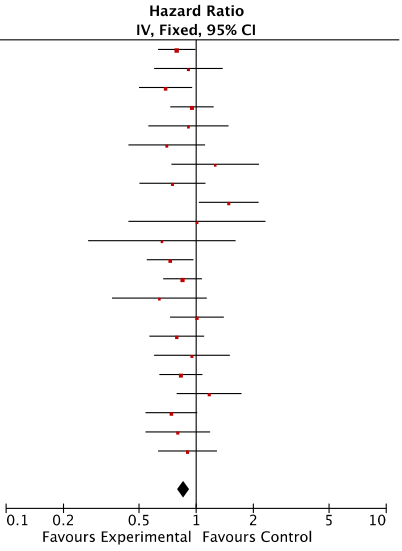
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 22.51$, df = 21 (P = 0.37); $I^2 = 7\%$
Test for overall effect: Z = 3.80 (P = 0.0001)



B

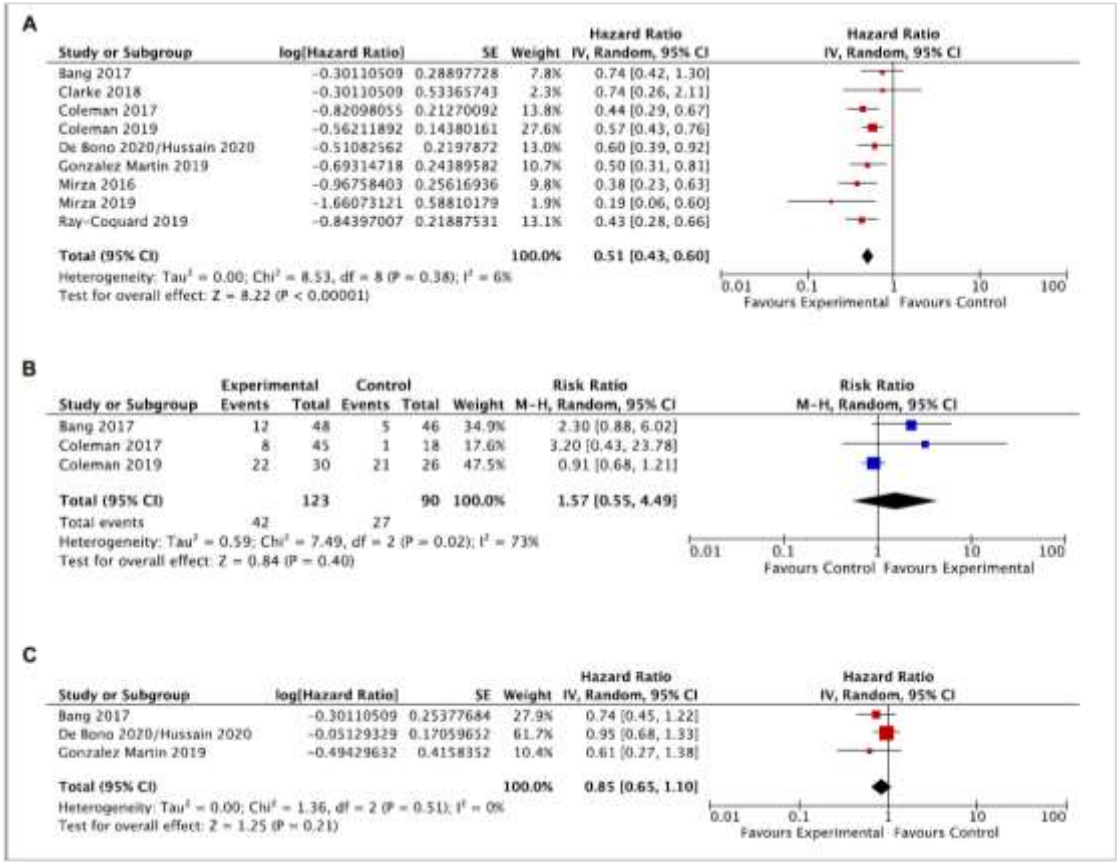
Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
Bang 2017	−0.23572233	0.11546588	10.8%	0.79 [0.63, 0.99]
Clarke 2018	−0.09431068	0.21250763	3.2%	0.91 [0.60, 1.38]
De Bono 2020/Hussain 2020	−0.37106368	0.16432832	5.3%	0.69 [0.50, 0.95]
Dieras 2020	−0.05129329	0.13439666	8.0%	0.95 [0.73, 1.24]
Golan 2019	−0.09431068	0.24770807	2.3%	0.91 [0.56, 1.48]
Gonzalez Martin 2019	−0.35667494	0.23689062	2.6%	0.70 [0.44, 1.11]
Gorbunova 2019	0.23111172	0.27153919	2.0%	1.26 [0.74, 2.15]
Han 2018	−0.28768207	0.20381788	3.5%	0.75 [0.50, 1.12]
Han 2018	0.39406706	0.18498388	4.2%	1.48 [1.03, 2.13]
Kaye 2011	0.00995033	0.42394433	0.8%	1.01 [0.44, 2.32]
Kaye 2011	−0.41551544	0.45602953	0.7%	0.66 [0.27, 1.61]
Ledermann 2012/2016	−0.31471075	0.14445217	6.9%	0.73 [0.55, 0.97]
Litton 2020	−0.16487464	0.12020557	10.0%	0.85 [0.67, 1.07]
Liu 2014/2019	−0.4462871	0.29355314	1.7%	0.64 [0.36, 1.14]
Middleton 2015	0.00895974	0.16653774	5.2%	1.01 [0.73, 1.40]
Middleton 2015	−0.23572233	0.16922125	5.0%	0.79 [0.57, 1.10]
Moore 2018	−0.05129329	0.23445527	2.6%	0.95 [0.60, 1.50]
Owonikoko 2018	−0.18632958	0.13263139	8.2%	0.83 [0.64, 1.08]
Oza 2015	0.15700375	0.20037045	3.6%	1.17 [0.79, 1.73]
Pujade-Lauraine/Poveda 2017/2020	−0.30110509	0.16075564	5.6%	0.74 [0.54, 1.01]
Ramalingam 2016	−0.22314355	0.20053193	3.6%	0.80 [0.54, 1.19]
Robson 2017/2019	−0.10536052	0.18197701	4.3%	0.90 [0.63, 1.29]
Total (95% CI)			**100.0%**	**0.86 [0.80, 0.92]**

Heterogeneity: $Chi^2 = 22.51$, df = 21 (P = 0.37); $I^2 = 7\%$
Test for overall effect: Z = 4.02 (P < 0.0001)



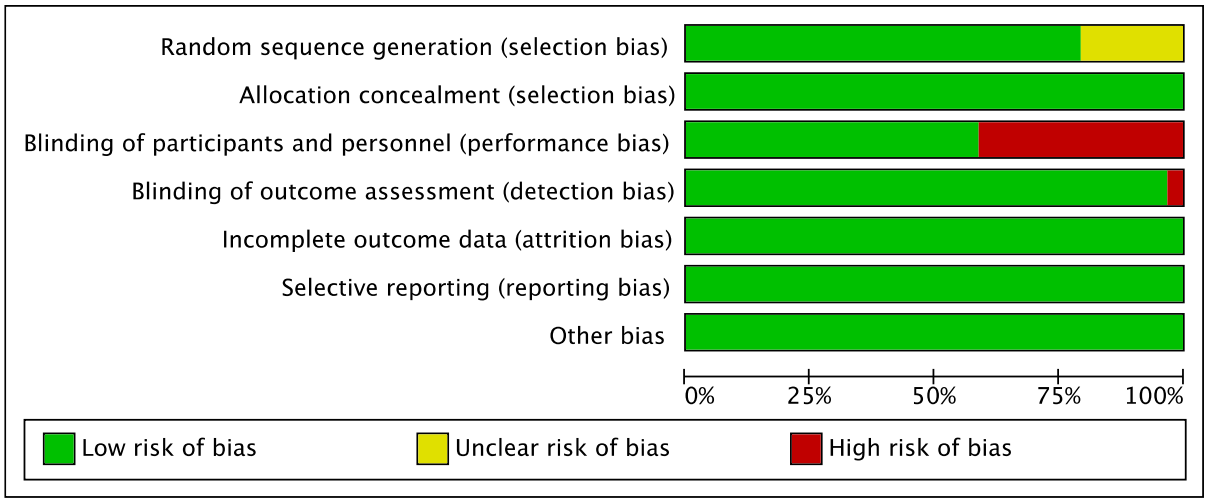
Legend. A: results under random-effect model; B: results under fixed-effect model; SE: standard error; HR: hazard ratio; IV: inverse variance method; Random: random-effect model; Fixed: fixed-effect model; CI: confidence interval.

Figure 5. Pooled results of HRD-positive tumours. (A) PFS result; (B) ORR result; (C) OS result.



Legend: OS: overall survival; PFS: progression-free survival; ORR: overall response rate; SE: standard error; IV: inverse-variance method; M-H: Mantel-Haenszel method; Random: random-effects model; CI: confidence interval; HRD: homologous recombination deficiency.

Figure 6. Risk of bias analysis



3.8 SUPPLEMENTARY METHODS

Data Analysis

The I2 statistic (0–100%) was used to assess the proportion of variability in the results that was attributable to heterogeneity between the trials (1). Being the trials included potentially heterogeneous, we preplanned to conduct the analyses using the random-effects (RE) model of Der-Simonian and Laird (2). In case of non-significant heterogeneity, a fixed-effects (FE) model was subsequently applied to confirm the result and perform the pre-specified subgroup analyses (2). To further investigate heterogeneity, we used a graphical method (Baujat plot) (3). To assess the stability of the pooled results, sensitivity analysis (influence analysis) were performed by sequential omission of individual studies (Leave-One-Out-method). Moreover, outliers and influential case diagnostics were identified generating the influence diagnostic plots (4). This latter method is based on the impact of excluding studies on various statistics such as the summary externally standardized residuals, DFFITS values, Cook's distances, covariance ratios, leave-one-out estimates of the amount of heterogeneity, hat values, and weights (4). As a rule of thumb, influential cases are studies with very extreme values (respect to a proposed cut-off), and in the graphs are displayed in red (4). The meta-analytic models were thus run both with and without influential effect sizes.

References

1. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat. Med. 2002; 21(11):1539–1558.

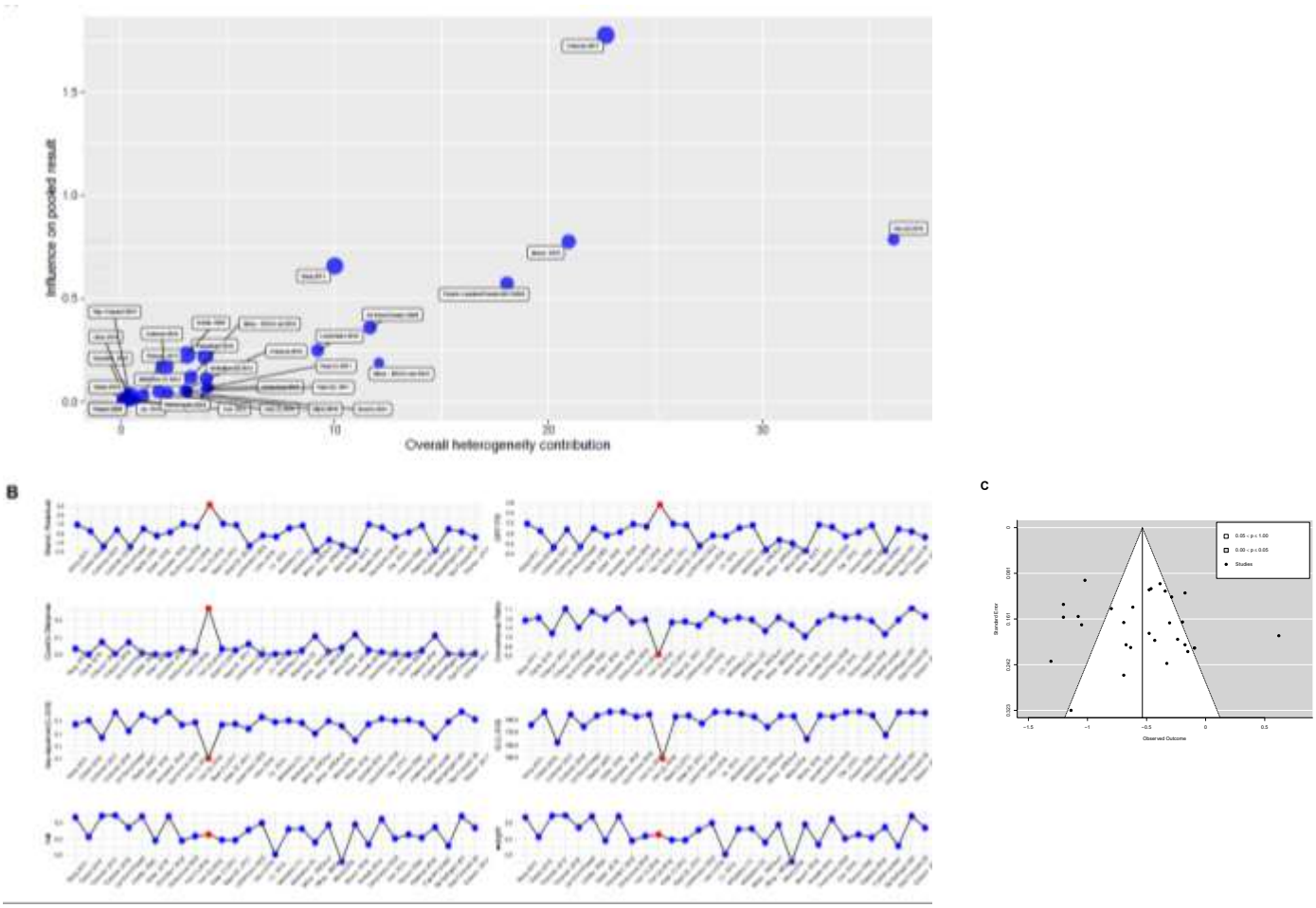
2. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res. Synth. Methods 2010; 1(2):97–111.

3. Baujat B, Mahé C, Pignon J-P, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat. Med. 2002; 21(18):2641–2652.

4. Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. Res. Synth

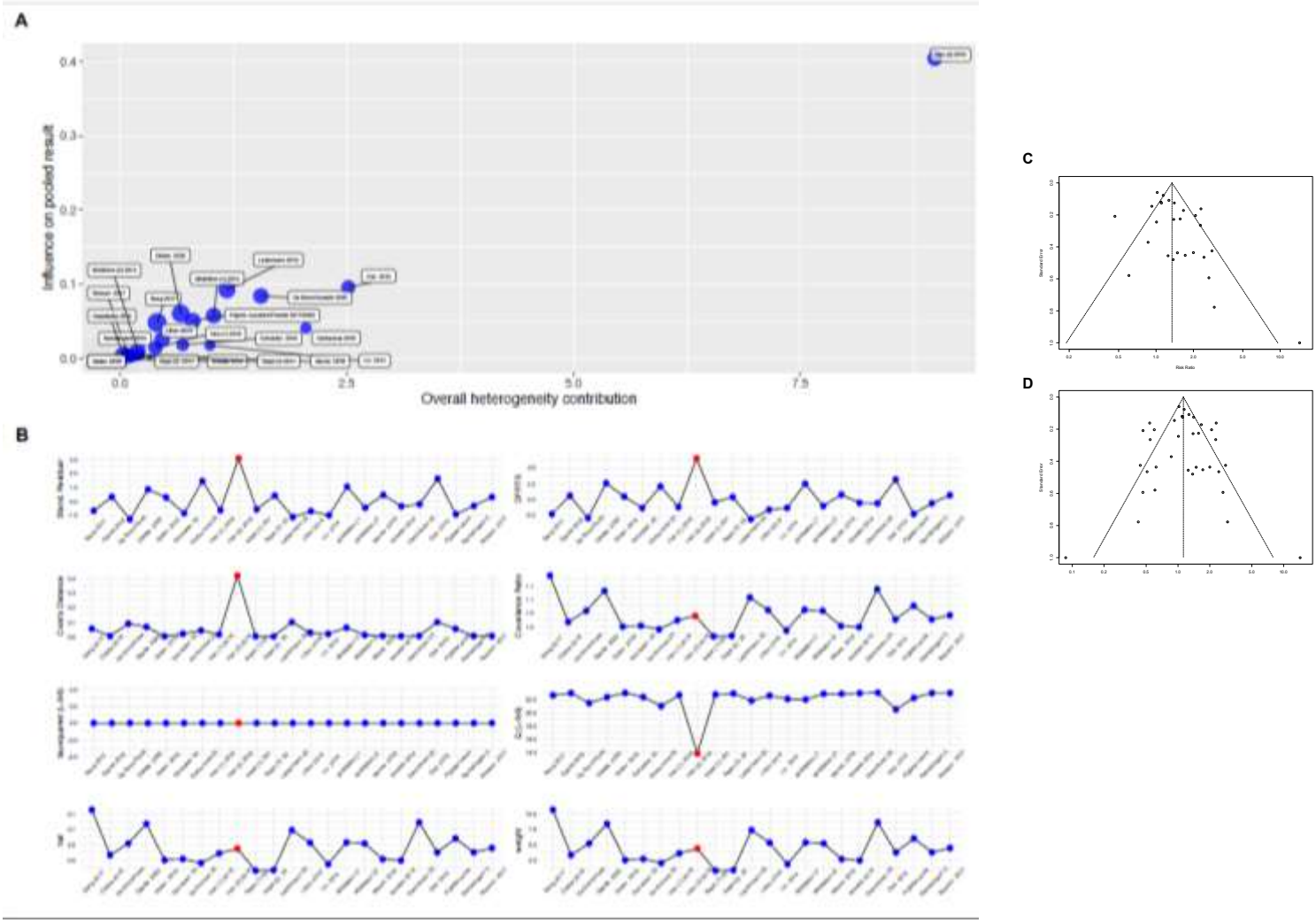
3.9 SUPPLEMENTARY FIGURES

Supplementary figure 1. Baujat plot, influence diagnostic plots and funnel plot for the PFS endpoint



Legend. A: Baujat plot; B: Influential analysis plots; C: funnel plot; Stand. Residual: Standardized Residuals, which are the residuals divided by the estimates of their standard errors. They test the hypothesis that the corresponding observation does not follow the regression model that describes the other observations; DFFITS: Studentized DFFIT, where Studentization is achieved by dividing by the estimated standard deviation of the fit at that point. DFFIT is the change in the predicted value for a point, obtained when that point is left out of the regression; Cook's Distance: an estimate of the influence of a data point when performing a least-squares regression analysis; Covariance Ratio: a parameter expressing the means of the change in the variance–covariance matrix of the parameter estimates. It indicates if the removal of the *i*th study can yields more precise estimates of the model coefficients; Tau-squared: the between-study variance, indicates how much residual heterogeneity exists which has not been explained by the covariate; Q: the statistic of the homogeneity test; hat: mathematical parameter indicating high/low leverage studies; weight: a mathematical parameter depending on the sampling variance and Tau-squared. Essential to calculate the predicted (average) effect size and 95% confidence interval, through its variance. (Ref. Viechtbauer W & Cheung MWL Res Synth Methods 2010; 1(2):112-125).

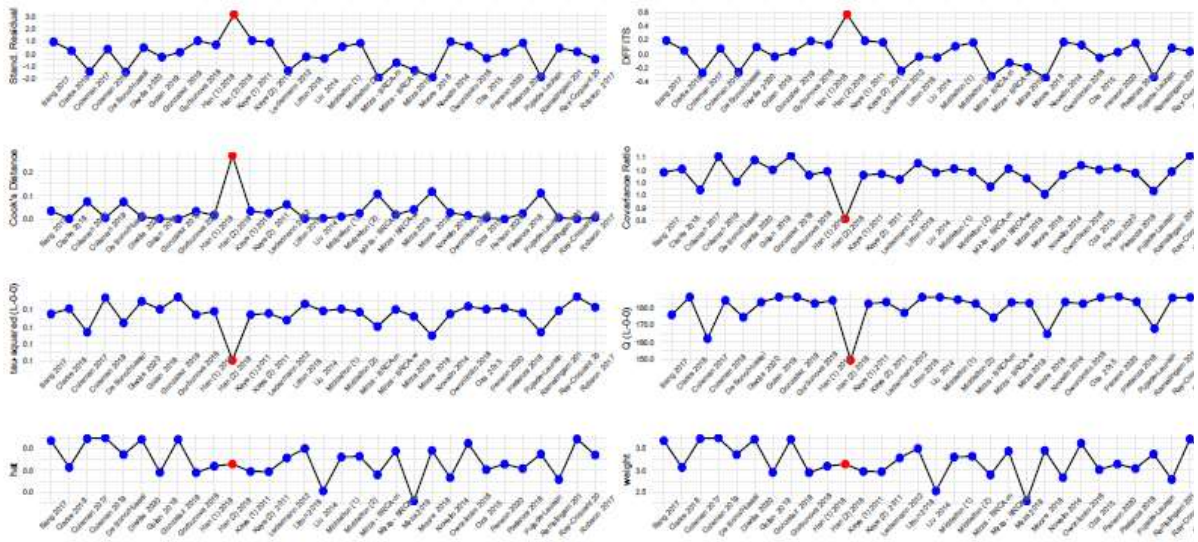
Supplementary figure 2. Baujat plot, influence diagnostic plots and funnel plots for the ORR endpoint



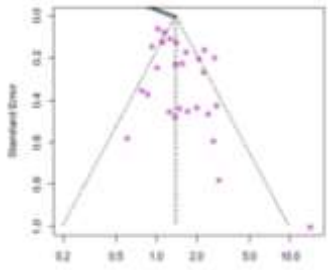
Legend. A: Baujat plot; B: Influential analysis plots; C: funnel plot; D: funnel plot with *trim-and-fill* analysis; Stand. Residual: Standardized Residuals, which are the residuals divided by the estimates of their standard errors. They test the hypothesis that the corresponding observation does not follow the regression model that describes the other observations; DFFITS: Studentized DFFIT, where Studentization is achieved by dividing by the estimated standard deviation of the fit at that point. DFFIT is the change in the predicted value for a point, obtained when that point is left out of the regression; Cook's Distance: an estimate of the influence of a data point when performing a least-squares regression analysis; Covariance Ratio: a parameter expressing the means of the change in the variance–covariance matrix of the parameter estimates. It indicates if the removal of the *i*th study can yields more precise estimates of the model coefficients; Tau-squared: the between-study variance, indicates how much residual heterogeneity exists which has not been explained by the covariate; Q: the statistic of the homogeneity test; hat: mathematical parameter indicating high/low leverage studies; weight: a mathematical parameter depending on the sampling variance and Tau-squared. Essential to calculate the predicted (average) effect size and 95% confidence interval, through its variance. (Ref. Viechtbauer W & Cheung MWL Res Synth Methods 2010; 1(2):112-125).

