UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Efficacy and Safety of Ibrutinib Alone or as Combination Treatment in Patients with non-Hodgkin Lymphoma: A Systematic Review and Meta-Analysis

Malai Nhim

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Tomasz BURZYKOWSKI

SUPERVISOR :
Ryan CRASS
Wonkyung BYON

2021
2022

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Efficacy and Safety of Ibrutinib Alone or as Combination Treatment in Patients with non-Hodgkin Lymphoma: A Systematic Review and Meta-Analysis

Malai Nhim

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Tomasz BURZYKOWSKI

SUPERVISOR :

Ryan CRASS

Wonkyung BYON

Acknowledgement

This master thesis was a project under collaboration between Hasselt University (UHasselt) and Amador Bioscience, and a requirement to complete the Master of Statistics and Data Science specialized in Biostatistics in UHasselt.

First of all, I would like to express my gratitude to Prof. Dr. Tomasz BURZYKOWSKI, my internal supervisor from UHasselt, for his guidance, comments and feedback on this thesis, and to Dr. Ryan CRASS and Dr. Wonkyung BYON, my external supervisors from Amador Bioscience and A2PG company, for initiating this project and guiding me through the pharmaceutical research context and methodologies used in meta-analysis from searching for trial reports and/or author manuscripts to the analyses. Your expertise in these areas, supports, constructive feedback and endless patience led and motivated me to keep learning and do my best for this research.

Furthermore, I want to acknowledge CenStat of UHasselt for his prestigious and worldwide recognized knowledge in statistics and his mission in preparing every students to be globally competent. Additionally, I am always grateful to a one-in-a-lifetime scholarship and financial support opportunity from VLIR-UOS to do this master.

Moreover, I am thankful to my seniors, batch mates, and friends for advices and good words to keep me on track in overcoming any obstacles to complete this master. Extra special thanks to my peers: Luigi, Marsha, Dubhe, Kedir, Evan and many more for studying and working together tirelessly to complete every project and task during this master. I wish everyone the greatest success in all his/her endeavors.

Last but the most important, I would like to express my respect and thankfulness to my parents, sister, brother-in-law, nephew and niece for always believing in me. Your love and supports toward me make who I am today. Without you, I could not come this far.

To all those whom I may have failed to mention, thank you so much!

Malai NHIM

Genk, June 2022

Abstract

Non-Hodgkin lymphoma (NHL) is a type of cancer that occurs in lymphocytes (white blood cells) which are the part of the immune system in the body. DLBCL and MCL are among the most common adult B cell neoplasms subtypes of NHL. Ibrutinib was the first in class of BTK inhibitors to be approved by FDA and EMA for treating patients with MCL, CLL, SLL, and WM. The objectives of this study are (1) to assess the association between endpoints (ORR, PFS, and OS) for DLBCL and MCL, separately; (2) to obtain the pooled estimated ORR of ibrutinib alone and ibrutinib in combination, and explored the influence of the combination therapies; and (3) to estimate the pooled estimate of hematological adverse events with ibrutinib, and assess the influence of subgroups of intervention based on the two disease types. The method for obtaining the included studies in the meta-analysis was to use electronic databases such as PubMed-Medline and ClinicalTrials.gov. The association between ORR or PFS and OS was assessed by using (weighted) Pearson correlation and adjusted R^2 obtained from (weighted) linear regression. In the meta-analysis, DerSimonian and Laird random-effects model was used. Cochrane Q test, I^2 statistic and τ^2 were used to assess the heterogeneity. There were total of 20 included studies. For DLBCL, ORR can be used to predict median OS with very strong correlation (weighted Pearson correlation was 0.932 and adjusted R^2 was 0.846); and there was no strong correlation between median PFS and median OS. For MCL, there was a relationship between ORR and probability of OS at 6, 9, 12, 15 and 18 months; and the probability of PFS and probability of OS showed very strong correlation (≥ 0.80) when measured at the same time point except at 3 months and when measured across 6 months to 18 months. In meta-analysis of DLBCL and MCL together, ibrutinib in combination with B-cell depleting antibody on average was associated with 53.73% higher ORR compared to ibrutinib in combination with checkpoint inhibitors and was associated with 16.28% less grade 3 or higher anemia compared to ibrutinib in combination with cytotoxic chemotherapy. In the meta-analysis of DLBCL alone, ibrutinib in combination with checkpoint inhibitors on average was associated with 30.92% and 44.43% less any grade anemia than ibrutinib in combination with B-cell depleting antibody and ibrutinib in combination with cytotoxic chemotherapy respectively. This suggested that ibrutinib in combination with checkpoint inhibitors had the least risk of any grade anemia in DLBCL.

Key Words: Association; Diffuse large B-cell lymphoma; Ibrutinib; Mantle cell lymphoma; Meta-analysis.

Contents

1 Introduction

Lymphomas are a type of hematologic malignancy, or blood cancer, that develop in white blood cells of lymphoid lineage (lymphocytes) via abnormal cell replication faster and live longer [1, 54, 55]. Lymphocytes are the white blood cells associated with adaptive immunity, which recognize and develop specific defenses against pathogens. Malignant lymphocytes, like normal lymphocytes, are present in the blood and lymphatic system, and can spread and grow in lymph nodes, spleen, bone marrow, and other organs. Lymphomas represent a diverse group of over 60 different subtypes, broadly grouped into Hodgkin and non-Hodgkin lymphomas [38, 42, 54]. Non-Hodgkin lymphomas (NHL) make up 90% of all lymphomas and arise from B cell precursors, mature B cells, T cell precursors, and mature T cells [37, 38]. NHL is categorized into different groups, each group with its own epidemiology, etiology, immunophenotypic, genetic, clinical, and therapeutic aspects. Follicular lymphoma, Burkitt lymphoma, diffuse large B-cell lymphoma, mantle cell lymphoma, marginal zone lymphoma, and primary central nervous system (CNS) lymphoma are the most common adult B cell lineage NHL. Adult T cell lymphoma, or Mycosis fungoides, is the most frequent mature T cell lymphoma [2, 37].

In 2009, 12,294 patients in the United Kingdom (U.K.) were diagnosed with NHL with 4,452 deaths from the disease in 2010. Similarly, in United Sates (U.S.), there were 65,540 new cases in 2007 and 20,210 deaths in 2008 [38]. More than two-thirds of patients with NHL were age 60 years or older with a median age of 67 years [37, 38]. The incidence of NHL has been increasing globally. The age-standardised incidence of NHL increased by 35% in England, Scotland, and Wales over the last 30 years (1988–2007) [38]. Similar trends were observed in the U.S., Brazil, India, Japan, Singapore, and western Europe over this time scale. Treatment of NHL depends on the type, stage, histopathological features of the specific cancer and corresponding patient symptoms. Chemotherapy, radiotherapy, immunotherapy, stem cell transplant, and in rare situations surgery, are the most common treatments [37].

Bruton's tyrosine kinase (BTK) inhibitors are oral, small molecule anti-cancer therapeutics indicated for the treatment of B-cell malignancies. These therapeutics represent novel agents for NHL that may afford patients the possibility of cytotoxic chemotherapy-free management [53]. Ibrutinib, the innovator, first in class BTK inhibitor, was approved by the European Medical Agency (EMA) and the U.S. Food and Drug Administration (FDA) since 2012 and in subsequent years: 560 mg taken orally once daily for treating patients with mantle cell lymphoma (MCL); and 420 mg taken orally once daily for treating patients with chronic lymphocytic leukemia (CLL) or small lymphocytic lymphoma (SLL), and Waldenström's macroglobulinaemia (WM) (also known as lymphoplasmacytic lymphoma) [24, 29]. In a phase II study, ibrutinib administered orally at 560 mg once daily demonstrated efficacy in 111 patients with previously-treated MCL with an overall response rate (ORR) of 68% and 21% complete responders (CR)

[51]. Additionally, in patients with relapsed and refractory CLL/SLL, ibrutinib 420mg once daily achieved an ORR of 71% in a 51 patient phase Ib/II study, and in a phase III trial, ibrutinib 420 mg once daily demonstrating significantly prolonged overall survival (OS) compared to chlorambucil with (hazard ratio [HR] 0.16, p-value < 0.001) [8, 9]. Finally, a 90.5% ORR was achieved with ibrutinib 420mg once daily in a phase II study of 63 patients with WM who were previously treated [45]. The safety profile was favorable, especially compared to cytotoxic chemotherapy, with the most common adverse reactions ($\geq 30\%$) of thrombocytopenia, diarrhea, fatigue, musculoskeletal pain, neutropenia, rash, anemia and bruising [25, 36].

Overall survival (OS) is defined as the time from randomization or initiation of treatment to the time of death from any causes and it is widely considered as good-standard indicator of benefit in clinical trials [10, 16, 31, 39]. However, it has a number of weaknesses. Indeed, OS results can often be confounded by the use of (and access to) successive lines of therapy, patient crossover, and/or access to the investigational agent for patients in control arms, challenges with patient follow-up, and increased postprogression survival. The median overall survival is not always reported and sometimes is reported as "not reached" due to insufficient follow up time in the study to observe this outcome. Early endpoints such as progression-free survival (PFS) and overall response rate (ORR) are surrogate endpoints for OS and appealing as primary clinical trial endpoints for a number of reasons. PFS is generally defined as the time from randomization to tumor progression or death resulting from any causes, and ORR is defined as the proportion of patients in a trial who have a partial or complete response to therapy [5, 31, 47]. These surrogate endpoints allow for shorter trial durations and smaller patient cohorts, and in the case of ORR, may also allow for single-arm trial designs [10, 60]. Surrogate endpoints, such as PFS and ORR, have been adopted in clinical studies as primary endpoints to demonstrate clinically significant increases in patient survival or quality of life; however, their validity as surrogate endpoints is still uncertain across cancer types [56]. Similarly, in single-arm studies, the relationship between ORR or PFS and OS are of a interest.

Over the past decade, many additional clinical trials have been conducted to evaluate the efficacy and safety of ibrutinib as monotherapy or in combination with other cytotoxic chemotherapeutics, immunotherapies, or targeted agents in various cancers [53]. These studies include many published and unpublished investigations of the efficacy and safety of ibrutinib as a monotherapy and/or in a combination with other therapies in patients with NHL outside included in approved product labeling, specifically DLBCL and MCL. These studies have demonstrated significant improvement of patient outcomes and safety of ibrutinib alone and as combination, however, the indications are not consistent. There is heterogeneity in studies characteristics, follow-up time, and reported outcomes of efficacy and safety endpoints. For example, in a phase I/II study, ibrutinib alone was used to treat 80 previously treated DLBCL patients achieving an ORR of

25% and median OS of 6.41 months with a median follow up of 11.53 months [58]. In a phase Ib/II study, an ORR of 62% was achieved in 26 patients with relapsed and/or refractory DLBCL treated with ibrutinib in combination with lenalidomide and DA-EPOCH-R with a median OS of 15.84 months [57]. Furthermore, in a phase II study in 50 patients with relapsed and/or refractory MCL treated with ibrutinib in combination with rituximab and lenalidomide, an ORR of 76% and 22 months of median overall survival were observed [27]. In another phase II study in 16 Japanese patients with relapsed and/or refractory MCL treated with ibrutinib alone with a median follow up of 22.5 months, the median OS was not reached and the ORR was 94% [32]. Adverse events of any grade thrombocytopenia, any grade anemia, and any grade neutropenia were reported ranging from 25% to 46%, from 18% to 46% and from 23%, respectively.

1.1 Rationale

The rationale of this study is to explore and improve understanding of relationships between different endpoints (OS, PFS, ORR) as well as the effect on efficacy and safety of ibrutinib administered as monotherapy or in combination to patients with non-Hodgkin lymphoma, specifically DLBCL and MCL. This information can inform development of novel BTK inhibitors for treatment of B-cell malignancies by informing the target profile in this class and indication [44]. The key stakeholders are pharmaceutical companies who are developing novel BTK inhibitors as this has significant implications for the design and execution of clinical trials intended to enable market authorization of novel therapeutics in this cancer type. Results from peer-reviewed manuscripts reporting clinical trials approved by Institutional Review Boards (IRB) and Ethics Committee (EC) will be used to ensure data quality.

This study addresses the following clinical questions: (1) is there a relationship between early endpoints (ORR and PFS) and standard endpoint (OS) with ibrutinib treatment in patients with non-Hodgkin lymphoma?; (2) what is the ORR of ibrutinib treatments in patients with non-Hodgkin lymphoma?; and (3) what is the probability of hematological adverse events with ibrutinib treatment and how is this influenced by combination therapy?

1.2 Objectives

To answer the clinical questions, the objectives of this study are the following: (1) to assess the association between early endpoints (ORR and PFS) and standard endpoint (OS) for the two subtypes of non-Hodgkin lymphoma, DLBCL and MCL, separately; (2) to obtain pooled estimate of ORR for DLBCL and MCL and combination of the two disease types, and to explore the influence of combination therapy of ibrutinib; and (3) to estimate the pooled probability of hematological adverse events: any grade, and grade 3 or higher anemia, thrombocytopenia, and neutropenia based on ibrutinib as a monotherapy and ibrutinib in combinations.

This paper is structured as follows. Section 2 contains the methodology where eligibility criteria, information sources and searches, study selection and data extraction, evaluation of study quality and publication bias, and statistical analysis are described. Section 3 dedicates to the results which eligible studies, characteristics of included studies, risk of bias assessment and publication bias, association between endpoints, meta-analysis are presented. The discussion about the results is given in Section 4. Finally, the conclusion is made in section 5.

2 Methodology

2.1 Eligibility Criteria

The inclusion criteria for trials to be included in this systematic review and meta-analysis are predetermined as following: (1) single-arm trials or randomized trials evaluating ibrutinib administered alone or as combination with other therapies in patients with diffuse large B-cell lymphoma (DLBCL) or mantle cell lymphoma (MCL), untreated or previously treated; (2) data were available on OS with either or both of ORR and PFS, or hematological adverse events including thrombocytopenia, anemia and neutropenia. All subtypes of DLBCL including germinal center B-cell like (GCB) and activated B-cell like (ABC) categorized by Gene expression profiling (GEP) or GCB and non-GCB by Hans algorithm were eligible for inclusion. Studies conducted in non-human species, long-term extension studies of previously reported clinical studies in humans, on-going trials where results were not available at the time of literature search, and studies not registered in the ClinicalTrials.gov database were excluded. The studies were grouped for synthesis or meta-analysis based on the availability of the outcomes such as OS with ORR, OS with PFS, and the adverse events.

2.2 Information Sources and Searches

PubMed-Medline database was used for a comprehensive literature search to identify all relevant published studies as of April 26, 2022. The following exact key terms were used: "(Mantle Cell Lymphoma) AND ((((Ibrutinib) OR (PCI-032765)) OR (CRA-032765)) OR (IMBRUVICA))", and "(Diffuse Large B-Cell Lymphoma) AND ((((Ibrutinib) OR (PCI-032765)) OR (CRA-032765)) OR (IMBRUVICA))". The results were filtered only for the clinical trials. Additionally, ClinicalTrials.gov was used to search for the trials with available results that have not been published in PubMed-Medline. The key words used in ClinicalTrials.gov were "Mantle Cell Lymphoma" and "Diffuse Large B Cell Lymphoma" in the condition or disease box, and "ibrutinib" in other terms box. Completed recruitment status, and with results study results filters were applied.

2.3 Study Selection and Data Extraction

The author independently reviewed the list of retrieved articles to identify potentially relevant articles. Any doubts about particular studies were consulted with supervisors. The author independently extracted the data from studies. Data extracted included: title of the study, the national clinical trial number (NCT), first author, year of publication, study design, percentage of male patients, mean age, median follow-up time, line of therapy, intervention, disease types and/or sub-disease types, and endpoints of interest. The primary endpoints of interest were

ORR, median PFS and median OS. In addition, probabilities of PFS and OS were extracted at 3 months, 6 months, 9 months, 12 months, 15 months, 18 months, 21 months and 24 months from the Kaplan-Meier curves by using WebPlotDigitizer. The secondary outcome of interests were hematological adverse events. The rates of thrombocytopenia, anemia, and neutropenia at any grade and grade 3 or higher were extracted.

2.4 Evaluation of Study Quality and Publication Bias

Systematic reviews are designed to gather and synthesize all studies that fulfill predetermined eligibility criteria while attempting to minimize bias. However, determining the amount to which biases influenced outcomes of a trial is often impossible [21]. To investigate sources of bias in included randomized trials, the Cochrane risk of bias assessment was used [21]. Six domains of bias were included in the risk bias tool such as selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. The following sources of bias were evaluated: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personal, (4) binding outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other sources of bias. These items were marked as low risk of bias, high risk of bias or unclear risk of bias for each randomized control trial. Single-arm studies were marked as higher risk of bias due to the absence of a control condition and randomization. Publication bias in the results of the meta-analysis was assessed by using funnel plot and its asymmetry was evaluated by Egger's test.

2.5 Statistical Analysis

2.5.1 Association Between Endpoints

Early endpoints such as ORR and PFS are of interest to predict for OS since they requires shorter period of time and smaller sample size. Association and regression can be used to assess their relationships and appropriateness of their predictability respectively [12]. In this meta-analyse of single-arm studies, the association between ORR and OS, and PFS and OS were explored using aggregate data. Initially, for the disease types DLBCL and MCL separately, the (weighted) Pearson correlations were calculated for the association between ORR and median OS (using inverse-variance weighting), ORR and probability of OS (using inverse-variance weighting), median PFS and median OS (using sample size weighting), median PFS and probability of OS (using sample size weighting) and probability of PFS and probability of OS (using sample size weighting). All the data were tested for normality by using Kolmogorov-Smirnov test to test their eligibility for performing Pearson correlation. Lastly, the association were quantified through (weighted) linear regression for ORR and median OS (using inverse-variance weighting) and median PFS and median PS (using sample size weighting). For MCL, since there were only

a few studies available which report both median PFS and median OS; therefore, the correlations of these variables with other variable were not evaluated.

Kolmogorov-Smirnov Test

The one-sample Kolmogorov-Smirnov test allows to test the normality of data. The null hypothesis is that the distribution of data for testing does not deviate from normality. The test works as follow: the cumulative distribution function (CDF) of the testing data is plotted against the CDF of the normal distribution with the same mean and standard deviation; then the test determines the largest difference between the two CDFs. One-sample two-sided Kolmogorov-Smirnov test is defined by:

$$D = \sup_x |F_{\text{data}}(x) - F_T(x)|$$

where $F_T(x)$ is the CDF of the normal distribution; $F_{\text{data}}(x)$ is the CDF of the data for testing; $\sup_x$ is the supremum of the set of distances. Reject the null hypothesis at the approximate level of significance α if D exceeds the $1-\alpha$ quantile which can be found in the Table A14 of [11].

