PSI workshop conference

Reweighting randomized controlled trial (RCT) evidence to better reflect real life – a case study of the Innovation in Medicine initiative using patients with non-small cell lung cancer (NSCLC)

Alan J M Brnabic
Principal Research Scientist
Eli Lilly
September 2018

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Eli Lilly.

Acknowledgements

Abrams KR[2]
Belger M[1]
Didden EM[5]
Efthimiou O[4]
Faries D[1]
Girvan A[1]
Happich M[1]
Jonsson P[3]
Johnston J[1]
Kadziola Z[1]
Ruffieux Y[4]
Winfree K[1]

[1]Eli Lilly, Indianapolis, IN, USA,
[2]University of Leicester, Leicester, UK,
[3]National Institute for Health and Care Excellence (NICE), Manchester, UK
[4]Insitute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland
[5]F.Hoffmann-La Roche Ltd, Basel, Switzerland

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no [115303], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.
www.imi.europa.eu

Abstract

The objective of the presentation will be to present a case study that assesses the generalizability of efficacy (overall survival [OS]) from the pivotal RCT (JMDB) comparing pemetrexed with gemicitabine to treat non-squamous non-small cell lung cancer using real-world data from a prospective observational study (FRAME) using a reweighting approach. Both inverse propensity scoring and entropy balancing were used to reweight the RCT data based on the real-world FRAME data in an attempt to mirror routine clinical practice in the trial setting.

Presentation contents

- Background to studies
- Summary of approach
- Findings
- Recommendations from IMI
- Further research/ challenges

Background

- Although the demonstration of improved patient and clinical outcomes within randomized controlled trials (RCTs) is widely accepted as foundational evidence of the efficacy of new treatments, concerns are frequently expressed that RCTs lack external validity
- Health technology assessment (HTA) are frequently questioning the generalisability of results from RCTS to routine clinical practice
- GetReal, a project under the umbrella of the Innovation in Medicine Initiative, has explored how “real-life” clinical data can be brought in earlier in drug development
- We describe a case study that considers lung cancer, the most common cancer worldwide and investigate the generalizability of overall survival

Case Study: Studies Included – RCT - JMDB

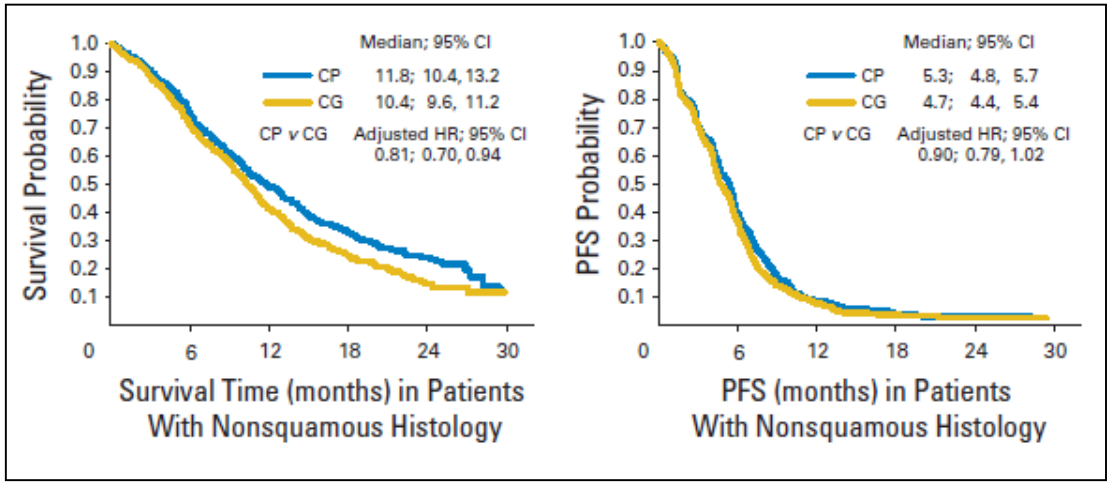
VOLUME 26 · NUMBER 21 · JULY 20 2008

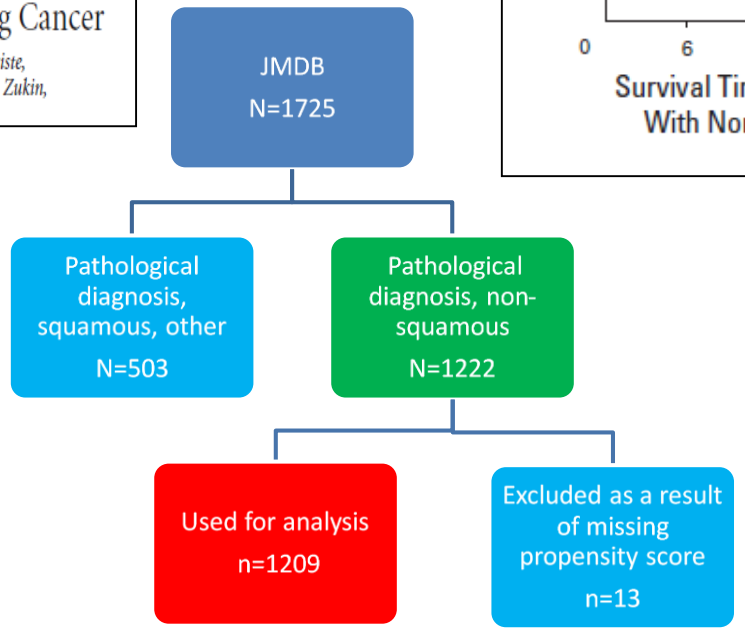
JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Phase III Study Comparing Cisplatin Plus Gemcitabine With Cisplatin Plus Pemetrexed in Chemotherapy-Naive Patients With Advanced-Stage Non–Small-Cell Lung Cancer

