



Assessment of Successful Randomization Through a Machine Learning and Visualization Tool for Pre-treatment Symptoms: Examples from CCTG/AGITG CO.17 and CO.20 Trials

Danielle Lilly Nicholls^{1,2}, Maria Xu², Lanujan Kanewaran^{1,2}, Benjamin Grant², M. Catherine Brown², Jeremy Shapiro³, Christos S. Karapetis⁴, John Simes⁵, Derek Jonker⁶, Dongsheng Tu⁷, Christopher O'Callaghan⁷, Geoffrey Liu^{1,2,8*}

¹ Temerty Faculty of Medicine, University of Toronto, Toronto M5S 1A8, Canada

² Medical Oncology and Hematology, Princess Margaret Cancer Centre, Toronto M5G 2C1, Canada

³ Department of Medicine, Cabrini Hospital and Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne 3800, Australia

⁴ Flinders Medical Centre and College of Medicine and Public Health, Flinders University, Adelaide 5042, Australia

⁵ NHMRC Clinical Trials Centre, University of Sydney, Sydney 2050, Australia

⁶ Ottawa Hospital Research Institute, University of Ottawa, Ottawa K1Y 4E9, Canada

⁷ Canadian Cancer Trials Group, Queen's University, Kingston K7L 2V5, Canada

⁸ Dalla Lana School of Public Health, University of Toronto, Toronto M5T 3M7, Canada

Corresponding Author Email: Geoffrey.Liu@uhn.ca

<https://doi.org/10.18280/ria.360612>

ABSTRACT

Received: 5 April 2022

Accepted: 31 October 2022

Keywords:

baseline, pre-treatment symptoms, randomized controlled trials, RCT, visualization

Patients in randomized controlled trials (RCTs) must be successfully randomized to reduce or eliminate bias. Because pre-treatment symptoms have prognostic significance in cancer patients, qualitative and quantitative tools were developed to assess similarity of baseline pre-treatment symptoms across different treatment arms of RCTs as one measure of randomization success. Clinician-reported symptom data from two colorectal cancer RCTs, CO.20 and CO.17, were used to demonstrate the utility of a qualitative visualization tool and quantitative machine learning K-means tool, which grouped patients into clusters using baseline symptoms. Qualitatively, reflection bar graphs (RBGs) visualized potential imbalances in baseline symptoms (i) across treatment arms and (ii) by corresponding patient clusters identified within each treatment arm. RBGs found that the treatment arms for both RCTs had similar symptom profiles, while the lack of significant differences in the proportions of patients in each cluster across treatment arms further confirmed successful randomization. This paper details the creation of visualization, machine-learning, and statistical tools to compare baseline symptoms across RCT treatment arms, demonstrating that the CO.20 and CO.17 trials were successfully randomized by baseline symptoms and are comparable. These tools can therefore be implemented easily to ensure an extra layer of quality assurance of the randomization process for study assessment.

1. INTRODUCTION

Randomized controlled trials (RCTs) provide one of the highest levels of evidence for evaluating the treatment efficacy of interventions in clinical research [1-10]. Adequate randomization minimizes selection bias, ensuring that outcomes are truly attributable to the intervention, rather than confounding factors [2-6]. Randomization ensures the formation of comparable and balanced intervention groups, equivalent in all known and unknown variables, except by chance [3-5]. Successful randomization is essential, as baseline imbalances of prognostic factors can significantly compromise the validity of study results [2-5, 7].

Success of patient randomization is often assessed by comparing baseline characteristics across treatment arms, as randomization itself does not always produce groups that are similar in all prognostic factors [2-5, 7, 8]. In practice, a given RCT can have one or more characteristics that are significantly different across treatment arms, especially in smaller sized studies [4, 6].

Assessment of baseline characteristics can identify latent imbalances in important prognostic variables that can be addressed through statistical modeling. Traditionally, only demographic features and a few hand-picked clinical features have been used to assess baseline characteristics between treatment arms [7, 9, 10]. A physician-assigned performance status is often the only feature included to account for patient well-being. However, unless the distributions of all key prognostic features are analyzed, comparison of clinical-demographic characteristics may be insufficient [7].