Supplementary figure 3. Influence diagnostic plots and funnel plot for the OS endpoint

A



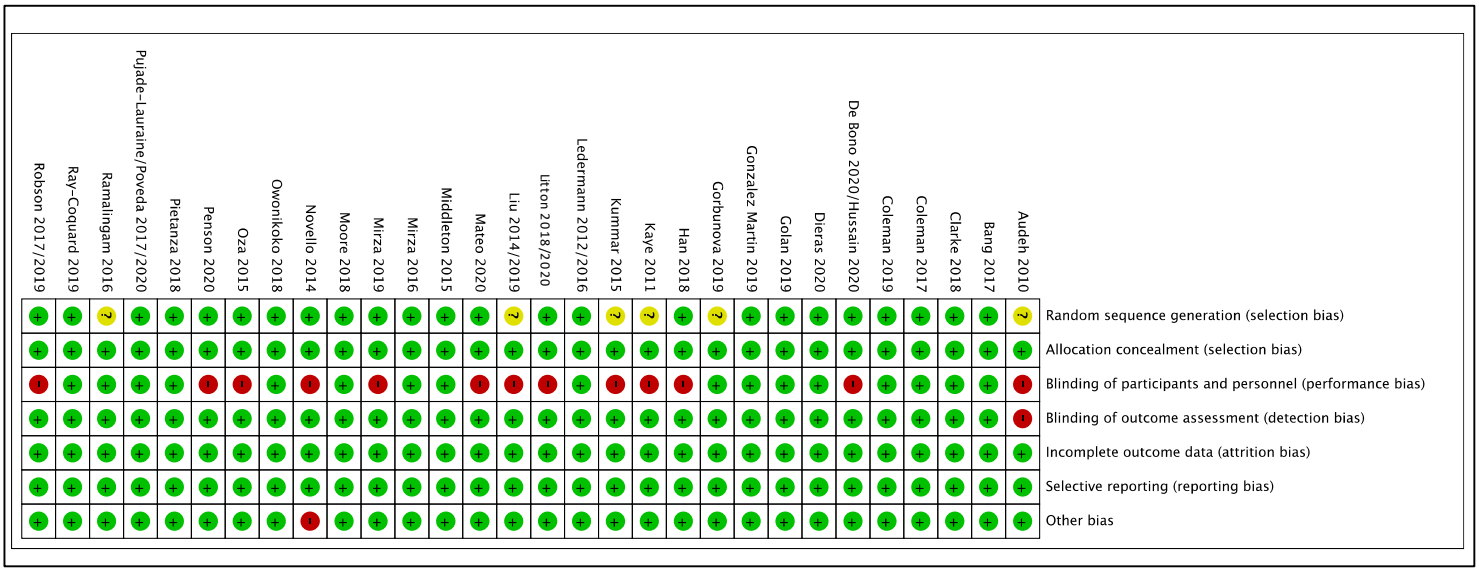
B



Legend. A: Influential analysis plots; B: funnel plot; Stand. Residual: Standardized Residuals, which are the residuals divided by the estimates of their standard errors. They test the hypothesis that the corresponding observation does not follow the regression model that describes the other observations; DFFITS: Studentized DFFIT, where Studentization is achieved by dividing by the estimated standard deviation of the fit at that point. DFFIT is the change in the predicted value for a point, obtained when that point is left out of the regression; Cook's Distance: an estimate of the influence of a data point when performing a least-squares regression analysis; Covariance Ratio: a parameter expressing the means of the change in the variance–covariance matrix of the parameter estimates. It indicates if the removal of the *i*th study can yields more precise estimates of the model coefficients; Tau-squared: the between-study variance, indicates how much residual heterogeneity exists which has not been explained by the covariate; Q: the statistic of the homogeneity test; hat: mathematical parameter indicating high/low leverage studies; weight: a mathematical parameter depending on the sampling variance and Tau-squared. Essential to calculate the predicted (average) effect size and 95% confidence interval, through its variance. (Ref. Viechtbauer W & Cheung MWL Res Synth Methods 2010; 1(2):112-125).

Supplementary figure 4. Detailed risk of bias analysis

Legend. Red circle: high risk; yellow circle: unknown risk; green circle: low risk.



4. ADJUSTING SURVIVAL OUTCOMES FOR TREATMENT SWITCHES: STATISTICAL METHODS AND A PRACTICAL APPLICATION IN BREAST CANCER

ABSTRACT

The intention-to-treat (ITT) analysis is the established method for evaluating the efficacy of a new treatment in randomized clinical trials. Although the ITT analysis is a valid test to compare two treatment strategies, its estimate of the treatment effect can differ considerably from the on-treatment effect estimates when there is a considerable non-adherence. In event-driven trials with long-term follow-up, non-adherence to study drug may be extensive, particularly in populations with substantial morbidities. If a large proportion of follow-up time and accumulation of events occur while patients are not taking randomized treatment, the on-treatment effect may be underestimated. It may be a clinically relevant question to estimate the efficacy that would have been observed if no patients had switched, for example, to estimate 'real-life' clinical effectiveness for a health technology assessment. Several commonly used statistical methods are available that try to adjust time-to-event data to account for treatment switching, ranging from naive exclusion and censoring approaches to more complex inverse probability of censoring weighting (IPCW) and rank-preserving structural failure time (RPSFT) models. These are described, along with their key assumptions, strengths, and limitations. Key considerations include having a clearly articulated rationale and research question and a well-designed trial with sufficient good quality data collection to enable robust statistical analysis. No analysis method is universally suitable in all situations, and each makes strong untestable assumptions. There is a need for further research into new or improved techniques.

I provide a practical application based on derived data from a two-arms breast cancer trial (1) to assess adjuvant (either delayed or not) tamoxifen treatment effect taking into account compliance to treatment in order to provide a robust and reliable estimate of treatment effect and predictions of survival probabilities.

Summary of Statistical Methods applied- Methods used to estimate survival time after switching: "Naive" methods (Exclude switchers, Censor at switch, Time varying covariate) and "Complex" methods (Inverse Probability of Censoring Weighting (IPCW; observational), Marginal Structural Model (MSM) and Rank Preserving Structural Failure Time model (RPSFT; randomisation based)

This chapter regards a project in which I'm involved during my abroad period (April-June 2021) at Gustave Roussy Institute (Onco-Stat Team). The collaboration is still going on and we are performing other sensitivity analysis: the draft is in preparation

Draft in preparation**: Long-term effect of adjuvant tamoxifen: adherence-based analysis**

Giudici F., Bardet A, Pistilli B., Vaz-Duarte-Luis IM, Micheils S.

4.1 INTRODUCTION

Treatment switch in a randomized controlled parallel group trial is when a patient randomized to one treatment arm changes to the alternative treatment during the study. This switch may be built into the trial design, for example, allowing placebo patients to switch to experimental treatment following occurrence of a shorter-term primary endpoint. Alternatively, switching can happen 'spontaneously' (at the choice of the patient and their treating physician) if the alternative treatment is available in clinical practice or through other clinical trials.

Unfortunately, treatment switching can introduce complexities in estimating treatment effects for longer-term outcomes, most notably overall survival (OS). Suppose an experimental treatment extends OS and that control group patients benefit from switching to the experimental treatment. In this case, the observed OS difference between the experimental and control arms would be smaller in magnitude than what would have been seen had switching not occurred. Whether this is problematic depends on the population parameter of interest. In health technology assessment (HTA), judgments around the cost-effectiveness of introducing experimental treatments into clinical practice typically rely on accurate OS comparisons with current standard care, where switching to the experimental treatment would not be possible. (2-3). Hence, for the purpose of HTA decision making, it is often desirable to adjust OS estimates to reflect what would have been observed had control group patients not switched treatments. It is worth noting that treatment switching in the opposite direction, from the experimental to the control treatment, does not usually pose the same problem for HTA decision making. Typically, no adjustment for treatment switching would be necessary provided the switches reflect what might occur with the introduction of the experimental treatment into clinical practice (eg, patients ceasing the experimental treatment and commencing existing (control) treatments because of disease progression or toxicity). If the switch therapy is effective, this will reduce the estimated treatment difference for long-term trial endpoints between the randomized arms. An intention-to-treat (ITT) analysis of the observed data will underestimate the treatment benefit that would have been seen without switch. Hence, if the relevant clinical question is to compare the long-term effectiveness of experimental treatment with a regimen without any experimental treatment, the ITT analysis will provide a biased answer.

A variety of statistical methods have been proposed to adjust or treatment switching in oncology trials, or equivalently, to estimate a switching-adjusted estimand. Crude approaches to adjust for this bias may be attempted (4): per-protocol analyses, for instance, would exclude patients who crossover from the analyses; on-treatment analyses would include these patients, but censor their OS time at crossover; as-treated analyses would account for the change in treatment at crossover using time-dependent variables in the analyses. Exclusion of patients or even portions of their follow-up time introduces selection bias and breaks the randomization, leading to biased results. More advanced statistical methods are therefore required to properly address this problem.

Rank-preserving structural failure time (RPSFT) models (5-6) and inverse probability of censoring weighted (IPCW) analyses (7-8) have been applied to adjust for the effect of crossover. These methods were developed to deal with complex confounding caused by non-adherence to randomized therapy driven by changes in the patients' condition. The two approaches differ in terms of how the effect of treatment is expressed, how it is estimated, and the assumptions invoked. Other approaches to deal with treatment switches have been proposed (9-11), but the RPSFT and IPCW methods remain the most commonly used to adjust for crossover bias, and have been applied in analyses of trials (12-14) and successfully incorporated in health technology assessments (15-17). Simulation studies have shown that these methods tend to produce more accurate estimates of the

switching-adjusted estimand than simple adjustment methods or a standard intention to treat (ITT) analysis, but their performance can be compromised when underlying assumptions are violated (18-20). A good understanding of these methods is important to be able to properly assess the validity and plausibility of the results from these approaches. Sound statistical advice is critical given the variety of potential methods. This project aims to provide descriptions of the RPSFT and IPCW approaches, highlighting their similarities and differences, and discussing their suitability for the crossover problem. Their application is illustrated with an example using data from a two-arms breast cancer trial (1).

4.2 METHODS

Naïve methods

These methods are fairly simple to implement but subject to large biases.

A) **Intention-to-treat (ITT).** The term ITT analysis is used for the comparison of observed data between randomized treatment groups—this may be in, for example, a modified ITT analysis set. This is the primary analysis of the trial and addresses the question of efficacy of the treatments as randomized within the circumstances of the trial. This is usually the primary question of interest for regulators, as described in the International Conference on Harmonisation (ICH) E9 guideline (21). However, another question of interest may be the efficacy of the treatments if switch had not occurred. If the experimental treatment is effective in later lines of therapy, the ITT estimate will be biased in favor of the control arm for this objective (19).

B) **Exclude switchers**. Simply removing switchers from the analysis and comparing the remaining control arm non-switchers to all patients in the observed experimental arm makes the assumption that the control arm switchers and non-switchers have the same prognosis. In other words, there are no confounders—variables that influence both survival and the decision to switch. This is highly unlikely to hold, leading to bias (19). Patients also have to live long enough to be able to switch, so longer-living individuals are removed. This approach also breaks randomization, and reduces the number in the control arm, which can be a particular problem with 2:1 randomization. Given the flaws with this method and the availability of alternatives, it is not recommended.

C)**Censor switchers.** In a standard survival analysis, it is assumed that censoring is independent of outcome. If the censoring is due to switch, then this is highly unlikely to hold as outlined earlier. So an analysis censoring patients at the time of switching also relies on the unlikely assumption of no confounders and is often biased (19)

D) **Time-varying covariate for treatment or switch**. A time-varying covariate for either exposure or switch to experimental treatment could be used in a survival model. However, this also relies on the 'no confounders' assumption. If there are confounding variables that influence both the time-varying treatment covariate and survival, the result will be biased (22)

Complex methods

These methods are technically harder to implement but try to reduce the bias seen with naive methods by not assuming that switch and prognosis are unrelated.

A) Rank-Preserving Structural Failure Time Models (RPSFT)

The RPSFT approach was proposed by Robins and Tsiatis (6) to deal with non-compliance in randomized trials when estimating the causal effect of a treatment had all patients followed the study protocol. A patient who switches treatment has, in theory, an unknown counterfactual event time: the time-to-event if no experimental treatment were received. The RPSFT method is a semiparametric approach that estimates the counterfactual event time of patients. The situation faced in oncology trials with crossover is slightly different in that the change in treatment may be part of the study protocol, or may occur after patients enter a new phase of the study. The underlying analytical issue of estimating an effect had all patients remained on the original treatment is, however, the same as that addressed by Robins and Tsiatis. This method is based on an accelerated failure time (AFT)model, which assume that exposure to treatment has a multiplicative effect e^{φ} on a patient's observed survival time (23.). The on-treatment effect can be estimated using a causal model to relate e^{φ} and patient's observed failure time to their counterfactual failure time (19). This approach aims to estimate the efficacy of the study drug as if patients maintained their randomized treatment for the entire study duration. As described in Morden's publication (19), the observed failure time T_{Ri} for the i-th patient is related to his or her counterfactual failure time T_{Li}, the time that would have been observed if no treatment had been received. $T_{Exposed}$ represents the underlying failure time of a patient exposed to the treatment and $T_{Unexposed}$ represents the underlying failure time if no treatment was given to the patient. For all patients, T_{Ri} was composed of the time when the patient was exposed and not exposed to the treatment:

$$\mathbf{T_{Ri} = T_{Exposed,i} + T_{Unexposed,i}}$$

Under an AFT model, T_{Li} can be derived for each patient using

$$\mathbf{T_{Li} = T_{Unexposed,i} + e^{-\varphi}\, T_{Exposed,i}}$$

A multiplicative effect of $\boldsymbol{e^{-\varphi} < 1}$ indicates a beneficial treatment effect while $\boldsymbol{e^{-\varphi} > 1}$ represents a detrimental treatment effect.

For patients randomized to placebo who were never exposed to treatment $T_{Li} = T_{Ri}$. In the RPSFTM framework, T_{Li} is a pre-randomization variable and is independent of randomization. Therefore, the treatment effect φ can be obtained from a grid search over a range of plausible values of φ until T_{Li} is equally distributed between the two treatment groups using a test-based method (i.e. log-rank) (4). This is an iterative process of searching a grid of possible φ values and the corresponding test statistic for the null hypothesis: T_{Li} is independent of randomized treatment. The test statistic could be taken from any standard survival analysis model, for example, log rank, Wilcoxon, and Cox, with or without covariates. It may be preferable to use the same model as the ITT analysis. The value of φ that satisfies the null hypothesis (test statistic = 0) is selected. Care should be taken to ensure this is a unique solution. If no unique solution can be found, the plausibility of the different values should be considered. An unweighted test statistic such as log rank can result in uncertainty in estimating φ.Weighting schemes (Wilcoxon, Tarone–Ware, and Peto- Peto tests) or unweighted and adjusted methods (such as multivariate Cox regression) can increase the power of the test statistics and increase the likelihood of achieving a unique estimate of φ. This process relies on the assumption that the treatment arms are balanced in terms of underlying prognostic factors so that the null hypothesis holds. This should be reasonable in a large trial with effective randomization. If the trial is small or the data are from a smaller subgroup where there are chance imbalances in observed baseline prognostic factors, they can be included as covariates to adjust for this, as in an

ITT analysis. Moreover, the model assumes that the accelerated factor owing to the experimental treatment is constant over time for all patients no matter when it was first received, which is known as the "**common treatment effect**" assumption (24) with the effect applying immediately upon commencement and ceasing immediately upon discontinuation of treatment.

Like any statistical analysis, the validity of the RPSFTM hinges on estimation performance and the suitability of underlying assumptions. G-estimation performance can be assessed by plotting potential values for the accelerated factor against the observed test statistic; if successful, the procedure should identify a unique solution where the test statistic equals 0. The success of g-estimation and the suitability of model assumptions can also be assessed by comparing counterfactual survival times between randomized groups using a Kaplan-Meier plot. Assuming randomization is successful in balancing prognostic variables, counterfactual survival times should be equivalently distributed across randomized groups. Given the untestable nature of the common treatment effect assumption, clinical input into its plausibility is also critical. If the beneficial effect of the experimental treatment is anticipated to be quite different between patients originally randomized to the experimental arm and patients who switched to the experimental treatment partway through the trial, then alternatives to the RPSFTM should be considered.

B) Inverse Probability of Censoring Weights

Unlike the RPSFTM, which attempts to recreate the distribution of survival times had treatment switching not occurred, the IPCW method involves adjusting for the effects of switching during estimation of the treatment effect. In the context of treatment switching from the control to the experimental treatment, the IPCW method involves (1) censoring patients at the time of switching and (2) addressing potential selection bias by reweighting remaining control group patients still at risk of death by the inverse of their probability of not switching. Higher weights are assigned to non-switching patients with similar characteristics to switching patients, allowing these patients to represent both themselves and switching patients in the analysis. (18) To satisfy an assumption of "no unmeasured confounders" (NUC), the weights should be calculated from a correctly specified model, which includes all baseline and time-varying characteristics predictive of both treatment switching and OS; in general, this necessitates extensive data collection. Another important requirement of the IPCW method is that the probability of treatment switching must always be less than 1 for all possible predictor combinations; otherwise, weights cannot be estimated. (25-26). The key principle of the IPCW method is to recreate the population

that would have been observed had patients remained on assigned study drug. It does so by censoring data at the time of study drug discontinuation for non-adherent patients and assigning weights that are proportional to the inverse of the probability of remaining on study drug given each individual patient's characteristics. The underlying assumption for IPCW is that censoring of events due to discontinuation of study drug is independent of failure time (i.e. missing at random) (27).

To derive the weights, the patients' follow-up time up until the time of study drug discontinuation was partitioned into several intervals (C(t), t= 1, 2, ….)). The probability of remaining on study drug at the end of each time interval pr ((c(t)=0) adjusted for baseline variables, and time-varying confounders were estimated using a Cox proportional hazards or pooled logistic regression model. To avoid possible extreme values when taking the inverse of these probabilities, these weights were stabilized by multiplying the probability of remaining on study drug, conditional only on baseline variables. The equation for calculating the stabilized weight is given as follows (28):

$$\text{Stabilized weight } w_i(t) = \prod_t \frac{pr[C(t) = 0 | C(t-1) = 0, R = r_i, V = v_i]}{pr[C(t) = 0 | C(t-1) = 0, R = r_i, V = v_i, L(t-1) = l_i(t-1)]}$$

where i is the patient index, t is the time interval index, C(t)=1 if patient was censored (i.e. stopped study drug) within the time interval t and 0 otherwise, R denotes randomized treatment (0 for placebo, 1 for treatment), V is a vector of baseline covariates and L(t-1) is a vector of time-varying confounders which are specified in the succeeding texts. All clinically relevant baseline covariates and time-varying confounders that were considered to be affected by prior exposure to study drug in both treatment groups were included in the calculation of the weights. The IPCW method requires the use of a marginal structural model (MSM) to describe the relationship between the treatment arm and the primary endpoint, that is overall survival. A Cox MSM is a Cox model that estimates marginal effects that would have been observed in the absence of switch or discontinuation of treatment. More specifically, assuming that all confounders have been observed, applying these weights to the Cox partial likelihood estimators creates a pseudo–population that would have been studied if the patients had complied with their assigned treatment arm (3). The "hypothetical" causal effect of the experimental treatment on the overall survival is obtained using these IPCW weights in a Cox marginal structural model. Specifically, the estimand represents the effect in a hypothetical setting where all patients would have continued to take the randomization treatment.

In Table 1, recommendations were provided on what should be reported after a switching-adjusted analysis (29). The list of recommendations includes items that apply to all switching-adjusted analyses and items specific to individual methods of adjustment. No single analysis method is 'best' in all situations. Each method has a set of strong assumptions that are often untestable, and the clinical and statistical plausibility of those will vary according to the disease and treatments. In some situations, some methods cannot be applied; others are known to have large biases (30). Therefore, a trial-specific assessment must be made to determine which, if any, methods are appropriate. The key assumptions and limitations to consider when making this assessment are provided in **Table 2** (3)

4.3. APPLICATION

An example from oncology is provided where methods to adjust Invasive Disease-Free Survival (iDFS) and Overall Survival (OS) for treatment switch have been applied.

Overview of the trial

Despite meaningful, incremental improvements in screening, in local treatment and in adjuvant systemic therapies for breast cancer, there remains a significant risk of late relapse in hormone receptor (HR)-positive disease. Tamoxifen adjuvant treatment in early breast cancer has proved to be efficient on increasing Invasive Disease-Free Survival (iDFS) and Overall Survival (OS) compared to placebo (31). Five years of tamoxifen or an aromatase inhibitor for all patients with HR-positive early breast cancer is considered standard; however, there are data to support an improvement of survival outcomes extending tamoxifen treatment up to 10 years (32)

This study aims at estimating the switching-adjusted treatment effect of tamoxifen for OS and iDFS using several causal inference methods, performing simulation study to assess the operational characteristics of each method.

The analysis was based on data from TAM1 trial (1). The original cohort includes 3973 women with breast cancer and 2 to 3 years of tamoxifen exposure at randomization, of whom 1882 were allocated to short term tamoxifen (ST) (tamoxifen was immediately stopped after randomization) and 1911 to long term tamoxifen (LT) (patients continued tamoxifen for further 10 years). Eligibility criteria were: age up to 75 years, Negative/Positive Lymph Nodes and estrogen status Positive /Negative. Endpoints evaluated: Invasive Disease Free interval (iDFI) defined as the time elapsed between inclusion and local or regional recurrence, metastases and death preceding one of the former events; Overall Survival i.e. the time elapsed between inclusion and death, regardless of its cause. Population characteristics were well balanced between the two arms as reported in table 3. Median duration of follow up was 111 months in the short term group and 116 months in the long duration group while Median tamoxifen duration at randomization was 27 months in both groups.