(Weighted) Pearson Correlation

The correlation coefficients are used to quantify the strength of a relationship between two variables. Pearson correlation illustrates the linear relationship that exists between two sets of data. For correlating aggregated data, the weighted correlation is appropriate. The degree of association is predetermined. The ranges of correlation coefficient are determined as follow: 0.00 to 0.19 (very weak), 0.20 to 0.39 (weak), 0.40 to 0.59 (moderate), 0.60 to 0.79 (strong) and 0.80 to 1.00 (very strong) [5]. Supposed $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two vectors, the weighted Pearson correlation is computed using the formula below [3]:

$$r_{Pearson} = \frac{\sum_{i=1}^{n}[w_i(x_i - \bar{x})(y_i - \bar{y}))]}{\sqrt{\sum_{i=1}^{n}[w_i(x_i - \bar{x})^2]\sum_{i=1}^{n}[w_i(y_i - \bar{y})^2]}}$$

where w_i is the weight; $\bar{x}$ and $\bar{y}$ is the weighted mean of $\boldsymbol{X}$ and $\boldsymbol{Y}$ variable respectively;

$$\bar{x} = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i x_i; \qquad \bar{y} = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i y_i;$$

and n is the number of elements in $\boldsymbol{X}$ and $\boldsymbol{Y}$. All weights are set to one for unweighted Pearson correlation.

The test of significance for the sample correlation coefficient was done by using the Student's t test with the t statistics where the null hypothesis was $H_0 : r_{Pearson} = 0$ and the alternative hypothesis was $H_A : r_{Pearson} \neq 0$ [33]:

$$t = \frac{r_{Pearson} \times \sqrt{n-2}}{\sqrt{1 - r^2_{Pearson}}}$$

which follows a t distribution with $n-2$ degrees of freedom.

The $100 \times (1-\alpha)\%$ confidence interval (CI) of the correlation coefficient was calculated from the following [22]:

1. Transform the Pearson correlation with the Fisher's transformation

$$r' = \frac{1}{2} \times \ln\left(\frac{1 + r_{Pearson}}{1 - r_{Pearson}}\right) = \text{arctanh}\,(r_{Pearson})$$

2. Calculate the standard deviation of the transformed correlation

$$S' = \frac{1}{\sqrt{n-3}}$$

3. Calculate the 95% confidence interval of the transformed correlation using the z statistic

$$[r' - z_{1-\alpha/2} \times S', r' + z_{1-\alpha/2} \times S']$$

4. Transform the lower and upper values back to the Pearson correlation scale

$$\left[\frac{\exp\left[2 \times (r' - z_{1-\alpha/2} \times S')\right] - 1}{\exp\left[2 \times (r' - z_{1-\alpha/2} \times S')\right] + 1}, \frac{\exp\left[2 \times (r' + z_{1-\alpha/2} \times S')\right] - 1}{\exp\left[2 \times (r' + z_{1-\alpha/2} \times S')\right] + 1}\right]$$

(Weighted) Linear Regression

Weighted linear regressions are used to quantify the study-level association between two endpoints. R-squared is used to assess the proportion of variance explained by the regression. An R-squared of 0.7 or above defines that the association is very strong [59]. The weighted linear regression is formulated as follow:

$$T_i = \mu_T + \gamma S_i + \varepsilon_i$$

where $i = 1, 2, ..., n$ (n is the number of trials); $\varepsilon_i \sim N(0, \frac{\sigma^2}{w_i})$; and w_i is the weights. T_i and S_i are the outcomes of the outcome endpoint and the predictive endpoints in trial i respectively. For unweighted linear regression, the weights are set to one.

The $100 \times (1-\alpha)\%$ percentile interval (PI) of adjusted R-squared was obtained by using non-parametric bootstrap [14]. Re-sample pairs was done without making any assumption on the distribution of outcome, predictor, and weight variables and the model was fitted and adjusted R-squared was estimated for 10000 bootstrap indexed samples. The `R codes` can be found in

the Appendix.

2.5.2 Meta-analysis

Meta-analysis is a statistical tool for combining the results of multiple studies, particularly those with small sample sizes or contradictory results [63]. DerSimonian and Laird random-effects models were used to conduct the analysis since some heterogeneity is expected. Random-effect model assumes the outcomes, θ_i, are estimated from different populations and it accounts for within- and between-study variability. Random-effects model is formulated as below:

$$\hat{\theta_i} = \theta + \mu_i + \varepsilon_i$$

where $i = 1, 2, ..., I$ and I is the number of studies; $\varepsilon_i \sim N(0, \nu_i)$; $\mu_i \sim N(0, \tau^2)$; and then therefore $\hat{\theta}_i \sim N(\theta, \nu_i + \tau^2)$.

The logit transformation method was used. The heterogeneity was assessed by Cochran χ^2 test, index of heterogeneity I^2, and evaluated heterogeneity (between-study variance) τ^2. Cochrane χ^2 which is used to test for the statistical heterogeneity. The null hypothesis is that $\tau^2 = 0$. The null hypothesis is rejected when the value of the Q-statistics is larger than the critical χ^2 value indicating that the random-effect model is favored. Index of heterogeneity, I^2, which is estimating the proportion of variability in meta-analysis that explained by variation across the trials rather than sample error [35]. When $I^2 = 0$, it indicates that all of heterogeneity is caused by the sampling error whereas when $I^2 = 1$, it means that the true variation between studies accounts for all of the overall heterogeneity. In this case subgroup analysis or meta-regression was advisable to account for potential moderators.

Subgroup analysis was implemented with random-effects model to test if any subgroups of intervention had significant different effects or proportions of adverse events. Since some subgroups have small number of units (less than 4), the assumption of common between-study variance component across subgroups were made. τ^2 across all studies was used to calculate a new overall summary proportion for each subgroup. The summary of effect sizes or proportions for each group and the within-subgroup heterogeneity were obtained. The summary effect size or proportion across subgroups was compared by using the Wald-type test. For the test that showed significant p-value (< 0.05), the contrast analysis was done by using Tukey method and Holm–Bonferroni adjusted method [7, 23].

Subgroups of intervention were fallen into five categories such as: (1) ibrutinib as monotherapy, (2) ibrutinib with cytotoxic chemotherapy (R-CHOP, DA-EPOCH, and DA-EPOCH-R), (3) ibrutinib with checkpoint inhibitors (nivolumab, and durvalumab), (4) ibrutinib with B-cell depleting antibody with or without other small molecule inhibitors without cytotoxic chemotherapy (obinutuzumab with or without venetoclax, and rituximab with or without lenalidomide),

and (5) ibrutinib with PI3K inhibitors (umbralisib, and buparlisib).

2.6 Software

All the analyses were performed using statistical software `R version 4.1.2` (https://www.r-project.org/) which is a free software environment for statistical computing and graphics. Function `ks.test()` of `R`'s statistical base-package was used for conducting Kolmogorov-Smirnov test. Weighted Pearson correlation was performed by using function `weightedCorr()` of `wCorr` package, and function `wtd.cor()` of `weights` package for its p-value, and function `CIr()` of `psychometric` package for its 95% CI. Function `cor()` and `cor.test()` of `R`'s statistical base-package was used for unweighted Pearson correlation and its p-value and 95% CI. (Weighted) linear regression was performed by using function `lm()` of `R`'s statistical base-package. For meta-analysis, package `metafor` was used for statistical analysis. More details can be found in the Appendix.

3 Results

3.1 Eligible Studies

There were total of 65 studies identified across both DLBCL and MCL in the literature search. After reviewing for duplication, 54 studies remained. 21 studied were excluded during the title and abstract review based on eligible criteria. 33 studies were continued for full article review and 2 studied were excluded due to 1) duplication and 2) study not registered in ClinicalTrials.gov. The screening continued for 31 studies to assess for eligibility. At this stage, total of 11 studies were removed (6 studies did not reported median OS or safety; 3 studies had different disease type, and 2 studies were ineligible for study design). Finally, 20 studies were included. This procedure is presented in Figure 1.

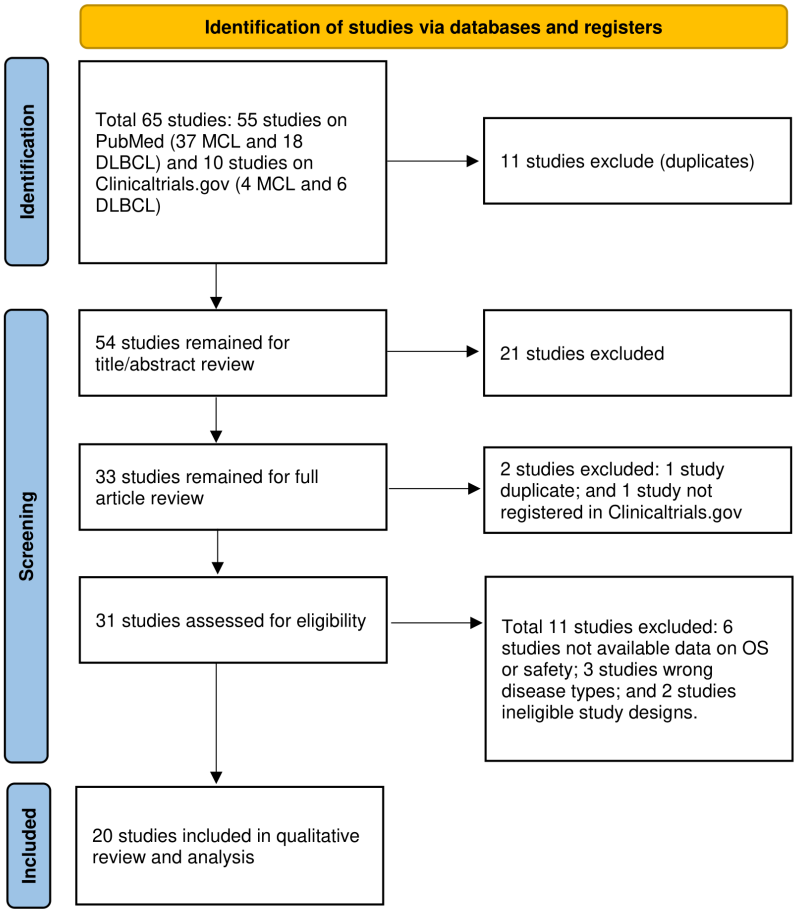


Figure 1: Flow diagram for study selection procedure

3.2 Characteristics of Studies

The summary of the characteristics of the included studies and the details of ORR, median PFS, and median OS for the intervention group are presented in Table 10 in the Appendix. Each trial was identified by its unique national clinical trial (NCT) number. There were 2 phase I trials, 6 phase I/II trials, 9 phase II trials, and 3 phase III trials. 17 trials were single-arm studies and 3 trials were randomized control trials. The interventions were as follows: ibrutinib monotherapy (4 trials), ibrutinib in combination with cytotoxic chemotherapy (2 trials), ibrutinib in combination with checkpoint inhibitors (2 trials), ibrutinib in combination with B-cell depleting antibody with or without other small molecule inhibitors without cytotoxic chemotherapy (9 trials), and ibrutinib in combination with IPI3K inhibitors (3 trials). One trial included patients with both DLBCL and MCL, 7 trials included patients with DLBCL but not MCL, and 12 trials included patients with MCL but not DLBCL. The number of patients enrolled in individual trials ranged from 9 and 419. Median age ranged between 56 and 74 years old. Males were more frequently enrolled than females in almost all trials and treatment groups ranging from 50% to 90% of enrolled subjects with the exception of one group that had 40% males. The median follow-up duration for all included studies had mean of 24.05 months (ranged between 11.53 and 42 months), 24.85 months (ranged between 11.53 to 40.8 months) in DLBCL studies and 23.68 months (ranged between 15.8 to 42 months) in MCL studies.

3.3 Risk of Bias Assessment and Publication Bias

There were three randomized trials included in this analysis. One study had high risk of selection bias, performance bias and detection bias. This study was open label and did not specify the randomization mechanism. Another study had high risk of performance bias since it unmasked patients and investigators. Last study had a lower risk in all domains (see Figure 7 in the Appendix). For meta-analysis of ORR, there was evidence of publication bias for MCL and combination of DLBCL and MCL but not for DLBCL given the p-value of the Egger test respectively 0.0253, 0.0016 and 0.2150. The funnel plots are presented in Figure 8 in the Appendix.

3.4 Association Between Endpoints

3.4.1 Diffuse Large B-Cell Lymphoma

The correlations between ORR and OS, and PFS and OS were explored. Kolmogorov-Smirnov test was conducted to test the normality of all data. All p-values from the test were statistically not significant (p-value > 0.05) meaning that the null hypothesis of normality cannot be rejected. Therefore, it can be assumed that the distribution of all data was approximately normality and the (weighted) Pearson correlation can be performed. To obtain more data points for the analysis,

subtypes of DLBCL were included where there existed more than one groups and DLBCL was not included since they shared some great amount of information.

The association between ORR and median OS was evaluated based on 6 trials [19, 20, 41, 57, 58, 62] that the median OS was available in total of 8 units (data points included in the analysis). Inverse-variances of ORR were used as weights. The weighted Pearson correlation was 0.932 (95% CI: $[0.660, 0.988]$; p-value $= 0.001$) and the unweighted Pearson correlation was 0.881 (95% CI: $[0.466, 0.978]$; p-value $= 0.004$). The adjusted R-squared obtained from the weighted linear regression is 0.846 (95% PI: $[0.137, 0.988]$). Figure 3 shows data points distributed near the predictive regression line. This showed very strong association between ORR and median OS.

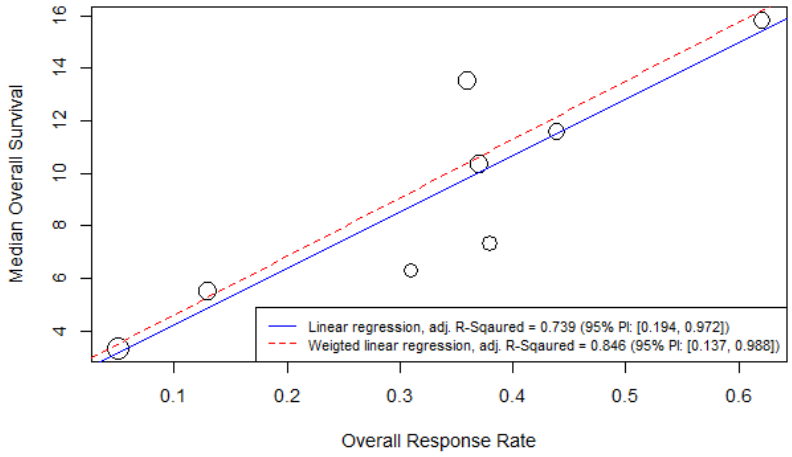


Figure 2: Scatter plot of the relationship of ORR with median OS

There were 4 trials [41, 20, 57, 58, 61], in total of 8 units, included for the calculation of (weighted) Pearson correlation between ORR and probability of OS at different time points. Inverse-variances of ORR were used as weights. The results were obtained and presented in Table 1. The unweighted correlations were presented in the brackets. There were very strong weighted correlations (≥ 0.80) and significant at 5% between ORR and probability of OS except at 24 months which showed very weak correlation.

Table 1: Weighted (Unweighted) Pearson correlation between ORR and probability of OS at different time points

Pearson correlation	Probability of OS							
	3 mths	6 mths	9 mths	12 mths	15 mths	18 mths	21 mths	24 mths
ORR	**0.848***	**0.942***	**0.966***	**0.910***	**0.939***	**0.902***	**0.902***	-0.086
	(0.755*)	(**0.923***)	(**0.950***)	(**0.899***)	(**0.939***)	(**0.885***)	(**0.879***)	(0.142)

mths: months; value of correlation in bold indicating very strong correlation;
*: significant at 5% level of significance.

There were 6 trials [19, 20, 41, 57, 58, 62], in total of 8 units, included in evaluating the association between median PFS and median OS. Sample sizes were used as weights in these analyses. The weighted Pearson correlation was 0.547 (95% CI: $[-0.257, 0.903]$; p-value = 0.161) and unweighted Pearson correlation was 0.618 (95% CI: $[-0.154, 0.921]$; p-value = 0.103). This did not show any strong correlation. Similarly, adjusted R-squared from the weighted linear regression was 0.182 (95% PI: $[-0.163, 0.832]$) shown in Figure 3 where points were not distributed near the predictive regression line.

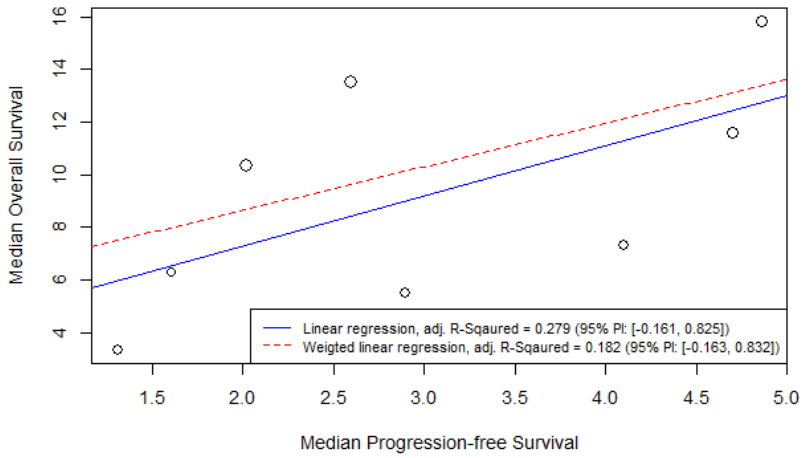


Figure 3: Scatter plot of the relationship of median PFS and median OS

The associations between median PFS and probability of OS were evaluated based on 5 trials [20, 41, 57, 58, 62] consisting of total of 7 units. The correlations were weighted by sample sizes. The weighted (unweighted) Pearson correlations are presented in Table 2. Median PFS had strong weighted correlation and significant at 5% with probability of OS only at 12 months and 15 months.

Table 2: Weighted (Unweighted) Pearson correlation between median PFS and probability of OS at different time points

Pearson correlation	Probability of OS							
	3 mths	6 mths	9 mths	12 mths	15 mths	18 mths	21 mths	24 mths
median PFS	0.645	0.645	0.733	**0.838***	**0.843***	0.741	0.704	0.379
	(0.636)	(0.749)	(**0.833***)	(**0.906***)	(**0.906***)	(**0.836***)	(0.774*)	(0.297)

mths: months; value of correlation in bold indicating very strong correlation;
*: significant at 5% level of significance.