In cancer patients, baseline symptoms are often prognostic of treatment response and survival [11-20]. Baseline symptoms in cancer patients have shown associations with therapeutic efficacy, toxicity, response, and survival, including disease progression [11-17]. Yet, there are often a myriad of baseline symptoms, which poses a significant analytical challenge. Nonetheless, comparison of baseline symptoms across treatment arms in cancer RCTs can lead to less misattribution and greater understanding of symptom and tolerability of the intervention studied [11, 12, 14, 18, 19].

Methods assessing the impact of baseline symptoms by the success of creating comparable RCT treatment arms are lacking. Further, clinicians typically desire simple, easily understandable methods, rather than overly complicated methods of comparison, given that the clinical-demographic comparisons between RCT treatment arms are presented in manuscripts within a single table [20]. If there are statistical comparisons, then significance testing by Fishers exact test or Chi-squared tests and Student t-tests are typically presented [21].

In contrast, patient symptoms at baseline are collected from multiple patient-reported and clinician-reported tools in most RCTs. The number of variables that describe potential patient symptoms often number in the dozens to hundreds. Analyses of these baseline symptoms have been hampered by this high dimensionality of variables, which can range from a long list by organ, such as nausea, vomiting, stomach upset, and diarrhea for the gastrointestinal system, to muscle pains (myalgias), joint pains (arthralgias) in the musculoskeletal symptoms, among others [22].

To deal with this high dimensionality issue of baseline symptoms, in this paper, visualization and machine learning tools were adapted to compare detailed baseline symptoms between RCT treatment arms. The utility of these tools was demonstrated by validating the adequacy of randomization of baseline symptoms in the Canadian Cancer Trials Group (CCTG) and the Australasian Gastro-Intestinal Trials Group (AGITG) RCT trials, CO.17 and CO.20 [23, 24].

2. METHODOLOGY

2.1 Population

Data was obtained from the CO.17 and CO.20 phase III RCTs [23, 24]. CO.17 was an open-label, multicentre RCT that studied 572 patients with chemotherapy-refractory, incurable colorectal cancer (CRC) from December 2003 to August 2005 [23]. Patients were stratified by Eastern Cooperative Oncology Group (ECOG) performance score and randomized at a 1:1 ratio into cetuximab (CO.17 CET; n=287) or best supportive care (CO.17 BSC; n=285) treatment arms [23]. In CO.20, 750 chemo refractory, incurable CRC patients between February 2008 to February 2011 were stratified by ECOG performance status and randomized in a double-blind, placebo controlled, 1:1 ratio to receive cetuximab plus placebo (CO.20 CET; n=374) or cetuximab and brivanib alaninate (CO.20 CET-BRIV; n=376) [24].

Detailed methods for both RCTs have been published prior [23-25]. Both studies obtained approval from the relevant institutional review boards and written informed consent from the patients.

2.2 Symptom assessment

At baseline, symptom data was collected using the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 2.0 (CO.17) or version 3.0 (CO.20) [23, 24], a widely-recognized standard for collecting clinician-reported symptoms and toxicities [22], grading each symptom/toxicity on a scale from 0-5; Grade 0 indicates absence of symptom and Grade 5 death related to symptom, while grades 1-4 represent increasing severity [22].

2.3 Data preparation

414 NCI-CTCAE graded symptoms experienced at baseline were extracted prior to the treatment start date for both RCTs. Symptom data dimensionality was reduced to 70 symptom categories using clinical knowledge, collapsing symptoms with similar pathophysiology and variance threshold to eliminate symptoms of negligible prevalence and clinical insignificance. When combining similar symptoms, the highest symptom grade was kept.

2.4 Assessing randomization success with respect to baseline symptoms

Reflection bar graphs (RBGs) were created to visualize each symptom prevalence of any grade > 0 across treatment arms at baseline. The graphs reflect across the x-axis, with one treatment arm above the axis, and the other inverted below it.

Smooth reflection indicates similar baseline symptom profiles across treatment arms. This process yields a rapid, visual qualitative assessment of baseline symptoms across RCT treatment arms that flags unusual symptom patterns for quality control purposes, before any quantitative analyses is performed.