Giorgio Vittorio Scagliotti, Purvish Parikh, Joachim von Pawel, Bonne Biesma, Johan Vansteenkiste, Christian Manegold, Piotr Serwatowski, Ulrich Gatzemeier, Raghunadharao Digumarti, Mauro Zukin, Jin S. Lee, Anders Mellemgaard, Keunchil Park, Shehkar Patil, Janusz Rolski, Tuncay Goksel,





Scagliotti et al J Clin Oncol. 2008 Jul 20;26(21):3543-51

Case Study: Studies Included - RWE- observational study FRAME

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Outcomes and resource use of non-small cell lung cancer (NSCLC) patients treated with first-line platinum-based chemotherapy across Europe: FRAME prospective observational study

Denis Moro-Sibilot[a,*], Egbert Smit[b], Javier de Castro Carpeño[c], Krzysztof Lesniewski-Kmak[d], Joachim Aerts[e,f], Rosa Villatoro[g], Kees Kraaij[h], Karim Nacerddine[i], Yulia Dyachkova[j], Karen T. Smith[k], Kaisa Taipale[l], Allicia C. Girvan[k], Carla Visseren-Grul[h], Philipp A. Schnabel[m]

- Only baseline data is needed from the RWD source

Table 1
Summary of patient and disease characteristics.

	Pem + Plat	Gem + Plat	Tax + Plat	Vin + Plat	Total
Any histology NSCLC and any platinum[a,b]	***n* = 569**	***n* = 360**	***n* = 295**	***n* = 300**	***n* = 1564**[c]
Median age (range) in years	62 (33–86)	65 (38–84)	65 (37–87)	64 (34–83)	64 (33–87)
≥70 yrs, %	23.4	34.7	36.3	28.0	29.3
Gender male, %	66.6	78.3	73.9	71.0	71.7
Never smoked[d], %	13.7	7.8	9.8	8.3	10.4
ECOG PS 2–3[e], %	17.9	11.1	22.7	15.0	16.8
Disease stage IV[f], %	85.8	73.6	74.6	67.0	76.7
Non-squamous histology, %	97.2	55.8	64.1	53.3	72.3
Non-squamous NSCLC and any platinum[b]	***n* = 553**	***n* = 201**	***n* = 189**	***n* = 160**	***n* = 1130**[c]
Median age (range) in years	62 (33–86)	66 (38–84)	64 (37–85)	63 (34–81)	63 (33–86)
≥70 yrs, %	23.0	31.8	33.3	25.0	26.5
Gender male, %	66.4	73.1	69.8	64.4	68.0
Never smoked, %	14.1	10.9	12.7	8.8	12.5
ECOG PS 2–3, %	18.1	9.5	21.7	17.5	17.3
Disease stage IV[f], %	85.9	80.6	79.4	69.4	81.0
Non-squamous NSCLC and cisplatin	***n* = 374**	***n* = 107**	***n* = 44**	***n* = 91**	***n* = 633**[c]
Median age (range) in years	61 (33–86)	63 (38–79)	60 (37–76)	59 (38–81)	60 (33–86)
≥70 yrs, %	12.8	14.0	9.1	15.4	13.0
Gender male, %	66.3	72.9	63.6	62.6	66.8
Never smoked, %	13.9	11.2	9.1	13.2	12.8
ECOG PS 2–3, %	13.6	5.6	22.7	13.2	13.1
Disease stage IV[f], %	86.6	80.4	77.3	67.0	81.5

Moro-Sibilot 2015 Lung Cancer. 2015 Dec;90(3):427-32

Case Study: Methods

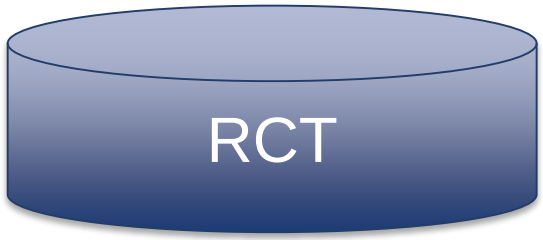
- 2-stage approach
  1. the baseline profile of patients in both studies was compared to determine common variables to be used in the weighting
  2. Weight the RCT outcome of interest (overall survival) based on the list of variables identified in the first step
- Patients were weighted in JMDB using two methods:
  1. the inverse propensity score (IPS)[1]
  2. entropy balancing (EB) method.[2]

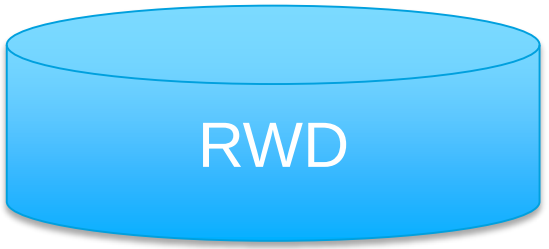
[1] Faries, Douglas, Andrew C. Leon, Josep Maria Haro, and Robert L. Obenchain. 2010. Analysis of Observational Health Care Data Using SAS®. Cary, NC: SAS Institute Inc. Analysis of Observational Health Care Data Using SAS® Copyright © 2010, SAS Institute Inc., Cary, NC, USA

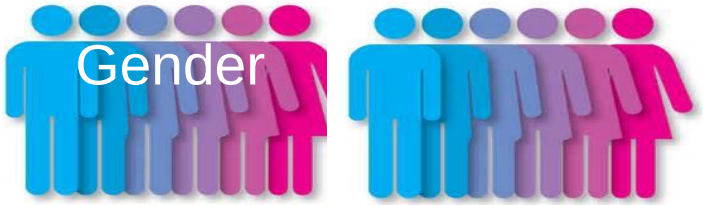
[2] Hainmueller *Political Analysis* 2012 20(1);25-46

Reweights the RCT patients so as to match the RW population

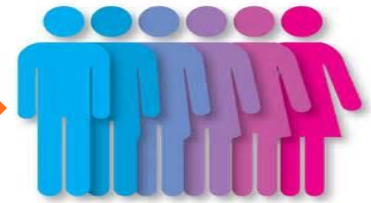
For example if the cohort patients were found to be significantly older than those participating in the RCT, then the method would tend to downweight the younger patients in the RCT.