The main aim of the TAM1 trial was to assess the effects of the Tamoxifen treatment duration on mortality and on recurrence in early breast cancer. In original trial were performed only intention to treat analysis: they compared OS/iDFS data for treatment (LT) vs control (ST) ignoring that some patients could stop tamoxifen during follow-up. The results were the following: OS did not differ between the two groups: 8 years OS was 79 % in both groups, while as regard as iDFS, Long Term treatment showed a 23% relative reduction in relapses (RR:0.77 95%CI 0.65 to 0.91). Moreover, significant risk reductions in disease free survival were observed in ER positive and node positive patients receiving long term tamoxifen but not in women who were node negative.

Adjusting for treatment switching

In TAM01 about 27% of patients in LT arm stop treatment during follow-up. In addition to the standard statistical methods (ITT, per protocol, censorship of switchers, exclusion of switchers and use of time-varying treatment variable, complex methods were applied: Inverse Probability of Censoring Weights (IPCW), Marginal Structural Models (MSM) and Rank-Preserving Structural Failure Time Models (RPSFTM) to account for confounding associated with treatment switching (20,24).

The switching process from the Tamoxifen exposure to stop treatment can be represented as in the Figure 1. The graph shows that the switching process cannot be considered at random because it may be affected by many variables under study (Age, tumor's patterns ..), and it may also affect the outcome of interest (Death /Relapse). The key question is related to the estimate the true benefit of experimental drug group on survival end points that would have been estimated if there were no treatment switches

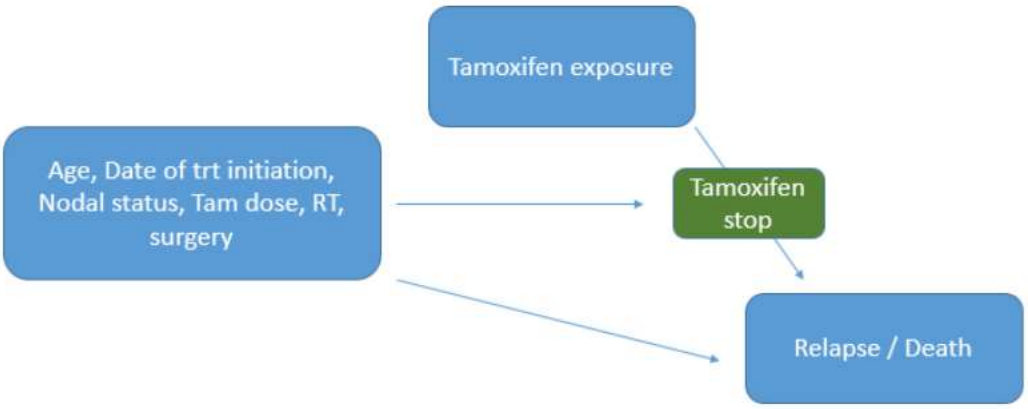


Figure 1: Switching process from the Tamoxifen exposure to stop treatment

Counterfactual methods aim to emulate the original randomization of the treatment by creating a pseudo-population. In particular Inverse Probability of Censoring Weights (IPCW) and Marginal Structural Models (MSM) create a "pseudo population" adjusted for the distortions that arise from the prognostic differences between switchers and non-switchers. Switchers are censored at the time point of cross over (stop tamoxifen), but remaining patients are weighted according to their probability to switch treatment. On the other hand, the Rank Preserving Structural Failure Time Model (RPSFTM) method estimate the counterfactual event times for patients who switch treatment as if they would not have switched. The method estimates the survival end points relative to a specific treatment, constructing a pseudo-population that hypothesizes what would have happened to the survival of the switchers, if they would not have switched to the alternative treatment

-IPCW method: in order to satisfy unmeasured confounding assumption, the variables in the weight calculation should fully capture all reasons for switching that are also linked to survival. For this reason, we perform logistic regression model to identify factors related to both switching and survival outcomes (death and recurrence). In TAM01 trial there were no time dependent covariates but only fixed covariates at time of randomization. We identified risk factors related to switching and death to include in the model for weights computations: Age, Year of randomization, Year of Initial Treatment, Nodal status, Dose of tamoxifen, Radiotherapy and previous Surgery. Excluding patients with missing data in covariates selected for weights computation (no missing data are let in IPCW model) we finally analyzed 3755 women (only 2% excluded). After the identification of baseline covariates in order to implement IPCW method it is fundamental proceed in this way: 1) create correct data base: i.e. split follow-up period in time intervals with matching patient status and covariates (the choice of time intervals is the key point of the method); 2) Determine IPCW weights via Cox Regression model;3) Apply resulting weights in the analysis of survival outcomes by means Cox regression weighted model. Analysis are performed using *ipcwswitch* R-package (33) Several sensitivity analyses were performed to evaluate robustness of IPCW (different time-intervals, excluding estrogen negative patients, excluding both estrogen negative women and patients with unknown estrogen status)

-RPSFTM: In contrast to the IPCW method which requires potential confounders to be collected over time, the RPSFTM only requires information on the randomized treatment group, observed event times, and treatment history in order to estimate a causal treatment effect. We adapted *rpsftm* R- package (34) to TAM01 trial (the package was created for situation in which patients in the

control group switch to experimental arm, while in TAM01, in ST arm (control) there are no crossover to LT arm). Accelerated failure parameter was calculated using several approaches (Log Rank test, Cox regression model and Weibull distribution) and we check the performance of the model assessing the counterfactual Kaplan-Meirer curves generating through psi parameter were similar. Sensitivity analyses were performed to evaluate robustness of RPSFTM (analysis with e without re-censoring and assuming that the treatment effect in switchers is % lower or higher than in experimental group to verify the validity of the "common treatment effect" assumption.

Results

Overall non-adherence rate in LT arm was 27 %. ITT analysis estimated a 6% reduction in the hazard of death and a 10% reduction of hazard of relapse with tamoxifen treatment (hazard ratio (HR) respectively, 0.94 (0.84-1.07) and 0.90 (0.81-0.99)). All causal inference methods adjusted for switching showed that a significant survival benefit would have been observed had there been no selective switching (see **Table 2**). Results from counterfactual and RPSFTM methods differ on iDFS endpoint, which underlines the need for careful assessment of underlying assumptions in each method. TAM01 original trial was an old trial in which Tamoxifen was administered to both estrogen positive and negative patients. According to actual adjuvant therapy protocols, Tamoxifen should be taken only by ER positive patients: sensitivity analysis excluding 1428 patients with ER receptor status negative or unknown showed that a greater treatment benefit for ER+ patients in Long Term arm (see Table 3)

4.4 CONCLUSION

The ITT method remains the established method to evaluate efficacy of a treatment; however, additional analyses should be considered to assess the on-treatment effect when substantial non-adherence to study drug is expected or observed. IPCW and RPSFTM are well-established methods to deal with this issue, however their use remains still uncommon, notably when dealing with a time-to-event outcome. This could be explained at least partially because of the difficulty in implementation given the need to adapt existing software programs to treatment stop, rather than to treatment crossover. Adjusting for treatment compliance reveals a significant higher protective effect of tamoxifene on both OS and IDFS compared as standard ITT analysis, Effect size is variable and related to assumption underlying causal model

In conclusion, in the presence of switching treatment, it is important to perform a detailed analysis as suggested by a recently review (29). No method is universally "best": results are sensitive to the assumptions associated with each adjustment method and their applicability depends on the characteristics of the trial in question. For this reason, assessment of the plausibility of assumptions' methods and implementation of a range of sensitivity analyses are fundamental.

4.5 REFERENCES

1. Delozier T., Spielmann M., Macé-Lesec'h J. et al.,Tamoxifen Adjuvant Treatment Duration in Early Breast Cancer: Initial Results of a Randomized Study Comparing Short-Term Treatment With Long-Term Treatment. J Clin Oncol., 2000, 18:3507-3512
2. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada 2017. https://www. cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf. Accessed January 15, 2018.
3. Watkins C, Huang X, Latimer N, Tang Y, Wright EJ. Adjusting overall survival for treatment switches: commonly used methods and practical application. Pharm Stat. 2013;12(6):348–357)
4. Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. Clin Trials. 2008;5(1):5–13
5. Hernan MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiol Drug Saf. 2005;14(7):477–91.
6. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods. 1991;20(8):2609–31
7. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000;11(5):561–70
8. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3):779–88.
9. Law MG, Kaldor JM. Survival analyses of randomized clinical trials adjusted for patients who switch treatments. Stat Med. 1996;15(19):2069–76
10. Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics. 2003;59(1):100–5
11. Walker AS, White IR, Babiker AG. Parametric randomization based methods for correcting for treatment changes in the assessment of the causal effect of treatment. Stat Med. 2004;23(4):571–90
12. Colleoni M, Giobbie-Hurder A, Regan MM, Thurlimann B, Mouridsen H, Mauriac L, et al. Analyses adjusting for selective crossover show improved overall survival with adjuvant letrozole compared with tamoxifen in the BIG 1-98 study. J Clin Oncol. 2011;29(9):1117–24
13. Latimer NR, Bell H, Abrams KR, Amonkar MM, Casey M. Adjusting for treatment switching in the METRIC study shows further improved overall survival with trametinib compared with chemotherapy. Cancer Med. 2016;5(5):806-815. doi:10.1002/cam4.643
14. Latimer NR, Abrams KR, Siebert U. Two-stage estimation to adjust for treatment switching in randomised trials: a simulation study investigating the use of inverse probability

weighting instead of re-censoring. BMC Med Res Methodol. 2019;19(1):69. Published 2019 Mar 29. doi:10.1186/s12874-019-0709-9)

15. NICE. Sunitinib for the treatment of gastrointestinal stromal tumours. Final appraisal determination; 2009. http://www.nice. org.uk/guidance/index.jsp?action=download&o=45125. Accessed Sep 2010
16. NICE. Everolimus for the second-line treatment of advanced renal cell carcinoma. Final appraisal determination; 2010. http://www.nice.org.uk/nicemedia/live/12044/49566/49566.pdf. Accessed Sep 2010
17. NICE. Pazopanib for the first-line treatment of advanced renal cell carcinoma. Final appraisal determination; 2010. http://www. nice.org.uk/nicemedia/live/12032/52263/52263.pdf
18. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting for treatment switching in randomised controlled trials - a simulation study and a simplified two-stage method. Stat Methods Med Res. 2017;26(2):724–751
19. Morden J, Lambert P, Latimer N, Abrams K, Wailoo A. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Medical Research Methodology 2011; 11(4):1–20
20. Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. Stat Methods Med Res. 2018;27(3):765–784
21. ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonized tripartite guideline. Statistics in Medicine 1999;18:1905–1942
22. Fewell Z, Hernan M, Wolfe F, Tilling K, Choi H, Sterne J. Controlling for time-dependent confounding using marginal structural models. The Stata Journal 2004; 4(4):402–420
23. Kleinbaum DG. Survival Analysis, a Self-Learning Text. Biometrical Journal 1998; 40:107–8.
24. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials–an economic evaluation context: methods, limitations, and recommendations. Med Decis Making. 2014;34(3):387–402
25. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J Am Stat Assoc. 2001;96(454):440–448
26. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran M, Berry D, eds. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York, NY: Springer; 1999:95–134
27. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000; 56:779–88.).
28. Robins JM, Hernan MA, Brumback B.Marginal structural models and causal inference in epidemiology. Epidemiology (Cambridge, Mass) 2000; 11:550–60):
29. Sullivan TR, Latimer NR, Gray J. et al., Adjusting for Treatment Switching in Oncology Trials: A Systematic Review and Recommendations for Reporting. Value Health. 2020 Mar;23(3):388-39
30. Morden JP, Akehurst RL, Campbell MJ. Adjusting survival time estimates to account for treatment switching in randomised controlled trials—a simulation study. University of

Sheffield, Health Economics and Decision Sciences Discussion Paper, March 2013. Available at http://www.shef.ac.uk/scharr/sections/heds/discussion-papers/1306-1.25948

31. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet. 1992 Jan 4;339(8784):1-15. PMID: 1345950. (No Author Listed)).

32. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, Abraham M, Medeiros Alencar VH, Badran A, Bonfill X, Bradbury J, Clarke M, Collins R, Davis SR, Delmestri A, Forbes JF, Haddad P, Hou MF, Inbar M, Khaled H, Kielanowska J, Kwan WH, Mathew BS, Mittra I, Müller B, Nicolucci A, Peralta O, Pernas F, Petruzelka L, Pienkowski T, Radhika R, Rajan B, Rubach MT, Tort S, Urrútia G, Valentini M, Wang Y, Peto R; Adjuvant Tamoxifen: Longer Against Shorter (ATLAS) Collaborative Group. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet. 2013 Mar 9;381(9869):805-16. doi: 10.1016/S0140-6736(12)61963-1

33. Grafféo N, Latouche A, Le Tourneau C, Chevret S. ipcwswitch: An R package for inverse probability of censoring weighting with an application to switches in clinical trials. Comput Biol Med. 2019 Aug;111:103339. doi: 10.1016/j.compbiomed.2019.103339. Epub 2019 Jun 24. PMID: 3144276

34. rpsftm: An R Package for Rank Preserving Structural Failure Time Models PDF download Annabel Allison, Ian R White and Simon Bond , The R Journal (2017) 9:2, pages 342-353)

4.6 TABLES

Table 1. Recommendations for the reporting of switching-adjusted analyses

Item	Recommendation
All adjustment methods	
1	Provide results from an analysis unadjusted for treatment switching for comparison
2	Describe the treatment-switching mechanism: who could switch and when
3	Detail the number of patients that switched, the number eligible to switch, and when switching occurred
4	Give an overview of the data available for adjustment: what predictors and how frequently measured
5	State whether the chosen adjustment approach, including all model fitting steps, was prespecified; if not, explain how the final model was selected∗
6	Provide a statement around the plausibility of key assumptions (eg, no unmeasured confounding for IPCW and common treatment effect for the RPSFTM)
7	Provide a visual comparison of observed and adjusted survival times
8	Report on sensitivity analyses showing the robustness of treatment effect estimates to violations of key assumptions
Inverse probability of censoring weights (IPCW)	
I.1	State whether unstabilized or stabilized weights were used
I.2	Detail the statistical procedure used to calculate weights (eg, pooled logistic regression†, Cox model)
I.3	State the portion of data used in the WD model including time-varying predictors (eg, post-progression data only)
I.4	Describe the extent of and the method used to address missing data on predictors in the WD model(s)
I.5	Present parameter estimates and associated measures of precision from the WD model(s)
I.6	Summarize the distribution of weights and state whether values were truncated
I.7	Detail the FO model, including the estimation method (eg, robust variance estimation) and the baseline variables adjusted for
Rank preserving structural failure time model (RPSFTM)	
R.1	State and justify the structural model assumed (eg, as treated, ever treated)
R.2	State the metric used for g-estimation (eg, log-rank test), including baseline variables for adjustment where applicable
R.3	State the grid-search algorithm used
R.4	Plot g-estimation results to show that the estimation process has worked well
R.5	Present the estimated acceleration factor and its confidence interval
R.6	Compare counterfactual survival times between randomized groups in a Kaplan-Meier plot
R.7	Detail the FO model, including method for calculating a CI around the estimated treatment effect (eg, retain ITT *P* value, bootstrapping) and baseline variables adjusted for
R.8	Present results both with and without re-censoring applied

AFT indicates accelerated failure time; CI, confidence interval; FO, final outcomes; IPCW, inverse probability of censoring weights; ITT, intention to treat; RPSFTM, rank preserving structural failure time model; WD, weight determining.

∗ Given the complexity of the methods, it may not always be feasible to fully prespecify without consideration of the actual data collected or the performance of the models.

Table 2. Key assumptions, strengths and limitations of some commonly used statistical methods to adjust for switch.

	Description	Key assumptions	Strengths	Limitations
Naive methods				
Intention-to-treat	Analyze all data as observed in randomized groups, regardless of switch, using standard survival analysis techniques.	Switch does not affect survival	As per RCT design	Assumption unlikely to hold if treatment is effective, leading to underestimation of overall survival benefit. Simulations show underestimation can be large.
Exclude switchers	Analyze control arm data in non-switchers only, experimental arm in all patients, using standard survival analysis techniques.	Switch is not influenced by any prior treatments or covariates that also influence survival, that is, *no confounders*. Only important difference between switchers and non-switchers is switch treatment.	Simple	Assumption unlikely to hold, introducing selection bias in treatment effect estimation. Simulations show bias can be large. Breaks randomization.
Censor at switch	Analyze all data with control arm patients censored at the time of switch, using standard survival analysis techniques.	Switch is not influenced by any prior treatments or covariates that also influence survival, that is, *no confounders*. Only important difference between switchers and non-switchers is switch treatment.	Simple	Assumption unlikely to hold, introducing selection bias in treatment effect estimation. Simulations show bias can be large.
Time-varying covariate for treatment exposure (or switch)	Analyze all data with a time-varying covariate for exposure (or switch) to experimental, using standard survival analysis techniques	Switch is not influenced by any prior treatments or covariates that also influence survival, that is, *no confounders*. Only important difference between switchers and non-switchers is switch treatment.	Fairly simple	Assumption unlikely to hold, introducing selection bias in treatment effect estimation. Simulations show bias can be large.
Complex methods				
Rank-preserving structural failure time method (Robins 1991, White 2002)	Counterfactual survival time in absence of switch to experimental treatment estimated for control arm and compared to observed experimental arm. Estimated using an accelerated failure time model with G-estimation.	Experimental *treatment effect is the same regardless of when it is given* (at randomization or at switch). Counterfactual survival time is balanced between treatment groups due to randomization.	Reduces selection bias compared with naive methods. Does not require assumption of no unmeasured confounders. Simulations show it performs well if constant treatment effect assumption holds. Preserves ITT *p*-value (conservative approach—see also limitation). Maintains randomization.	Assumption of constant treatment effect may not be realistic in many diseases. Simulations show this can then lead to large biases. Method and results can be difficult for non-experts to understand. Unstable if amount of experimental treatment is similar on both arms. Treatment effect will not change much if the ITT HR is already close to 1. Preserves ITT *p*-value (do not regain lost power due to switch—see also strengths).
Inverse probability of censoring weighting (Robins 2000, Howe 2001)	Time-varying weights estimated for control arm patients that have not switched to reflect how 'similar' their characteristics are to switched patients using propensity score methods. Treatment arms then compared using weighted Cox model.	All factors that influence both switch and survival are included in the model, that is, *no unmeasured confounders*. Only important differences between switchers and non switchers are switch treatment and variables included in weight calculation.	Reduces selection bias compared to naive methods. Does not require assumption of constant treatment effect. Can recover some lost power due to switch (stronger *p*-value than ITT—see also limitation).	Assumption of no unmeasured confounders unlikely to hold, especially if little data collected post-randomized treatment. Simulations show this can lead to large biases. Method and results can be difficult for non-experts to understand. Cannot be used if all patients switch or a perfect predictor of switch, and large biases if a high or low proportion switch. Can recover some lost power due to switch (anti-conservative—see also strengths). Can lose some power due to loss of events in switching patients (weaker *p*-value than ITT).

Table 3: Clinical and Pathologic Characteristics of Eligible Patients

	Treatment Group	
Variables	**Short-Term**	**Long-Term**
Randomized patients	1,882	1,911
Eligible patients	1,863	1,894
Mean age, years	62.8	62.8
Surgery		
No surgery	62 (3.3%)	70 (3.7%)
Lumpectomy	929 (49,9%)	887 (46.8%)
Mastectomy	869 (46.7%)	934 (49.3%)
UnKnownn	3 (0.1%)	3 (0.1%)
Pathologicnodalstatus		
Negative	552 (29.6%)	565 (29.8%)
Positive	1224 (65.7%)	1232 (64.0%)
UnKnownn	87 (4.7%)	97 (4.9%)
ER status*		
Positive	1215 (65.2%)	1220 (64.4%)
Negative	174 (9.3%)	182 (9.6%)
UnKnownn	474 (25.4%)	492 (26.0%)
Tamoxifen dosage		
10 mg	2 (0,1%)	0 (0.0%)
20 mg	855 (45.9%)	862 (45.5%)
30 mg	647 (34.7%)	693 (36.6%)
40 mg	313 (16.8%)	303 (16.0%)
70 mg	1 (0.1%)	0 (0.0%)
Unknown	45 (2.4%)	36 (1.9%)
Adjuvant chemotherapy		
No	1306 (70.1%)	1310 (69.2%)
Yes	557 (29.9%)	584 (30.8%)
UnKnownn	0 (0.0%)	0 (0.0%)
Relapses	348 (18.7%)	285 (15.0%)
Deaths	218 (11.7%)	228 (12.0%)

Table4: IPCW and RPSFTM estimates of overall survival and invasive disease free survival treatment effect

Outcome: Overall Survival			**Outcome: Invasive Disease Free Survival**		
NAIVE METHODS			**NAIVE METHODS**		
Method	**HR (95% CI)**	**p-value**	**Method**	**HR (95% CI)**	**p-value**
ITT ANALYSIS (unadjusted)	0.94 (0.84-1.07)	0.352	ITT ANALYSIS (unadjusted)	0.90 (0.81-0.99)	0.045
PP ANALYSIS	1.14 (1.00-1.30)	0.005	PP ANALYSIS	0.99 (0.89-1.18)	0.95
Censoring switchers*	0.96 (0.84-1.09)	0.506	Censoring switchers*	0.94 (0.84-1.05)	0.283
Switching as time-dependent covariate*	0.95 (0.84-1.08)	0.459	Switching as time-dependent*	1.08 (0.97-1.21)	0.159
COMPLEX-METHODS			**COMPLEX METHODS**		
Method	**HR (95% CI)**	**p-value**	**Method**	**HR (95% CI)**	**p-value**
IPCW	0.73 (0.63-0.84)	<0.001	IPCW	0.45 (0.38-0.51)	<0.001
RPSFTM (with re-censoring)	0.75 (0.57-0.99)	//	RPSFT (with re-censoring)	0.70 (0.53-0.92)	//
RPSFT (without re-censoring)	0.80 (0.65-0.99)	//	RPSFT (without re-censoring)	0.83 (0.72-0.96)	//

* HR adjusted for age, year of initial treatment, surgery, nodal status, radiotherapy and dose of Tamoxifen.