Table 3 shows the correlation between probability of PFS and OS at different time points. These correlation coefficients were calculated based on 5 trials [20, 41, 57, 58, 62] consisting of 7 units in total. These correlations were weighted by sample sizes. There were very strong weighted correlation (≥ 0.80) between probability of PFS and probability of OS at the early stage. At 3 months, probability of PFS showed strong weighted Pearson correlation with probability of OS at 9 months, 12 months, 15 months, 18 months and 21 months. At 6 months, probability of PFS showed strong weighted Pearson correlation with probability of OS at various time points such as 3 months, 6 months, 12 months, 15 months, 18 months and 21 months. All very strong correlation were significant at 5%. There were no very strong correlations at other time points.

Table 3: Weighted (Unweighted) Pearson's correlation between probability of PFS and probability of OS at different time points

Pearson correlation		Probability of OS							
		3 mths	6 mths	9 mths	12 mths	15 mths	18 mths	21 mths	24 mths
Probability of PFS	3 mths	0.774*	0.757*	**0.807***	**0.894***	**0.898***	**0.836***	**0.855***	0.118
		(0.748)	(**0.840***)	(**0.889***)	(**0.937***)	(**0.943***)	(**0.896***)	(**0.891***)	(-0.002)
	6 mths	**0.900***	**0.835***	0.783*	**0.954***	**0.945***	**0.955***	**0.965***	0.305
		(**0.910***)	(**0.894***)	(**0.843***)	(**0.942***)	(**0.936***)	(**0.954***)	(**0.978***)	(0.139)
	9 mths	0.545	0.460	0.304	0.499	0.542	0.618	0.635	0.112
		(0.571)	(0.438)	(0.281)	(0.386)	(0.340)	(0.511)	(0.582)	(-0.009)
	12 mths	0.440	0.326	0.159	0.467	0.401	0.578	0.531	0.359
		(0.429)	(0.248)	(0.084)	(0.272)	(0.204)	(0.397)	(0.414)	(0.298)
	15 mths	0.250	0.180	-0.022	0.195	0.152	0.338	0.322	0.140
		(0.214)	(0.034)	(-0.171)	(-0.053)	(-0.089)	(0.089)	(0.151)	(-0.002)
	18 mths	0.345	0.353	0.145	0.270	0.251	0.419	0.402	0.306
		(0.298)	(0.177)	(-0.042)	(0.02)	(0.004)	(0.165)	(0.220)	(0.168)
	21 mths	0.212	0.007	-0.109	-0.388	-0.344	-0.331	-0.114	-0.513
		(0.273)	(-0.071)	(-0.263)	(-0.388)	(-0.356)	(-0.324)	(-0.102)	(-0.431)
	24 mths	-0.206	-0.098	-0.246	0.049	0.016	0.138	-0.051	0.580
		(-0.455)	(-0.364)	(-0.462)	(-0.270)	(-0.298)	(-0.200)	(-0.353)	(0.488)

mths: months; value of correlation in bold indicating very strong correlation;
*: significant at 5% level of significance.

3.4.2 Mantle Cell Lymphoma

Similarly, the correlation between ORR and OS, and PFS and OS was evaluated. All data were confirmed to be normality by Kolmogorov-Smirnov where p-value were greater than 0.05. For MCL, since not many trials reported the median PFS and median OS, the correlation between these variables and other variables were not assessed.

The association of ORR and probability of OS were evaluated based on 9 trials [15, 18, 27, 28, 41, 43, 48, 49, 51] consisting of 11 units. Inverse-variances of ORR were used as weights. The weighted and unweighted Pearson were obtained and presented in Table 4. ORR showed very strong weighted correlation and significant at 5% with probability of OS at 6 months, 9 months, 12 months, 15 months and 18 months.

Table 4: Weighted (Unweighted) Pearson's correlation between ORR and probability of OS at different time points

Pearson	Probability of OS							
correlation	3 mths	6 mths	9 mths	12 mths	15 mths	18 mths	21 mths	24 mths
ORR	0.781*	**0.867***	**0.915***	**0.898***	**0.910***	**0.907***	0.663*	0.649
	(0.458)	(0.592)	(0.730*)	(0.689*)	(0.757*)	(0.772*)	(0.086)	(0.040)

mths: months; value of correlation in bold indicating very strong correlation;
*: significant at 5% level of significance.

There were 10 trials [13, 15, 18, 27, 28, 41, 43, 48, 49, 51], in total of 12 units, included evaluating the correlation between probability of PFS and OS at specified time points. Sample sizes were used as weights. At 6 months, 9 months, 12 months, 15 months and 18 months, there were very strong weighted correlations (≥ 0.80) and across the time points and for all the same time points except at 3 months. All very strong correlations were significant at 5%.

Table 5: Weighted (Unweighted) Pearson's correlation between probability of PFS and probability of OS at different time points

Pearson correlation		Probability of OS							
		3 mths	6 mths	9 mths	12 mths	15 mths	18 mths	21 mths	24 mths
Probability of PFS	3 mths	0.787*	**0.972***	**0.944***	**0.936***	**0.912***	**0.901***	0.631*	0.685*
		(0.675*)	(**0.940***)	(**0.879***)	(**0.854***)	(0.791*)	(0.744*)	(0.045)	(0.144)
	6 mths	0.649*	**0.947***	**0.985***	**0.962***	**0.965***	**0.953***	0.707*	0.772*
		(0.527)	(**0.926***)	(**0.966***)	(**0.919***)	(**0.922***)	(**0.904***)	(0.235)	(0.318)
	9 mths	0.713*	**0.949***	**0.984***	**0.989***	**0.984***	**0.978***	0.726*	0.771*
		(0.588*)	(**0.892***)	(**0.939***)	(**0.969***)	(**0.959***)	(**0.949***)	(0.263)	(0.337)
	12 mths	0.740*	**0.930***	**0.955***	**0.992***	**0.992***	**0.990***	0.721*	0.749*
		(0.616*)	(**0.825***)	(**0.854***)	(**0.978***)	(**0.979***)	(**0.972***)	(0.259)	(0.308)
	15 mths	0.797*	**0.948***	**0.962***	**0.981***	**0.986***	**0.988***	0.713*	0.730*
		(0.678*)	(**0.854***)	(**0.893***)	(**0.958***)	(**0.972***)	(**0.969***)	(0.220)	(0.257)
	18 mths	0.752*	**0.948***	**0.971***	**0.984***	**0.987***	**0.987***	0.702*	0.732*
		(0.629*)	(**0.871***)	(**0.914***)	(**0.963***)	(**0.970***)	(**0.968***)	(0.207)	(0.257)
	21 mths	0.579	0.746*	0.796*	**0.803***	0.797*	0.789*	**0.967***	**0.976***
		(0.129)	(0.189)	(0.270)	(0.328)	(0.310)	(0.298)	(**0.964***)	(**0.973***)
	24 mths	0.656*	0.705*	0.737*	0.746*	0.739*	0.733*	**0.967***	**0.978***
		(0.111)	(0.142)	(0.176)	(0.244)	(0.232)	(0.221)	(**0.964***)	(**0.976***)

mths: months; value of correlation in bold indicating very strong correlation;
*: significant at 5% level of significance.

3.5 Meta-analysis of ORR

3.5.1 Combination of the Two Disease Types

A meta-analysis for all 20 trials (n=1343) with DLBCL and MCL disease types combined was conducted. Summary measure for pooled estimate for ORR for overall and each subgroup of intervention were presented in Figure 4. The overall summary measure of the meta-analysis showed the pooled ORR equals 0.72 (95% CI: $[0.26, 0.81]$). The Cochran Q test was 244.25 (p-value < 0.1) showing that variation in outcomes exists between studies, and this variation was due to heterogeneity rather than chance. In support, the $I^2 = 91.0\%$ meaning that there was heterogeneity. The estimated variation, $\tau^2 = 1.35$, indicating that the effect sizes varied across studies. The global test for subgroup analysis was significant (p-value = 0.03) meaning that at least two intervention groups had significant different ORR at 5% level of significance.

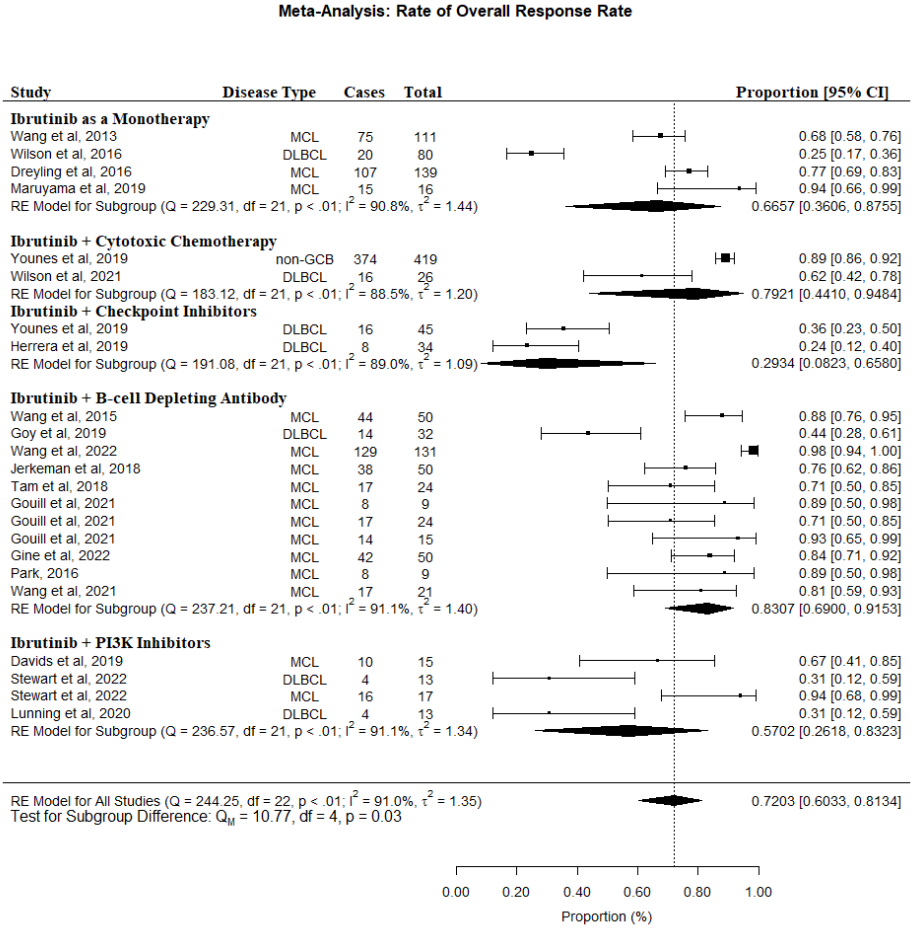


Figure 4: (DLBCL and MCL) Meta-Analysis on Overall Response Rate

The contrast analysis was done by using Holm-Bonferroni correction. Among the 10 pairs, 1 pair showed significant difference shown in Table 6. Ibrutinib in combination with checkpoint inhibitors was significantly different from ibrutinib in combination with B-cell depleting antibody. The difference of logit scale is -2.44 (SE: 0.83). It can be said that ibrutinib in combination with B-cell depleting antibody is associated with 53.73% higher ORR on average compared to ibrutinib in combination with checkpoint inhibitors.

Table 6: Contrast estimate (difference of logit scale) of significant pair(s) for subgroup model

	Estimate	**Std. Error**	z **value**	**Pr**$(>\|z\|)$
ICI − IBCDA	-2.4397	0.8328	-2.930	0.0339

- IC: Ibrutinib + Checkpoint Inhibitors
- IBCDA: Ibrutinib + B-cell Depleting Antibody

3.5.2 Diffuse Large B-Cell Lymphoma

For the DLBCL, a meta-analysis was based on 8 trials (n=662). The forest plot presenting summary measure pooled estimated for ORR for overall and each subgroup of intervention is shown in Figure 5. The pooled estimated for ORR in overall was 0.44 (95% CI: $[0.20, 0.71]$). The Cochran Q test was 186.59 (p-value < 0.1) indicating that there was variation caused by heterogeneity. $I^2 = 96.2\%$ and $\tau^2 = 2.64$ were also suggesting between-study variation. The global test for subgroup analysis was significant (p-value $= 0.03$) meaning that at least two intervention subgroups had significantly different ORR at 5% level of significance. However, the contrast analysis using Tukey method and Holm-Bonferroni correction did not show any significant differences among the 10 pairs.

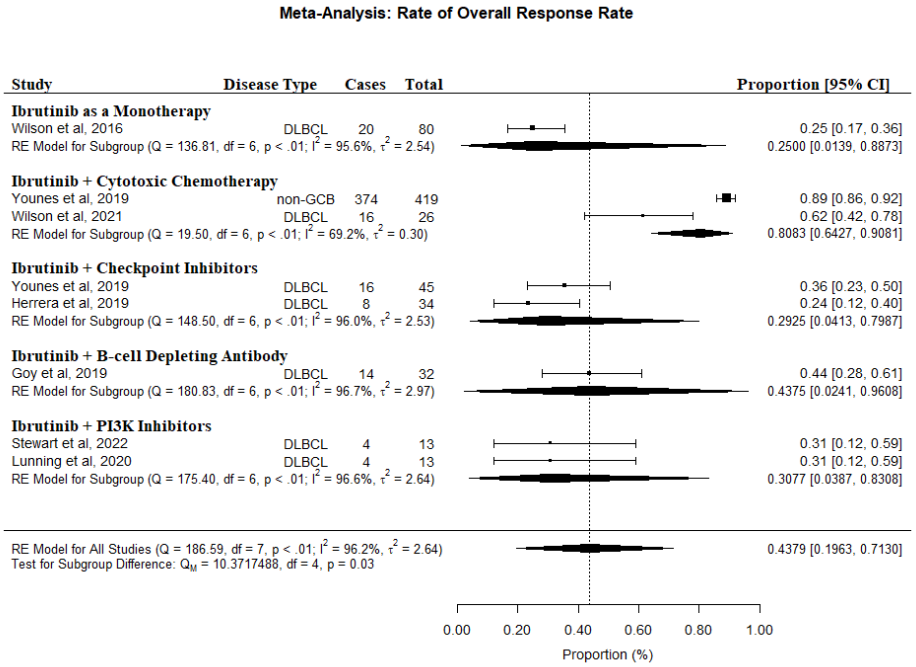


Figure 5: (DLBCL) Meta-Analysis on Overall Response Rate

3.5.3 Mantle Cell Lymphoma

A meta-analysis for patients with MCL was done based on 13 trials (n=681). The forest plot is presented in Figure 6. The overall pooled estimated for ORR was 0.82 (95% CI: $[0.75, 0.87]$). The Cochran Q test was 37.04 (p-value < 0.01), the I^2 was 62.2% and τ^2 was 0.30 indicating between-study heterogeneity. The global test for subgroup analysis was not significant (p-value $= 0.51$) meaning there was no intervention subgroup that had significant different ORR at 5% level of significance.

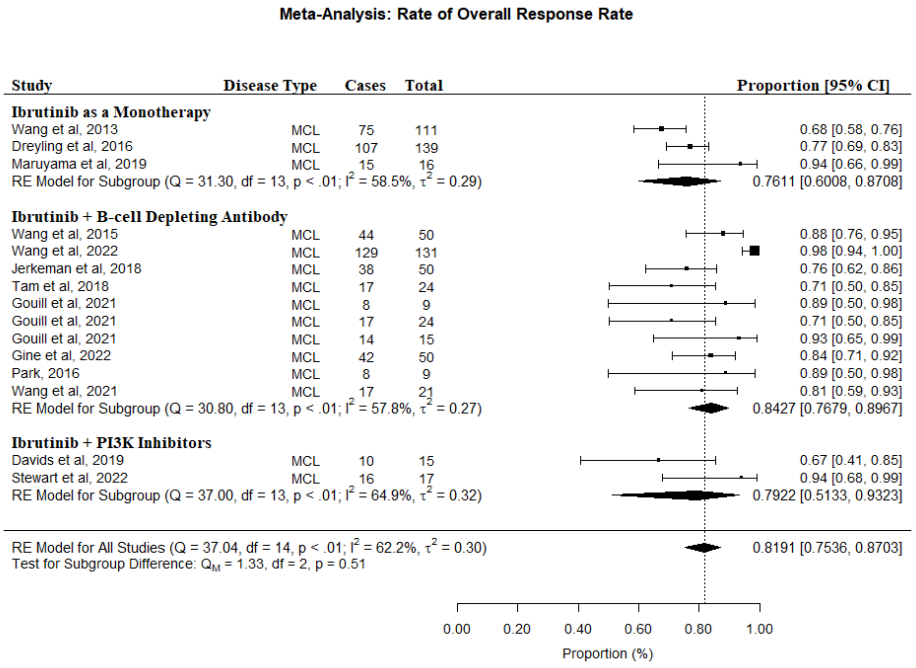


Figure 6: (MCL) Meta-Analysis on Overall Response Rate

3.6 Meta-analysis of the Rate of Hematologic Adverse Events

The meta-analysis was done for DLBCL, MCL and combination of the two disease types. For each disease type and combination, the meta-analysis of 6 different outcomes and subgroups of intervention was performed separately. Those 6 outcomes were any grade anemia, any grade thrombocytopenia, any grade neutropenia, grade 3 or higher anemia, grade 3 or higher thrombocytopenia, and grade 3 or higher neutropenia. Table 9 in the Appendix shows the details of theses outcomes.

3.6.1 Combination of the Two Disease Types

For both disease types combined, the pooled estimated proportion for any grade anemia, thrombocytopenia and neutropenia were 0.25 (95% CI: $[0.18, 0.35]$), 0.28 (95% CI: $[0.22, 0.36]$) and 0.22 (95% CI: $[0.16, 0.34]$) respectively (see Figure 9, 10, 11 in the Appendix). The tests for subgroup difference among the 5 subgroups of treatment for these adverse events were not significant (p-value = 0.10 for anemia, p-value = 0.55 for thrombocytopenia, and p-value = 0.24 for neutropenia) meaning that there were no subgroups that were significantly difference at 5% level of significance.