Weight RCT outcome using an *algorithm*
(IPS or Entropy Balancing)
to *match* with RWD data on selected baseline characteristics



Case Study: IPS

- The IPS method utilizes a logistic model with study membership as the outcome and the covariates identified as independent variables. $Y_i = 1$ if patient i was an RCT patient, $Y_i = 0$ otherwise
- Using the results of this logistic regression, we estimated a propensity score $\hat{p}_i$ for each patient i .
- This propensity score corresponds to the probability of this patient having participated in the RCT rather than in the observational studies, given the set of balancing covariates.
- We then assigned the following weight to patient i in the RCT:

$$w_i = \frac{1 - \hat{p}_i}{\hat{p}_i}$$

- We kept in our analyses only the reweighted RCT patients.

[1] Faries, Douglas, Andrew C. Leon, Josep Maria Haro, and Robert L. Obenchain. 2010. Analysis of Observational Health Care Data Using SAS®. Cary, NC: SAS Institute Inc. Analysis of Observational Health Care Data Using SAS® Copyright © 2010, SAS Institute Inc., Cary, NC, USA

[2] Hainmueller *Political Analysis* 2012 20(1);25-46

Case Study: Methods

- The propensity adjusted balance between the studies (JMDB vs FRAME) and between the weighted treatments cohorts within JMDB were examined using standardized differences
- *Total patient characteristics* in the FRAME study were used to reweight the total non-squamous population in the JMDB trial. *(We applied the reweighing independently of the treatment, thereby preserving treatment randomization)*
- Overall survival was estimated using Cox proportional hazard models and implemented via the weight statement in both PROC PHREG and PROC LIFETEST.
- Because of the uncertainty around the calculated standard errors, 1000 Bootstrap samples with replacement of each of the datasets (using sample size from original datasets) were created
- Weights described above were calculated for each dataset and applied to each patient in JMDB for the analysis using the weight statement.
- A distribution of hazard ratios and associated 95% confidence intervals (CIs) was obtained using the percentile approach (ie, selecting the 2.5 and 97.5 percentiles of the bootstrap distribution).
- Different scenarios for weighting were considered:
  - 1. Standardised weights
  - 2. Trimming of the weights

[1] Faries, Douglas, Andrew C. Leon, Josep Maria Haro, and Robert L. Obenchain. 2010. Analysis of Observational Health Care Data Using SAS®. Cary, NC: SAS Institute Inc. Analysis of Observational Health Care Data Using SAS® Copyright © 2010, SAS Institute Inc., Cary, NC, USA

[2] Hainmueller *Political Analysis* 2012 20(1);25-46

Case Study: Entropy balancing

- Entropy balancing relies on a maximum entropy re-weighting scheme, which matches the empirical moments of the two data sources
- For this study constraints were applied to the first moments only. We reweight the RCT patients so that the covariate moments match those observed in the cohort study defining $x_{RCT,ik}$ (resp. $x_{COH,jk}$) as the value of the k-th covariate for patient i in the RCT (resp. patient j in the cohort study) , and setting N_{RCT} and N_{COH} as the number of patients in the RCT and cohort respectively, the entropy weights $\omega = \{\omega_i, i = 1, \dots, N_{RCT}\}$ satisfy the following optimization problem:

$$\min_{\omega} H(\omega) = \sum_{i=1}^{N_{RCT}} \omega_i \log(N_{RCT}\omega_i)$$