For a quantitative assessment of baseline symptoms, a method was needed that further reduced the 70 symptoms into a format that could be displayed easily in tabular format, similar to how baseline clinic-demographical data is presented across RCT arms to show adequacy of randomization. Thus, machine learning was used to cluster symptoms into subgroups: namely, K-Means clustering, a commonly used non-parametric hill-climbing algorithm in the expectation-maximization class, was used to group each individual arm into subpopulations based on baseline symptom prevalence of any grade>0 [26].

Initially, each observation is assigned to a cluster, and optimal clustering is reached by alternating between the expectation phase, where centres of each cluster are computer, and the maximization phase, where each observation is assigned to its nearest cluster, until no further changes occur. K-means has been implemented in various health care settings to successfully identify patient subpopulations with respect to multiple joint variables, which makes it a useful tool for the high dimensionality of baseline toxicity-symptom data present in the CO.20 and CO.17 studies [27].

The Calinski-Harabatz index was used to identify the optimal number of subpopulations, or “clusters”, per treatment arm (2-6 clusters were obsessed, and 2 was chosen based on this index) on the basis of well-defined separation between the clusters; several studies have found that the Calinski-Harabatz index performs the best in identifying the correct number of clusters in pre-defined datasets [28, 29].

Adequately randomized treatment should produce similar proportions of each subpopulation between arms; thus, by using K-means as a quantitative partitioning tool on the CO.20 and CO.17 high dimensionality baseline toxicity-symptom data, these clusters can be produced to assess randomization success.

RBGs were also created again as a means of rapid quality control assessment to compare the different subpopulations identified between and across the treatment arms of each trial. This cross-plot format was developed to visualize smooth reflection of subpopulations with similar baseline symptom profiles across arms over the x-axis, and to compare the

differences of baseline symptom profiles in subpopulations/clusters within the same arm across the y-axis.

2.5 Statistical analysis

The proportions of patients in the K-means partitioning algorithm-based subpopulations for each of the four treatment arms (CO.20 CET-BRIV, CO.20 CET, CO.17 CET, CO.17 BSC) were summarized using counts and percentages in a clinical-demographic table. Fisher's exact test was used to assess the correlations between the K-means generated clusters and the two arms in each trial.

The K-means clustering package was used in R. Visualizations were created using matplotlib in Python. All other data processing and analyses were performed in Microsoft Excel.

3. RESULTS

3.1 Reflection bar graphs by treatment arm

The RBGs created to compare overall baseline symptom profiles across treatment arms within each trial are shown in Figure 1, by depicting the 70 symptoms for each treatment arm flipped over the horizontal axis, the RBGs allows for easy verification of symptom balance across trial arms.

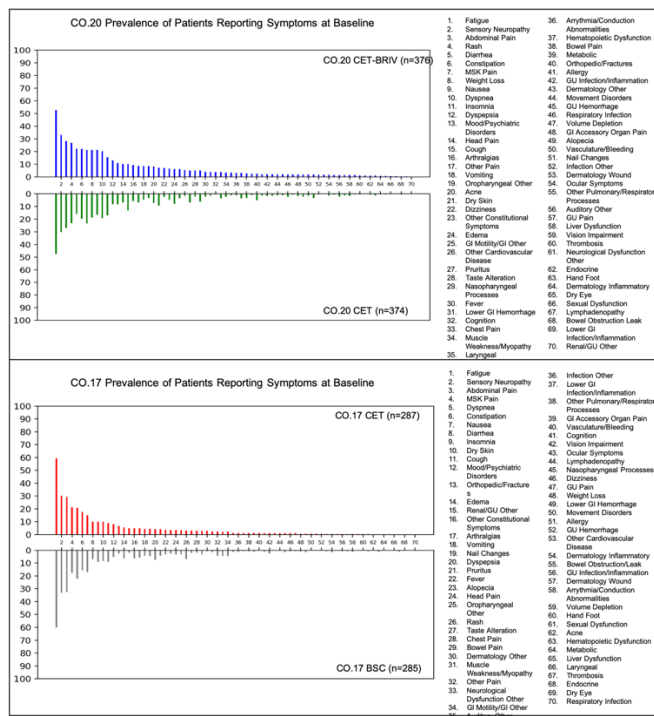


Figure 1. Reflection bar graphs of baseline symptom prevalence of any grade >0. (A) Symptoms are shown in order of most to least prevalent for the CO.20 CET-BRIV (top blue) and CO.17 CET (bottom red) arms respectively. (B) Only the 70 clinically relevant and prevalent symptom categories identified are shown