The pooled estimated proportion for grade 3 or higher anemia, thrombocytopenia and neutropenia were 0.09 (95% CI: $[0.06, 0.15]$), 0.11 (95% CI: $[0.07, 0.16]$) and 0.13 (95% CI: $[0.07, 0.21]$) respectively (see Figure 12, 13, 14 in the Appendix). Only grade 3 or higher anemia, the test

for subgroup difference (p-value = 0.02) was significant meaning that at least two subgroups of treatment were significantly difference at 5% level of significance. The contrast analysis for grade 3 or higher anemia was conducted. After Holm-Bonferroni correction for multiplicity was applied, in total of 10 pairs, 1 pair showed significant p-value presented in Table 7. The significant difference of logit scale between ibrutinib in combination with cytotoxic chemotherapy and ibrutinib in combination with B-cell depleting antibody was 1.92 (SE: 0.60). It can be concluded that ibrutinib in combination with B-cell depleting antibody on average is associated with a lower rate of grade 3 or higher anemia by 16.28% compared to ibrutinib in combination with cytotoxic chemotherapy.

Table 7: Contrast estimate (difference of logit scale) of significant pair(s) for subgroup model

	Estimate	**Std. Error**	z **value**	**Pr**$(>\|z\|)$
ICC − IBCDA	1.9236	0.6045	3.182	0.0146

- ICC: Ibrutinib + Cytotoxic Chemotherapy
- IBCDA: Ibrutinib + B-cell Depleting Antibody

3.6.2 Diffuse Large B-Cell Lymphoma

For DLBCL, the meta-analysis was based on 4 trials (n=510) for any grade anemia, thrombocytopenia and neutropenia and the pooled estimated proportion were obtained and equaled to 0.37 (95% CI: $[0.24, 0.53]$), 0.29 (95% CI: $[0.20, 0.41]$) and 0.38 (95% CI: $[0.22, 0.56]$) respectively (see Figure 15, 16, 17 in the Appendix). Only any grade anemia, the test for subgroup difference was statistically significance at 5% (p-value = 0.01) indicating that among the 3 subgroups of intervention, at least 2 were significantly difference. The contrast analysis was done for any grade anemia. Holm-Bonferroni correction was done for the 3 pairs and 2 pairs showed significant differences. Ibrutinib in combination with checkpoint inhibitors was significantly different from ibrutinib in combination with B-cell depleting antibody and ibrutinib in combination with cytotoxic chemotherapy with the difference of logit scale of -0.50 (SE: 1.47) and -3.97 (SE: 1.43) respectively. The interpretation can be made that on average, ibrutinib with checkpoint inhibitors on average associated with lower rate of any grade anemia by 30.92% and 44.43% compared to ibrutinib in combination with B-cell depleting antibody and ibrutinib in combination with cytotoxic chemotherapy respectively.

Table 8: Contrast estimate (difference of logit scale) of significant pair(s) for subgroup model

	Estimate	**Std. Error**	z **value**	**Pr**$(>|z|)$
ICI − IBCDA	-3.4965	1.4708	-2.377	0.0349
ICI − ICC	-3.9703	1.4277	-2.781	0.0163

- ICC: Ibrutinib + Cytotoxic Chemotherapy
- ICI: Ibrutinib + Checkpoint Inhibitors
- IBCDA: Ibrutinib + B-cell Depleting Antibody

There were 6 trials (n=568) included in the meta-analysis of grade 3 or higher anemia, thrombocytopenia, and neutropenia and the pooled estimated proportion was obtained respectively 0.19 (95% CI: $[0.10, 0.33]$), 0.11 (95% CI: $[0.05, 0.23]$) and 0.15 (95% CI: $[0.05, 0.36]$) (see Figure 18, 19, 20 in the Apppendix). For all the three outcomes, the tests for subgroup differences were statistically not significant at 5% level of significance.

3.6.3 Mantle Cell Lymphoma

For MCL, due to availability of data, different number of trials were included in the meta-analysis of any grade anemia (12 trials; n=680), thrombocytopenia (12 trials; n=647) and neutropenia (11 trials; n=630). The pooled estimated proportion was 0.22 (95% CI: $[0.13, 0.34]$) for any grade anemia, 0.27 (95% CI: $[0.19, 0.39]$ for any grade thrombocytopenia and 0.22 (95% CI: $[0.16, 0.30]$) for any grade neutropenia (see Figure 21, 22, 23 in the Appendix). No p-values were less than 0.05 for the test of subgroup difference for these three outcomes.

All 12 trials (n=680) with MCL were included in the meta-analysis of grade 3 or higher anemia, thrombocytopenia, and neutropenia and obtained the pooled estimated proportion respectively 0.07 (95% CI: $[0.05, 0.11]$), 0.10 (95% CI: $[0.06, 0.017]$) and 0.14 (95% CI: $[0.09, 0.36]$). The tests for subgroup differences were not significant for each outcome. Therefore, no treatment subgroups were significantly different at 5% level of significance.

4 Discussion

There has been great interest in assessing the relationship between early endpoints (such as ORR and PFS) and endpoint of interest (OS) in oncology studies, including studies of patients with Non-Hodgkin lymphoma (NHL), to facilitate shorter trial durations and smaller sample sizes. For DLBCL, the present analysis found that the study-level correlation between ORR and median OS was very strong. The weighted Pearson correlation was 0.932. From the regression analysis, adjusted R-squared was 0.846. Similarly, the correlations between ORR and probability of OS from 3 months to 21 months were very strong; however, at 24 months the correlation between ORR and probability of OS was very weak. This may be due to a high event rate prior to this follow up time. These results indicate that there is a consistent positive relationship between ORR and probability of OS. Batlevi and Younes (2018) [4] agreed that ORR is the most widely used surrogate for OS in single-arm study. On the other hand, the median PFS and median OS did not show any strong correlation. Moreover, the median PFS showed strong correlation with probability of OS only at 12 months and 15 months. However, probability of PFS at early stage (3 months and 6 months) showed strong correlation with probability of OS at later time points from 12 months to 21 months but not at 24 months. This finding is in contrast with a conclusion made by meta-analysis study by Zhu et al (2017) [65] which suggested that probability of PFS at 6 months is a potential surrogate endpoint for probability of OS at 24 months in newly diagnosed DLBCL and MCL. This is maybe due to the study by Zhu et al (2017) had longer median follow-up time for DLBCL 33 months and ranged from 4.6 to 84.3 months compared to the average of 24.85 months ranged from 15.8 to 40.8 in the present study.

For MCL, there were only a few trials that had available data on median PFS and median OS, so the correlations between these two variables and other variables were not evaluated. The correlations between ORR and probability of OS were calculated and it can be seen that there were strong correlation from 6 months to 18 months. Meanwhile, there were no very strong correlations between probability of PFS and OS when probability for OS was measured at three months. This may be due to not enough events were observed yet for OS endpoint. Nevertheless, the correlations were very strong when the probabilities of survival were measured at the same time but not too early in the study preferably from 6 months. Additionally, across 6 months to 18 months there were very strong correlations. The finding from the study indicates no very strong correlation between probability of PFS at 6 months and probability of OS at 24 months which is in contrast with the findings of Zhu et al (2017) [65]. The same reason can be made that in their study, there was longer median follow-up time.

Meta-analysis quantitatively combined the results of several studies to provide a pooled estimate on a question of interest. Additionally, subgroup analysis of meta-analysis allows to provide overall summary statistics within subgroup and compare them. Meta-analysis of ORR was done

for DLBCL, MCL and the combination of the two disease types. When combining the two diseases types, the pooled estimated of ORR were high (0.72) with great heterogeneity ($I^2 = 91\%$). After conducting analysis for DLBLC and MCL separately, there was still great heterogeneity in respective subgroups ($I^2 = 96.2\%$ for DLBCL and $I^2 = 62.2\%$ for MCL). Therefore, there is suspicion of other factors such as median follow-up time, median age, proportion of male, and line of therapy that may play role as moderators. The subgroup analysis showed significant results which led to a conclusion that ibrutinib in combination with B-cell depleting antibody had higher ORR compared to ibrutinib in combination with checkpoint inhibitors. Noticing that B-cell depleting antibodies are approved with demonstrated efficacy in treatment of NHL whereas checkpoint inhibitors are investigational [26]. For DLBCL, there were very high heterogeneity, $I^2 = 96.2\%$, and the pooled estimate of proportion is 0.44. It can be observed that an RCT by Younes et al (2019) [61] had larger ORR with great difference compared to the rest. This study had the large sample size and was influential (with z-value of 5.29). This may lead to high between-study variance. The subgroup analysis showed significant p-value which indicated that there was difference in subgroups of intervention. However, the contrast analysis by Tuky method with Holm-Bonferroni correction did not show any significant pairs. This maybe due to the differences of the subgroups were significant borderline and Holm-Bonferroni was too conservative for the 10 comparisons. For MCL, the pooled estimate for proportion was 0.82. MCL had more consistency between the studies but still suffered from heterogeneity ($I^2 = 62.2\%$).

Meta-analysis of the safety was done on DLBCL, MCL and combination of the two disease types with any grade and grade 3 or higher of anemia, thrombocytopenia, and neutropenia. For combination of the two disease types, it was found that ibrutinib in combination with B-cell depleting antibody led to fewer rate of grade 3 or higher anemia than ibrutinib in combination with cytotoxic chemotherapy. Again, ibrutinib in combination with B-cell depleting antibody were mostly used to treat patients with MCL whereas ibrutinib in combination with cytotoxic chemotherapy were used to treat only patient with DLBCL. Notably, DLBCL is more aggressive than MCL and ibrutinib was approved by FDA and EMA for MCL but not for DLBCL. For DLBCL, it was indicated that ibrutinib with checkpoint inhibitors associated with lower rate of any grade anemia compared to ibrutinib with B-cell depleting antibody and ibrutinib in combination with cytotoxic chemotherapy. In this meta-analysis, there were only 4 trials included and two subgroups had only one trial to base on. This could lead to reduced statistical power. The between-study variance was estimated from all 4 trials since common between-study variance was made. However, Cochrane handbook suggested at least 10 studies for each subgroup or moderator; and Fu et all (2011) [17] recommended for categorical subgroup that the minimum number of studies in each group should be 4 to have clinically meaningful result.

There were some limitations in the analyses. For the correlation and regression analyses,

first, the correlations were based only on the trial-level. The individual patient data were not available and cannot be obtained from the publicly available trial-level information, therefore the patient-level correlation cannot be evaluated. Secondly some correlations were conducted based on less than 10 paired data points (7 or 8) which may be small to obtain reliable result. Lastly, there were limited randomized controlled trials in the source, and the single-arm studies were said to be at high risk of bias by their nature. This may also impose the obtained correlation to be unreliable. For future studies, if possible, patient-level association should also be done; including 10 or more data points in the analyses; and including more RCTs. For meta-analysis, the first limitation was that only few number of RCTs included in the study and large number of single-arm studies. By its nature single-arm studies have higher risk of bias. Second limitation is that there is great amount of heterogeneity in the analysis. Thirdly, the publicly available information did not allow to assess for individual-patient data, which is gold standard for meta-analysis. Lastly, there were evidence for publication bias in some meta-analysis for example meta-analysis of ORR for MCL and combination of DLBCL and MCL. For future studies, bias-correction methods for meta-analysis, for example: the method by Begg & Pilote (1991) [6], the method by Zhang et al (2019) [40, 64], or the method by Verde (2021) [46], should be considered. However, the mentioned methods are for meta-analysis of treatment effect, the adaptations may need to be applied for meta-analysis of proportion or prevalence. Moderators such as median follow-up time, median age, proportion of male, and line of therapy should also be considered. Finally, if possible, meta-analysis on individual-data should be done. Additional limitation for the subgroup analysis is that there was small number of studies in the subgroup analysis. There should be at least four studies in each subgroup. Moreover, differences reported in subgroup analysis between studies cannot be interpreted as causal evidence [52]. It is likely that the difference in effect sizes among subgroups is due to unidentified factors that are not assessed in such moderator analyses. This problem, unfortunately, does not have a solution .

5 Conclusion

In conclusion, for DLBCL, the findings in this study found a positive relationship between ORR and probability of OS consistently from 3 months to 21 months for DLBCL, suggesting that ORR was might be considered adequate to predict median OS; whereas PFS has an inconsistent relationship with OS suggesting that there was not enough evidence to show that PFS can be used to predict OS for DLBCL. For MCL, the relationship between ORR and probability OS was determined from 6 months to 18 months and similarly the across 6 months to 18 months the relationship between probability of PFS and probability of OS was determined. When DLBCL and MCL were analysed together, the meta-analyses suggested that the ORR ans probability of adverse events was impacted by the type of combination therapy. Ibrutinib in combination with B-cell depleting antibody was associated with higher ORR than ibrutinib in combination with checkpoint inhibitors and was associated with less grade 3 or higher anemia than ibrutinib in combination with cytotoxic chemotherapy. When DLBCL was analysed alone, the meta-analysis showed that ibrutinib in combination with checkpoint inhibitors associated with less rate of any grade anemia than ibrutinib in combination with B-cell depleting antibody and ibrutinib in combination with cytotoxic chemotherapy. This suggested that ibrutinib in combination with checkpoint inhibitors had the least risk of any grade anemia in DLBCL.

References

[1] *About lymphoma*. Sept. 2021. URL: https://lymphoma.org/aboutlymphoma/.

[2] James O Armitage and Dennis D Weisenburger. "New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project." In: *Journal of Clinical Oncology* 16.8 (1998), pp. 2780–2795.

[3] Paul Bailey et al. *wCorr Formulas*. May 2021. URL: https://cran.r-project.org/web/packages/wCorr/vignettes/wCorrFormulas.pdf.

[4] CL Batlevi and A Younes. "Surrogate end points in lymphoma". In: *Annals of Oncology* 29.8 (2018), pp. 1622–1623.

[5] C Beauchemin et al. "Relationship between progression-free survival and overall survival in chronic lymphocytic leukemia: a literature-based analysis". In: *Current oncology* 22.3 (2015), pp. 148–156.

[6] Colin B Begg and Louise Pilote. "A model for incorporating historical controls into a meta-analysis". In: *Biometrics* (1991), pp. 899–906.

[7] Frank Bretz, Torsten Hothorn, and Peter Westfall. *Multiple comparisons using R*. Chapman and Hall/CRC, 2016.

[8] Jan A Burger et al. "Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia". In: *New England Journal of Medicine* 373.25 (2015), pp. 2425–2437.

[9] John C Byrd et al. "Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia". In: *New England Journal of Medicine* 369.1 (2013), pp. 32–42.

[10] Jeffrey M Clarke, Xiaofei Wang, and Neal E Ready. "Surrogate clinical endpoints to predict overall survival in non-small cell lung cancer trials—are we in a new era?" In: *Translational Lung Cancer Research* 4.6 (2015), p. 804.

[11] William Jay Conover. *Practical nonparametric statistics*. Vol. 350. john wiley & sons, 1999.

[12] Sybil L Crawford. "Correlation and regression". In: *Circulation* 114.19 (2006), pp. 2083–2088.

[13] Matthew S Davids et al. "Umbralisib in combination with ibrutinib in patients with relapsed or refractory chronic lymphocytic leukaemia or mantle cell lymphoma: a multicentre phase 1–1b study". In: *The Lancet Haematology* 6.1 (2019), e38–e47.

[14] Thomas J DiCiccio and Bradley Efron. "Bootstrap confidence intervals". In: *Statistical science* 11.3 (1996), pp. 189–228.

[15] Martin Dreyling et al. “Ibrutinib versus temsirolimus in patients with relapsed or refractory mantle-cell lymphoma: an international, randomised, open-label, phase 3 study”. In: *The Lancet* 387.10020 (2016), pp. 770–778.

[16] F Fiteni et al. “Endpoints in cancer clinical trials”. In: *Journal of visceral surgery* 151.1 (2014), pp. 17–22.

[17] Rongwei Fu et al. “Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program”. In: *Journal of clinical epidemiology* 64.11 (2011), pp. 1187–1197.

[18] Eva Giné et al. “Ibrutinib in combination with rituximab for indolent clinical forms of mantle cell lymphoma (IMCL-2015): a multicenter, open-label, single-arm, phase II trial”. In: *Journal of Clinical Oncology* (2022), JCO–21.

[19] Andre Goy et al. “Ibrutinib plus lenalidomide and rituximab has promising activity in relapsed/refractory non–germinal center B-cell–like DLBCL”. In: *Blood, The Journal of the American Society of Hematology* 134.13 (2019), pp. 1024–1036.

[20] Alex F Herrera et al. “Safety and activity of ibrutinib in combination with durvalumab in patients with relapsed or refractory follicular lymphoma or diffuse large B-cell lymphoma”. In: *American journal of hematology* 95.1 (2020), pp. 18–27.

[21] Julian PT Higgins et al. “The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials”. In: *Bmj* 343 (2011).

[22] Harold Hotelling. “New light on the correlation coefficient and its transforms”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 15.2 (1953), pp. 193–232.

[23] Torsten Hothorn et al. “Package ‘multcomp’”. In: *Simultaneous inference in general parametric models. Project for Statistical Computing, Vienna, Austria* (2016).

[24] *Imbruvica*. Nov. 2014. URL: `https://www.ema.europa.eu/en/medicines/human/EPAR/imbruvica`.

[25] *IMBRUVICA safely and effectively.* May 2022. URL: `https://www.imbruvica.com/files/prescribing-information.pdf`.

[26] *Immunotherapy for Non-Hodgkin Lymphoma.* URL: `https://www.cancer.org/cancer/non-hodgkin-lymphoma/treating/immunotherapy.html`.

[27] Mats Jerkeman et al. “Ibrutinib, lenalidomide, and rituximab in relapsed or refractory mantle cell lymphoma (PHILEMON): a multicentre, open-label, single-arm, phase 2 trial”. In: *The Lancet Haematology* 5.3 (2018), e109–e116.

[28] Steven Le Gouill et al. "Ibrutinib, obinutuzumab, and venetoclax in relapsed and untreated patients with mantle cell lymphoma: a phase 1/2 trial". In: *Blood, The Journal of the American Society of Hematology* 137.7 (2021), pp. 877–887.

[29] Liang Li and Olanrewaju Okusanya. *Clinical Pharmacology and Biopharmaceutics Review(s), Application Number: 210563Orig1s000, 210563Orig2s000O*. Feb. 2018.

[30] Matthew Lunning. *TGR-1202 and Ibrutinib in treating patients with relapsed or refractory diffuse large B-cell lymphoma - full text view*. Feb. 2020. URL: `https://clinicaltrials.gov/ct2/show/NCT02874404`.

[31] Naveen Mangal et al. "Relationship between response rates and median progression-free survival in non-Hodgkin's lymphoma: a meta-analysis of published clinical trials". In: *Hematological oncology* 36.1 (2018), pp. 37–43.