such that

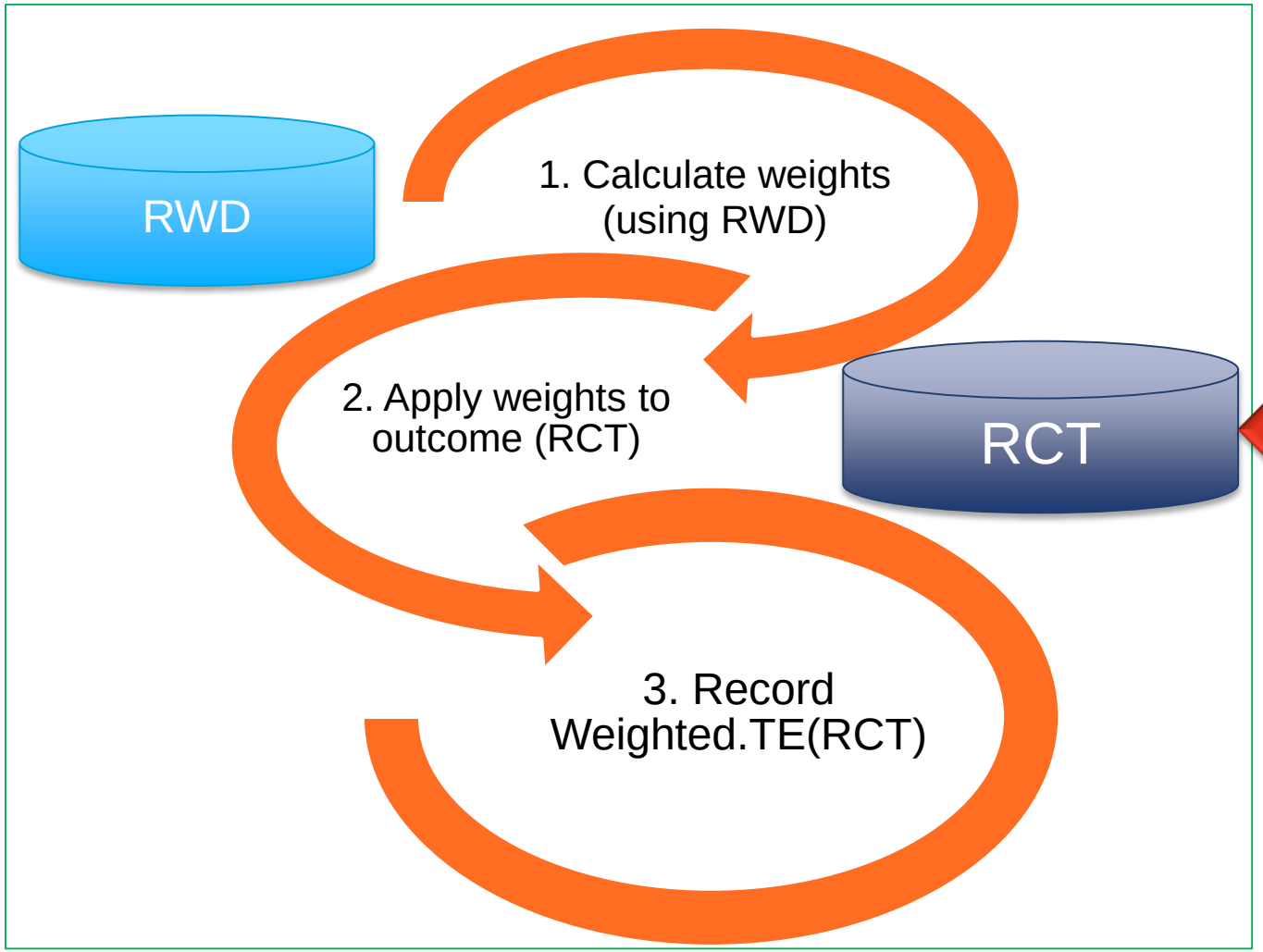
$$\sum_{i=1}^{N_{RCT}} \omega_i \, (x_{RCT,ik})^r = \sum_{j=1}^{N_{COH}} (x_{COH,jk})^r \quad \text{for all } k, \text{ and for } r = 1, \dots, R \,,$$

$$\sum_{i=1}^{N_{RCT}} \omega_i = N_{COH} \,, \qquad \text{and } \omega_i \geq 0 \text{ for all } i = 1, \dots, N_{RCT} \,.$$

- The number R represents the highest order of moments with respect to which we balance the two samples.
- This optimization program above describes a specific case of the entropy balancing scheme.
- For estimating treatment effect and confidence intervals, we replicate the approach for IPS

Hainmueller J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Polit Anal. 2012;20(1):25–46.

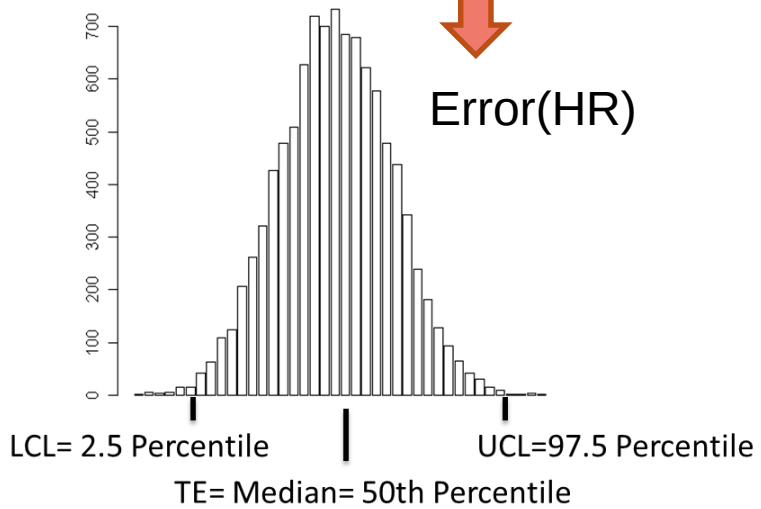
Case Study: Methods



Uncertainty estimated using bootstrapping:

1. Take 1000 samples; $n=n_{RCT}$ from RCT with replacement
2. Repeat the calculation of weights for each sample

Repeat 1000x



Variables including in balance

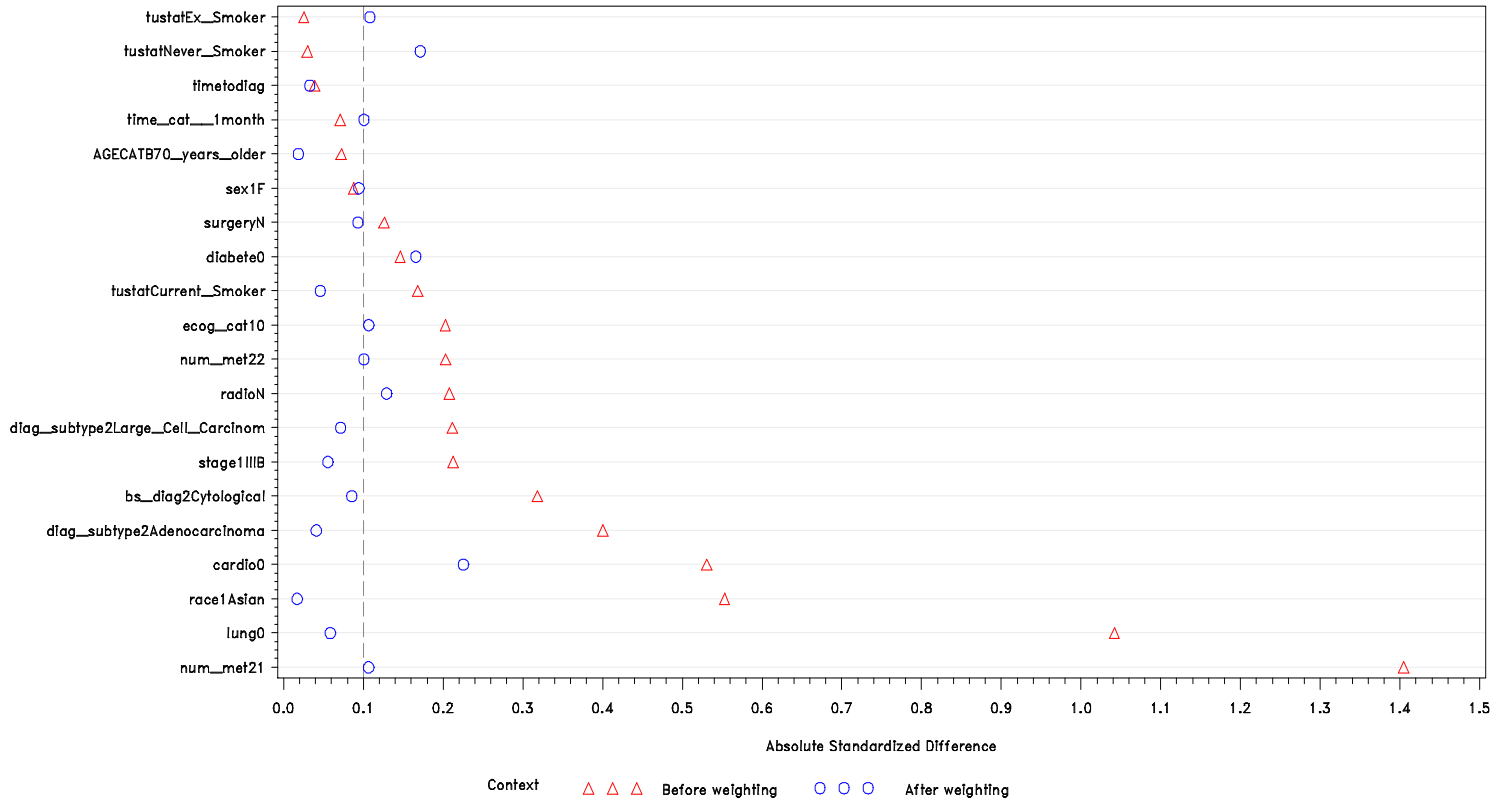
- Variables selected based on availability in both data sets & *some* were considered effect modifiers
  - Age (70 years older vs younger 70 years)
  - Gender (female vs male)
  - Race (Asian vs non-Asian)
  - Smoking status (ex-smoker vs current smoker, never smoker vs current smoker)
  - Basis for diagnosis (histologic or cytological)
  - Time since diagnosis of NSCLC to study entry
  - Diagnosis subtype (adenocarcinoma vs large cell)
  - Stage of disease at study entry (IIIB vs IV)
  - ECOG performance status (0 vs 1)
  - Number of metastatic sites
  - Prior surgery (yes or no)
  - Prior radiotherapy (yes or no)
  - Presence of cardiovascular condition
  - Presence of lung conditions
  - Diabetes

Results: Baseline characteristics for included patients in JMDB (RCT) & FRAME (RWE)

- majority of patients in the FRAME study had 0-1 metastases compared to the JMDB study, which had sicker patients with more than half having 3+ metastases

			FRAME (N=948)	JMDB (N=1209)	p-Value*
Age in years – mean (SD)			62.3 (9.86)	59.7 (9.34)	<.001
Time since diagnosis – month (SD)			2.8 (12.61)	1.9 (7.75)	<.001
Female n (%)			293 (31%)	405 (33.5%)	0.211
Non-Asian n (%)			930 (98%)	997 (83%)	<.001
Smoking status	Current smoker		297 (31%)	277 (23%)	<.001
	Ex-smoker		484 (51%)	585 (48%)	
	Never smoker		121 (13%)	195 (16%)	
	Unknown		46 (5%)	152 (13%)	
Diagnosis n (%)	Cytological		242 (26%)	453 (38%)	<.001
	Histopathological		706 (74%)	756 (62%)	
Diagnosis subtype n (%)	Adenocarcinoma		725 (77%)	861 (71%)	0.005
	Large Cell Carcinoma		77 (8%)	145 (12%)	
	other		146 (15%)	203 (17%)	
Stage n (%)	IIIB		206 (22%)	272 (23%)	0.676
	IV		742 (78%)	937 (77%)	
ECOG n (%)	0		275 (29%)	446 (37%)	<.001
	1		673 (71%)	763 (63%)	
Number of Metastatic Sites n (%)		0-1	771 (81%)	288 (24%)	<.001
		2	157(17%)	296 (24%)	
		3+	20 (2%)	625 (52%)	
Prior radiotherapy n (%)	Yes		111 (12%)	63 (5%)	<.001
Prior Surgery n (%)	Yes		92 (10%)	94 (8%)	0.122
CV History n (%)	Yes		390 (41%)	723 (60%)	<.001
Diabetes n (%)	Yes		107(11%)	78 (7%)	<.001
Lung History n (%)	Yes		117 (19%)	738 (61%)	<.001