Table 5: IPCW and RPSFTM estimates of overall survival and invasive disease free survival treatment effect. Sensitivity Analysis excluding ER negative patients

OUTCOME OS	**EXCLUDING ER- (n=3402)**	**ONLY ERPOSITIVE (n=2402)**
ITT ANALYSIS	0.99 (0.87-1.09)	0.93 (0.80-1.09)
IPCW (shortintervals)	0.73 (0.62-0.86)	0.68 (0.56-0.82)
RPSFT(with re-censoring)	0.76 (0.54-1.05)	0.74 (0.51-1.09)
RPSFT(without re-censoring)	0.90 (0.78-1.02)	0.85 (0.69-1.04)
OUTCOME IDFS	**EXCLUDING ER- (n=3402)**	**ONLY ERPOSITIVE (n=2402)**
ITT ANALYSIS	0.93 (0.83-1.04)	0.90 (0.79-1.03)
IPCW (shortintervals)	0.49 (0.43-0.57)	0.44 (0.37-0.53)
RPSFT(with re-censoring)	0.70 (0.50-0.98)	0.66 (0.46-0.96)
RPSFT(without re-censoring)	0.85 (0.73-0.99)	0.80 (0.67-0.98)

5. BREAST CANCER RECURRENCE ESTIMATION FROM POPULATION-BASED CANCER REGISTRY LINKED WITH ADMINISTRATIVE DATA.

Abstract

Background: Cancer registries capture complete information at a population level at the time of cancer diagnosis and also provide active follow-up status of patients in the long term (i.e., 10 years or more after diagnosis). On the other hand, rarely longitudinal follow-up evaluations regarding recurrence and treatment have been collected. Patterns of event after initial treatment such as re-operation and receipt of subsequent chemotherapy or radiotherapy may indicate recurrence.
In recent years, electronic medical records (EMRs) and population-based cancer registries increasingly contain information on cancer outcomes and treatment that can be used synergistically. The aim of this study is to develop a claims-based algorithm to identify breast cancer recurrences during a 10 years' follow-up through a record-linkage of two data sources, the Friuli Venezia Giulia population based-cancer registry (CR) and the administrative individual-record regional database.

Methods: We conducted a retrospective analysis of linked CR-EMRs data. All patients in the FVG population-based CR who were identified with non-metastatic breast cancer during 2004-2010 were followed through death, disenrollment, or study end date (December 31, 2017, last update of FVG CR). Hospital discharge and the outpatient medical claims were used to identify treatment for recurrence. Incidence of recurrence was calculated using individual person-time at risk (after 18 or 24 months from breast cancer diagnosis according to breast cancer molecular subtype) and taking into account competing events (secondary tumor or death for all causes).

Results: In total, 5420 non-metastatic (stage I-III) patients with breast cancer were included in analyses. 5268 women (97.2%) were eligible for the surveillance period of recurrence. After 18 months after breast cancer diagnosis, 1406 (26.7%) had at least a procedure suggesting a recurrence. In particular, 14.7% received a chemotherapy, 8.4% radiotherapy, 8.2% another breast surgery, and 5.5% were hospitalized for a secondary malignant neoplasm. The overall recurrence rate in the cohort during the 45775 person-years (py) of observation was 30.7 per 1000 person-years (95%CI: 29.2-32.4). Five and ten-year cumulative recurrence were respectively 15.9% (95%CI: 15.0%-16.9%) and 26.0% (95%CI: 24.8%-27.0%). The recurrence rates were higher for women aged <50 years (41 per 1000 py), diagnosed at stage III (72 per 1000 py) or with triple negative subtype (47 per 1000 py).

Conclusion:

Our results using 10 years of follow-up data, yielded pertinent information on recurrence in women with breast cancer. The method we reported for ascertaining breast cancer recurrence at population level can be used to investigate the real-world impact of specific treatments. This study also provides a potential framework for constructing similar algorithms to identify recurrences for other cancers using administrative data from a health system.

Summary of Statistical Methods applied- person-time analysis, cumulative incidence function, competing risks analysis

This chapter (the draft is in preparation) regards a project in which I am involved during my last year of the Ph.D. with the *Cancer Epidemiology Unit, CRO (Centro di Riferimento Oncologico) Aviano National Cancer Institute*

5.1 INTRODUCTION

In 2020, about 2.2 million women were diagnosed with Breast Cancer (BC) worldwide (1).

In Italy, among women, breast cancer is always the most frequent neoplasm, with approximately 55,000 new diagnoses estimated for 2020 (2): 2.6% of all Italian women (834,000-0.8 millions) were alive after a breast cancer diagnosis (3) and the net 5-years survival was 87%, one of the highest recorded in Europe (4).

Among women living in Friuli Venezia Giulia, a region of the north-east of Italy, breast cancer continues to represent the most frequent malignancy, with an incidence rate equal to 168.3 cases / year / 100,000 women, slightly higher than what observed in other regions of northern Italy (161.8 cases / year / 100,000). The 5-year survival after diagnosis is 89% (5).

These data confirmed that a considerable number of women are living after a BC diagnosis and this number is projected to increase with the aging population and advances in breast cancer treatment, with most women living for many years after a breast cancer diagnosis (6-7).

In the era of precision medicine, overall survival alone is not an adequate endpoint for assessing healthcare quality, comparing treatment efficacy, or informing decision making for patients with cancer, especially for cancers with long survival times such as breast cancer.

Knowing the risks of recurrences are important to improve patients' quality of life (8), allow patients taking more informed decisions about their treatment and to let cancer control experts identifying research priorities and to efficiently plan public health care policies (9). Consequently, the importance of studying long-term outcomes in breast cancer patients is growing: breast cancer recurrences (i.e., loco-regional, metastases and second primary breast cancers) are of interest in these studies, and efficient methods of identifying and collecting data on the occurrence of second breast cancer events are needed (10).

Although population-based cancer registries data are useful in tracking and reporting the evolving burden of cancer in the population, the information they recorded reflects the outcomes of diagnosis and death but do not routinely collect information on cancer progression or recurrence (11). This is mainly because continuous follow-up by any registry to assess for recurrence would be costly and manual case detection to identify metastatic cohorts is prohibitively laborious.

Lack of recurrence information has led to explorations of other approaches to determine the risk and frequency of cancer recurrence. For aggregate summaries, Mariotto et al. (12) decomposed disease-specific survival from diagnosis for non-distant metastatic cases into the time from diagnosis to metastasis and the time from metastasis to death. By using external estimates of survival from metastasis to death, they were able to estimate the distribution of the time to disease recurrence in populations of stage I, II, and III patients.
An alternative method recently spreading is the use of the growing bulk of population-based administrative data from hospitals and other health care-related institutions as proxy for patients follow-up. Such data offer new possibilities for the generations of disease models for health evaluation: more specifically, records of breast cancer patients from administrative data can be used for identifying recurrences. For individual-level data, statistical learning and data mining

approaches have been harnessed to predict recurrence events from claims histories. Chubak et al. (13) used classification and regression tree (CART) analysis to predict whether and when a patient had experienced a breast cancer recurrence or second breast cancer diagnosis. Ritzwoller et al. (14) used a combination of logistic regression and change-point detection to identify the presence and timing of recurrence events. In 2020, Izci H. et al. (15) published the first meta-analysis on the topic of detection of breast cancer recurrence using administrative data. Analyzing the 17 studies included in this meta-analysis (period: 2003-2019), only two conducted in Europe and fifteen in United States, the authors discussed the main pros and cons of using this methodology. The meta-analysis reported a high accuracy overall, which indicates algorithms as promising tools to identify breast cancer recurrence at the population level.
However, high heterogeneity among algorithms as regard as data source (16-17), recurrence definition (18-19), approaches to construct the algorithm and a small dataset for training and validation (20) demonstrated that these algorithms are not generalizable to different health systems, highlighting the need for more standardization and exploration of new methods. A promising alternative are machine learning technology that analyze unstructured clinical text in electronic medical records and have shown higher sensitivity and specificity (21-22). However, their limitations include a high cost of initial development, difficulty in adapting to new systems, and most significantly, the requirement for a prohibitively large amount of manually annotated training data. Complex models have been developed to extract breast cancer recurrence information from clinical notes, electronic health records and population-based cancer registries with varying degrees of success in generating labels: Natural Language Processing (NLP) models (21,23), neural network-based approach (24), key-word based search with Bayesian inference methods (25) and sequential deep learning models using large sample of free-text clinic notes (26). Core limitations of these advanced models are that they were trained using single institutional data that contains biases regarding syntactic style of clinical narratives, patient populations and treatment planning.

Despite full potential of administrative databases, this overview reveals that prediction of breast cancer recurrence from them remains a challenging problem. The main issue concerns the lack of a univocally definition of gold standard.

The availability of population-based data can overcome the problem of no well-defined gold standard and in particular may extended knowledge on long-term survival outcomes. Indeed, data on long term survival and long-term risk of breast cancer recurrence which were derived from population-based samples of patients are scant: most studies were restricted to 5 years of follow-up only and as regard as recurrence, the vast majority of published studied used selected patient's cohorts (hospital-based and clinical trials). Accurate cancer survival statistics are necessary for describing population-level survival patterns, measuring advancements in cancer care and for estimate cancer prognosis. This population-based study aims to:

1) estimate 5 and 10-year relative survival (RS) of FVG breast cancer women diagnosed in 2004-2010 and followed up to 2017, by major prognostic factors for breast cancer (i.e. stage, molecular subtype, and grade) (27,28)

2) among the still living women, to identify those who have relapse from those relapse free for the whole period of follow-up, developing a claims-based algorithm for breast cancer recurrences,

through a record-linkage of two data sources- i) Friuli Venezia Giulia population based-cancer registry and ii) administrative individual-record database in the same region. Such an algorithm has the potential to be implemented in future data repositories to facilitate studies of disease surveillance, monitoring, and quality assessment. Moreover, being able to accurately identify recurrences may help to recruit high-risk breast cancer patients for clinical trials on time and can guide more tailored and precise treatment strategies. Such predictions could also inform patients about their future risks, which may guide their life decisions.

5.2 MATERIALS AND METHODS

Source of Data

Two datasets were used in the present project:
1) The first was extracted from the ***FVG population-based cancer registry (CR).*** A population-based cancer registry has been registering all incident cases diagnosed in people living in the whole region since 1995. The FVG cancer registry is member of the Italian Association of Cancer Registries (AIRTUM) and it is accredited according to the quality standard required by the International Agency for Research on Cancer (IARC) and by the International Association of Cancer Registries (IACR). For this study, incident breast cancer cases from 2004 to 2010 have been selected from the FVG cancer registry, including age at diagnosis, vital status, stage at diagnosis calculated according to TNM VI edition or TNM VII edition (29), respectively 84% and 16% of the breast cases (% of completeness: 95%), tumor markers (hormonal receptors (estrogens/progesterone) and human epidermal growth factor receptor 2 (HER2): data available for at least 80% of BC cases). In cancer registry, proliferation index (Ki67) was collected only for a small number of cases.

2) Administrative Individual regional database. In order to obtain information regarding disease progression, records for incident breast cancer cases extracted from the FVG cancer registry were linked to multiple administrative individual-record regional databases that were hosted in a single data warehouse of FVG health information system. The system covers the entire regional population and aimed to be complete since hospitals are reimbursed only if the procedure were registered. The administrative regional database includes various electronic health administrative databases that can be linked with one another on an individual basis through a unique encrypted ID identifier modified periodically. The hospital discharge data (HDD) and the outpatient services database were considered for our study. The HDD includes records from all the regional hospitals (both public and private accredited to the public health system) and those regarding admissions of regional residents to extra-regional hospitals. In particular, the inpatient claims provide information on date of admission and discharge as well as diagnostic and procedure codes using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9 CM) for each admission. Each diagnostic code (1 for primary diagnosis and 5 secondary diagnoses) and each procedure code (1 for primary procedure and 10 secondary procedures) were flagged with 1 if the corresponding code was related to breast cancer treatment, according to EPI-COST project (30). Outpatient database contain information on date, type of service and the flag 1 if the code is related

to breast cancer treatment (according to Busco et al. (30)). Detailed tables of administrative database extractions are available in **APPENDIX A**

Study Cohort

Data extraction

Individual-level data were obtained from the FVG population-based cancer registry (CR).
All cases of female breast cancer diagnosed between January 1, 2004 and December 31, 2010 were identified. All other previous or subsequent cancers to date of breast cancer incidence were also extracted (excluding non-melanoma skin cancer, ICD10: C44). Patients were followed through death, change of residence (0.9 %), or study end date (December 31, 2017, last update of FVG CR). In total, 10162 tumors related to 8940 women were obtained. All women included in the analysis had a minimum of 7 years of follow-up after their initial cancer diagnosis.

Eligibility Criteria

We identified 8940 women with invasive breast cancer diagnosed during 2004-2010 (ICD-9CM site code C50.0-50.9). Cases were followed until December, 31,2017. In constructing study cohort, we applied exclusion criteria typically used with CR data for survival analysis: (i) breast cancer women aged less than 15 years (n=0), (ii) breast cancers diagnosed on death certificate only (DCO) or by autopsy (n=36) and (ii) alive with no survival times (n=22). We further excluded women with previous (n=213 in which breast cancer was not the initial primary tumor) or synchronous cancer within 90 days from incidence (n=59). Eligibility criteria were stage I–III breast cancer (n= 225 de novo metastatic breast cancer –stage IV- were excluded) with availability of stage of disease (n=741 cases were excluded because with unknown stage at diagnosis). We included only women younger than 74 years to guarantee an adequate surveillance period in which patients could be develop a recurrence. The final cohort enrolled was constituted of 5420 women with primary invasive breast cancer diagnosed. **Figure 1** shows the detailed flow-chart on the inclusion and exclusion criteria aimed at defining the cohort under study.

Breast Cancer subtypes

Surrogate definitions based on immunohistochemical measurements of the expression of hormone receptors (HR) and HER2 are used to classify breast cancer and correlate well with genetically different breast cancer subtypes (31), which are associated with the risk of recurrence and outcome in addition to classic prognostic factors (32).
Based on the available measurements of HR expression and HER2 expression, tumors were classified as follows:

1) Hormone Receptor-positive (HR+)/HER2-negative (HER2-), characterized by the expression of Estrogen receptor (ER) and/or progesterone receptor (PR), with no amplification or overexpression of the HER2 gene (33);
2) HER2-positive (HER2+), characterized by HER2 overexpression/amplification, irrespective of HR status;
3) Triple Negative (TN) breast cancers, showing no expression of both HR and HER2.

It has been shown, that these immunohistochemical definitions largely describe the major intrinsic BC subtypes, namely Luminal A, Luminal B, HER2-Enriched and Basal-like.

Time-to-event end-points

As recommended by the CONSORT statement (35) each time-to-event (TTE) end-point should be precisely defined. It implies specifying the date of origin, the list of events to be considered, such as failures, and the censoring process. The use of standard definitions for breast cancer clinical trial or observational studies events and end points will help to reduce the inconsistencies that currently confound the analysis and interpretation of results across studies. The consensus-based definitions for recurrence event were the following (36):

- local recurrence (LR): any epithelial breast cancer or ductal carcinoma in situ (DCIS) in ipsilateral breast tissue, or in skin and subcutaneous tissue of the ipsilateral thoracic wall;
- regional recurrence (RR): breast cancer in ipsilateral lymph nodes or contralateral lymph nodes if axillary lymph node dissection was performed;
- distant metastases (DM): breast cancer in any other body location.

Potential Indicators of Recurrence

With the collaboration of epidemiologists, an oncologist and a surgeon we identified procedures, diagnoses and medications that could indicate that a second breast cancer event had occurred. In particular, they assigned to every claims (ICD9-CM) the value 1 if an appropriate claim related to relapse was found, 0 otherwise and value of 2 if uncertain association. Potential indicators of recurrence included: standardized diagnosis, procedure, chemotherapy, radiation therapy, breast conserving surgery, mastectomy and secondary malignant neoplasm codes. Variable classification codes were noted in **APPENDIX B**. The cancer recurrence surveillance period should begin after a period without register-based evidence of ongoing disease, to ensure that the patients were in remission. Codes were considered potential indicators of recurrence if they fell after this period and before the end of follow up. In particular, the model we have developed apply only to events occurring at least 18 months (for HER2 negative BC) or 24 months (for HER2 positive BC) after the date of diagnosis of primary breast cancer treatment. This time-window of 18/24 months was the maximum length of initial treatment according the therapeutic guidelines in the period (see adjuvant treatment schemes as reported in **APPENDIX C**). This choice has been agreed with breast oncologists, in particular as regard as HER positive tumour, based on the duration of standard adjuvant chemotherapy and anti-HER2-based regimens and a review of contemporary patterns-of-care. (37)

Recurrence Algorithm

We hypothesized that the timing, frequency, and combinations of procedures, diagnoses, and

medication fills could all potentially be important in identifying breast cancer recurrence. **(Figure 2)**

We aim to identify recurrence from claims data using this approach: a chemotherapy/radiotherapy/hospitalization 18/24 months after initial diagnosis were those who experience a disease recurrence (any type). In details:

(1) T*he presence of any treatment or procedure codes that indicated restarting new chemotherapy or radiation*. We assumed that breast cancer patients who underwent a second round of chemotherapy or radiotherapy after primary treatment were more likely a recurrent case than those who did not.

(2) Second local breast surgery treatment
Based on the assumption that patients who undergo a second breast surgery have a higher risk of recurrence than those who did not, we built indicator variables for second local treatment (e.g., mastectomy or conservative surgery).

(3*) The presence of any hospital admission code indicating a malignant neoplasm of female breast or a secondary malignant neoplasm without any other chemotherapy, radiotherapy or breast surgery code.*
We assumed that diagnosis codes indicating a malignant neoplasm of breast (174.x), a secondary cancer in the breast (198.81 or 198.2) or lymph node (196.x) could be associated to local recurrence. Distant recurrence could be identified by ICD9-CM diagnosis codes for secondary cancers (197.xx-198.xx): we omitted the code 198.89 "secondary malignant neoplasm of unspecified site" because of its high false-positive rate (20;38).

Women were classified as having had a recurrence if, after the period of curative treatment (18/24 months), a therapy code (chemotherapy or radiotherapy) or breast surgery code or an hospital admission code indicating a secondary malignant neoplasm (without any other chemotherapy, radiotherapy or breast surgery code) were found during follow-up period.

If a second primary cancer diagnosis other than C50 (breast code) was present in the CR before a recurrence diagnosis code, the recurrence diagnosis code was disregarded as it could be related to the second primary cancer. The recurrence date estimated by the algorithm was defined as the date with a registration of an indicator of recurrence. If the same patient had more than one indicator of cancer recurrence, the first registered date was regarded as the date of cancer recurrence.

Since recurrence events are strongly related to breast cancer subtypes (39) and stage (40), we identify potential recurrence using the algorithm in all cohort and stratifyng respect to these two breast cancer characteristics.

Timing to recurrence
Person-years at risk of recurrence were computed between 18/24 months after breast cancer diagnosis and the first of subsequent events (occurred > 18/24 months after diagnosis):
1. the date of second primary cancer;
2. death or residence outside FVG region;

3. 31/12/2017 (last date of FU available);
4. first chemotherapy;
5. first radiotherapy;
6. first breast surgery;
7. first hospital admission without any of the above points 4-6

Statistical Analysis

We calculated the frequency and proportion of patients according to demographic, tumor, and treatment characteristics. Surveillance of cancer recurrence began 18/24 months after breast cancer diagnosis and continued to the outcome of interest as specified in the "Time to recurrence" section. Cumulative recurrence proportions were calculated using the number of events of interest as the numerator over a denominator that included all women at risk during a given time period, with time of follow-up specified (e.g., 3,5,10 years). We calculated cumulative incidence rates (CumIR) as the first occurrence of a study outcome, accounting for competing risks: i) secondary primary cancer (this event may interfer with the observation of a breast cancer recurrence because does not let to distinguish the origin of a subsequent recurrence); ii) death for breast cancer (due to the lack of "gold standard definition of breast cancer recurrence, we considered the death for breast cancer as an event that precludes the observation of the recurrence and not an event in itself) and iii) death for other or unknown cause (after 5 years of follow-up, most patients are at a clinically relevant risk of non-breast cancer death which forecloses to observe a recurrence). CumIR are calculated for all cohort and stratified by age, molecular profile, grading, nodal status and stage. The used estimator of the CumIR is based on a generalization of the Kaplan-Meier estimator and quantifies the probability that the event under study will occur before any specified time in the presence of competing risks (41). Along with point estimates, 95% confidence intervals of the CumIR were derived. The R Language and Environment for Statistical Computing (release 4.0.2) (42) and the cmprsk extension package were used for data preparation, statistical analyses and visualization. P-values of a two sided Gray's test of the equality of CumIR curves across subsamples (43) were derived. Moreover, we estimated breast cancer relative survival (RS) by age, molecular profile and stage at diagnosis. RS is used to summarize the excess mortality the cancer patients have in comparison with a corresponding general population group. RS is defined as the ratio of the proportion of observed survivors in breast cancer patients to the proportion of expected survivors in a comparable set of cancer free individuals. RS describes the probability of surviving a cancer diagnosis in the absence of competing causes of death (4). RS was calculated with the Pohar-Perme method (44-45) using the SEER * Stat software (46) provided by the US cancer surveillance and control program of the American National Cancer Institute. RS were reported with 95% confidence intervals. All statistical analysis were performed in STATA version 14.1 (StataCorp LP) and in R software.

5.3 RESULTS

Patients' characteristics

In total, 5420 patients with stage I to III breast cancer were included in analyses. Clinical and pathological features of the whole population are shown in **Table 1**. Median age at diagnosis was 60 years (interquartile range [IQR]: 50–66). The majority of patients had primary tumors ≤1cm (69%), grade II disease (52%), ductal invasive histology (77%) and no nodal involvement (60%). Advanced stage (II-III) breast cancer was identified in 48.8% women.
Full immunohistochemistry data were available for 4795 women, allowing the classification of their cancer as HR+/HER2- (3576, 66.0%), HER2+ (799, 14.7%) and TN breast cancers (420, 7.7%). Median follow-up was 10.5 years (IQR: 8.6–11.9) during which 17.5% of the patients died.