Both CO.20 and CO.17 RBGs show a smooth reflection across the x-axis, with no major variance between the symptom prevalence experienced at baseline between the two treatment arms. The similar baseline symptom profiles across treatment arms of the same trial support the verification of an

adequate patient randomization. RBGs helped us identify and fix coding errors for the “fatigue” symptom and resolve concerns over baseline differences across CO.17 and CO.20 trials for the baseline “rash” symptom due to initial imbalances between trial arms.

3.2 Subpopulations identified through K-means Clustering

Each treatment arm was clustered into two distinct subpopulations: Cluster A had fewer patients who experienced a higher prevalence of symptoms at baseline, while Cluster B had more patients who experienced a lower prevalence of baseline symptoms. The cross-plot RBGs created to visualize and compare the symptom profiles of the subpopulations are shown in Figure 2, which provides a graphical representation of the symptom profiles of the patients in each cluster to assess cluster reproducibility.

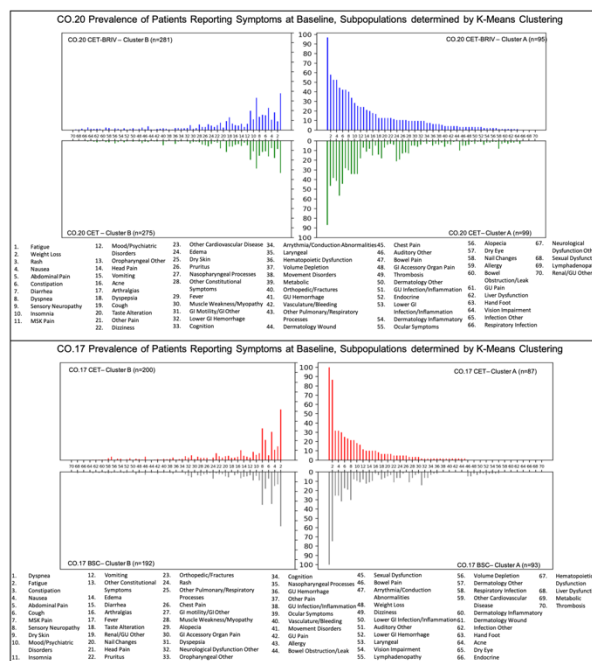


Figure 2. Reflection bar graphs of baseline symptom prevalence of any grade >0 for CO.20 & CO.17 subpopulations. (A) Clustering produced two subpopulations for each of the four treatment arms: a smaller subpopulation with higher symptom prevalence (the right graphs) and a larger subpopulation with lower symptom prevalence (left graphs). (B) Symptoms are ordered from greatest to least prevalence for the CO.20 CET-BRIV (top panel) and CO.17 CET (bottom panel) arms respectively. (C) Bars are sorted according to the symptom prevalence of the top right (Cluster A) sub-panel of each panel

For the most part, there was symmetrical reflection across the x-axis (showing that each arm had corresponding clusters) and poor reflection across the y-axis (showing that different clusters in each arm differed in prevalence of symptoms). Minor differences in the reflection across the x-axis did not change the fact that different clusters clearly corresponded to each other, a critical step that RBG visualization provides effectively. Table 1 further demonstrates that the proportions of each subpopulation are similar across arms, serving as evidence of adequate randomization producing trial arms with patients with similar baseline symptom profiles.

Table 1. Summary statistics of K-means clustering demonstrating proportion of patients in Cluster A & Cluster B for each of the four treatment arms

Cluster ^a	CO.20 trial ^b		CO.17 trial	
	CET-BRIV arm n=376	CET arm n=374	CET arm n=287	BSC arm n=285
Cluster A ^c	95 (25%)	99 (26%)	87 (30%)	93 (33%)
Cluster B	281 (75%)	275 (74%)	200 (70%)	192 (67%)

^aClustering was performed separately for each trial arm separately.