[32] Dai Maruyama et al. "Final analysis of a phase II study of ibrutinib in Japanese patients with relapsed/refractory mantle cell lymphoma". In: *Journal of clinical and experimental hematopathology* 59.2 (2019), pp. 98–100.

[33] Manoranjan Pal and Premananda Bharati. "Introduction to correlation and linear regression analysis". In: *Applications of Regression Techniques*. Springer, 2019, pp. 1–18.

[34] Byung Park. *Obinutuzumab in combination with Ibrutinib in treating patients with relapsed mantle cell lymphoma - full text view*. Jan. 2022. URL: `https://clinicaltrials.gov/ct2/show/NCT02736617`.

[35] Richard D Riley, Julian PT Higgins, and Jonathan J Deeks. "Interpretation of random effects meta-analyses". In: *Bmj* 342 (2011).

[36] Thomas D Rodgers and Patrick M Reagan. "Targeting the B-cell receptor pathway: a review of current and future therapies for non-Hodgkin's lymphoma". In: *Expert Opinion on Emerging Drugs* 23.2 (2018), pp. 111–122.

[37] Sumina Sapkota and Hira Shaikh. "Non-Hodgkin Lymphoma". In: (2020).

[38] Kate R Shankland, James O Armitage, and Barry W Hancock. "Non-hodgkin lymphoma". In: *The Lancet* 380.9844 (2012), pp. 848–857.

[39] Beth Sherrill et al. "Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology". In: *OncoTargets and therapy* 5 (2012), p. 287.

[40] Janharpreet Singh, Keith R Abrams, and Sylwia Bujkiewicz. "Incorporating single-arm studies in meta-analysis of randomised controlled trials: a simulation study". In: *BMC medical research methodology* 21.1 (2021), pp. 1–13.

[41] Caitlin M Stewart et al. “Phase I/Ib study of the efficacy and safety of buparlisib and ibrutinib therapy in MCL, FL, and DLBCL with serial cell-free DNA monitoring”. In: *Clinical Cancer Research* 28.1 (2022), pp. 45–56.

[42] Steven H Swerdlow et al. “The 2016 revision of the World Health Organization classification of lymphoid neoplasms”. In: *Blood, The Journal of the American Society of Hematology* 127.20 (2016), pp. 2375–2390.

[43] Constantine S Tam et al. “Ibrutinib plus venetoclax for the treatment of mantle-cell lymphoma”. In: *New England journal of medicine* 378.13 (2018), pp. 1211–1223.

[44] *Target product profiles*. URL: https://www.who.int/observatories/global-observatory-on-health-research-and-development/analyses-and-syntheses/target-product-profile/who-target-product-profiles#:~:text=A%20target%20product%20profile%20(TPP,safety%20and%20efficacy%2Drelated%20characteristics..

[45] Steven P Treon et al. “Long-term follow-up of ibrutinib monotherapy in symptomatic, previously treated patients with Waldenström macroglobulinemia”. In: *Journal of Clinical Oncology* 39.6 (2021), pp. 565–575.

[46] Pablo Emilio Verde. “A bias-corrected meta-analysis model for combining, studies of different types and quality”. In: *Biometrical Journal* 63.2 (2021), pp. 406–422.

[47] Liza C Villaruz and Mark A Socinski. *The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement*. 2013.

[48] Michael Wang et al. “Concurrent ibrutinib plus venetoclax in relapsed/refractory mantle cell lymphoma: the safety run-in of the phase 3 SYMPATICO study”. In: *Journal of hematology & oncology* 14.1 (2021), pp. 1–8.

[49] Michael L Wang et al. “Ibrutinib in combination with rituximab in relapsed or refractory mantle cell lymphoma: a single-centre, open-label, phase 2 trial”. In: *The Lancet Oncology* 17.1 (2016), pp. 48–56.

[50] Michael L Wang et al. “Ibrutinib–rituximab followed by R-HCVAD as frontline treatment for young patients (65 years) with mantle cell lymphoma (WINDOW-1): a single-arm, phase 2 trial”. In: *The Lancet Oncology* (2022).

[51] Michael L Wang et al. “Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma”. In: *New England Journal of Medicine* 369.6 (2013), pp. 507–516.

[52] Naike Wang. “Conducting meta-analyses of proportions in R”. In: *Research Gate: College Station, TX, USA* (2017).

[53] Tingyu Wen et al. “Inhibitors targeting Bruton’s tyrosine kinase in cancers: drug development advances”. In: *Leukemia* 35.2 (2021), pp. 312–332.

[54] *What is lymphoma?* Mar. 2020. URL: https://lymphoma-action.org.uk/about-lymphoma/what-lymphoma.

[55] *What is Non-Hodgkin lymphoma?* July 2018. URL: https://www.cancer.org/cancer/non-hodgkin-lymphoma/about/what-is-non-hodgkin-lymphoma.html.

[56] Michelle K Wilson, Katherine Karakasis, and Amit M Oza. "Outcomes and endpoints in trials of cancer treatment: the past, present, and future". In: *The Lancet Oncology* 16.1 (2015), e32–e42.

[57] Wyndham H Wilson et al. "Phase 1b/2 study of ibrutinib and lenalidomide with dose-adjusted EPOCH-R in patients with relapsed/refractory diffuse large B-cell lymphoma". In: *Leukemia & Lymphoma* 62.9 (2021), pp. 2094–2106.

[58] Wyndham H Wilson et al. "Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma". In: *Nature medicine* 21.8 (2015), pp. 922–926.

[59] Wanling Xie et al. "A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses". In: *JNCI cancer spectrum* 3.1 (2019), pkz002.

[60] Jiabu Ye et al. "Relationship Between Progression-Free Survival, Objective Response Rate, and Overall Survival in Clinical Trials of PD-1/PD-L1 Immune Checkpoint Blockade: A Meta-Analysis". In: *Clinical Pharmacology & Therapeutics* 108.6 (2020), pp. 1274–1288.

[61] Anas Younes et al. "Randomized phase III trial of ibrutinib and rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in non–germinal center B-cell diffuse large B-cell lymphoma". In: *Journal of Clinical Oncology* 37.15 (2019), p. 1285.

[62] Anas Younes et al. "Safety and activity of ibrutinib in combination with nivolumab in patients with relapsed non-Hodgkin lymphoma or chronic lymphocytic leukaemia: a phase 1/2a study". In: *The Lancet Haematology* 6.2 (2019), e67–e78.

[63] Tao Zeng and Ke-Qin Xie. "Cytochrome P4502E1 Gene Polymorphisms and the Risks of Ethanol-Induced Health Problems in Alcoholics". In: *Molecular Aspects of Alcohol and Nutrition*. Elsevier, 2016, pp. 231–245.

[64] Jing Zhang et al. "Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies". In: *Statistical methods in medical research* 28.5 (2019), pp. 1293–1310.

[65] Rui Zhu et al. "Assessment of correlation between early and late efficacy endpoints to identify potential surrogacy relationships in non-Hodgkin lymphoma: a literature-based meta-analysis of 108 phase II and phase III studies". In: *The AAPS journal* 19.3 (2017), pp. 669–681.

Appendix

Tables and Figures

	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Dreyling et al, 2016	(+)	(+)	(-)	(+)	(+)	(+)	(+)
Younes et al, 2019	(+)	(+)	(+)	(+)	(+)	(+)	(+)
Wang et al, 2021	(?)	(-)	(-)	(-)	(+)	(+)	(+)

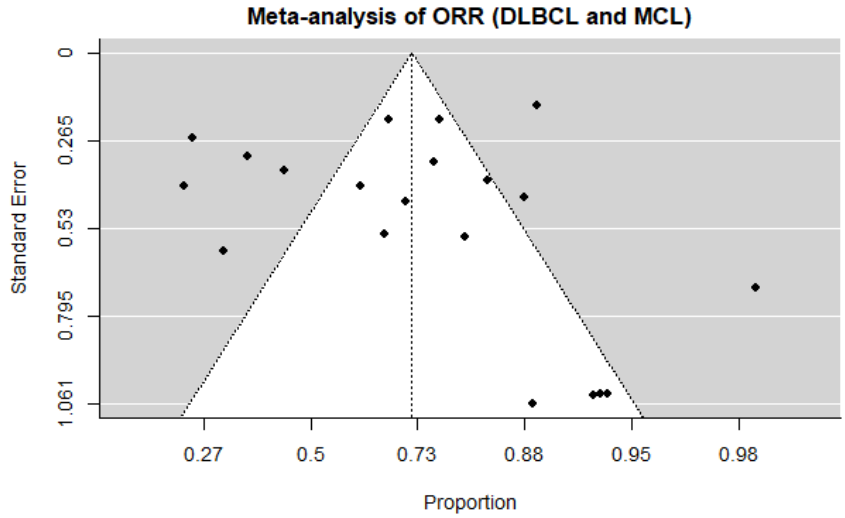
Key

(+) Low risk of bias

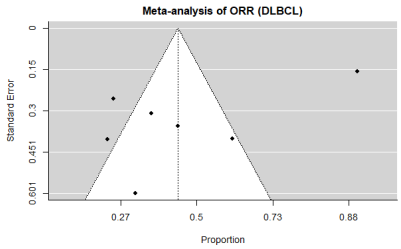
(-) High risk of bias

(?) Unclear risk of bias

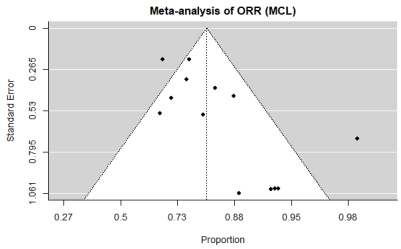
Figure 7: Cochrane risk of bias assessment for 3 included randomized trials



(a) Combination of DLBCL and MCL



(b) DLBCL



(c) MCL

Figure 8: Funnel plots for meta-analysis of ORR

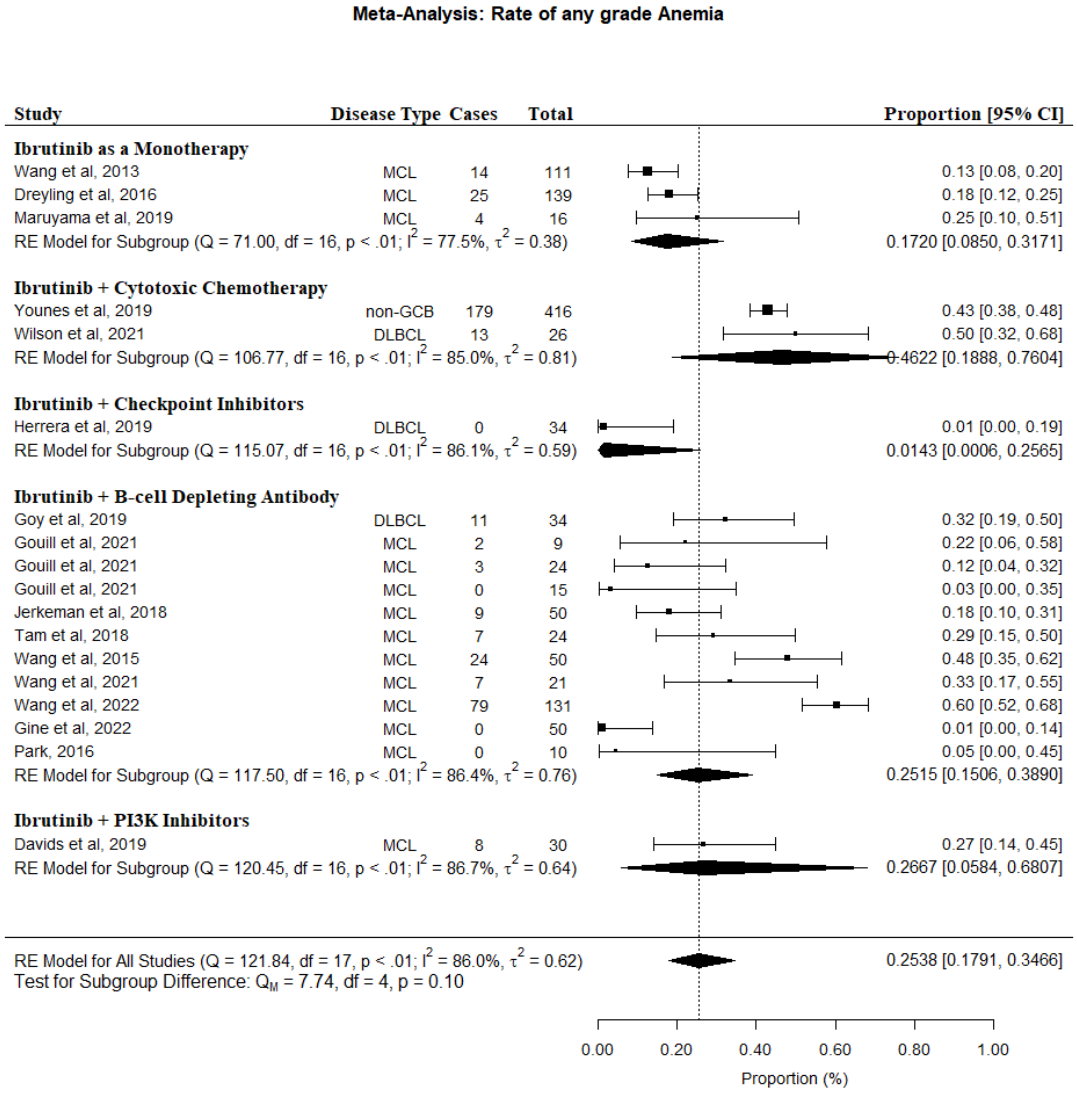


Figure 9: (DLBCL and MCL) Meta-analysis on the rate of any grade anemia

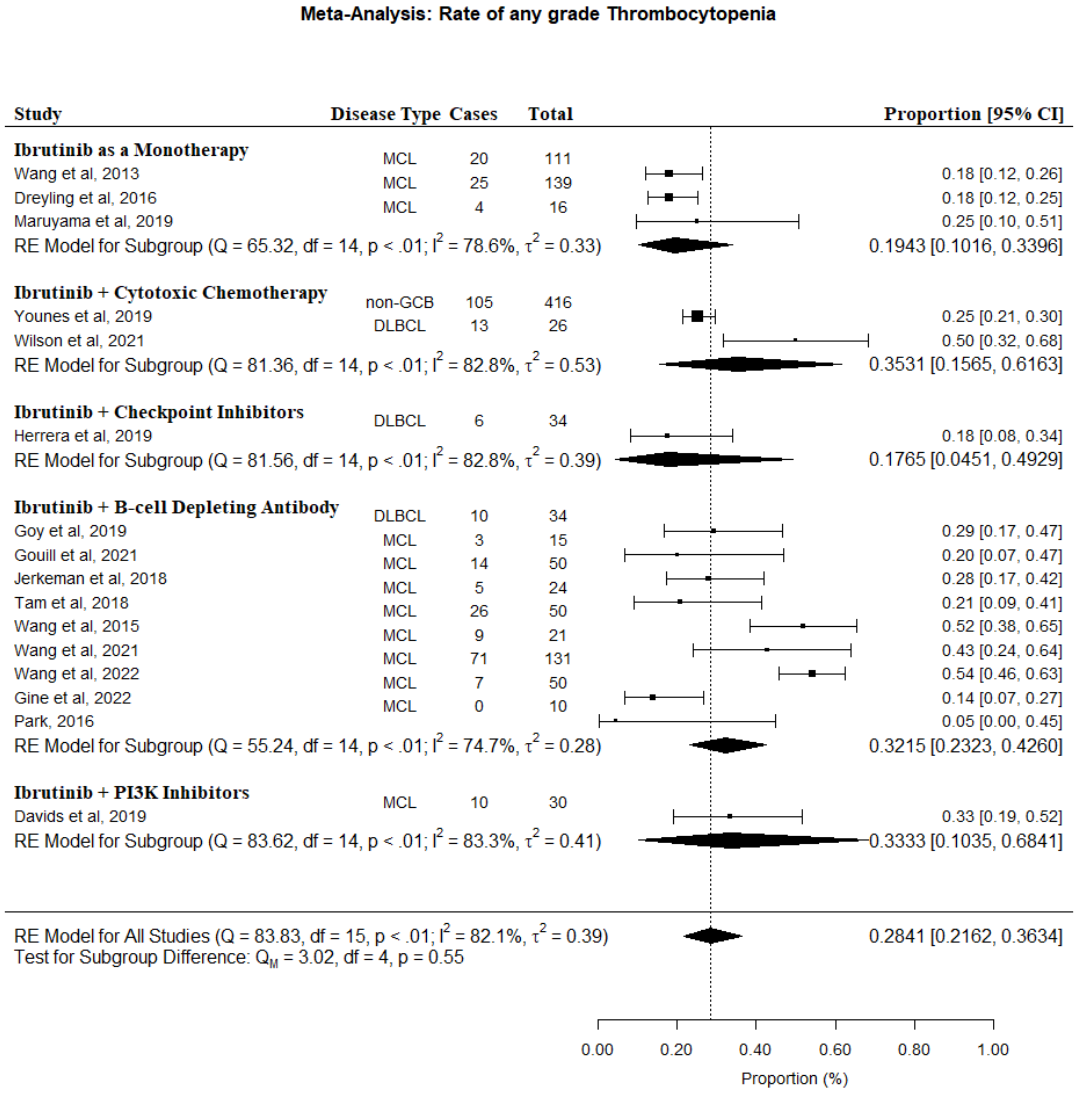


Figure 10: (DLBCL and MCL) Meta-analysis on the rate of any grade thrombocytopenia