Breast cancer recurrence identified by the algorithm

Figure 3 summarized the observed breast recurrence of study patients. Overall, 97.2% (5268 out of 5420) of women were eligible for the surveillance of recurrence. Of 5268 women with primary invasive stage I-III breast cancer, 1406 (26.7%) had at least a chemotherapy (14.7%) or radiotherapy (8.4%) or breast surgery (8.2%) or an hospital admission code indicating a secondary malignant neoplasm (5.5%) (without any other chemotherapy, radiotherapy or breast surgery code) indicating recurrence. This meant that women who underwent a new round of therapy, or a breast surgery or an hospital admission for a secondary neoplasm at 18 months or more after the primary treatment were classified as a recurrent case. We tested different timeframes: 12 months, 18 months, 24 months and found that the 18 months for HER2 negative BC and 24 months for HER2 positive BC, worked best in according to clinical treatment guidelines. After a median follow-up of 10.5 years, 236 patients (5.0%) developed a second tumor and then censored at this date; 3599 (68.3%) were relapse-free at the end of follow-up (of whom only the 0.6% dead for breast cancer). At the end of follow-up period, 83.9% of the women were alived, while 9.2% and 6.8% dead respectively for breast cancer and for other or unknown causes.
Stratifying respect to molecular profile (**Table 2**) and breast cancer stage **(Table 3**) we observed substantial variations in the identification of recurrence: as expected, in TN and HER2+ breast cancer, the percentage of administrative codes potentially related to recurrence, were higher than that found in HR+ breast cancer (35.6%, 28.8% and 24.1% respectively). Similarly, for women with stage III, the codes indicating a possible recurrence are significantly higher than for stages II and I (53.1%, 30.6% and 18.1% respectively).

Cumulative Recurrence

Table 4 and **Table 5** reported respectively recurrence rate per person-years (py) and cumulative recurrence at 3, 5 and 10 years of follow-up. The overall recurrence rate in the cohort during the 45775 person-years (py) of observation was 30.7 per 1000 person-years (95%CI: 29.2-32.4), i.e. women with primary breast cancer, stage I-III, developed recurrent breast cancer at a rate of 3.1 per year in the period between 18/24 months post-diagnosis (2004-2010) with follow-up through 31/12/2017. Five and ten-year cumulative recurrence were respectively 15.9% (95%CI: 15.0%-16.9%) and 26.0 % (95%CI: 24.8%-27.0%) (**Figure 4**). Women with node-positive tumors at diagnosis had significantly higher annual recurrence rates than women who had node-negative tumors (4.9% vs 2.3%, $p < 0.001$). Recurrence rates were also higher for women aged less than 50 years compared with screening group and older (4.1% vs 2.6 vs 3.4 % per year, $p < 0.001$). The

cumulative incidences and recurrence rates were particularly increased if tumors were advanced, of high histopathologic grade or classified as TN subtype (**Figure 5)**.

Relative Survival

The Relative Survival (RS) among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 was showed in **Figure 6a.** Women with breast cancer presented a high 5-year relative survival, equal to 95.8% (95%CI: 95.0%-96.5%). This means that 96% of those who have breast cancer diagnosis did not die for their disease. Ten -year RS was 89.8% (unless otherwise stated, the 10-year estimate of the RS will be reported subsequently). Relative survival was very similar among age subgroups **(Figure 6b).** Stratifying respect to molecular profile, the best survival was observed among women with HR+ subtype (RS: 92.1%), followed by HER2+ (RS: 86.3%) and TN subtype (RS: 76.6%) (**Figure 6c, Table 6)** Although molecular subtype affected survival, stage at diagnosis seemed to be a more powerful prognostic factor: 5-year RS was 99.8%, 95.2% and 81.5% for stages I, II and III, respectively and 10-year RS was 97.9%, 88.8% and 61.9% (**Figure 6d, Table 6)**. In particular, the 5- year RS for the TN subtype (known for poor prognosis) was 94.6% among stage I disease, while decreased to 42.7 % among those with stage III disease (**Figure 7c, Table 6**).

5.4 DISCUSSION

Currently, chart review is the only reliable way to obtain recurrence status, but this approach is time-consuming and inefficient. However, investigators may be interested in questions relating to subsequent relapsed disease. Linking administrative data to registry cancer data can provide the ability to infer the occurrence of relapse in selected situations. Identification of recurrence from population-based registry and administrative data using clinical algorithms is feasible for cancers where a majority of patients receive treatment for relapse, without a "watch and wait" strategy, and where that treatment is with a modality that can be detected in billing data (i.e. intravenous chemotherapy, radiation, surgery or hospital admission). This combined approach has the potential to greatly reduce the resources needed to identify recurrences in a large population and may impact future health services research, including the facilitation of quality improvement or effectiveness studies in addition to population health studies. In this context we have developed a claims-based algorithm to identify breast cancer recurrences during a 10 years' follow-through a record-linkage of the Friuli Venezia Giulia population based-cancer registry and administrative individual-record regional database.

Performance of previous algorithms with gold standard

Prior studies that evaluated cancer recurrence algorithms using administrative data found moderate to high sensitivities and specificities but have several important limitations. Lamont et al. (20) used medicare claims data to measure disease-free survival in individuals ≥65 years of age diagnosed

with breast cancer (N = 45, 12 recurrences and 2 deaths). Algorithm sensitivity and specificity were 83% and 97% respectively. Rasmussen et al. (46) used national data in Denmark to identify breast cancer recurrence (n = 471, 149 recurrences). Sensitivity and specificity were 97% and PPV was 94%. Xu et al. (47) developed algorithms to identify breast cancer recurrence among women ≤40 years of age or those who received neoadjuvant chemotherapy in Alberta (N = 598, 121 recurrences). Sensitivity values ranged from 75% to 94% and specificity values ranged from 94 to 98%. Chubak et al. (13) developed several algorithms to determine second breast cancer events and recurrence only among women diagnosed with stage I or II breast cancer (n = 3152, 407 breast cancer events). Sensitivity values ranged from 69 to 99% and specificity values ranged from 81 to 99%. In the recent published study of Lambert et al. (48) sensitivity was 68.5% and specificity was 97.0%. This high heterogeneity of algorithm's performance is mainly related to heterogeneity of databases used to very different sample size, different periods of surveillance of cancer recurrence, some studies did not distinguish between recurrence and second breast cancer primary (i.e., a new primary cancer unrelated to the prior cancer), and above all there was not a standardization in the use of administrative codes as potential indicator of recurrence.

Breast Cancer Recurrence: Main Findings and comparison with literature

We evaluated a simple-intuitive algorithm for the identification of cancer recurrence based on the presence of selected ICD-9-CM codes. Using multiple routinely collected health datasets, we found that women diagnosed with primary breast cancer developed any recurrence at a rate of 3.1% per year in the period between 18 months and 10 years' post-diagnosis. This result was very similar to that of an Australian study (49) which identified a rate of 3.3% per year in the period between 18 and 72 months. Our finding that recurrence risk is greater for women with node-positive (50), with higher stage and histologic grading (51) and with non-luminal tumors (52) was consistent with existing international evidence. In particular, the different breast cancer subtypes showed different times to disease progression: as expected, the HR+ subgroup had the longest disease-free survival (10-year recurrence rate equal to 23.5%), whereas patients with HER2+ or TN disease had worse prognoses (10-years recurrence rates respectively: 26.9% and 36.0%).

As reported in APPENDIX B, the standard adjuvant treatment for HER2+ positive patients was chemotherapy for 4-6 months in combination with trastuzumab, followed by trastuzumab alone to complete a 1-year treatment. In our study period (2004-2017), adjuvant trastuzumab treatment was available for breast cancer patients (first introduction of trastuzumab was in 2001 (53). Our choice to consider the period of surveillance of cancer recurrence after 24 months from diagnosis was related to this adjuvant scheme and let us to find similar results to a recent meta-analysis, in terms of recurrence rate (54). In our HER2+ cohort, the 5 and 10-cumulative recurrence were 17.5% and 26.9%, while in the meta-analysis these rates were respectively equal to 17.0% and 22.9%. As regards as Stage at diagnosis, 10-year cumulative recurrence was 17.7%, 29.4% and 49.3% for Stage I, II and III, respectively. Our findings were lower than that obtained by Cheng L. (55) with a similar algorithm: their estimates were in fact 35%, 44% and 56%. This discrepancy could be explained considering that Cheng L. enrolled only women over than 65 years: the exclusion of women over 75 is certainly questionable and in future projects it could be decided to consider women at least up to 79 years of age. Moreover, differently from our analysis where, deaths for all causes happened before evidence of recurrence were treated as competing risk, the other authors

evaluated cumulative recurrence with only breast cancer-specific death as competing risk.
Several issues hampered the comparison of literature' findings with the results presented in this study. In general, recurrence estimates were likely higher in our work than previously reported studies probably due to our definition of recurrence which capture all recurrence, local, regional and distant metastases. In the **APPENDIX D** we provided a list of studies that 1) had developed algorithms for identify breast cancer recurrences using administrative databases and 2) population - based estimates of recurrences. The heterogeneous clinical definition of recurrence used as gold standard and very different median follow-up, did not let direct comparison of estimates. For example, several studies (56-58) considered only metastases as recurrence event not allowing a homogeneous comparison with the studies that also consider the loco-regional relapses. Moreover, all studies in which were used algorithms (except for Cronin-Fenton et al. (16) and Kemp-Casey et al. (49)), recurrence rates were computed as relative frequency, instead to considered as denominator the total person-years. Also the comparisons with published population-based studies were difficult: most of them used 'classical' Kaplan-Meier estimators, reported proportions, or did not mention whether competing risks had been taken into account (see in **APPENDIX D:** Geurts (10), Minicozzi (59), Fredholm (60), Schaffar (61), van Maaren (39), Lao (62), Stokes (38)).

Relative Survival: Main Findings and comparison with literature

We evaluated relative survival for overall cohort and by stage and molecular subtype. 5-year relative survival values for breast cancer patients with stage I at diagnosis (99.8%) and stage II (95.2 %) were very high, while they decrease for stage III (81.5%). Similar results were described by a recent population-based Italian study of Mangone et al. (5-years RS: 100% (Stage I), 91.9% (Stage II) and 78.8% (Stage III)) (28) by Siegel et al. (63) (98%, 92% and 75% for stages I, II and III respectively) and by SEER data (64).
No recent data regarding 10-year relative survival in European countries have been published. In Australia the relative year survival rate for breast cancer was estimated equal to 91% in 2018 (65) and the relative survival rate 10 years after diagnosis of breast cancer was 86% similar but slightly lower than that estimated in our study (5-years RS: 95.8% and 10-years RS: 89.8%).
Our study was the first in Italy, to our knowledge, to use population-based cancer registry data to examine survival by the major molecular breast cancer subtypes, although considering their immunohistochemistry-defined surrogated. In United States study (66) using SEER cancer registry data, 4-year relative survival by molecular subtype, has been assessed among women diagnosed during 2010-2013 and followed thorough December 31, 2014. They estimated a 4-year relative survival of 92.5% for HR+/HER2-, of 90.3% for HR+/HER2+, of 82.7 % for HR-/HER2+ and of 77.0 % for triple-negative subtype. Our corresponding values at 4 years were better for all subtypes: for HR+/HER2-, HER2+ and triple negative FVG breast cancer women, the 4-year RS was equal to 98.5%, 94.7% and 82.4% respectively.

Strength e Limitations

To our knowledge, this is the first study in Italy to investigate the long-term health outcomes beyond death and survival in a population-based cohort of breast cancer patients using

administrative database. The strengths of this study include the high quality of data at diagnosis available at the FVG Cancer Registry and the possibility to merge it with additional databases. A novel aspect of this study is ascertainment of recurrence using datasets and variables that are widely available to researchers, allowing ongoing monitoring of recurrence rates over time in the absence of routinely collected or available recurrence data from FVG's cancer registry. We used health records for women diagnosed with primary stage I-III breast cancer, for which all hospital admissions and outpatient services had been captured. Multidisciplinary approach is another important point of strength of this study: involving several professional figures expertise in breast cancer surveillance and treatment, it contributes reducing the existing gap between basic research and the daily clinical practice. Epidemiologists, two breast oncologists and one breast surgeon identified procedures and diagnosis specifically indicating that a recurrence had occurred, with corresponding ICD-9-CM codes. Moreover, treatment drug-therapy scheme provided by the oncologists was fundamental to deal with the issue of surveillance period of breast cancer recurrence. The strength of our study concerned also the use of appropriate statistical methods to estimate recurrence rate over time: 1) we corrected for the "survival bias", considering the follow-up period after 18/24 months' time window, i.e. time 0 for the follow up was fixed at 18/24 months after breast cancer diagnosis (67). 2) Moreover, since in our setting competing-risks, i.e. events (second tumors or death) that preclude the occurrence of the outcome of interest (recurrence) were present, we showed results computing cumulative incidence of recurrence instead the standard Kaplan-Meier approach, which biased recurrence rates especially in long-term follow-up studies (68).

As regard as relative survival estimates, our results were population based with no selection bias. FVG cancer registry had complete follow-up information for 99% of cases, so reporting of survival was reliable.

We acknowledged several limitations in our present study. We were not able to directly assess breast cancer recurrence by patients' chart review because of the unavailability of such data. For this reason, it was not possible to ascertain the accuracy of the algorithm by means of the calculation of sensitivity and specificity even because there is no reliable gold standard. Moreover, we could not differentiate between local, regional recurrence and distant metastasis. Only the examination of the woman's clinical records could allow a more precise attribution. Because of the importance of this aspect, for the further studies, it may try to select the key information that can be extracted from the clinical charts and integrated with that of the administrative archives. Despite these limitations, our recurrence rate results and patterns observed respect to molecular profile and stage at diagnosis were in trend with the literature data. The main limitation of the study concerned the risk of misclassification of recurrence and recurrence dates from missing or incorrect registrations. Missing data are of less concern as patients with BC recurrence were unlikely to occur, without contact with regional health system. However, comorbid and frail women may be at higher risk of being missed by the algorithm as their delicate state may contraindicate cancer treatment by chemotherapy, radiotherapy or surgery.

The model we have developed apply only to recurrence occurring at least 18/24 months after the breast cancer diagnosis, i.e. using a cross-sectional approach. This period was chosen as it is likely to reduce the risk of designating a patient as having recurrent disease based on continuing initial treatment. For this reason, the model cannot be used to identify synchronous second primaries or early recurrences. We chose this model to avoid confusion with treatment for the initial primary. However, the time of 18/24 months from diagnosis, after which the surveillance period of

recurrence was began, was an arbitrary cut-off that could have anyway led to misclassification of disease status. In particular, false positives cases could be identified by chemotherapy or radiotherapy codes that appeared shortly after this cut-off and were expected to concern women with delayed initiation of adjuvant therapy. The issue to determine the time of recurrence is very complex. In fact, only two studies Ritzwoller et al (14) and A'mar et al. (69) tried to identify individual-level recurrence status and predicted its timing by identifying the month of greatest change in the count of each code grouping, and reconciling the months so identified across the groupings. However, advances are needed because single-source data sets do not encompass all providers and have limited generalizability.

Requiring an interval with no claims for treatment might be one solution to avoid these potential false positive cases taking into account also that in the clinical practice delay in initiation of adjuvant treatment can occur (70). Lastly, administrative database considered (hospital discharge data and the outpatient services database) did not capture usage of hormone therapies and this could have led to an underestimation of recurrence events for those women with positive hormone receptor status.

Further studies, at multicentre level, are needed to enhance this initial work: it will be essential to increase the follow-up of breast cancer women since late recurrences after many years of disease-free survival (>10 years) remain an open question (71)

5.5 CONCLUSION

In summary, our results using 10 years of follow-up data, yield original information on frequency of recurrence in women with breast cancer by stage of disease and molecular profile. The method we report for ascertaining breast cancer recurrence can be used in other Italian areas to investigate the real-world impact of population screening and specific treatments on breast cancer outcomes. This study also provides a potential framework for constructing similar algorithms to identify recurrences for other cancers using administrative data from a health system. Continued efforts to improve and validation of algorithms that identifies breast cancer recurrence using administrative database is worthwhile, because while success may not be assured, the alternative, asking registries to track recurrence status, would likely be cost prohibitive. A claims-based algorithm could provide valuable information about the experiences and outcomes of the many patients with recurrent breast cancer and help realize the full potential of administrative databases for comparative effectiveness research. These results provide information helpful to patients and clinicians in order to disentangle, not only frequency of breast cancer patients that will die or be cured from their cancer (7), but also who will remain Disease Free from those who will face some recurrence but still are alive 10 years or more since diagnosis.

5.6 REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021. https://doi.org/10.3322/caac.21660
2. ECIS European Cancer Information System, European Union, 2021: https://ecis.jrc.ec.europa.eu)
3. Guzzinati S, Virdone S, De Angelis R, et al. Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020. BMC Cancer. 2018;18(1):169. Published 2018 Feb 9. doi:10.1186/s12885-018-4053-y
4. I numeri del Cancro 2020, Airtum Working Group, https://www.registri-tumori.it/cms/pubblicazioni/i-numeri-del-cancro-italia-2020
5. Bidoli E, Toffolutti F, Zanier L, Serraino D: "I tumori in Friuli Venezia Giulia. Dati di incidenza e sopravvivenza: aggiornamento al 2018". Regione Autonoma Friuli Venezia Giulia, Udine, 2020.
6. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and sur-vivorship statistics, 2019. CA Cancer J Clin. 2019; 69:363- 385
7. Dal Maso L, Panato C, Tavilla A, Guzzinati S, Serraino D, Mallone S, Botta L, Boussari O, Capocaccia R, Colonna M, Crocetti E, Dumas A, Dyba T, Franceschi S, Gatta G, Gigli A, Giusti F, Jooste V, Minicozzi P, Neamtiu L, Romain G, Zorzi M, De Angelis R, Francisci S; EUROCARE-5 Working Group. Cancer cure for 32 cancer types: results from the EUROCARE-5 study. Int J Epidemiol. 2020 Oct 1;49(5):1517-1525. doi: 10.1093/ije/dyaa128.
8. Ferreira AR, Di Meglio A, Pistilli B, et al. Differential impact of endocrine therapy and chemotherapy on quality of life of breast cancer survivors: a prospective patient-reported outcomes analysis. Ann Oncol. 2019;30(11):1784–95. https://doi.org/10.1093/annonc/mdz298.
9. Shimozuma K. Quality of life assessment. Breast Cancer. 2002;9(2):100–6. https://doi.org/10.1007/bf02967573
10. Y. M. Geurts, A. Witteveen,R. Bretveld, P. M. Poortmans, G. S. Sonke, L. J. A. Strobbe, S. Siesling, Patterns and predictors of first and subsequent recurrence in women with early breast cancer. Breast Cancer Res Treat (2017) 165:709–720
11. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov), released December 2020
12. Mariotto AB, Zou Z, Zhang F, Howlader N, Kurian AW, Etzioni R. Can we use survival data from cancer registries to learn about disease recurrence? The Case of Breast Cancer. Cancer Epidemiol Biomark Prev. 2018;27(11):1332–41. https://doi.org/10.1158/1055-9965.EPI-17-1129
13. Chubak J, Yu O, Pocobelli G, Lamerato L, Webster J, Prout MN, et al. Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. J Natl Cancer Inst. 2012;104(12):931–40. https://doi.org/10.1093/jnci/djs233

14. Ritzwoller DP, Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, et al. Development, validation, and dissemination of a breast cancer recurrence detection and timing informatics algorithm. J Natl Cancer Inst. 2018;110(3):273–81. https://doi.org/10.1093/jnci/djx200

15. Izci H, Tambuyzer T, Tuand K, Depoorter V, Laenen A, Wildiers H, Vergote I, Van Eycken L, De Schutter H, Verdoodt F, Neven P. A Systematic Review of Estimating Breast Cancer Recurrence at the Population Level with Administrative Data. J Natl Cancer Inst. 2020 Oct 1;112(10):979-988. doi: 10.1093/jnci/djaa050. PMID: 32259259; PMCID: PMC7566328

16. Cronin-Fenton D, Kjærsgaard A, Nørgaard M, et al. Breast cancer recurrence, bone metastases, and visceral metastases in women with stage II and III breast cancer in Denmark. Breast Cancer Res Treat. 2018;167(2):517–528

17. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. Med Care. 2014; 52(10):e65–e73.

18. Whyte JL, Engel-Nitz NM, Teitelbaum A, Gomez Rey G, Kallich JD. An Evaluation of Algorithms for Identifying Metastatic Breast, Lung, or Colorectal Cancer in Administrative Claims Data. Med Care. 2015 Jul; 53(7): e49-57

19. Kroenke CH, Chubak J, Johnson L, Castillo A, Weltzien E, Caan. Enhancing Breast Cancer Recurrence Algorithms Through Selective Use of Medical Record Data. BJJ Natl Cancer Inst. 2016 Mar; 108(3)

20. Lamont EB, Herndon JE, Weeks JC, et al. Measuring disease-free survival and cancer relapse using Medicare claims From CALGB breast cancer trial participants (Companion to 9344.). J Natl Cancer. 2006;98(18):1335–1338

21. Carrell DS, Halgrim S, Tran DT, Buist DSM, Chubak J, Chapman WW, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: The case of breast cancer recurrence. Am J Epidemiol. 2014;179(6):749–58.