^bFisher's exact tests were non-significant (p=0.71 for CO.20 arms and p=0.44 for CO.17 arms).

^cClusters were assigned alphabetically in ascending order of number of patients in each cluster.

4. DISCUSSION

The purpose of this manuscript is to describe methods that can be used to determine the success of the randomization process in a RCT, based on similarities and differences in the baseline symptoms of each RCT treatment arm. Tools that can generate data that are easily understood by clinicians, in a way that can be incorporated into the current clinician framework for comparing clinical-demographic variables between RCT treatment arms are essential for acceptability of clinical use. Using RBG visualization and patient clustering by baseline symptoms in both the CO.20 and CO.17 trial datasets, reproducible clustering was demonstrated in this paper across treatment arms into similarly sized subpopulations with similar pre-treatment symptom pattern distribution. These results provide evidence that the treatment groups were well-balanced with respect to important symptom-based baseline prognostic factors. Furthermore, the proportions of patients in each cluster can easily be added to the typical RCT table that compares clinical-demographic information by treatment arm, whilst the RBG visualizations can easily be added to the supplementary tables of RCT publications. This simplicity of summarizing complex high dimensionality sets of symptoms in these qualitative and quantitative fashion is a prime example of how to translate standard machine learning approaches into clinical applications.

Hypothesis testing to compare baseline characteristics in RCTs is generally disapproved of, and the preferred method evaluates the prognostic strength of the measured variables and the magnitude of chance imbalances [3, 4, 8, 30]. RBGs allow for simple visualization of the magnitude of imbalances by identifying regions of poor reflection, and data can be summarized quantitatively as proportions of patients in baseline symptom-based patient clusters.

Both the RBGs and clustering algorithms are straightforward to generate and to interpret. Inclusion of baseline symptom comparison through such visualization tools thus adds a layer of assurance to the randomization process by ensuring that no known or unknown prognostic features are neglected. While the primary analyses would remain unchanged in instances where chance imbalances are identified, sensitivity analyses could also then be applied [31].

Of the various scales and indices developed to assess the quality of RCTs, each method asks a series of binary questions to assess aspects of RCT validity including randomization, allocation concealment, baseline characteristics, blinding, co-interventions, compliance, participant withdrawal, and incomplete outcome data [9, 10].

Pre-treatment symptoms have been identified as having a strong prognostic association (i.e., greater disease progression at follow-up, higher risk of hospitalization, poorer progression-free and overall survival) in patients with advanced CRC, including nausea and vomiting, pain, dyspnea, sleep disturbances, fatigue, lack of appetite, depression,

anxiety, diarrhea, and constipation [12-14]. The authors suggest incorporating baseline symptom assessment into these scales and indices because of their strong relevance to patient treatment outcome; ensuring adequate baseline symptom balance across trial arms through successful randomization is then necessary to significantly reduce the risk of bias.

The quantitative K-means method described in this paper is especially important to sufficiently summarize baseline symptom data that oftentimes has high dimensionality due to many recorded symptoms/toxicities. K-means has proven its utility for large-scale data through analysis of Internet text data; it has also been used effectively in oncological research in microarray breast cancer data clustering [32, 33].

While patient-reported symptoms in this instance were recorded through the NCI-CTCAE adverse event grading, Eastern Cooperative Oncology Group (ECOG) performance status and patient quality of life have been noted to have strong prognostic associations as well, suggesting that they can be used with this approach to assess study randomization success [13, 17, 34, 35].

There are several limitations to this approach. Symptom data was recorded using the NCI-CTCAE and was designed to assess toxicities resulting from treatment effects [12]. Its use and efficacy in the reporting of baseline symptoms unrelated to the treatment of interest has not been widely studied.

Furthermore, *patient-reported* symptoms at baseline may be more predictive of survival outcomes and disease progression than *clinician-reported* scales, perhaps due to under-reporting of symptoms by clinicians, thus reducing sensitivity of this approach in detecting different patient clusters [12, 18]. This approach, however, is pragmatic, as it compares across treatment arms, regardless of the many causative factors for patient symptoms being present or absent prior to treatment initiation, including disease burden, comorbidities, prior treatments, and concomitant medications [18]. The current design of the RBGs is best suited to RCTs with two treatment arms and two clusters, though the design could be adapted to accommodate more treatment arms and multiple clusters per arm. Though symptom prevalence was used primarily in this paper, analyses were also performed using mean grade with similar results.