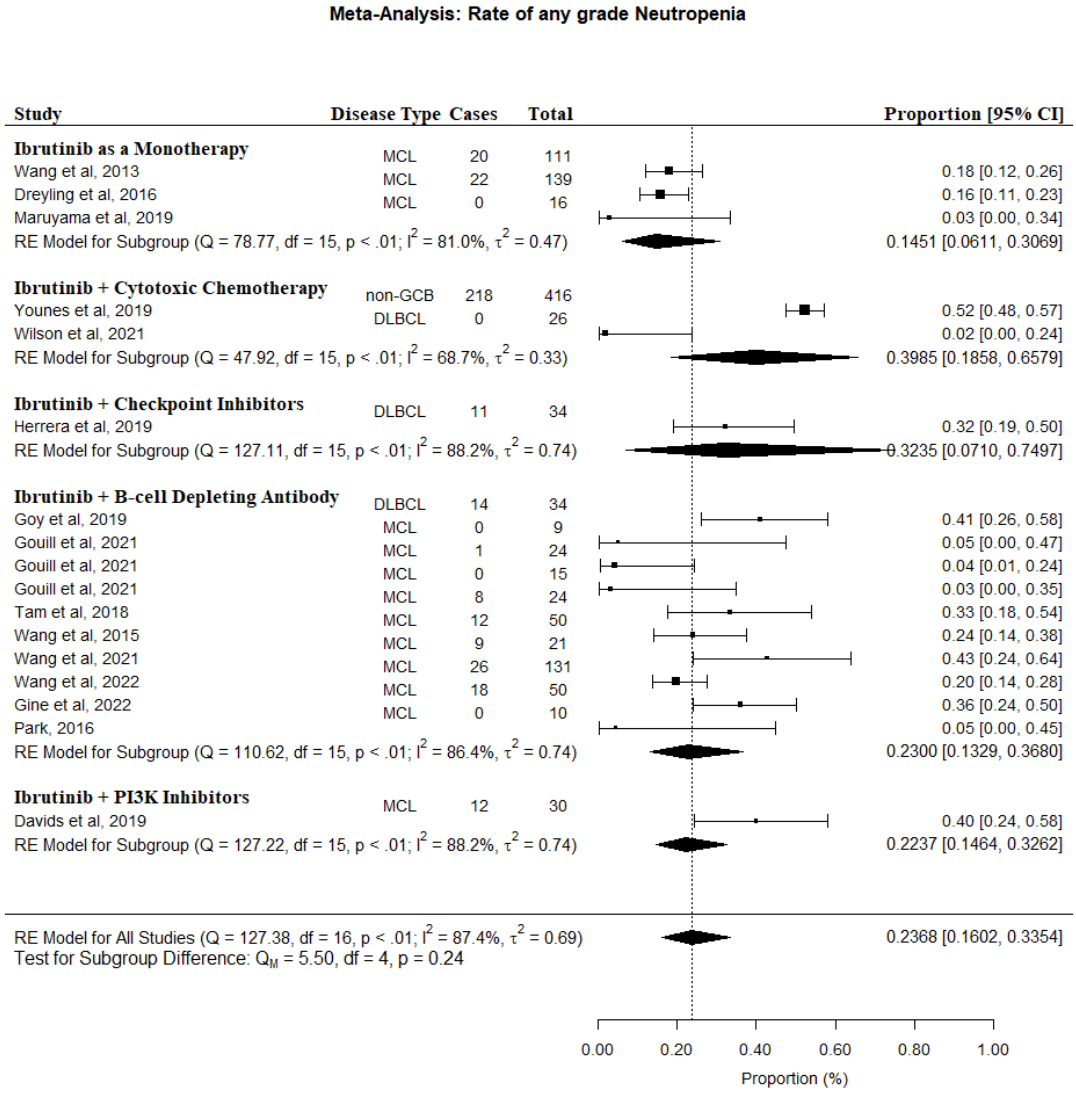


Figure 11: (DLBCL and MCL) Meta-analysis on the rate of any grade neutropenia

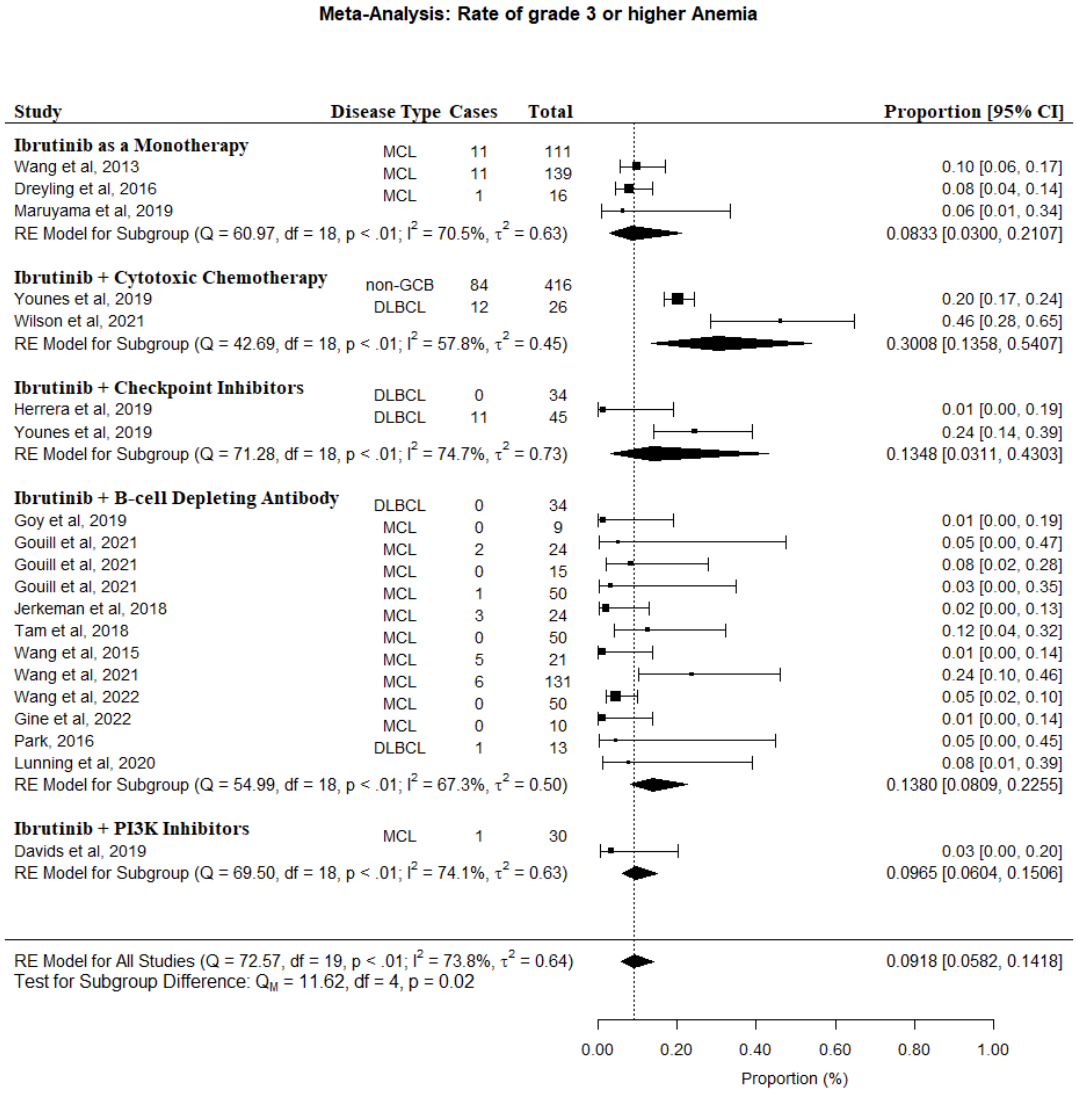


Figure 12: (DLBCL and MCL) Meta-analysis on the rate of grade 3 or higher anemia

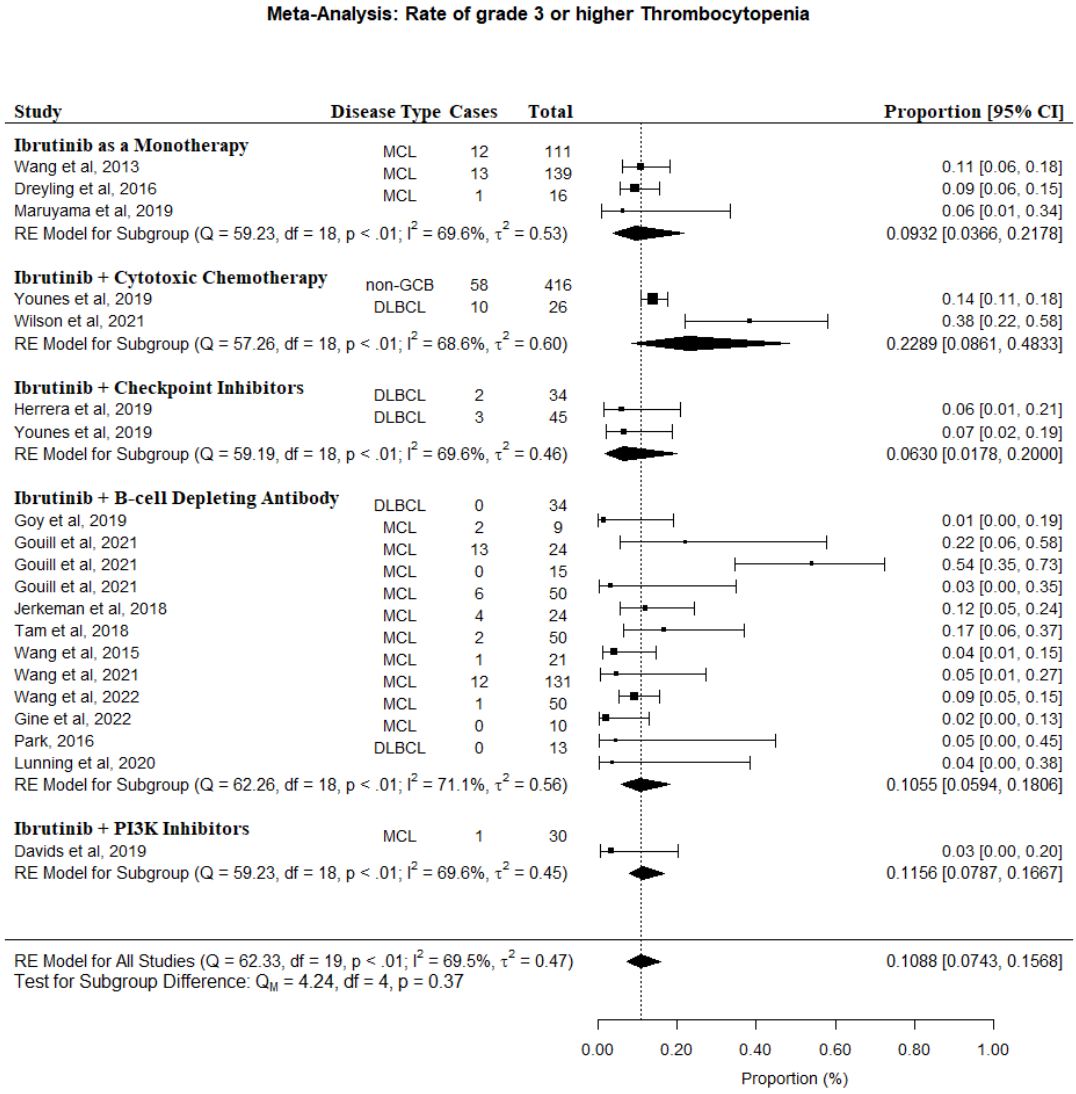


Figure 13: (DLBCL and MCL) Meta-analysis on the rate of grade 3 or higher thrombocytopenia

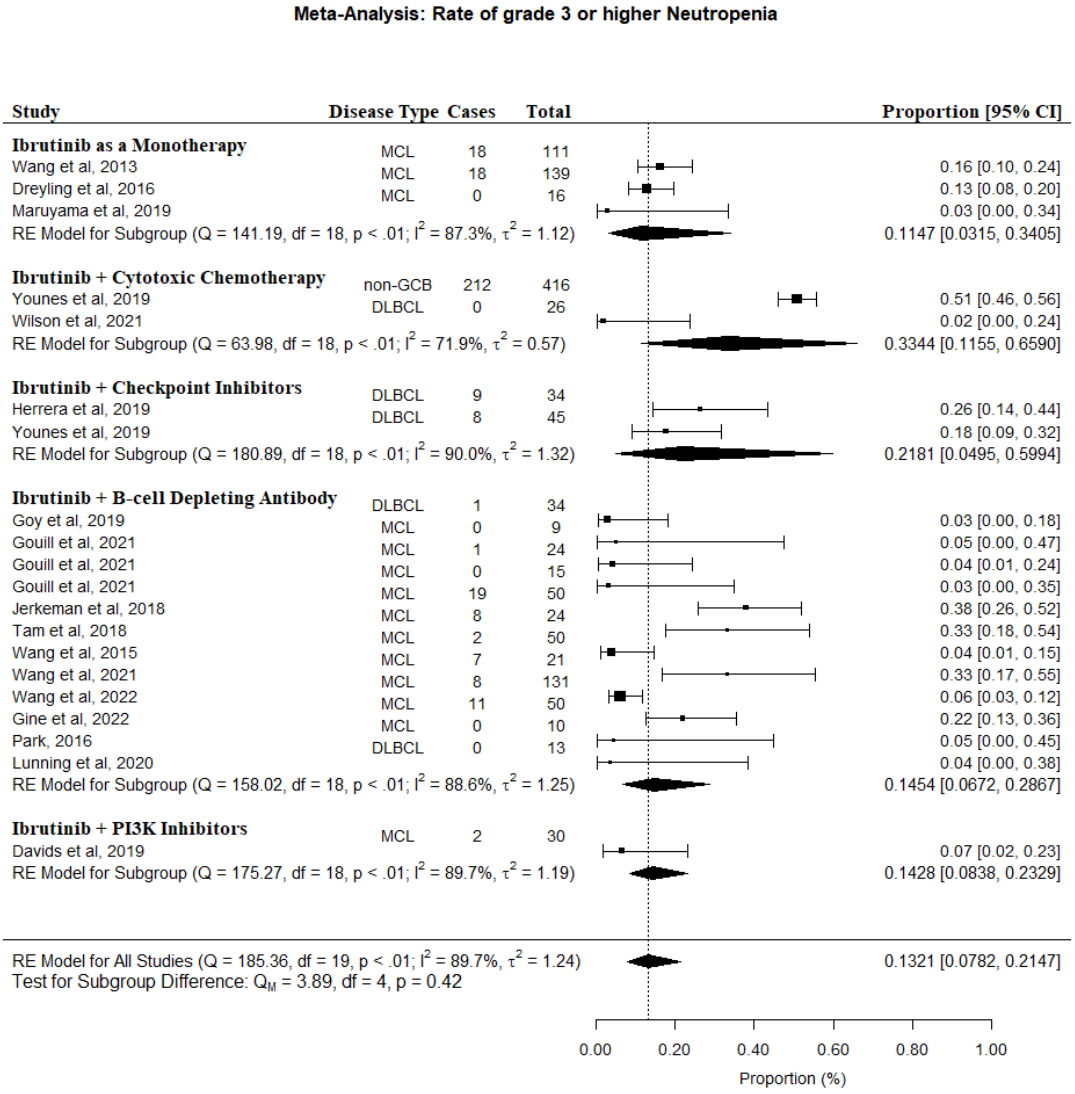


Figure 14: (DLBCL and MCL) Meta-analysis on the rate of grade 3 or higher neutropenia