22. Strauss JA, Chao CR, Kwan ML, Ahmed SA, Schottinger JE, Quinn VP. Identifying primary and recurrent cancers using a SAS-based natural language processing algorithm. J Am Med Informatics Assoc. 2013;20(2):349–55

23. Soysal, E., Warner, J. L., Denny, J. C. & Xu, H. Identifying metastases-related information from pathology reports of lung cancer patients. AMIA Jt. Summits Transl. Sci. Proc. 2017, 268–277 (2017)

24. Banerjee, I., Bozkurt, S., Caswell-Jin, J. L., Kurian, A. W. & Rubin, D. L. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. JCO Clin. Cancer Inform. https://doi.org/10.1200/CCI.19.00034 (2019)

25. Zhao, D. & Weng, C. Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. J. Biomed. Inform. 44, 859–868 (2011)

26. Sanyal, J., Tariq, A., Kurian, A.W. et al. Weakly supervised temporal model for prediction of breast cancer distant recurrence. Sci Rep 11, 9461 (2021). https://doi.org/10.1038/s41598-021-89033-6)

27. Crocetti E, Bossard N, Uhry Z, et al; GRELL EUROCARE-5 Working Group. Trends in net survival from 15 cancers in six European Latin countries: the SUDCAN populationbased study material. Eur J Cancer Prev 2017; 26: S3–S8

28. Mangone L, Mancuso P, Bisceglia I, et al. Five-year relative survival by stage of breast and colon cancers in Italy. Tumori Journal. 2021;107(4):318-324. doi:10.1177/0300891620964565).
29. Sobin L, Gospodarowicz M and Wittekind C. TNM Classification of Malignant Tumours, 7th edition. Milan: Raffaello Cortina Editore; 2009
30. Busco S, Tavilla A, Gigli A, Lopez T, Pierannunzio D, Mallone S, Guzzinati S, Capodaglio G, Giusti F, Angiolini C, Francisci S. A direct method for the identification of patterns of care using administrative databases: the case of breast cancer. Eur J Health Econ. 2021 Jul 26. doi: 10.1007/s10198-021-01327-8.
31. Park S, Koo JS, Kim MS, Park HS, Lee JS, Lee JS, Kim SI, Park BW. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry. Breast. 2012 Feb;21(1):50-7. doi: 10.1016/j.breast.2011.07.008. Epub 2011 Aug 23. PMID: 21865043
32. Lowery AJ, Kell MR, Glynn RW, Kerin MJ, Sweeney KJ. Locoregional recurrence after breast cancer surgery: a systematic review by receptor phenotype. Breast Cancer Res Treat. 2012;133:831–841. doi: 10.1007/s10549-011-1891-6
33. Campbell EJ, Tesson M, Doogan F, Mohammed ZMA, Mallon E, Edwards J (2016) The combined endocrine receptor in breast cancer, a novel approach to traditional hormone receptor interpretation and a better discriminator of outcome than ER and PR alone. Br J Cancer 115(8):967–973
34. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26(Suppl 5):v8–30. doi: 10.1093/annonc/mdv298
35. Schulz KF., Altman DG., Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010; 152(11): 726-732)
36. Moossdorff M, et al.Maastricht Delphi consensus on event definitions for classification of recurrence in breast cancer research. J Natl Cancer Inst. 2014 Nov 7;106(12)
37. The AIOM (Italian Association of Medical Oncology) practice guidelines for breast neoplasms, Edizione 2020, last update 16/02/2021 https://www.aiom.it/linee-guida-aiom-2020-neoplasie-della-mammella)
38. Stokes ME, Thompson D, Montoya EL, Weinstein MC, Winer EP, Earle CC. Ten-year survival and cost following breast cancer recurrence: estimates from SEER-medicare data. Value Health. 2008 Mar-Apr;11(2):213-20. doi: 10.1111/j.1524-4733.2007.00226.x. PMID: 18380633.)
39. van Maaren, M.C., de Munck, L., Strobbe, L.J., Sonke, G.S., Westenend, P.J., Smidt, M.L., Poortmans, P.M. and Siesling, S. (2019), Ten-year recurrence rates for breast cancer subtypes in the Netherlands: A large population-based study. Int. J. Cancer, 144: 263-272. https://doi.org/10.1002/ijc.31914
40. Lee, Sae Byul et al. "A retrospective prognostic evaluation analysis using the 8th edition of the American Joint Committee on Cancer staging system for breast cancer." Breast cancer research and treatment vol. 169,2 (2018): 257-266. doi:10.1007/s10549-018-4682-5.
41. Haesook T.Kim "Cumulative Incidence in Competing Risks Data and Competing Risk Regression Analysis",Clin Cancer Res 2007;13(2)
42. R Development Core. Team: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015)

43. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat. 1988;16:1141–54
44. Pohar Perme M, Stare J, Estève J. On estimation in relative survival. Biometrics 2012; 68: 113-20.;
45. Pohar Perme, M, Estève J, Rachet B. Analysing population-based cancer survival—settling the controversies. BMC Cancer 2016; 16: 933
46. Aagaard Rasmussen L, Jensen H, Flytkjær Virgilsen L, Jellesmark Thorsen LB, Vrou Offersen B, Vedsted P. A validated algorithm for register-based identification of patients with recurrence of breast cancer-Based on Danish Breast Cancer Group (DBCG) data. Cancer Epidemiol. 2019 Apr; 59:129-134. doi: 10.1016/j.canep.2019.01.016. Epub 2019 Feb 8. PMID: 30743224.
47. Xu, Y., Kong, S., Cheung, W.Y. et al. Development and validation of case-finding algorithms for recurrence of breast cancer using routinely collected administrative data. BMC Cancer 19, 210 (2019). https://doi.org/10.1186/s12885-019-5432-8
48. Lambert, P., Pitz, M., Singh, H. et al. Evaluation of algorithms using administrative health and structured electronic medical record data to determine breast and colorectal cancer recurrence in a Canadian province. BMC Cancer 21, 763 (2021). https://doi.org/10.1186/s12885-021-08526-9
49. Kemp-Casey A, Roughead EE, Saunders C, Boyle F, Lopez D, Bulsara M, Preen DB. Breast cancer recurrence following active treatment: determining its incidence in the NSW population. Public Health Res Pract. 2016;26(1):e2611607
50. Rack B, Janni W, Gerber B, Strobl B, Schindlbeck C, Klanner E, et al. Patients with recurrent breast cancer: does the primary axillary lymph node status predict more aggressive tumor progression? Breast Cancer Res Treat. 2003;82(2):83−92)
51. Holleczek B, Stegmaier C, Radosa JC, Solomayer EF, Brenner H. Risk of loco-regional recurrence and distant metastases of patients with invasive breast cancer up to ten years after diagnosis - results from a registry-based study from Germany. BMC Cancer. 2019 May 30;19(1):520. doi: 10.1186/s12885-019-5710-5. PMID: 31146706; PMCID: PMC6543576.)
52. Esserman LJ, Moore DH, Tsing PJ, et al. Biologic markers determine both the risk and the timing of recurrence in breast cancer. Breast Cancer Research and Treatment. 2011 Sep;129(2):607-616. DOI: 10.1007/s10549-011-1564-5. PMID: 21597921; PMCID: PMC4324750
53. Molina MA, Codony-Servat J, Albanell J, Rojo F, Arribas J, Baselga J. Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. Cancer Res 2001; 61: 4744-9.)
54. Rosie et al. Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials Bradley,.The Lancet Oncology, Volume 22, Issue 8, 1139 – 1150
55. Cheng L, Swartz MD, Zhao H, Kapadia AS, Lai D, Rowan PJ, Buchholz TA, Giordano SH. Hazard of recurrence among women after primary breast cancer treatment--a 10-year follow-up using data from SEER-Medicare. Cancer Epidemiol Biomarkers Prev. 2012 May;21(5):800-9. doi: 10.1158/1055-9965.EPI-11-1089. Epub 2012 Mar 16. Erratum in: Cancer Epidemiol Biomarkers Prev. 2012 Sep;21(9):1604-5. PMID: 22426147
56. Nordstrom BL, Simeone JC, Malley KG, et al. Validation of Claims Algorithms for Progression to Metastatic Cancer in Patients with Breast, Non-small Cell Lung, and Colorectal Cancer. *Front Oncol*. 2016; 6:18. Published 2016 Feb 1. doi:10.3389/fonc.2016.00018

57. Ling, A.Y., Kurian, A., Caswell-Jin, J., Sledge, G., Shah, N., & Tamang, S.R. (2019). A Semi-Supervised Machine Learning Approach to Detecting Recurrent Metastatic Breast Cancer Cases Using Linked Cancer Registry and Electronic Medical Record Data. *ArXiv, abs/1901.05958*.

58. Lord Sarah J, Belinda E Kiely, Sallie-Anne Pearson, Benjamin Daniels, Dianne L O'Connell, Jane Beith, Max K Bulsara, Nehmat Houssami, Metastatic breast cancer incidence, site and survival in Australia, 2001–2016: a population-based health record linkage study protocol, BMJ Open, 10.1136/bmjopen-2018-026414, **9**, 2, (e026414), (2019

59. Minicozzi P, Bella F, Toss A, Giacomin A, Fusco M, Zarcone M, Tumino R, Falcini F, Cesaraccio R, Candela G, La Rosa F, Federico M, Sant M. Relative and disease-free survival for breast cancer in relation to subtype: a population-based study. J Cancer Res Clin Oncol. 2013 Sep;139(9):1569-77. doi: 10.1007/s00432-013-1478-1. Epub 2013 Jul 28. PMID: 23892409.

60. Fredholm, H., Magnusson, K., Lindström, L.S. et al. Long-term outcome in young women with breast cancer: a population-based study. Breast Cancer Res Treat 160, 131–143 (2016). https://doi.org/10.1007/s10549-016-3983-9

61. Schaffar R, Bouchardy C, Chappuis PO, Bodmer A, Benhamou S, Rapiti E. A population-based cohort of young women diagnosed with breast cancer in Geneva, Switzerland. PLoS One. 2019 Sep 6; 14(9):e0222136. doi: 10.1371/journal.pone.0222136

62. Lao, C., Kuper-Hommel, M., Elwood, M. et al. Metastatic relapse of stage I–III breast cancer in New Zealand. Cancer Causes Control 32, 753–761 (2021). https://doi.org/10.1007/s10552-021-01426-0

63. Siegel, R, Miller, K, Jemal, A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70: 7–30

64. Mariotto AB, Noone AM, Howlader N, Cho H, Keel GE, Garshell J, Woloshin S, Schwartz LM. Cancer survival: an overview of measures, uses, and interpretation. J Natl Cancer Inst Monogr. 2014 Nov;2014(49):145-86. doi: 10.1093/jncimonographs/lgu024. Erratum in: J Natl Cancer Inst Monogr. 2015 May;2015(51):97. PMID: 25417231; PMCID: PMC4829054.

65. Source: AIHW Australian Cancer Database https://nbcf.org.au/about-breast-cancer/breast-cancer-stats/)

66. Howlader N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. Cancer Epidemiol Biomarkers Prev. 2018 Jun;27(6):619-626. doi: 10.1158/1055-9965.EPI-17-0627. Epub 2018 Mar 28. PMID: 29593010.

67. Zheng Zhou, Elham Rahme, Michal Abrahamowicz, Louise Pilote, Survival Bias Associated with Time-to-Treatment Initiation in Drug Effectiveness Evaluation: A Comparison of Methods, American Journal of Epidemiology, Volume 162, Issue 10, 15 November 2005, Pages 1016–1023, https://doi.org/10.1093/aje/kwi307

68. Saleh, R.R., Nadler, M.B., Desnoyers, A. et al. Influence of Competing Risks on Estimates of Recurrence Risk and Breast Cancer-specific Mortality in Analyses of the Early Breast Cancer Trialists Collaborative Group. Sci Rep 10, 4091 (2020). https://doi.org/10.1038/s41598-020-61093-0.

69. A'mar T, Beatty JD, Fedorenko C, et al. Incorporating Breast Cancer Recurrence Events Into Population-Based Cancer Registries Using Medical Claims: Cohort Study [published correction appears in JMIR Cancer. 2020 Sep 24;6(2):e23821]. *JMIR Cancer*. 2020;6(2):e18143. Published 2020 Aug 17. doi:10.2196/18143

70. He X, Ye F, Zhao B, Tang H, Wang J, Xiao X, et al. (2017) Risk factors for delay of adjuvant chemotherapy in non-metastatic breast cancer patients: A systematic review and

meta-analysis involving 186982 patients. PLoS ONE 12(3): e0173862. https://doi.org/10.1371/journal.pone.0173862

71. Pedersen RN, Esen BÖ, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, Nørgaard M, Cronin-Fenton D. The Incidence of Breast Cancer Recurrence 10-32 Years after Primary Diagnosis. J Natl Cancer Inst. 2021 Nov 8:djab202. doi: 10.1093/jnci/djab202. Epub ahead of print. PMID: 34747484

5.7 TABLES

Table 1: Characteristics of eligible women diagnosed with invasive breast cancer between January 1, 2004, and December 31, 2010

Characteristic	All women (N = 5420)
Age, Median (IQR), year	60 (50–66)
Study follow-up, Median (IQR), year	10.5 (8.6-11.9)
Year of Diagnosis	N (%)
2004	747 (13.8)
2005	792 (14.6)
2006	1053 (19.4)
2007	769 (14.2)
2008	647 (11.9)
2009	566 (10.4)
2010	846 (15.6)
Stage T	
T1	3726 (68.8)
T2	1324 (24.4)
T3	93 (1.7)
T4	166 (3.06)
Unknown	111 (2.1)
Stage N	
N0	3225 (59.5)
N1	1408 (25.9)
N2	368 (6.8)
N3	225 (4.2)
Unknown	194 (3.6)
Stage	
I	2765 (51.0)
II	1934 (35.7)
III	721 (13.3)
Tumor grade	
G1	790 (14.6)
G2	2798 (51.6)
G3	1636 (30.2)
Unknown	196 (3.6)
Histology	
Ductal	4187 (77.3)
Lobular	650 (12.0)
Ductal lobular mixed	149 (2.7)
Other	434 (8.0)
Focality	
Unifocality	4234 (78.3)

Multifocality/Multicentric	985 (18.2)
Unknowm	192 (3.5)
Estrogen Receptor (ER) status	
Positive	4475 (82.6)
Negative	798 (14.7)
Unknown	147 (2.7)
Progesteron Receptor (PR) status	
Positive	3725 (68.7)
Negative	1522 (28.1)
Unknown	173 (3.2)
HER2 status	
Positive	799 (14.7)
Negative	4009 (74.0)
Unknown	612 (11.3)
Molecular Profile [a]	
Hormonal Receptors + (HR+)	3576 (66.0)
HER2 Positive	799 (14.7)
Triple Negative	420 (7.7)
Unknown	625 (11.5)
Status of patient	
Dead	950 (17.5)
Alive	4421 (81.6)
Lost at follow-up	49 (0.9)
Cause of death	
Death due to breast cancer	539 (56.7)
Death due to other causes	382 (40.2)
Unknown	29 (3.1)

[a] HR: ER+ or PR+ or both + and HER-; HER2 positive: independently by HR status; unknown: a least one among ER/PR/HER2 missing

Table 2: Stage I-III Invasive Breast Cancer Cases in Friuli Venezia Giulia Region diagnosed between 2004-2010: identification of breast cancer recurrence in all cohort and stratifying respect to molecular profile using ICD9-CM administrative codes

	Start of Surveillance Period		**ICD9-CM code that could indicate a second breast cancer event**				**Breast cancer event identified by ICD9-CM code**	**Status at the end of follow-up (31/12/2017)**		
Molecular Profile N (%)	**Alive after 18 months**[a] N (%)[b]	**Second primary cancer before any recurrence** N (%)[c]	**CT code** N (%)[c]	**RT code** N (%)[2]	**Breast surgery code** N (%)[2]	**Hospital admission without any CT/RT/Breast Surgery claim** N (%)[2]	**Any Recurrence** N (%)[2]	**Alive at the end of FU** N (%)[2]	**Death from BC** N (%)[2]	**Deaths from other or unknown causes** N (%)[2]
All Cohort 5420	5268 **(97.2%)**	263 (5.0%)	768 (14.2%)	441 (8.4%)	433 (8.2%)	294 (5.5%)	**1406 (26.7%)**	4423 (83.9%)	486 (9.2%)	359 (6.8%)
HR+ 3576 (66.0%)	3491 (97.6%)	164 (4.7%)	421 (12.1%)	267 (7.6%)	262 (7.5%)	186 (5.3%)	**842 (24.1%)**	2997 (85.8%)	256 (7.3%)	238 (6.8%)
HER2+ 799 (14.7%)	767 (96.0%)	47 (6.1%)	131 (17.0%)	81 (10.7%)	77 (10.0%)	41 (5.3%)	**221 (28.8%)**	631 (82.2%)	91 (11.9%)	45 (5.9%)
TN 420 (7.7%)	399 (95.0%)	22 (5.5%)	91 (22.8%)	43 (10.5%)	30 (7.5%)	31 (7.8%)	**142 (35.6%)**	298 (74.7%)	80 (20.0%)	21 (5.2%)
Unknown 625 (11.5%)	611 (97.8%)	30 (4.9%)	125 (20.5%)	50 (8.2%)	64 (10.5%)	36 (5.9%)	**201 (32.3%)**	497 (81.3%)	59 (9.7%)	55 (9.0%)

[a]24 months for HER2 positive BC ; [b] of enrolled ; [c] of Alive at 18/24months

Table 3: Stage I-III Invasive Breast Cancer Cases in Friuli Venezia Giulia Region diagnosed between 2004-2010: identification of breast cancer recurrence in all cohort and stratifying respect to Stage using ICD9-CM administrative codes

	Start of Surveillance Period		**ICD9-CM code that could indicate a second breast cancer event**				**Breast cancer event identified by ICD9-CM code**	**Status at the end of Follow-up (31/12/2017)**		
Stage N (%)	**Alive after 18 months**[a] N (%)[b]	**Second primary cancer before any recurrence** N (%)[c]	**CT code** N (%)[c]	**RT code** N (%)[2]	**Breast surgery code** N (%)[2]	**Hospital admission without any CT/RT/Breast Surgery claim** N (%)[2]	**Any Recurrence N (%)[2]**	**Alive at the end of FU** N (%)[2]	**Death from BC** N (%)[2]	**Deaths from other or unknown causes** N (%)[2]
All Cohort 5420	5268 **(97.2%)**	263 (5.0%)	768 (14.2%)	441 (8.4%)	433 (8.2%)	294 (5.5%)	**1406 (26.7%)**	4423 (83.9%)	486 (9.2%)	359 (6.8%)
I 2765 (51.0%)	2705 (97.8%)	147 (5.4%)	207 (7.7%)	146 (5.4%)	227 (8.4%)	99 (3.7%)	**490 (18.1%)**	2460 (90.9%)	82 (3.0%)	163 (6.0%)
II 1934 (35.7%)	1887 (97.6%)	81 (4.3%)	322 (17.1%)	186 (9.9 %)	154 (8.2%)	132 (7.0%)	**578 (30.6%)**	1555 (82.4%)	194 (10.3%)	138 (7.3%)
III 721 (13.3%)	676 (93.8%)	35 (5.2 %)	239 (35.4%)	109 (16.1%)	52 (7.9%)	63 (9.3%)	**338 (50.0%)**	408 (60.4%)	210 (31.1%)	58 (8.5%)

[a]24 months for HER2 positive BC ; [b] of enrolled ; [c] of Alive at 18/24months

Table 4: Total Person Years (PY) and Recurrence rate of breast cancer per 1000 PY with associate 95% confidence interval (CI) for all cohort and by Age, Molecular Subtype and Stage

Outcome variable	Persons at risk at 18/24 months	Number of Any Recurrence (N, %)	Total Person Years (PY)	Recurrence Rate per 1000 PY (95% CI)	p-value (Gray's Test)
All Cohort	5268	1406 (27.0%)	45775	30.7 (29.2-32.4)	
Age					
<50	1309	451 (34.4%)	10915	41.3 (37.7-45.3)	
50-69	3262	767 (23.5%)	29255	26.2 (24.4-28.1)	
>=70	697	188 (27.0%)	5605	33.5 (29.1-38.7)	<0.001
Molecular Subtype					
HR+	3491	842 (24.1%)	30970	27.2 (25.4-29.1)	
HER2+	767	221 (28.8%)	6679	33.1 (29.0-37.8)	
TN	399	142 (35.6%)	3038	46.7 (39.7-55.1)	<0.001
Nodal Status					
Negative	3526	737 (20.9%)	32075	23.0 (21.4-24.7)	
Positive	1562	607 (38.9%)	12354	49.1 (45.4-53.2)	<0.001
Stage					
I	2705	490 (18.1%)	24901	19.7 (18.0-21.5)	
II	1887	578 (30.6%)	16195	35.7 (32.9-38.7)	
III	676	338 (50.0%)	4680	72.2 (64.9-80.3)	<0.001
Histopatologic Grade					
G1	767	121 (15.8%)	7180	16.9 (14.1-20.1)	
G2	2740	690 (25.2%)	24358	28.3 (26.3-30.5)	
G3	1575	526 (33.4%)	12727	41.3 (37.9-45.2)	<0.001

Table 5: Three, Five and Ten-Year Cumulative Recurrence of Breast Cancer with associate Confidence Interval (95%CI) for all cohort and by Age, Molecular Subtype and Stage. Cumulative recurrence was estimate taking into account mortality for all causes and second tumors as the competing risks

Category	Persons at risk at 18/24 month	3-years Cumulative Recurrence (95%CI)	5-years Cumulative Recurrence (95%CI)	10-years Cumulative Recurrence (95%CI)
All Cohort	5268	10.2% (9.2%-10.9%)	15.9% (15.0%-16.9%)	26.0% (24.8%-27.0%)
Age				
<50	1309	15.6% (13.7%-17.6%)	22.3 % (20.1%-24.6%)	33.5% (30.9%-36.2%)
50-69	3262	7.7% (6.8%-8.7%)	13.1% (12.0%-14.3%)	22.7% (21.1%-24.2%)
>=70	697	10.5% (8.3%-12.9%)	17.1% (14.4%-20.0%)	27.6% (24.1%-31.1%)
Molecular Subtype				
HR+	3491	7.4% (6.6%-8.3%)	13.1% (12.0%-14.3%)	23.5% (22.0%-25.0%)
HER2+	767	10.6% (8.5%-12.9%)	17.5% (14.9%-20.2%)	26.9% (23.8%-30.1%)
TN	399	23.6% (19.5%-27.9%)	29.9% (25.4%-34.4%)	36.0% (31.3%-40.8%)