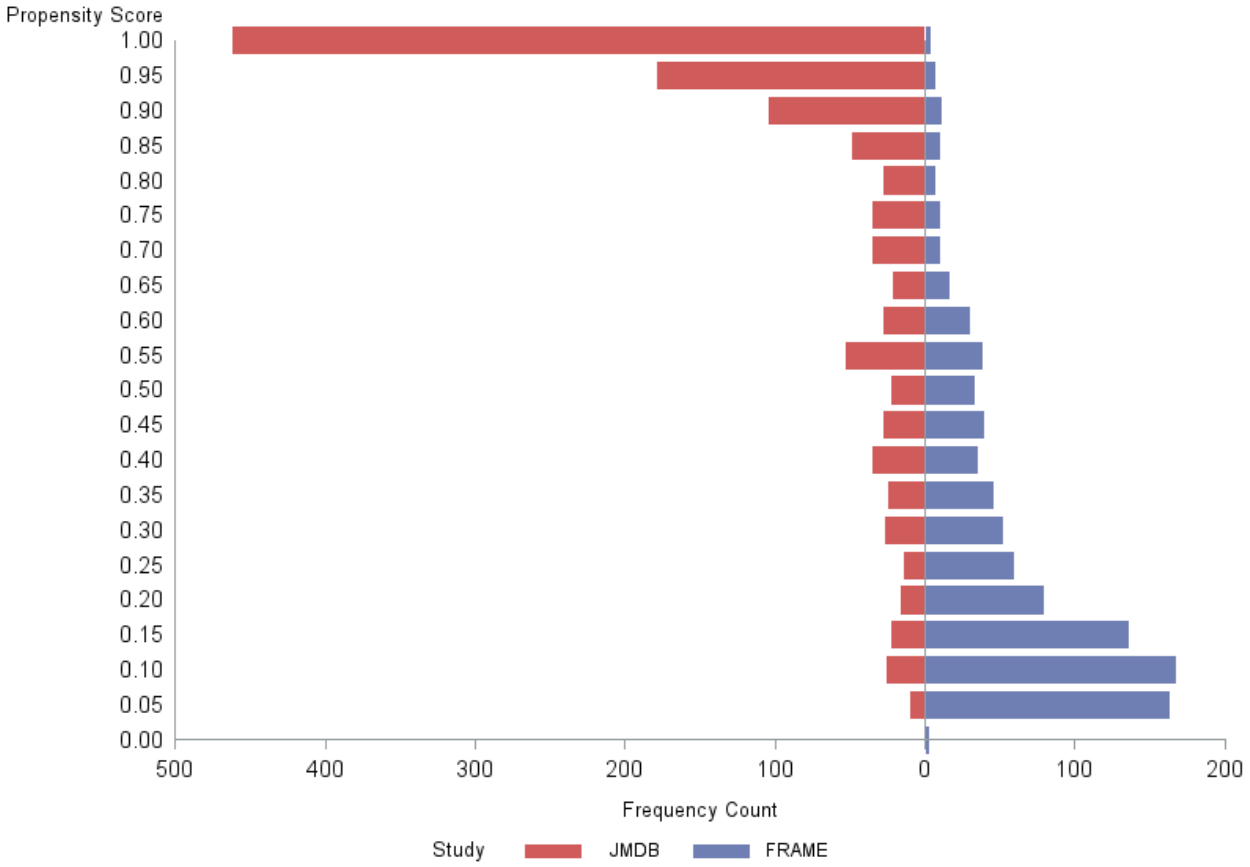
Results: Standardised Difference Plot: JMDB (RCT) & FRAME (RWE)