Finally, a simpler future approach may be to stratify by performance status during randomization, since performance status is often correlated with baseline symptoms; nonetheless, the methods that we have developed would still be helpful to verify the success of randomization.

5. CONCLUSIONS

In summary, machine learning and visualization tools were adapted to assess the adequacy of RCT randomization with respect to prognostically significant baseline symptom. These concepts were then applied to two colorectal cancer RCTs,

CO.17 and CO.20. This study also demonstrates how to summarize and interpret such data within a clinical-demographic table.

These tools could be applied in clinical trial reporting particularly the inclusion of patient baseline symptom cluster data in tables of baseline characteristics of RCTs, in situations where there are known prognostic effects of baseline symptoms data, to ensure an extra layer of necessary quality assurance of the randomization process.

ACKNOWLEDGMENT

The authors would like to acknowledge the clinical contributions of Lillian Siu. This work was supported by the Alan Brown Chair in Molecular Genomics, the Lusi Wong Family Fund, and the Posluns Family Fund, all through the Princess Margaret Cancer Foundation.

REFERENCES

- [1] He, J., Du, L., Liu, G., Fu, J., He, X., Yu, J., Shang, L. (2011). Quality assessment of reporting of randomization, allocation concealment, and blinding in traditional Chinese medicine RCTs: A review of 3159 RCTs identified from 260 systematic reviews. *Trials*, 12(1): 1-9. <http://dx.doi.org/10.1186/1745-6215-12-122>
- [2] Spieth, P.M., Kubasch, A.S., Penzlin, A.I., Illigen, B.M.W., Barlind, K., Siepmann, T. (2016). Randomized controlled trials—A matter of design. *Neuropsychiatric Disease and Treatment*, 12: 1341-1349. <http://doi.org/10.2147/NDT.S101938>
- [3] Bolzern, J.E., Mitchell, A., Torgerson, D.J. (2019). Baseline testing in cluster randomised controlled trials: should this be done? *BMC Medical Research Methodology*, 19(1): 1-5. <http://dx.doi.org/10.1186/s12874-019-0750-8>
- [4] Schulz, K.F., Chalmers, I., Altman, D.G., Grimes, D.A., Doré, C.J. (1995). The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. *The Online Journal of Current Clinical Trials*. <http://doi.org/10.1515/9783110907919.2.81>
- [5] Suresh, K.P. (2011). An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences*, 4(1): 8-11. <http://dx.doi.org/10.4103/0974-1208.82352>
- [6] Zabor, E.C., Kaizer, A.M., Hobbs, B.P. (2020). Randomized controlled trials. *Chest*, 158: S79-S87. <http://dx.doi.org/10.1016/j.chest.2020.03.013>
- [7] Chalmers, T.C., Smith Jr, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials*, 2(1): 31-49. [http://dx.doi.org/10.1016/0197-2456\(81\)90056-8](http://dx.doi.org/10.1016/0197-2456(81)90056-8)
- [8] Schulz, K.F., Chalmers, I., Grimes, D.A., Altman, D.G. (1994). Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Jama*, 272(2): 125-128. <http://dx.doi.org/10.1001/jama.272.2.125>
- [9] Berger, V.W., Alpers, S.Y. (2009). A general framework for the evaluation of clinical trial quality. *Reviews on Recent Clinical Trials*, 4: 79-88. <http://dx.doi.org/10.2174/157488709788186021>