Nodal Status				
Negative	3526	6.7% (5.9%-7.5%)	11.2% (10.1%-12.2%)	20.4% (19.0%-21.8%)
Positive	1562	16.8% (15.0%-18.7%)	25.9% (23.7%-28.1%)	37.6% (35.1%-40.0%)
Stage				
I	2705	5.1% (4.3%-5.9%)	9.1 %(8.1%-10.2%)	17.7% (16.2%-19.3%)
II	1887	12.0% (10.6%-13.5%)	18.6%(16.8%-20.3%)	29.4% (27.3%-31.6%)
III	676	24.6% (21.4%-27.9%)	36.0% (32.4%-39.6%)	49.3% (45.4%-53.1%)
Histopathologic Grade				
G1	767	4.8% (3.5%-6.5%)	7.3% (5.6%-9.3%)	15.8% (13.2%-18.7%)
G2	2740	7.7% (6.7%-8.7%)	13.8% (12.5%-15.1%)	23.4% (21.4%-25.6%)
G3	1575	16.4% (14.6%-18.3%)	23.4% (21.4%-25.6%)	32.4% (30.1%-34.8%)

Table 6: Five and Ten-Year Relative survival among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017: a) all cohort, b) by age, c) by Molecular Subtype and d) by Stage at diagnosis

	n	**5-year RS (95%CI)**	**10-year RS (95%CI)**
Breast Cancer			
All Cohort	5420	95.8% (95.0%-96.5%)	89.8% (88.6%-90.0%)
Age			
<49	1332	96.4% (95.2%-97.4%)	91.3% (89.4%-92.9%)
50-69	3358	95.6% (94.6%-96.4%)	90.2% (88.7%-91.6%)
70-74	730	95.3% (92.0%-97.3%)	84.5% (79.5%-88.4%)
Molecular Subtype			
HR+	3576	97.9% (97.0%-98.5%)	92.1% (90.6%-93.4%)
HER2+	799	94.2% (91.9%-95.9%)	86.3% (82.9%-89.1%)
TN	420	80.5% (76.1%-84.3%)	76.6% (71.5%-80.9%)
Stage			
I	2765	99.8% (96.4%-100.0%)	97.9% (96.1%-98.9%)
II	1934	95.2% (93.8%-96.3%)	88.8% (86.7%-90.6%)
III	721	81.5% (78.2%-84.3%)	61.9% (57.6%-65.9%)
Stage and Molecular Profile			
Stage I-HR+	1894	100.0%	99.1% (96.3%-99.8%)
Stage I-HER2+	346	98.9% (94.3%-99.8%)	95.6% (90.3%-98.0%)
Stage I-TN	170	94.6% (88.2%-97.6%)	91.7% (83.7%-95.9%)
Stage II-HR+	1245	97.1% (95.3%-98.2%)	89.8% (87.2%-91.9%)
Stage II-HER2+	301	95.2% (91.1%-97.5%)	88.7% (83.2%-92.5%)
Stage II-TN	186	80.4% (73.5%-85.7%)	78.8% (71.5%-84.4%)
Stage III-HR+	437	87.7% (83.7%-90.7%)	67.9% (62.5%-72.6%)
Stage III-HER2+	150	79.9% (72.0%-85.7%)	60.2% (51.1%-68.1%)
Stage III-TN	64	42.7% (30.0%-54.8%)	29.8% (18.4%-42.1%)

RS: relative survival

5.8 FIGURES

Figure 1: Flow-chart of inclusion and exclusion criteria

%Cum	Incident Breast Cancer 2004-2010 (FVG-CR)		N° Excluded	% Excluded	Criteria of Exclusion
100.0%	**8,940**				
	↓	→	36	0.4%	Death Certificate Only (DCO) (Data diagnosis =Data of Death)
99.6%	8,904				
	↓	→	22	0.2%	No follow-up (0 days between data of diagnosis and end of study)
99.4%	8,882				
	↓	→	2223	25.0%	Age >74 years
74.5%	6,659				
	↓	→	213	3.2%	Previous cancers (< 90 days from incidence BC data)
72.1%	6,446				
	↓	→	59	0.9%	Synchronous cancers (< 90 days from incidence BC data)
71.4%	6,387				
	↓	→	1	0.0%	In situ breast carcinoma
71.4%	6,386				
	↓	→	741	11.6%	Missing Stage
63.1%	5,645				
	↓	→	225	4.0%	De Novo Metastatic Breast Cancers
60.6%	5,420				

BC= Breast Cancer

Figure 2. Schematic overview of the algorithm

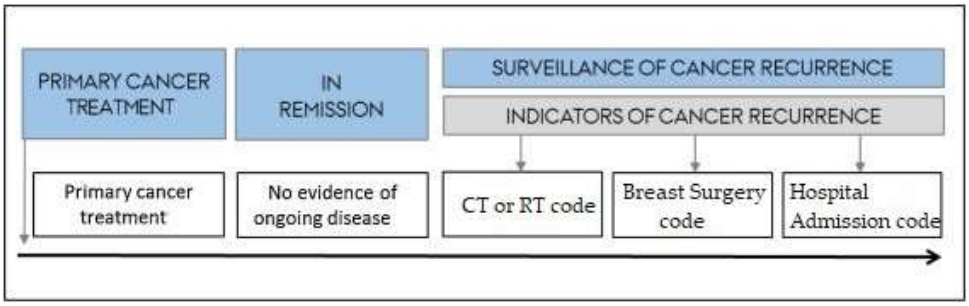


Figure 3. Overall summary of Recurrence among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017 (* 24 months for HER2 positive breast cancers)

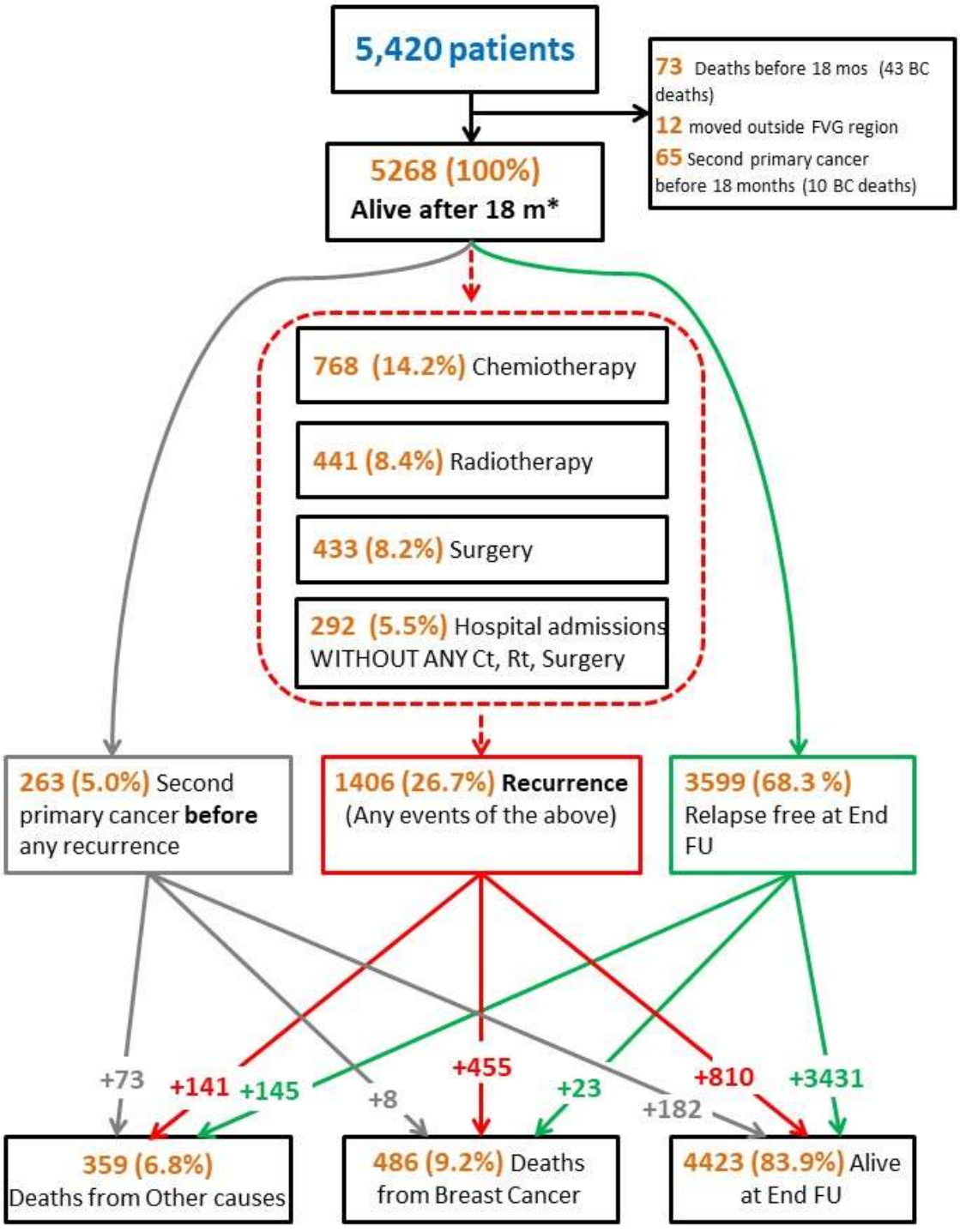


Figure 4. Cumulative Incidence (CI) of Recurrence among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017. CI was estimate taking into account mortality for all causes and second tumors as the competing risks.

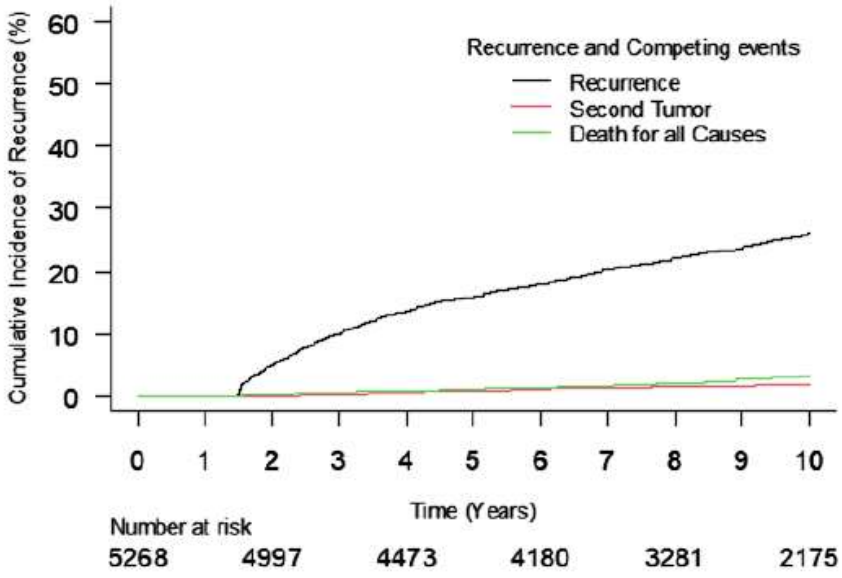


Figure 5: Cumulative Incidence (CI) of Recurrence among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017, a) by Age, b) Nodal Status, c) Molecular Subtype, d) Stage and histopathologic Grade. Cumulative recurrence was estimate taking into account mortality for all causes and second tumors as the competing risks.

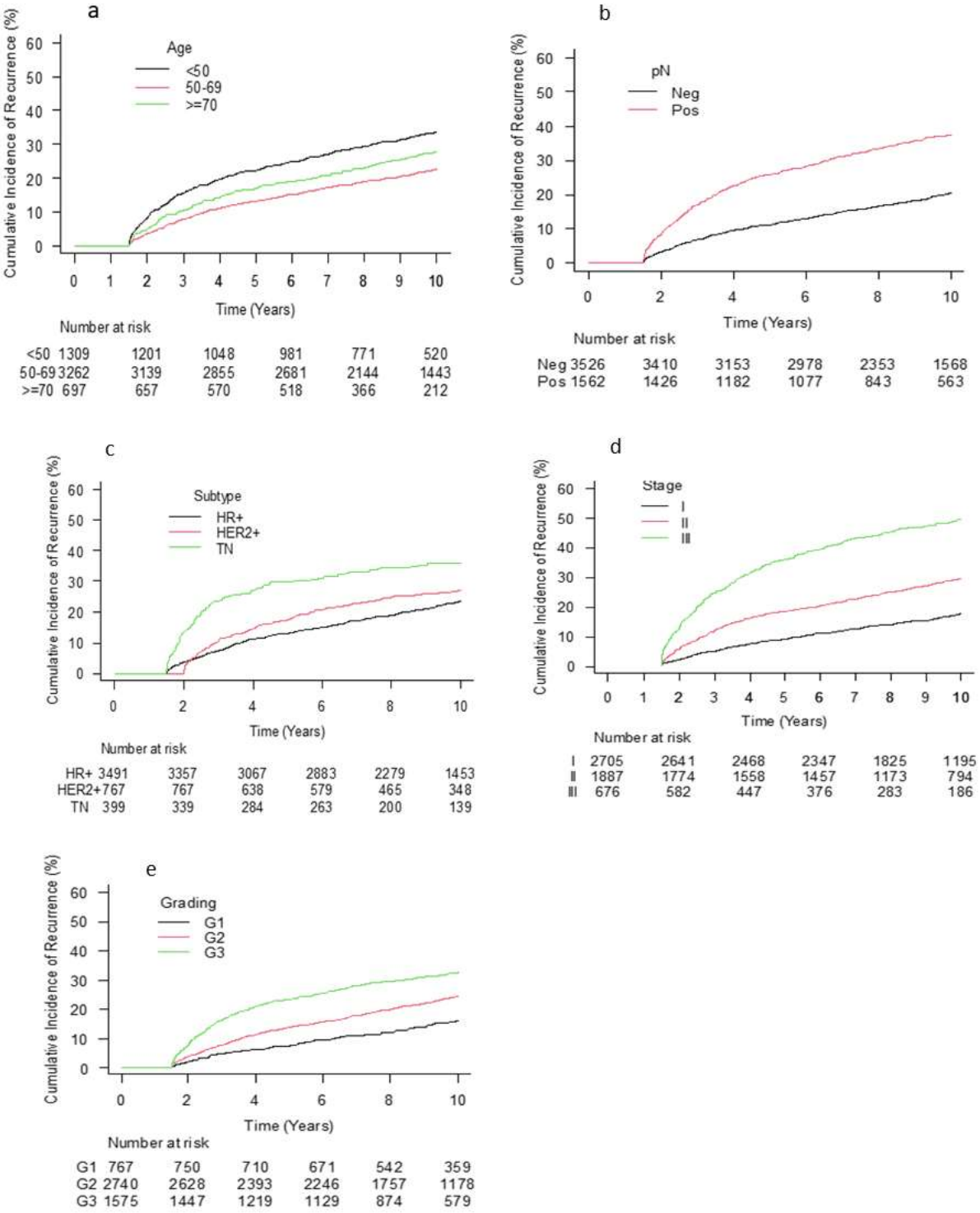
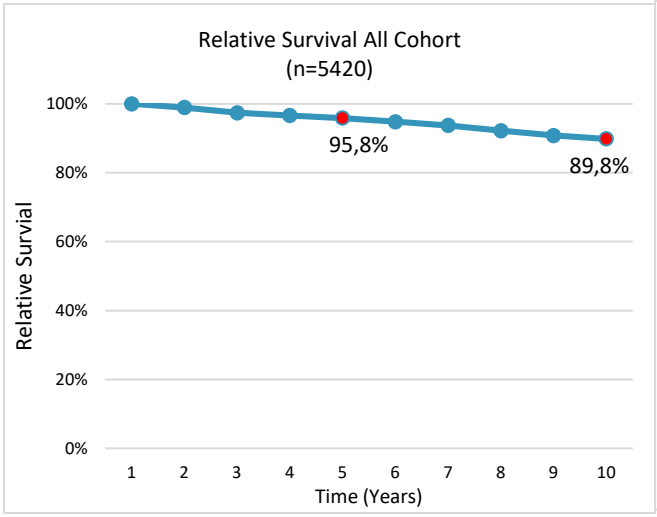
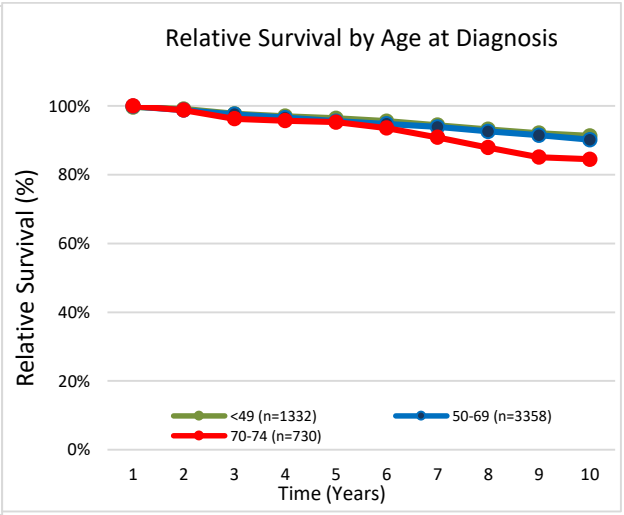


Figure 6: Relative survival among stage I-III breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017: a) all cohort, b) by age, c) by Molecular Subtype and d) by Stage at diagnosis

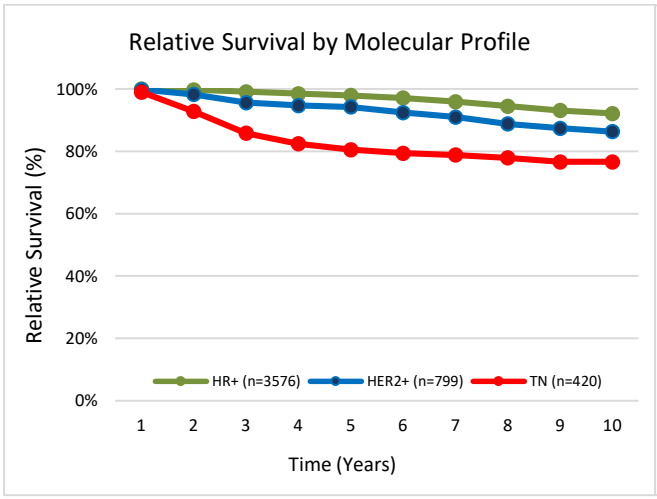
a



b



c



d

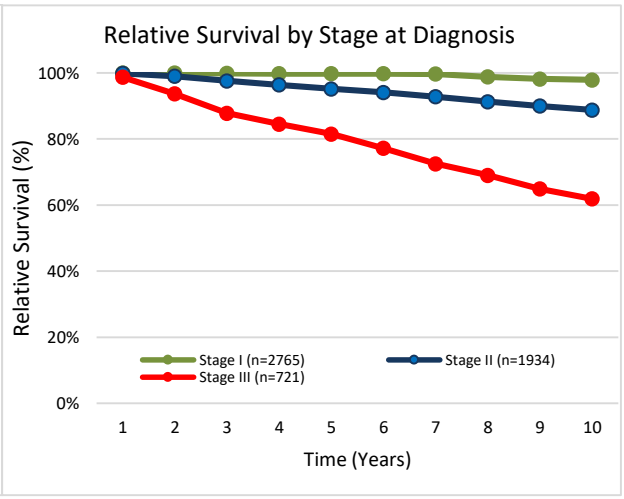
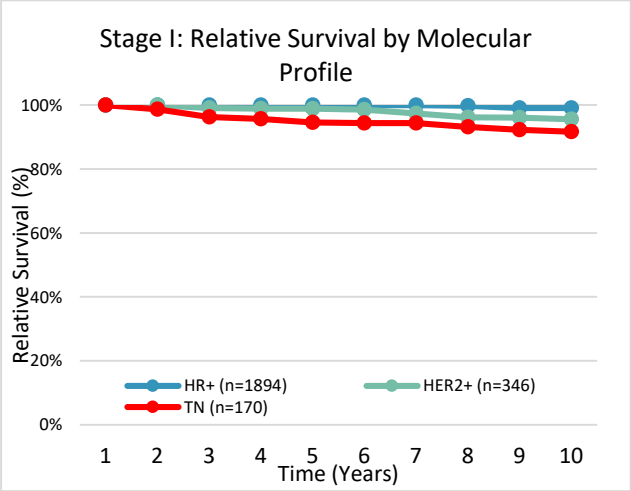
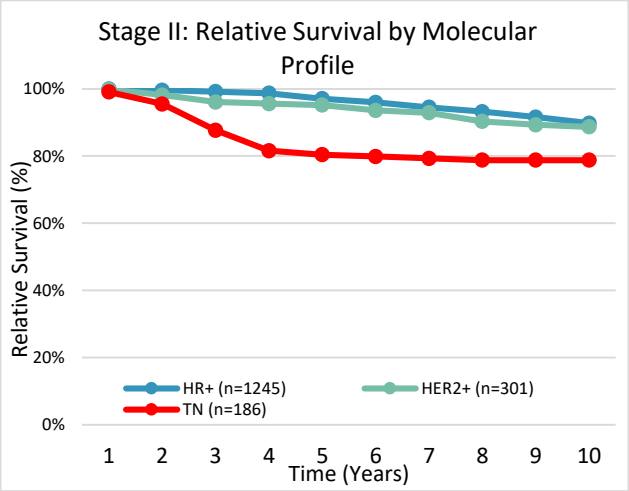


Figure 7: Relative survival by stage and molecular profile among breast cancer women diagnosed in Friuli Venezia Giulia region during 2004-2010 with follow-up through 31/12/2017: a) Stage I, b) Stage II and c) Stage III