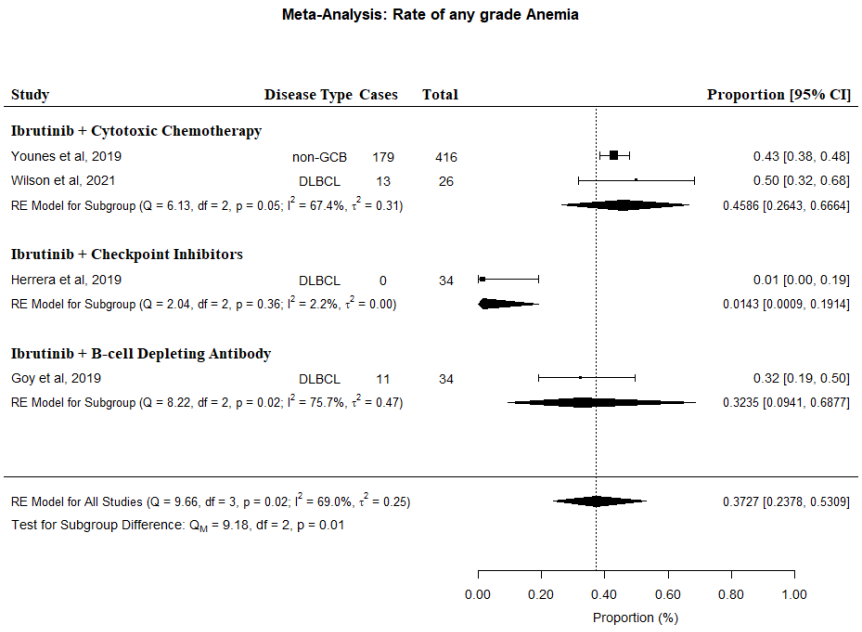


Figure 15: (DLBCL) Meta-analysis on the rate of any grade anemia

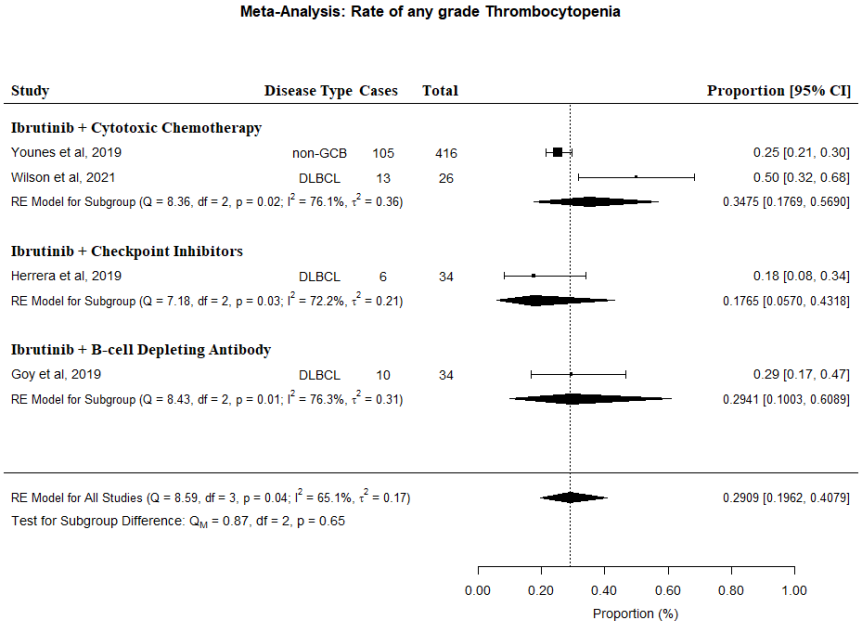


Figure 16: (DLBCL) Meta-analysis on the rate of any grade thrombocytopenia

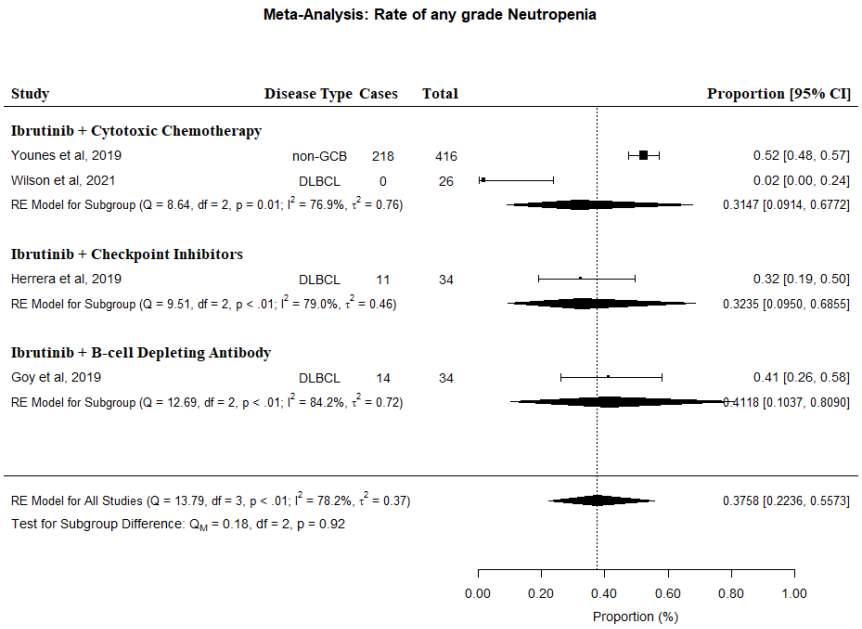


Figure 17: (DLBCL) Meta-analysis on the rate of any grade neutropenia

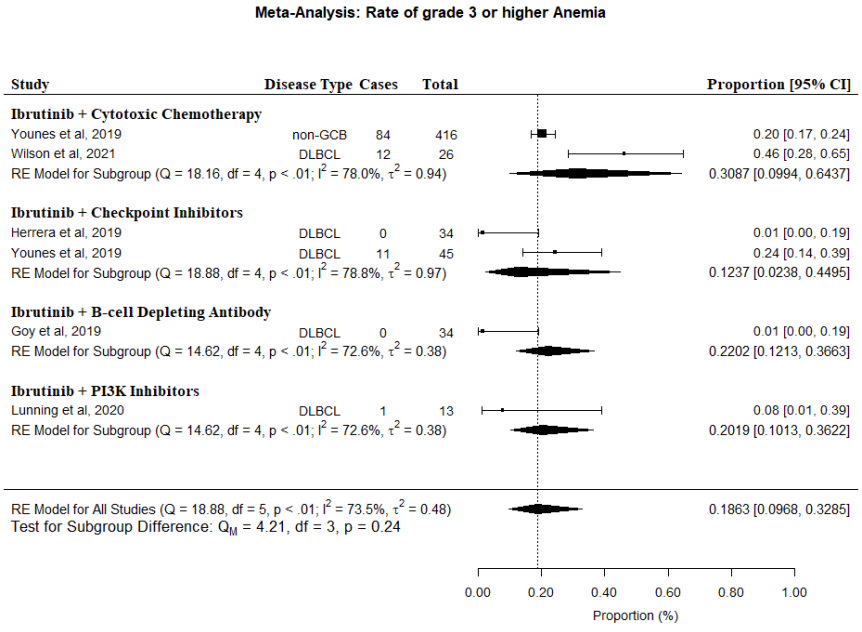


Figure 18: (DLBCL) Meta-analysis on the rate of grade 3 or higher anemia

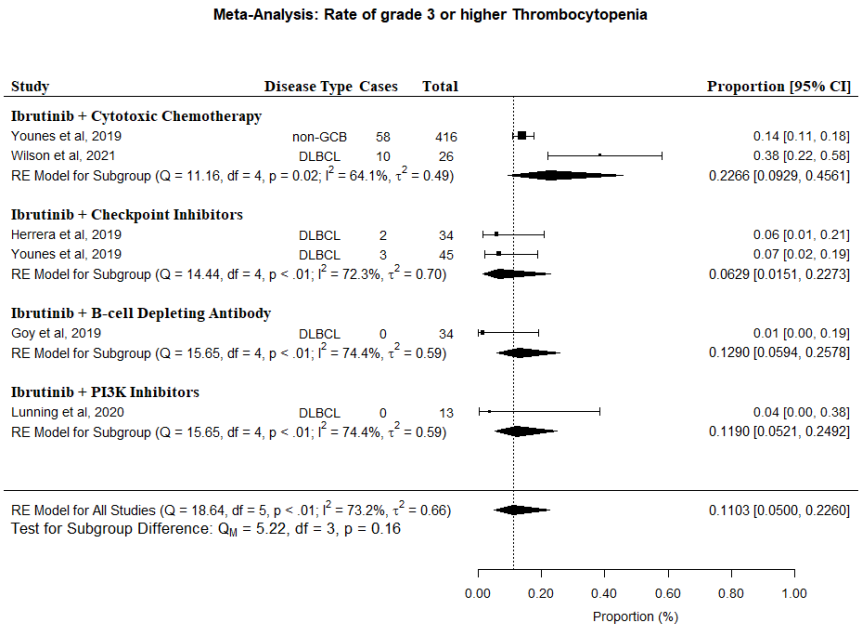


Figure 19: (DLBCL) Meta-analysis on the rate of grade 3 or higher thrombocytopenia

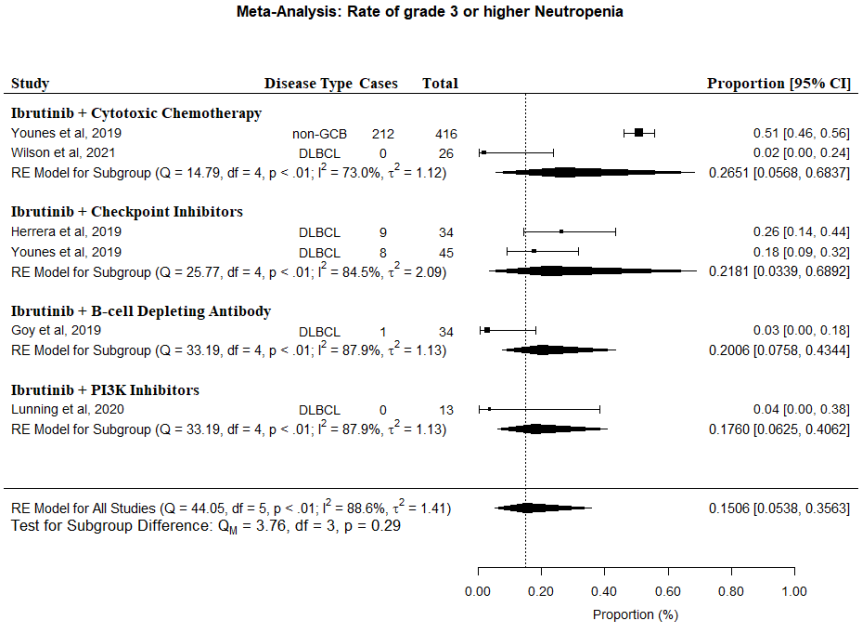


Figure 20: (DLBCL) Meta-analysis on the rate of grade 3 or higher neutropenia

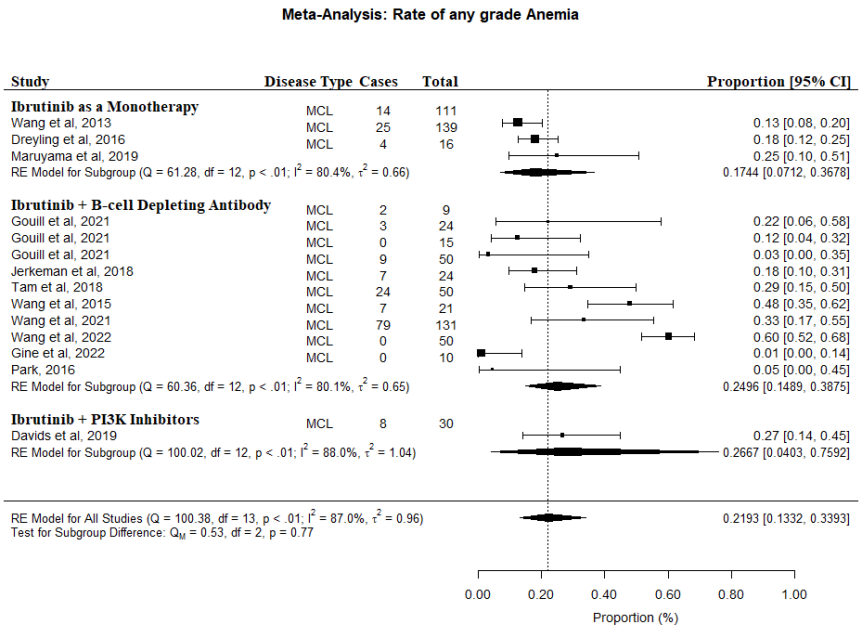


Figure 21: (MCL) Meta-analysis on the rate of any grade anemia

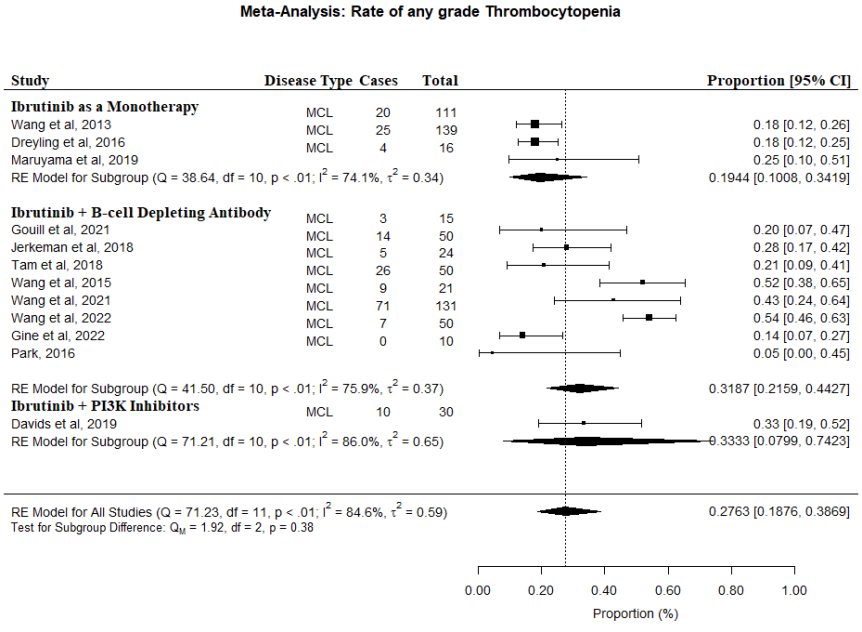


Figure 22: (MCL) Meta-analysis on the rate of any grade thrombocytopenia

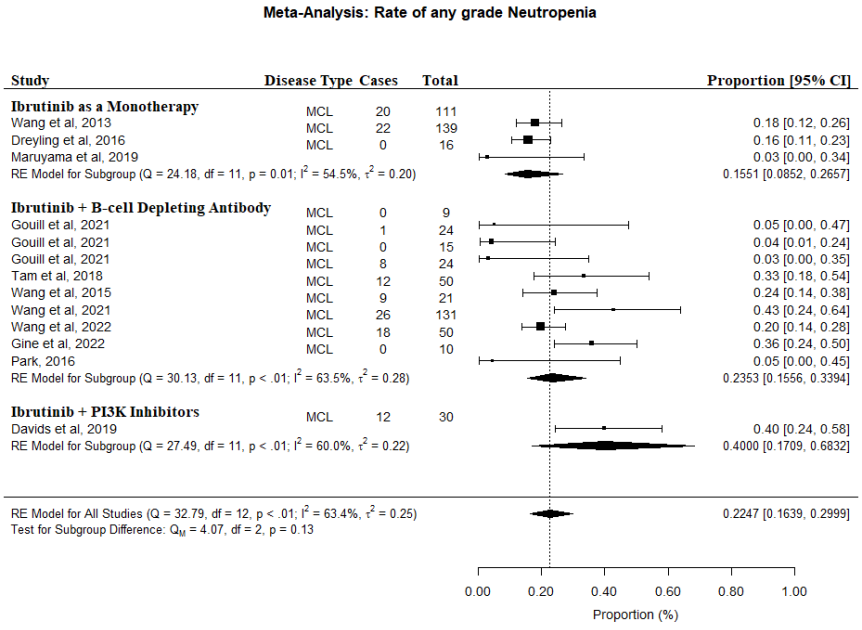


Figure 23: (MCL) Meta-analysis on the rate of any grade neutropenia

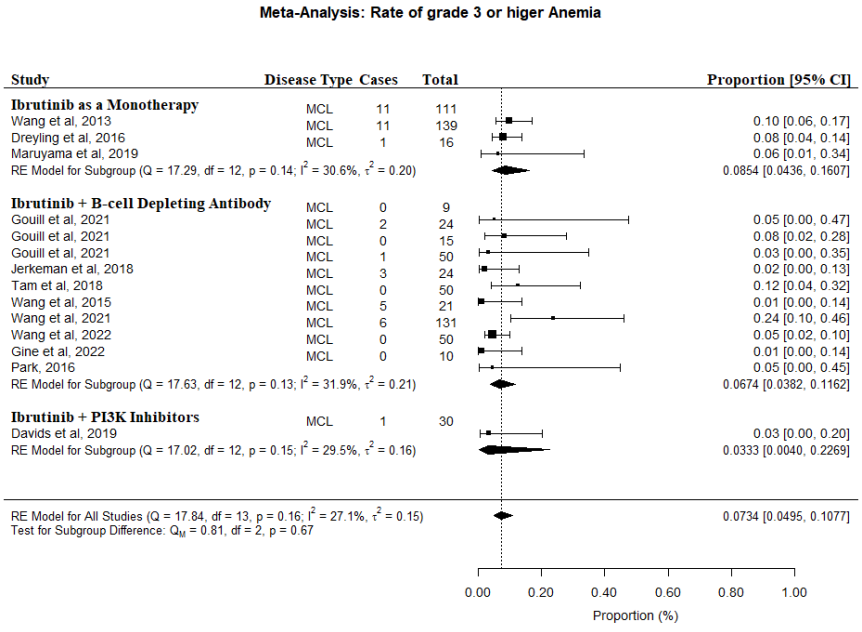


Figure 24: (MCL) Meta-analysis on the rate of grade 3 or higher anemia

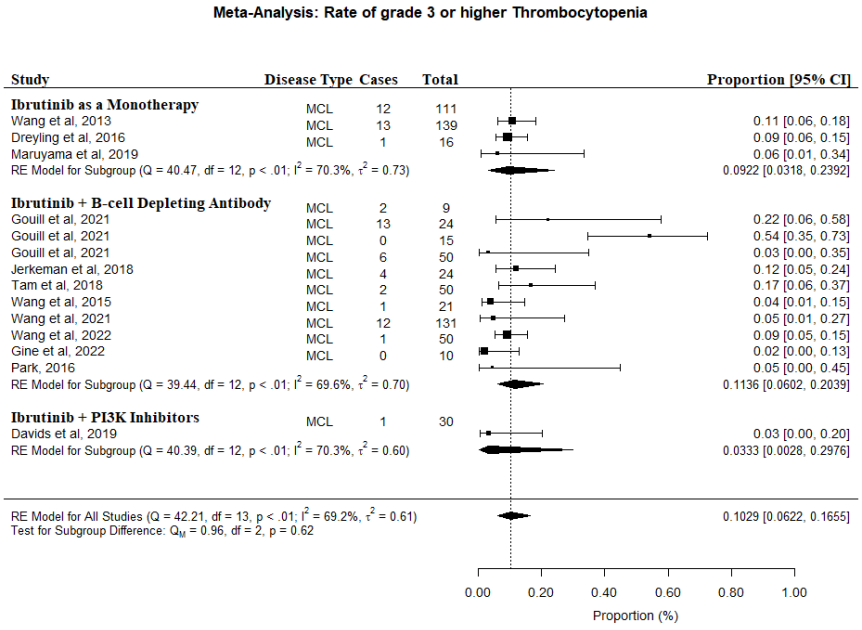


Figure 25: (MCL) Meta-analysis on the rate of grade 3 or higher thrombocytopenia

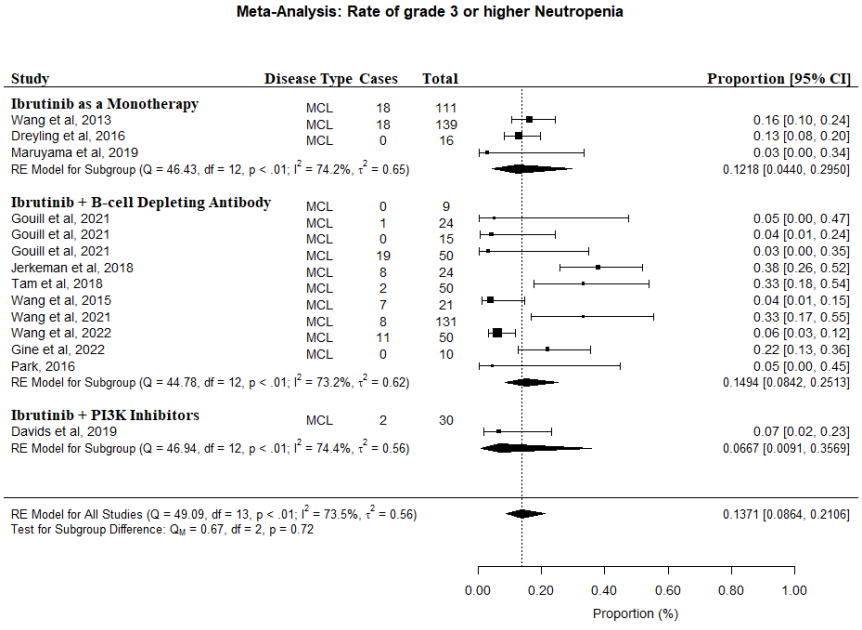


Figure 26: (MCL) Meta-analysis on the rate of grade 3 or higher neutropenia

Table 9: Trials included in meta-analysis of hematologic adverse events

Reference	Intervention Group	Disease Type	n (Safety)	Any Grade			Grade 3 or Higher		
				Ane.	Thr.	Neu.	Ane.	Thr.	Neu.
Younes et al, 2019	ICC	non-GCB	416	179	105	218	84	58	212
Wilson et al, 2021	ICC	DLBCL	26	13	13	0	12	10	0
Herrera et al, 2019	ICI	DLBCL	34	0	6	11	0	2	9
Younes et al, 2019	ICI	DLBCL	45	NA	NA	NA	11	3	8
Goy et al, 2019	IBCDA	DLBCL	34	11	10	14	0	0	1
Lunning et al, 2020	IPI3K	DLBCL	13	NA	NA	NA	1	0	0
Wang et al, 2013	IMONO	MCL	111	14	20	20	11	12	18
Dreyling et al, 2016	IMONO	MCL	139	25	25	22	11	13	18
Maruyama et al, 2019	IMONO	MCL	16	4	4	0	1	1	0
Gouill et al, 2021	IBCDA	MCL	9	2	NA	0	0	2	0
Gouill et al, 2021	IBCDA	MCL	24	3	NA	1	2	13	1
Gouill et al, 2021	IBCDA	MCL	15	0	3	0	0	0	0
Jerkeman et al, 2018	IBCDA	MCL	50	9	14	NA	1	6	19
Tam et al, 2018	IBCDA	MCL	24	7	5	8	3	4	8
Wang et al, 2015	IBCDA	MCL	50	24	26	12	0	2	2
Wang et al, 2021	IBCDA	MCL	21	7	9	9	5	1	7
Wang et al, 2022	IBCDA	MCL	131	79	71	26	6	12	8
Gine et al, 2022	IBCDA	MCL	50	0	7	18	0	1	11
Park, 2016	IBCDA	MCL	10	0	0	0	0	0	0
Davids et al, 2019	IPI3K	MCL	30	8	10	12	1	1	2

Ane.: Anemia; Thr.: Thrombocytopenia; Neu.: Neutropenia; NA: Not available.