- [10] Chung, J.H., Kang, D.H., Jo, J.K., Lee, S.W. (2012). Assessing the quality of randomized controlled trials published in the *Journal of Korean Medical Science* from 1986 to 2011. *Journal of Korean medical science*, 27(9): 973-980. <http://dx.doi.org/10.3346/jkms.2012.27.9.973>
- [11] Mendoza, T.R., Kehl, K.L., Bamidele, O., Williams, L.A., Shi, Q., Cleeland, C.S., Simon, G. (2019). Assessment of baseline symptom burden in treatment-naïve patients with lung cancer: an observational study. *Supportive Care in Cancer*, 27(9): 3439-3447. <http://dx.doi.org/10.1007/s00520-018-4632-0>
- [12] Ooki, A., Morita, S., Iwamoto, S., Hara, H., Tanioka, H., Satake, H., ... & Yamaguchi, K. (2020). Patient-reported symptom burden as a prognostic factor in treatment with first-line cetuximab plus chemotherapy for unresectable metastatic colorectal cancer: Results of Phase II QUACK trial. *Cancer Medicine*, 9(5): 1779-1789. <http://dx.doi.org/10.1002/cam4.2826>
- [13] Maisey, N.R., Norman, A., Watson, M., Allen, M.J., Hill, M.E., Cunningham, D. (2002). Baseline quality of life predicts survival in patients with advanced colorectal cancer. *European Journal of Cancer*, 38(10): 1351-1357. [http://dx.doi.org/10.1016/S0959-8049\(02\)00098-9](http://dx.doi.org/10.1016/S0959-8049(02)00098-9)
- [14] van Seventer, E.E., Fish, M.G., Fosbenner, K., Kanter, K., Mojtahed, A., Allen, J.N. et al. (2021). Associations of baseline patient-reported outcomes with treatment outcomes in advanced gastrointestinal cancer. *Cancer*, 127(4): 619-627. <http://dx.doi.org/10.1002/cncr.33315>
- [15] Batra, A., Yang, L., Boyne, D.J., Harper, A., Cheung, W.Y., Cuthbert, C.A. (2021). Associations between baseline symptom burden as assessed by patient-reported outcomes and overall survival of patients with metastatic cancer. *Supportive Care in Cancer*, 29(3): 1423-1431. <http://dx.doi.org/10.1007/s00520-020-05623-6>
- [16] Luoma, M.L., Hakamies-Blomqvist, L., Sjöström, J., Pluzanska, A., Ottoson, S., Mouridsen, H., Bengtsson, N.O., Bergh, J., Malmström, P., Valvere, V., Tennvall, L., Blomqvist, C. (2003). Prognostic value of quality of life scores for time to progression (TTP) and overall survival time (OS) in advanced breast cancer. *European Journal of Cancer*, 39(10): 1370-1376. [http://dx.doi.org/10.1016/S0959-8049\(02\)00775-X](http://dx.doi.org/10.1016/S0959-8049(02)00775-X)
- [17] Braun, D.P., Gupta, D., Grutsch, J.F., Staren, E.D. (2011). Can changes in health related quality of life scores predict survival in stages III and IV colorectal cancer?. *Health and Quality of Life Outcomes*, 9(1): 1-8. <http://dx.doi.org/10.1186/1477-7525-9-62>
- [18] Atkinson, T.M., Dueck, A.C., Satele, D.V., Thanarajasingam, G., Lafky, J.M., Sloan, J.A., Basch, E. (2020). Clinician vs patient reporting of baseline and postbaseline symptoms for adverse event assessment in cancer clinical trials. *JAMA oncology*, 6(3): 437-439. <http://dx.doi.org/10.1001/jamaoncol.2019.5566>
- [19] Mierzynska, J., Piccinin, C., Pe, M., Martinelli, F., Gotay, C., Coens, C., ... & Bottomley, A. (2019). Prognostic value of patient-reported outcomes from international randomised clinical trials on cancer: a systematic review. *The Lancet Oncology*, 20(12): e685-e698. [http://dx.doi.org/10.1016/S1470-2045\(19\)30656-4](http://dx.doi.org/10.1016/S1470-2045(19)30656-4)
- [20] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence.