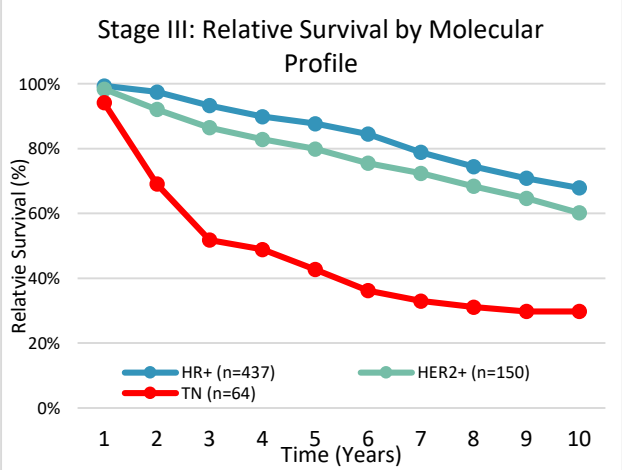
a



b



c



5.9 APPENDIX A

ADMINISTRATIVE DATABASE EXTRACTIONS

- **FVG-CR**

RECORD LAYOUT FVG-CR

Nome variabile	Descrizione variabile/codifica
ID_PAZIENTE	Identifica in modo univoco la persona.
IDCASO	Identifica in modo univoco il tumore.
ETA_DIAGNOSI	Età alla diagnosi in anni compiuti
DATA_DIAGNOSI	Data di incidenza
TOPOGRAFIA	Codice topografia ICD-O3
MORFOLOGIA	Morfologia e comportamento biologico ICD-O3
LATERALITA	Lato della lesione 1= monolaterale 2=bilaterale 3=destro 4=sinistro 9=ignoto
ICD-10	
GRADO	Grading
BASE_DIAGNOSI	Base della diagnosi: 0=DCO 1=Clinica 2=Indagini cliniche 4=Marker tumorali 5=Citologica 6=Istologica su metastasi 7=Istologica su tumore primitivo 8=Autopsia con istologia 9=Ignota
PT	Stadio patologico T
PN	Stadio patologico N
PM	Stadio patologico M
STADIO	Raggruppamento in stadi TNM
VERSIONETNM	Versione stadio TNM
LINFONODI	Numero di linfonodi analizzati
LINFONODI_POSITIVI	Numero di linfonodi positivi
LINFONODO_SENTINELLA	Linfonodo sentinella 1=eseguito 2=non eseguito 9=ignoto
FOCALITA	Focus 1=unifocale 2=multifocale 3=multicentrico 4=multifocale e multicentrico 9=ignoto
DISSEZIONE	Dissezione ascellare 1=effettuata 2=non effettuata 9=ignota
ER	Recettori estrogeni

	POS=positivo NEG=negativo XXX=ignoto
PR	Recettori progesterone POS=positivo NEG=negativo XXX=ignoto
HER2_NEU	Her2/neu POS=positivo NEG=negativo XXX=ignoto
STATO_VITA	Stato in vita 1 = vivo 2 = deceduto 3 = perso al follow-up
DATA_FOLLOW_UP	Data di follow-up della persona. Rappresenta la data di decesso per i deceduti e la data di trasferimento in altra regione per i persi al follow-up.
CAUSA_MORTE	Causa di morte, codifica ICD-9

- **HOSPITAL DISCHARGE DATA (HDD)**

RECORD LAYOUT ***HDD***

Nome variabile	Descrizione variabile/codifica
ID_PAZIENTE	Identificativo della persona.
ID_RICOVERO	Identificativo del ricovero.
REGIME_RICOVERO	1=ricovero ordinario 2=ricovero diurno (day hospital)
DATA_RICOVERO	In caso di ricovero diurno è indicata la data del 1° giorno del ciclo di contatti con la struttura
GIORNI_DEGENZA	Numero giorni di degenza nel caso di ricoveri ordinari; numero di accessi nell'arco dello stesso ciclo assistenziale nel caso di ricoveri DH
DATA_DIMISSIONE	In caso di ricovero diurno la data di dimissione corrisponde alla data dell'ultimo accesso presso la struttura in cui si è svolto il ciclo assistenziale
diagnosi_1	Diagnosi principale (classificazione ICD9-CM)
Diagnosi_lateralita_1	Ove applicabile specificare se la diagnosi principale si riferisce al lato destro, sinistro o bilaterale
Diagnosi_stadio_1	Indica lo stadio della neoplasia maligna riportata come diagnosi di dimissione principale
Diagnosi_nota_1	Indica se la diagnosi principale rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_1	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
diagnosi_2	Prima diagnosi secondaria (classificazione ICD9-CM)

Diagnosi_lateralita_2	Ove applicabile specificare se la prima diagnosi secondaria si riferisce al lato destro, sinistro o bilaterale
Diagnosi_stadio_2	Indica lo stadio della neoplasia maligna riportata come prima diagnosi secondaria di dimissione
Diagnosi_nota_2	Indica se la prima diagnosi secondaria rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_2	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
diagnosi_3	Seconda diagnosi secondaria (classificazione ICD9-CM)
Diagnosi_lateralita_3	Ove applicabile specificare se la seconda diagnosi secondaria si riferisce al lato destro, sinistro o bilaterale
Diagnosi_stadio_3	Indica lo stadio della neoplasia maligna riportata come seconda diagnosi secondaria di dimissione
Diagnosi_nota_3	Indica se la seconda diagnosi secondaria rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_3	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
diagnosi_4	Terza diagnosi secondaria (classificazione ICD9-CM)
Diagnosi_lateralita_4	Ove applicabile specificare se la terza diagnosi secondaria si riferisce al lato destro, sinistro o bilaterale
Diagnosi_stadio_4	Indica lo stadio della neoplasia maligna riportata come terza diagnosi secondaria di dimissione
Diagnosi_nota_4	Indica se la terza diagnosi secondaria rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_4	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
diagnosi_5	Quarta diagnosi secondaria (classificazione ICD9-CM)
Diagnosi_lateralita_5	Ove applicabile specificare se la quarta diagnosi secondaria si riferisce al lato destro, sinistro o bilaterale
Diagnosi_stadio_5	Indica lo stadio della neoplasia maligna riportata come quarta diagnosi secondaria di dimissione
Diagnosi_nota_5	Indica se la quarta diagnosi secondaria rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_5	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
diagnosi_6	Quinta diagnosi secondaria (classificazione ICD9-CM)
Diagnosi_lateralita_6	Ove applicabile specificare se la quinta diagnosi secondaria si riferisce al lato destro, sinistro o bilaterale

Diagnosi_stadio_6	Indica lo stadio della neoplasia maligna riportata come quinta diagnosi secondaria di dimissione
Diagnosi_nota_6	Indica se la quinta diagnosi secondaria rilevata alla dimissione era presente anche al momento del ricovero, oppure se è stata individuata attraverso l'anamnesi o diagnosticata successivamente all'ammissione, ma comunque preesistente nel paziente e non insorta durante il ricovero.
cod_correlato_diagnosi_6	Se ha valore 1, vuol dire che il codice della diagnosi corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_1	Codice intervento principale (classificazione ICD9-CM)
data_intervento_1	Data intervento principale
cod_correlato_intervento_1	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_2	Codice primo intervento secondario (classificazione ICD9-CM)
data_intervento_2	Data primo intervento secondario
cod_correlato_intervento_2	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_3	Codice secondo intervento secondario (classificazione ICD9-CM)
data_intervento_3	Data secondo intervento secondario
cod_correlato_intervento_3	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_4	Codice terzo intervento secondario (classificazione ICD9-CM)
data_intervento_4	Data terzo intervento secondario
cod_correlato_intervento_4	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_5	Codice quarto intervento secondario (classificazione ICD9-CM)
data_intervento_5	Data quarto intervento secondario
cod_correlato_intervento_5	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_6	Codice quinto intervento secondario (classificazione ICD9-CM)
data_intervento_6	Data quinto intervento secondario
cod_correlato_intervento_6	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_7	Codice sesto intervento secondario (classificazione ICD9-CM)
data_intervento_7	Data sesto intervento secondario
cod_correlato_intervento_7	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_8	Codice settimo intervento secondario (classificazione ICD9-CM)
data_intervento_8	Data settimo intervento secondario
cod_correlato_intervento_8	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.

intervento_9	Codice ottavo intervento secondario (classificazione ICD9-CM)
data_intervento_9	Data ottavo intervento secondario
cod_correlato_intervento_9	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_10	Codice nono intervento secondario (classificazione ICD9-CM)
data_intervento_10	Data nono intervento secondario
cod_correlato_intervento_10	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.
intervento_11	Codice decimo intervento secondario (classificazione ICD9-CM)
data_intervento_11	Data decimo intervento secondario
cod_correlato_intervento_11	Se ha valore 1, vuol dire che il codice dell'intervento corrispondente è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.

- **OUTPATIENT SERVICES**

RECORD LAYOUT OUTPATIENT SERVICES

Nome variabile	Descrizione variabile/codifica
ID_PAZIENTE	Identificativo della persona.
FONTE	Fonte del flusso
DATA_PRESTAZIONE	Data di erogazione prestazione. Nel caso di un ciclo di prestazioni è riportata la data di chiusura ciclo.
CODICE_ESENZIONE	Codice esenzione ricetta
CODICE_PRESTAZIONE	Codice prestazione secondo il Nomenclatore Tariffario regionale, da linkare con il dizionario delle prestazioni
QUANTITA	Quantità=001 di default Se si tratta di cicli di prestazioni indicare il numero effettivo di prestazioni erogate
CODICE_BRANCA	Codice della branca della prestazione, da linkare con il dizionario delle branche.
COD_SPECIFICO	Se ha valore 1, vuol dire che il codice della prestazione è correlato al trattamento del tumore della mammella secondo l'elenco utilizzato nello studio sui costi.

DIZIONARIO_PRESTAZIONI_AMB *(Ad ogni codice corrispondono più descrizioni)*

Nome variabile	Descrizione variabile/codifica
CODICE_PRESTAZIONE	Codice prestazione secondo il Nomenclatore Tariffario regionale
DESCRIZIONE_PRESTAZIONE	Descrizione della prestazione

DIZIONARIO_BRANCA_AMB

Nome variabile	Descrizione variabile/codifica
CODICE_BRANCA	Codice della branca
BRANCA	Branca della prestazione

5.10 APPENDIX B

Breast Cancer Adjuvant Treatment Times and protocols according to molecular profile

- **Triple Negative Adjuvant Scheme:**

Diagnostic Biopsy →within max 2 months→ surgery →1-2 months→Adjuvant Chemotherapy for 4-6 months→Follow up

- **HER2+ Adjuvant Scheme:**

Diagnostic Biopsy →within max 2 months→ surgery →1-2 months→Adjuvant Chemotherapy for 4-6 months→ Adjuvant Trastuzumab for 8-9 months (every 3 weeks)→Follow up

- **HR+ Adjuvant Scheme:**

Diagnostic Biopsy →within max 2 months→ surgery →1-2 months→Adjuvant Chemotherapy for 4-6 months→Follow up + Adjuvant Hormonotherapy (daily tablets that the patient takes at home)

Low-risk HR+ patients are not treated with adjuvant chemotherapy but only with adjuvant hormonotherapy post-surgical treatment.

Radiotherapy

Radiotherapy generally lasts one month in all cases and it is performed, if indicated, after 1-2 months from the end of the adjuvant chemotherapy.

5.11 APPENDIX C

Hospital and Outpatient Administrative Codes (ICD9-CM) used to Identify Breast Cancer Recurrence

Intervention	ICD9-CM CODE	DESCRIPTION	Database
Chemotherapy	**PROCEDURE**		-Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera) -Outpatient Services Database (Prestazioni Ambulatoriali)
	99.25	Injection or infusion of cancer chemotherapeutic substance	
	99.28	Injection or infusion of biological response modifier ÝBRM¨ as an antineoplastic agent	
	DIAGNOSIS		Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	V58.1	Encounter for chemotherapy and immunotherapy for neoplastic conditions	
	V58.11	Encounter for antineoplastic chemotherapy	
	ICD9-CM CODE	**DESCRIPTION**	
Radiotherapy	**PROCEDURE**		-Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera) -Outpatient Services Database (Prestazioni Ambulatoriali)
	92.23	Radioisotopic teleradiotherapy	
	99.24	Teleradiotherapy using photons	
	92.25	Teleradiotherapy using electrons	
	92.26	Teleradiotherapy of other particulate radiation	
	92.27	Implantation or insertion of radioactive elements	
	92.28	Injection or instillation of radioisotopes	
	92.29	Other radiotherapeutic procedure	
	DIAGNOSIS		
	V580	Radiotherapy	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	ICD9-CM CODE	**DESCRIPTION**	
	PROCEDURE		
	85.20	Excision or destruction of breast tissue, not otherwise specified	
	85.21	Local excision of lesion of breast	
	85.22	Resection of quadrant of breast	
	85.23	Subtotal mastectomy	
	85.24	Excision of ectopic breast tissue	

Surgery	85.25	Excision of nipple	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	85.33	Unilateral subcutaneous mammectomy with synchronous implant	
	85.34	Other unilateral subcutaneous mammectomy	
	85.35	Bilateral subcutaneous mammectomy with synchronous implant	
	85.36	Other bilateral subcutaneous mammectomy	
	85.41	Unilateral simple mastectomy	
	85.42	Bilateral simple mastectomy	
	85.43	Unilateral extended simple mastectomy	
	85.44	Bilateral extended simple mastectomy	
	85.45	Unilateral radical mastectomy	
	85.46	Bilateral radical mastectomy	
	85.47	Unilateral extended radical mastectomy	
	85.48	Bilateral extended radical mastectomy	
	ICD9-CM CODE	**DESCRIPTION**	
Malignant neoplasm of female Breast	174.0	Nipple and areola	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	174.1	Central portion	
	174.2	Upper-inner quadrant	
	174.3	Lower-inner quadrant	
	174.4	Upper-outer quadrant	
	174.5	Lower-outer quadrant	
	174.6	Axillary tail	
	174.8	Other specified sites of female breast	
	174.9	Breast (female), unspecified	
Secondary and unspecified malignant neoplasm of lymph nodes	196.0	Lymph nodes of head, face, and neck	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	196.1	Intrathoracic lymph nodes	
	196.2	Intra-abdominal lymph nodes	
	196.3	Lymph nodes of axilla and upper limb	
	196.5	Lymph nodes of inguinal region and lower limb	
	196.6	Intrapelvic lymph nodes	
	196.8	Lymph nodes of multiple sites	
	196.9	Site unspecified Lymph nodes NOS	
Secondary malignant	197.0	Secondary malignant neoplasm of the lung	
	197.1	Secondary malignant neoplasm of the mediastinum	

neoplasm of respiratory and digestive systems	197.2	Secondary malignant neoplasm of the pleura	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	197.3	Secondary malignant neoplasm of other respiratory organs	
	197.4	Secondary malignant neoplasm of the small intestine, including duodenum	
	197.5	Secondary malignant neoplasm of the large intestine and rectum	
	197.6	Secondary malignant neoplasm of the retroperitoneum and peritoneum	
	197.7	Secondary malignant neoplasm of the liver	
	197.8	Secondary malignant neoplasm of the other digestive organs and spleen	
Secondary malignant neoplasm of other specified sites	198.0	Secondary malignant neoplasm of the kidney	Hospital Discharge Database (SDO-Schede di Dimissione Ospedaliera)
	198.1	Secondary malignant neoplasm of other urinary organs	
	198.2	Secondary malignant neoplasm of the skin	
	198.3	Secondary malignant neoplasm of the brain and spinal cord	
	198.4	Secondary malignant neoplasm of the other parts of the nervous system	
	198.5	Secondary malignant neoplasm of the bone and bone marrow	
	198.6	Secondary malignant neoplasm of the ovary	
	198.7	Secondary malignant neoplasm of the adrenal gland	
	198.8	Secondary malignant neoplasm of other sites	
	198.82	Secondary malignant neoplasm of the genital organs	

ICD9-CM = International Classification of Diseases 9th Revision codes: Centers for Disease Control and Prevention. International Classifcation of Diseases, 9th Revision, Clinical Modifcation (ICD9-CM). https://www.cdc.gov/nchs/icd/icd9cm.htm. Published 2016. Accessed 4 Nov 2020

5.12 APPENDIX D

List of previous published study that evaluated the risk of breast cancer recurrence in patients with invasive breast cancer cited in the text presenting results using administrative database or presenting results derived from population-based registries.

Study (First Name and Year of Pubblication)	Country	Cancer type	Data source for development of algorithm	If Algorithm: Source of Administrative codes	Algorithm Method	Period of patient inclusion, incidence	No. of BC patients	Age (years)	Stage	Median follow-up period (years)	Clinical definition of recurrences used for gold standard	Reference Standard to validate algorithm	% of Recurrence (n° events) and/or Cumulative Incidence (CI) of Recurrence	Reference
Strokes, 2008	United States	Breast	SEER-Medicare	ICD-9	Rules-based	1991-1993	10,790	>=65 y	I-III	8.1	Recurrences (local and distant) or contralateral breast cancer	Yes	25% (2674)	[38]
Chubak, 2012	United States	Breast	Group Health Research Institute	ICD-9	CART-model based	1993–2006	3152	>=18 y	I-II	6.2	Recurrences (local, regional, and distant) or second breast primary	Yes	12.9% (407)	[13]
Cheng, 2012	United States	Breast	SEER-Medicare	ICD-9	Rules-based	1991-1997	20,027	65-79 y	I-III	10	Recurrence not specified	No	36.8% (7372); 10-years CI according to Stage: 35% (Stage I); 44% (Stage II); 56% (Stage III)	[55]
Lord JS, 2012	Australia	Breast	New South Wales Central CancerRegistry (CCR)	no algorithm: population-based	//	2001-2002	6644	>=18 y	I-III	not reported	Metastases	no algorithm: population-based	10.1% (673)-5-years CI: 10% (5.3% for N- and 18.1% for N+)	[58]
Minicozzi, 2013	Italy	Breast	Italian Association of Cancer Registries (AIRTUM) database	no algorithm: population-based	//	2003-2005	3203	>=15y	I-III	not reported	Recurrences (local, regional and distant)	no algorithm: population-based	10.8% (345)-5 years DFS according to molecular subtype Luminal A: 93.0%; Luminal B: 87.4%; HER2+: 77.9%; TN: 80.5%	[59]
Nordstrom, 2016	United States	Breast and others	Geisinger Health system	ICD-9	Random forests-model based	2004–2011	502	>=18y	I-III	3	Metastases	Yes	3.4% (17)	[56]
Kemp-Casey, 2016	Australia	Breast	New South Wales Central CancerRegistry (CCR)	Admitted Patient Data Collection; PBS;MBS	Rules-based	2003-2008	2416	>=18	I-III	3	Second breast cancer event (SBCE): recurrence or new breast cancer	No	9% (217); 5-years CI: 11.9%	[49]

Fredholm, 2016	Sweden	Breast	Sweden Breast Cancer registers	no algorithm: population-based	//	1992-2005	1120	18-69y	I-III	10	Recurrences (local, regional and distant)	no algorithm: population-based	LR recurrence: 15.7% (176); Distant Metastases: 28.3% (317)	[60]
Geurts, 2017	Netherland	Breast	Netherlands Cancer Registry	no algorithm: population-based	//	2003	9342	>=20y	I-III	not reported	Recurrences (local, regional and distant) excluded controlateral breast cancer	no algorithm: population-based	19.8% (1853)	[10]
Cronin-Fenton, 2018	Denmark	Breast	Danish National Patient Register	ICD-10	Rules-based	1991–2011	23478	>=18y	II-III	not reported	Bone metastases, visceral metastases, and breast cancer recurrence (included local, regional, and distant)	Yes	18.4% (4314)-5-years CI: 18.4%	[16]
Ritzwoller, 2018	United States	Breast	Kaiser Permanente regions of Colorado and Northwest	ICD-9 and ICD-10	Logistic Regression-model based	2000–2011	3370	>=21y	I-III	4.3	Recurrence not specified	Yes	7.2% (241)	[14]
Rasmussen, 2019	Denmark	Breast	Danish National Patient Register	ICD-10	Rules-based	2003–2007	471	18-69 y	I-III	7.5	Recurrence or second breast primary	Yes	32.0% (149)	[46]
Schaffar,2019	Switzerland	Breast	Geneva Cancer Registry	no algorithm: population-based	//	1970-2012	1586	22-45	I-III	10.2	Recurrences (loco-regional and distant)	no algorithm: population-based	33.7% (535); 10-years DFS: 68%	[61]
Xu, 2019	Canada	Breast	Alberta provincial administrative registry	NACRS-; DAD	CART-model based	2007–2010 and 2012–2014	598	<=40y	I-III	4	Recurrences (local, regional, and distant) or second breast primary	Yes	20.2% (121)	[47]
van Maaren, 2019	Netherland	Breast	Netherlands Cancer Registry	no algorithm: population-based	//	2005	8062	>=18y	I-III	11.3	Recurrences excluded controlateral breast cancer according to Maastricht Delphi consensus[a]	no algorithm: population-based	10-years DFS: 81.9%	[39]
Ling, 2019	United States	Breast	EMRs of Stanford Health Care (SHC)	text-mining	Semi-Supervised Machine Learning Approach	2000-2014	8892	all ages: Mean age 53.0y	I-III	7.8	Metastases	Yes	14.6% (1302)	[57]
Holleczeck,2019	Germany	Breast	Saarland Cancer Registry	no algorithm: population-based	//	1999-2005	9359	all ages:Mean age 62.7y	I-III	10. 3	Recurrences (local, regional and distant) or death	no algorithm: population-based	5-years CI 10.4%; 10-years CI:15.9%;	[51]
A'mar, 2020	United States	Breast	Puget Sound SEER cancer registry	ICD-9	gradient - boosting-model based	1993-2006	3152	>=18y	I-II	not reported	Second breast cancer event (SBCE): recurrence or new breast cancer	Yes	14-years net probability of SBCE: 25%	[69]

Lambert,2021	Canada	Breast and others	CancerCare Manitoba	ICD-O-3	CART-model based	2004-2012	993	all ages Mean age 60.3y	I-III	not reported	Recurrence not specified	Yes	19.9% (186)	[48]
Lao, 2021	New Zeland	Breast	Waikato and Auckland Breast Cancer Register	population-based	//	1991-2017 (Waikato Register) 2000-2017 (Auckland Register)	17,543	all ages	I-III	not reported	Metastases	no algorithm: population-based	5-years CI:11.2% ;10 years CI:16.5%;	[62]
Current Study, 2021	**Italy**	**Breast**	**FVG population-based cancer registry**	**ICD-9**	**Rules-based**	**2004-2010**	**5420**	**22-74 y**	**I-III**	**10.3**	**Recurrences (local, regional and distant)**	**No**	1406 (27.7%) 5-years CI:15.9% ;10 years CI: 26.0%;	

a-Moossdorff M, et al.Maastricht Delphi consensus on event definitions for classification of recurrence in breast cancer research. J Natl Cancer Inst. 2014 Nov

CART = classification, regression and decision tree; LACE = Life After Cancer Epidemiology; PBS: Pharmaceutical Benefits Scheme; MBS: Medical Benefits Schedule; NACRS: National Ambulatory care reporting system; DAD: discharge abstract data