- IMONO: Ibrutinib as a monotherapy
- ICC: Ibrutinib in combination with Cytotoxic Chemotherapy
- ICI: Ibrutinib in combination with Checkpoint Inhibitors
- IBCDA: Ibrutinib in combination with B-cell Depleting Antibody
- IPI3K: Ibrutinib in combination with PI3K inhibitors

Table 10: Main characteristics of all included studies

NCT Number	Reference	Phase	Type of Study	Intervention	Comparator	Disease Type	Line of Therapy	Intervention Group						
								Sample Size	Median Age (years)	Male (%)	Follow-up Time (Months)	ORR	Median PFS (months)	Median OS (months)
NCT01236391	Wang et al, 2013 [51]	II	SA	ibrutinib	None	MCL	PT	111	68	77	15.80	0.68	13.90	NR
NCT01325701	Wilson et al, 2016 [58]	I/II	SA	ibrutinib	None	DLBCL	PT	80	NA	70	11.53	0.25	1.64	6.41
						ABC	FL	38	60	66	10.12	0.37	2.02	10.32
						GCB	PT	20	65	70	17.05	0.05	1.31	3.35
NCT01646021	Dreyling et al, 2016 [15]	III	RCT	ibrutinib	temsirolimus	MCL	PT	139	67	72	20	0.77	15.6	30.3
NCT01855750	Younes et al, 2019 [61]	III	RCT	ibrutinib + R-CHOP	placebo + R-CHOP	non-GCB	FL	419	63	53	34.8	0.89	48.56	NR
NCT01880567	Wang et al, 2016 [49]	II	SA	ibrutinib + rituximab	None	MCL	PT	50	67	0.76	16.5	0.88	NR	NR
NCT02077166	Goy et al, 2019 [19]	Ib/II	SA	ibrutinib + rituximab + lenalidomide	None	DLBCL	PT	32	NA	NA	NA	0.44	4.7	11.6
NCT02142049	Wilson et al, 2021 [57]	Ib/II	SA	ibrutinib + lenalidomide + DA-EPOCH-R	None	DLBCL	PT	26	57.5	0.69	NA	0.62	4.86	15.84
						non-GCB	PT	21	60	0.67	NA	0.71	6.51	NR
						ABC	PT	14	65	0.64	NA	0.64	4.86	15.84
NCT02169180	Maruyama et al, 2019 [32]	II	SA	ibrutinib	None	MCL	PT	16	72	0.75	22.5	0.938	NR	NR
NCT02268851	Davids et al, 2019 [13]	I/Ib	SA	ibrutinib + umbralisib	None	MCL	PT	15	68	0.76	26	0.67	NA	NR
NCT02329847	Younes et al, 2019 [62]	I/IIa	SA	ibrutinib + nivolumab	None	DLBCL	PT	45	64	0.64	19.6	0.36	2.6	13.5
NCT02401048	Herrera et al, 2020 [20]	Ib/II	SA	ibrutinib + durvalumab	None	DLBCL	PT	34	67	0.62	17.5	0.24	2.6	5.1
						non-GCB	PT	16	67	0.50	17.5	0.38	4.1	7.3
						GCB	PT	16	68	0.75	17.5	0.13	2.9	5.5
NCT02427620	Wang et al, 2022 [50]	II	SA	ibrutinib + rituximab	None	MCL	FL	131	56	0.79	42	0.98	NR	NR
NCT02460276	Jerkeman et al, 2018 [27]	II	SA	ibrutinib + rituximab + lenalidomide	None	MCL	PT	50	69	0.72	17.8	0.76	16	22
NCT02471391	Tam et al, 2018 [43]	II	SA	ibrutinib + venetoclax	None	MCL	PT	24	68	0.88	15.9	0.71	NR	NR
NCT02558816	Gouill et al, 2021 [28]	I/II	SA	ibrutinib + obinutuzumab	None	MCL	PT	9	64	0.67	17	0.89	NR	NR
						MCL	PT	24	66	0.79	17	0.71	NR	NR
						MCL	FL	15	65	0.40	14	0.93	NR	NR
NCT02682641	Gine et al, 2022 [18]	II	SA	ibrutinib + rituximab	None	MCL	FL	50	65	0.66	36	0.84	NR	NR
NCT02736617	Park et al, 2022 [34]	II	SA	ibrutinib + obinutuzumab	None	MCL	PT	9	72	0.90	NA	0.89	14	NA
NCT02756247	Stewart et al, 2022 [41]	I/Ib	SA	ibrutinib + buparlisib	None	DLBCL	PT	13	74	0.50	40.8	0.31	1.6	6.3
						MCL	FL	17	71	0.89	42	0.94	33	NR
NCT02874404	Lunning et al, 2020 [30]	IIa	SA	ibrutinib + umbralisib	None	DLBCL	PT	13	NA	0.69	NA	0.31	8	NA
NCT03112174	Wang et al, 2021 [48]	III	RCT	ibrutinib + venetoclax	ibrutinib + placebo	MCL	PT	21	68	0.62	31	0.81	35	35

SA: Single arm; RCT: Randomized control trial; PT: Previously treated; FL: First-line; NA: Not available; NR: Not reach.

R Codes

```
# Kolmogorov-Smirnov Test
x<-ORRpOS$ORR
ks.test(x, "pnorm", mean=mean(x), sd=sd(x))

# Association Between Endpoints
library(wCorr)
library(weights)
library(psychometric)
library(dplyr)

## DLBCL - ORRmOS
ORRmOS<-mutate(ORRmOS, w.ORR=n/(ORR*(1-ORR)))

### (Weighted) Pearson Correlation
cor.test(ORRmOS$ORR, ORRmOS$mOS, method="pearson",
         alternative="two.sided")
weightedCorr(ORRmOS$ORR, ORRmOS$mOS, weights=ORRmOS$w.ORR,
             method="pearson")
CIr(r=0.9315429, n=8, level = 0.95)

### Adjusted R-squared from (Weighted) Linear Regression
ORRmOS.lm<-lm(mOS ~ ORR, data=ORRmOS)
ORRmOS.wlm<-lm(mOS ~ ORR, weights=w.ORR, data=ORRmOS)
summary(ORRmOS.lm)$adj.r.squared
summary(ORRmOS.wlm)$adj.r.squared

### Non-parametric Bootstrap
B.Data<- dplyr::select(ORRmOS, ORR, mOS, w.ORR)
set.seed(170622)
n<-nrow(B.Data)
index <- c(1:n)
B<-10000
adj.R2<-c(1:B)
for (i in 1:B) {
  index.n <-sample(index,n,replace=TRUE)
  B.Data.n<-B.Data[index.n,]
  fit.lm.n<-lm(mOS~ORR, weights=w.ORR, data=B.Data.n)
  adj.R2[i]<-summary(fit.lm.n)$adj.r.squared
}
quantile(adj.R2, probs = c(0.025,0.975))

## DLBCL - ORRpOS
ORRpOS<-mutate(ORRpOS, w.ORR=n/(ORR*(1-ORR)))

### (Weighted) Pearson Correlation
x = ORRpOS$ORR
y = ORRpOS$pOS3
w = ORRpOS$w.ORR
cor.test(x, y, method="pearson")
weightedCorr(x, y, weights=w, method="pearson")
wtd.cor(x, y, weight=w)
```

```
## DLBCL - mPFSmOS
### (Weighted) Pearson Correlation
cor.test(mPFSmOS$mPFS, mPFSmOS$mOS, method="pearson")
weightedCorr(mPFSmOS$mPFS, mPFSmOS$mOS, weights=mPFSmOS$n,
             method="pearson")
wtd.cor(mPFSmOS$mPFS, mPFSmOS$mOS, weight=mPFSmOS$n)
CIr(r=0.5469176, n=8, level = 0.95)

### Adjusted R-squared from (Weighted) Linear Regression
mPFSmOS.lm<-lm(mOS ~ mPFS, data=mPFSmOS)
mPFSmOS.wlm<-lm(mOS ~ mPFS, weights=n, data=mPFSmOS)
summary(mPFSmOS.lm)$adj.r.squared
summary(mPFSmOS.wlm)$adj.r.squared

### Non-parametric Bootstrap
B.Data<- dplyr::select(mPFSmOS, mPFS, mOS, n)
set.seed(170622)
n<-nrow(B.Data)
index <- c(1:n)
B<-10000
adj.R2<-c(1:B)
for (i in 1:B) {
  index.n <-sample(index,n,replace=TRUE)
  B.Data.n<-B.Data[index.n,]
  fit.lm.n<-lm(mOS~mPFS, weights=n, data=B.Data.n)
  adj.R2[i]<-summary(fit.lm.n)$adj.r.squared
}
quantile(adj.R2, probs = c(0.025,0.975))

## DLBCL - mPFSpOS
x=mPFSpOS$mPFS
y=mPFSpOS$pOS3
w=mPFSpOS$n
cor.test(x, y, method="pearson")
wtd.cor(x, y, weight=w)

## DLBCL - pPFSpOS
x<-pPFSpOS$pPFS9
y<-pPFSpOS$pOS21
w<-pPFSpOS$n
wtd.cor(x, y, weight=w)
cor.test(x, y, method="pearson")

## MCL - ORRpOS
x<-ORRpOS$ORR
y<-ORRpOS$pOS24
w<-ORRpOS$w.ORR
cor.test(x, y, method="pearson")
wtd.cor(x, y, weight=w)

## MCL - pPFSpOS
x<-pPFSpOS$pPFS24
y<-pPFSpOS$pOS3
w<-pPFSpOS$n
```

```
wtd.cor(x, y, weight=w)
cor.test(x, y, method="pearson")

# Meta-analysis of ORR
library(metafor)

## Combination of the Two Disease Types
ies.logit = escalc(xi=ORR, ni=n, measure="PLO", data=orr.dat, slab=Reference)
ies.summary=summary(ies.logit, transf=transf.ilogit, ni=ies.logit$n)

### Subgroup Summary, and Subgroup Analysis
subganal.moderator=rma(yi, vi, data=ies.logit, mods=~gTherapy, method="DL")
pes.logit.IMONO=rma(yi, vi, data=ies.logit, mods=~gTherapy=="IMONO",
                    method="DL")
pes.logit.ICC=rma(yi, vi, data=ies.logit, mods=~gTherapy=="ICC", method="DL")
pes.logit.ICI=rma(yi, vi, data=ies.logit, mods=~gTherapy=="ICI", method="DL")
pes.logit.IBCDA=rma(yi, vi, data=ies.logit, mods=~gTherapy=="IBCDA",
                    method="DL")
pes.logit.IPI3K=rma(yi, vi, data=ies.logit, mods=~gTherapy=="IPI3K",
                    method="DL")

pes.IMONO<-predict(pes.logit.IMONO, transf = transf.ilogit)
pes.ICC<-predict(pes.logit.ICC, transf = transf.ilogit)
pes.ICI<-predict(pes.logit.ICI, transf = transf.ilogit)
pes.IBCDA<-predict(pes.logit.IBCDA, transf = transf.ilogit)
pes.IPI3K<-predict(pes.logit.IPI3K, transf = transf.ilogit)

pes.subg.moderator=predict(subganal.moderator, transf=transf.ilogit)
dat.samevar=data.frame(estimate=c((pes.logit.IMONO$b)[1],
                                  (pes.logit.ICC$b)[1],
                                  (pes.logit.ICI$b)[1],
                                  (pes.logit.IBCDA$b)[1],
                                  (pes.logit.IPI3K$b)[1]),
                       stderror=c((pes.logit.IMONO$se)[1],
                                  (pes.logit.ICC$se)[1],
                                  (pes.logit.ICI$se)[1],
                                  (pes.logit.IBCDA$se)[1],
                                  (pes.logit.IPI3K$se)[1]),
                       tau2=subganal.moderator$tau2)

pes.logit.moderator=rma(estimate, sei=stderror, method="FE", data=dat.samevar)
pes.moderator=predict(pes.logit.moderator, transf=transf.ilogit)

print(pes.subg.moderator)
predict(subganal.moderator)
subganal.moderator=rma(yi, vi, data=ies.logit, mods=~factor(gTherapy)-1,
                       method="DL")
pes.subg.moderator=predict(subganal.moderator, transf=transf.ilogit)
anova.rma(subganal.moderator)
subganal.moderator=rma(yi, vi, data=ies.logit,
                       mods=~relevel(factor(gTherapy), ref="IMONO"),
                       method="DL")

### Contrast Analysis
```

```
subganal.moderator=rma(yi, vi, data=ies.logit, mods=~factor(gTherapy) - 1,
                       method="DL")
library(multcomp)
p.wise <- glht(subganal.moderator,
               linfct=contrMat(c("IBCDA"=1, "ICC"=1, "ICI"=1, "IMONO"=1,
               "IPI3K"=1),type="Tukey"))
summary(p.wise, test=adjusted("holm"))

### Forest Plot
mlabfun <- function(text, res) {
   list(bquote(paste(.(text),
      " (Q = ", .(formatC(res$QE, digits=2, format="f")),
      ", df = ", .(res$k - res$p),
      ", p ", .(metafor:::.pval(res$QEp, digits=2, showeq=TRUE,
                                sep=" ")), "; ",
      I^2, " = ", .(formatC(res$I2, digits=1, format="f")), "%, ",
      tau^2, " = ", .(formatC(res$tau2, digits=2, format="f")), ")")))}

par(cex=1, font=6)
forest(ies.summary$yi,
       ci.lb = ies.summary$ci.lb, ci.ub = ies.summary$ci.ub,
       ylim=c(-8,36),
       xlim=c(-1.5,1.5  ),
       slab=paste(ies.logit$Reference),
       ilab=cbind(data=ies.logit$Disease.Type,
                  ies.logit$ORR, ies.logit$n),
       ilab.xpos=c(-0.5,-0.3, -0.1),
       #ilab.pos=3,
       rows=c(32:29, 25:24, 21:20, 16:6, 2:-1),
       at=c(seq(from=0, to=1, by=0.2)),
       refline=pes$pred,
       main="Meta-Analysis: Rate of Overall Response Rate",
       xlab="Proportion (%)",
       digits=2)

par(cex=1.2, font=7)
text(-1.5, 34.5, pos=4, "Study")
text(c(-0.8, -0.4, -0.2), 34.5, pos=4,
     c("Disease Type", "Cases", "Total"))
text(0.9, 34.5, pos=4, "Proportion [95% CI]")
text(-1.5, c(33, 26, 22, 17, 3), pos=4,
     c("Ibrutinib as a Monotherapy",
       "Ibrutinib + Cytotoxic Chemotherapy",
       "Ibrutinib + Checkpoint Inhibitors",
       "Ibrutinib + B-cell Depleting Antibody",
       "Ibrutinib + PI3K Inhibitors"))

par(cex=1.5, font=7)
addpoly(pes.IMONO[1]$pred,
        ci.lb=pes.IMONO[1]$ci.lb, ci.ub=pes.IMONO[1]$ci.ub,
        row=28, digits=4,
        mlab=mlabfun("RE Model for Subgroup", pes.logit.IMONO))
addpoly(pes.ICC[5]$pred,
        ci.lb=pes.ICC[5]$ci.lb, ci.ub=pes.ICC[5]$ci.ub,
```

```
        row=23, digits=4,
        mlab=mlabfun("RE Model for Subgroup", pes.logit.ICC))
addpoly(pes.ICI[7]$pred,
        ci.lb=pes.ICI[7]$ci.lb, ci.ub=pes.ICI[7]$ci.ub,
        row=19, digits=4,
        mlab=mlabfun("RE Model for Subgroup", pes.logit.ICI))
addpoly(pes.IBCDA[9]$pred,
        ci.lb=pes.IBCDA[9]$ci.lb, ci.ub=pes.IBCDA[9]$ci.ub,
        row=5, digits=4,
        mlab=mlabfun("RE Model for Subgroup", pes.logit.IBCDA))
addpoly(pes.IPI3K[20]$pred,
        ci.lb=pes.IPI3K[20]$ci.lb, ci.ub=pes.IPI3K[20]$ci.ub,
        row=-2, digits=4,
        mlab=mlabfun("RE Model for Subgroup", pes.logit.IPI3K))
addpoly(pes$pred,
        ci.lb=pes$ci.lb, ci.ub=pes$ci.ub,
        row=-6, digits=4,
        mlab=mlabfun("RE Model for All Studies", pes.logit))

text(-1.5, -7, pos=4, cex=0.75, bquote(paste("Test for Subgroup Difference: ",
     Q[M], " = ", .(formatC(subganal.moderator$QM, digits=2, format="f")), ",
     df = ", .(subganal.moderator$p - 1), ", p = ",
     .(formatC(subganal.moderator$QMp, digits=2, format="f")))))

abline(h=-5)

### Funnel Plot
funnel(pes.logit, atransf = transf.ilogit, xlab="Proportion",
       main="Meta-analysis of ORR (DLBCL and MCL)")

### Egger's regression test
regtest(pes.logit, model="rma", predictor = "sei")

#Note: same procedure for DLBCL, MCL, and meta-analysis of adverse events
```