- BMC Medicine, 17(1): 1-9.
<https://doi.org/10.1186/s12916-019-1426-2>
- [21] Du Prel, J.B., Röhrig, B., Hommel, G., Blettner, M. (2010). Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, 107(19): 343-348. <https://doi.org/10.3238/arztebl.2010.0343>
- [22] Trotti, A., Colevas, A.D., Setser, A., Rusch, V., Jaques, D., Budach, V., Rubin, P. (2003). CTCAE v3. 0: development of a comprehensive grading system for the adverse effects of cancer treatment. In *Seminars in Radiation Oncology*, 13: 176-181. [http://dx.doi.org/10.1016/S1053-4296\(03\)00031-6](http://dx.doi.org/10.1016/S1053-4296(03)00031-6)
- [23] Jonker, D.J., O'Callaghan, C.J., Karapetis, C.S., Zalberg, J.R., Tu, D., Au, H.J., Moore, M.J. (2007). Cetuximab for the treatment of colorectal cancer. *New England Journal of Medicine*, 357(20): 2040-2048. <http://dx.doi.org/10.1056/NEJMoa071834>
- [24] Siu, L.L., Shapiro, J.D., Jonker, D.J. et al. (2013). Phase III randomized, placebo-controlled study of cetuximab plus brivanib alaninate versus cetuximab plus placebo in patients with metastatic, chemotherapy-refractory, wild-type K-RAS colorectal carcinoma: The NCIC clinical trials group and AGITG CO.20 Trial. *Journal of Clinical Oncology*, 31: 2477-2484. <http://dx.doi.org/10.1200/JCO.2012.46.0543>
- [25] Karapetis, C.S., Khambata-Ford, S., Jonker, D.J., O'Callaghan, C.J., Tu, D., Tebbutt, N.C. et al. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *New England Journal of Medicine*, 359(17): 1757-1765. <http://dx.doi.org/10.1056/NEJMoa0804385>
- [26] Genolini, C., Pingault, J.B., Driss, T., Côté, S., Tremblay, R.E., Vitaro, F., Arnaud, C., Falissard, B. (2013). KmL3D: A non-parametric algorithm for clustering joint trajectories. *Computer Methods and Programs in Biomedicine*, 109(1): 104-111. <http://dx.doi.org/10.1016/j.cmpb.2012.08.016>
- [27] Grant, R.W., McCloskey, J., Hatfield, M., Uratsu, C., Ralston, J.D., Bayliss, E., Kennedy, C.J. (2020). Use of latent class analysis and k-means clustering to identify complex patient profiles. *JAMA Network Open*, 3(12): e2029068-e2029068. <http://dx.doi.org/10.1001/jamanetworkopen.2020.29068>
- [28] Shim, Y., Chung, J., Choi, I.C. (2005). A comparison study of cluster validity indices using a nonhierarchical clustering algorithm. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, Vienna, Austria, pp. 199-204. <http://dx.doi.org/10.1109/CIMCA.2005.1631265>
- [29] Milligan, G.W., Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50: 159-179. <http://dx.doi.org/10.1007/BF02294245>
- [30] Dettori, J. (2010). The random allocation process: two things you need to know. *Evidence-based Spine-care Journal*, 1(3): 7-9. <http://dx.doi.org/10.1055/s-0030-1267062>
- [31] Coskinas, X., Schou, I.M., Simes, J., Martin, A. (2021). Reacting to prognostic covariate imbalance in randomised controlled trials. *Contemporary Clinical Trials*, 110: 106544. <http://dx.doi.org/10.1016/j.cct.2021.106544>
- [32] Wang, H., Zhou, C., Li, L. (2019). Design and application of a text clustering algorithm based on parallelized K-Means clustering. *Revue d'Intelligence Artificielle*, 33(6): 453-460. <http://dx.doi.org/10.18280/ria.330608>
- [33] Thottathyl, H., Pavan, K.K., Panchadula, R.P. (2020). Microarray breast cancer data clustering using map reduce based k-means algorithm. *Revue d'Intelligence Artificielle*, 34(6): 763-769. <http://dx.doi.org/10.18280/ria.340610>
- [34] Chen, R.C., Clark, J.A., Talcott, J.A. (2009). Individualizing quality-of-life outcomes reporting: How localized prostate cancer treatments affect patients with different levels of baseline urinary, bowel, and sexual function. *Journal of Clinical Oncology*, 27: 24. <http://dx.doi.org/10.1200/JCO.2008.18.6486>
- [35] Johansen, J., Boisen, M.K., Mellemegaard, A., Holm, B. (2013). Prognostic value of ECOG performance status in lung cancer assessed by patients and physicians. *Journal of Clinical Oncology*, 31: 8103-8103. http://dx.doi.org/10.1200/jco.2013.31.15_suppl.8103