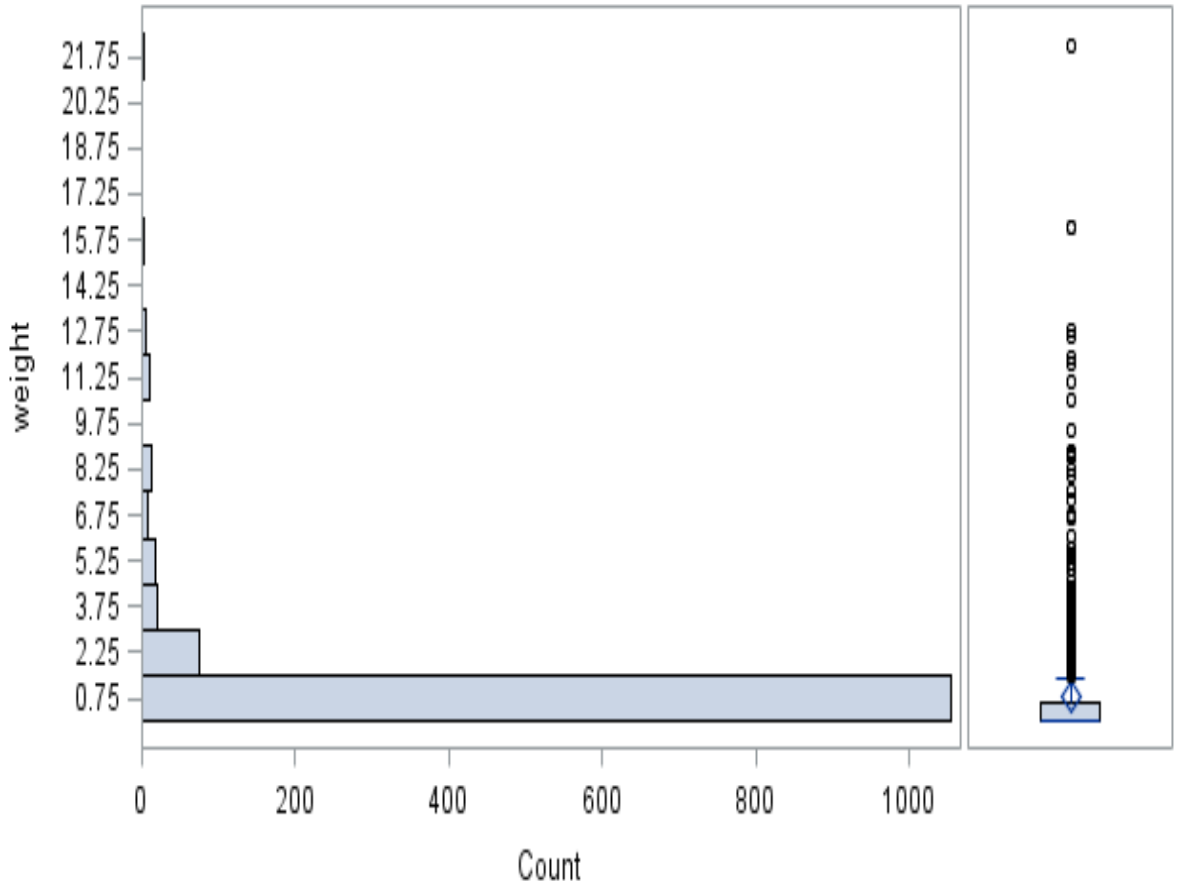


- Standardized difference plot for differences between studies
- JMDB and FRAME – propensity score weighting

Results:Distributions of propensity scores* for JMDB RCT (Red) and FRAME RWE study (Blue)

- While the distributions largely overlap, it is clear that the populations are substantially different – probably driven by the pronounced differences in number of metastases





- This imbalance leads to a highly skewed distribution of weights

*unstandardized and untrimmed weights for primary cohort (generated including all main effects

Results: Propensity weighted* overall survival re-analysis of JMDB RCT using FRAME RWE study for primary cohort

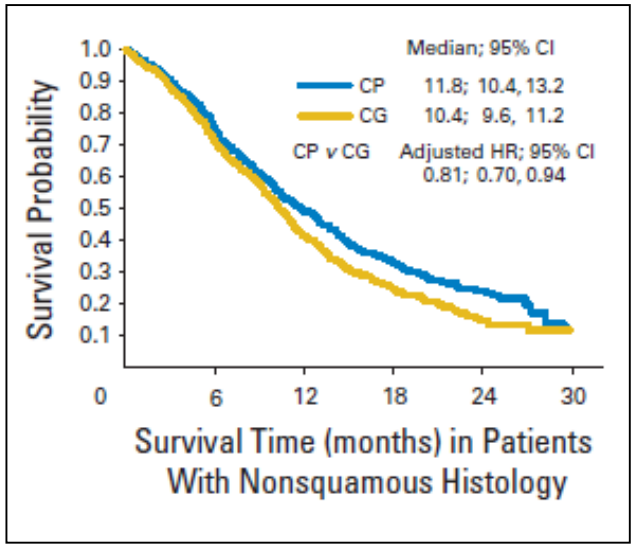
No weighting Unadjusted

Treatment	N	Median Time to OS (months)	Hazard Ratio	95% LCL	95% UCL
No weighting*					
Gemcitabine	608	10.15			
Pemetrexed	614	11.14	0.851	0.746	0.972

Weighted analysis: Unadjusted

Treatment	N	Median time to OS (months)	Hazard Ratio	Bootstrap 2.5 percentile	Bootstrap 97.5 percentile
Propensity score weighting					
Gemcitabine	593	10.15			
Pemetrexed	616	15.57	0.915	0.599	1.333

➔ As reference JMDB** (RCT) results:



*including all main effects - unstandardized untrimmed weights

**differences to Scagliotti et al (2008) are due to slightly different patient population included for analysis

Results: Propensity weighted overall survival re-analysis of JMDB RCT using FRAME RWE study for primary cohort

- ➔ fairly consistent results for HR
- ➔ better balance comes at the expense of higher variability

Weighted analysis: Unadjusted

Treatment	N	Median time to OS (months)	Hazard Ratio	Bootstrap 2.5 perc	Bootstrap 97.5 perc
Propensity score weighting					
Gemcitabine	593	10.15			
Pemetrexed	616	15.57	0.915	0.599	1.333
Propensity score weighting (standardized weights)					
Gemcitabine	593	10.12			
Pemetrexed	616	14.09	0.878	0.705	1.086
Propensity score weighting (excluding number of metastasis)					
Gemcitabine	593	9.26			
Pemetrexed	616	11.24	0.797	0.622	1.20
Entropy balancing weighting					
Gemcitabine	593	10.12			
Pemetrexed	616	14.95	0.904	0.602	1.293
Entropy balancing weighting (excluding number of metastasis)					
Gemcitabine	593	9.30			
Pemetrexed	616	11.24	0.789	0.629	1.004

Limitations

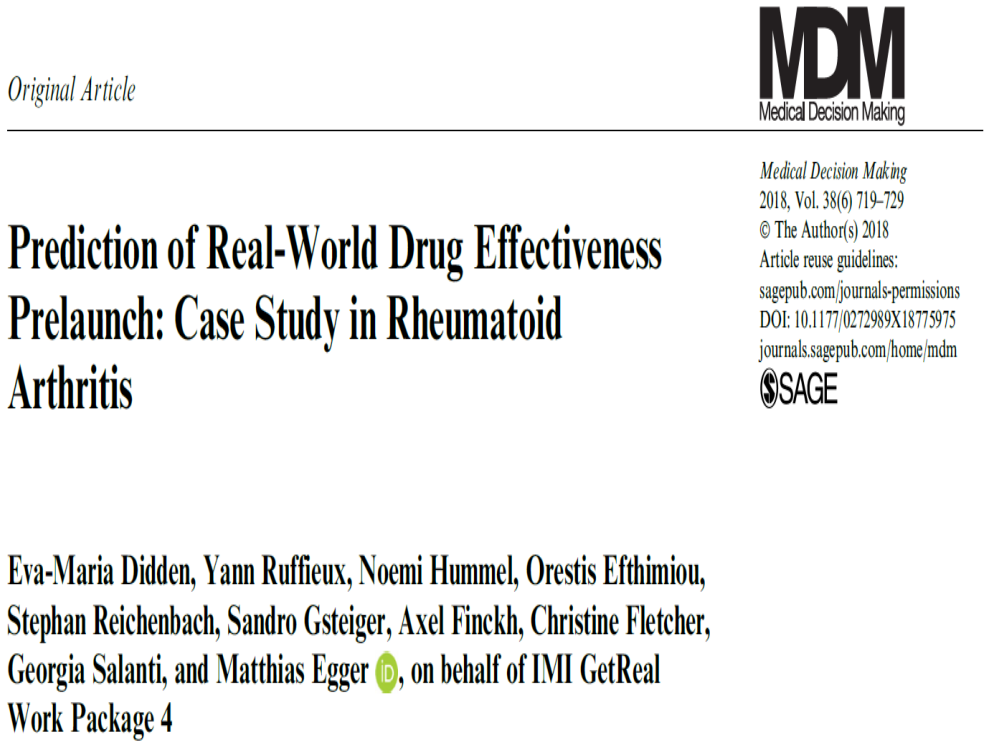
- Definitions of variables can be different between RCT and RWE studies
  - Baseline characteristics
  - Outcome measures
- Unmeasured confounders
- Non-overlapping propensity scores
- Extreme weights
- Specific categories of a variable are not available in RCT

Conclusions

- The key objective of this case study was to assess the generalizability of RCT results for the treatment of non-squamous NSCLC when projected to a real-world population.
- Tested reweighting efforts did not seem to invalidate findings from the original RCT.
- Representatives from Health Technology Assessment and Regulatory authorities as NICE, HAS, CADTH and EMA, as well as renowned academic institutions favourably commented on this approach with regard to impact on health care decision making.
- It can be used to answer the question,

  “what would the trial results look like if they were conducted in a population that is reflective of the local population in which this new treatment is being launched into.”

Future: Applying a Survival Outcome to Didden et al approach



Original Article

MDM Medical Decision Making

Medical Decision Making
2018, Vol. 38(6) 719–729
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X18775975
journals.sagepub.com/home/mdm
SAGE

Prediction of Real-World Drug Effectiveness Prelaunch: Case Study in Rheumatoid Arthritis

Eva-Maria Didden, Yann Ruffieux, Noemi Hummel, Orestis Efthimiou, Stephan Reichenbach, Sandro Gsteiger, Axel Finckh, Christine Fletcher, Georgia Salanti, and Matthias Egger, on behalf of IMI GetReal Work Package 4

AIM: To develop a method to predict drug effectiveness prelaunch and to apply it in a case study in rheumatoid arthritis (RA).

The approach :

1. identifies a market-approved treatment (S) currently used in a target population similar to that of the new drug (N);
2. quantifies the impact of treatment, prognostic factors, and effect modifiers on clinical outcome;
3. determines the characteristics of patients likely to receive N in routine care;
4. predicts treatment outcome in simulated patients with these characteristics.

- Sources of evidence include expert opinion, RCTs, and observational studies.
- The framework relies on generalized linear models.
- **Has not been tested on survival outcomes which presents new challenges**

Challenges with Survival data

- How to specify the ‘correct’ model?
- How to simulate survival time?
- How to simulate censoring?
- What do to in the case of proportional hazards assumption being violated?
- How practical is this approach if you are unable to validate the chosen model in other data sources?

Questions?

Back up

Case Study: NSCLC

Non-small cell lung cancer accounts for 85% of all lung cancer[1]

Non-small cell lung cancer histology subtypes	
Non-squamous (76%)[1]	Squamous (24%)[1]
Major subtype: adenocarcinoma[1,2] Less common subtypes: large cell, sarcomatoid, and adenosquamous carcinoma[3]	Subtypes: basaloid, clear cell type, papillary, small cell non-keratinizing[8]
Cells form recognizable glandular patterns[4]	Characterized by high keratin production[8] Sheets or islands of large polygonal malignant cells and intercellular bridges[8]
Usually develops in lung periphery[2,5]	Generally located centrally and in the larger bronchi[2,8]
Relatively more common in women[1]	Relatively more common in men[1]
Affects smokers and non-smokers[6]	Linked more strongly with smoking[6]
Incidence increasing in recent years[7]	Incidence decreasing in recent years[7]

1. SEER Cancer Statistics Review 1975-2011. National Cancer Statistics. Accessed January 2015
2. Merck Manuals. http://www.merckmanuals.com/professional/pulmonary_disorders/tumors_of_the_lungs/lung_carcinoma.html. Accessed January 2015
3. American Cancer Society. Detailed guide on NSCLC. Available at http://www.cancer.org Accessed February 2014
4. Zugazagoitia J et al. *J Thorac Dis* 2014;6:S526–S536
5. Travis WD et al. *Cancer* 1995;75:191–202
6. Pesch B et al. *Int J Cancer* 2012;131:1210–1219
7. Lortet-Tieulent J et al. *Lung Cancer* 2014;84:13–22
8. Jain D. Pathologyoutlines.com. Accessed January 2015