Overcoming the Straw Man Effect in Oncology: Visualization and Ranking of Chemotherapy Regimens Using an Information Theoretic Approach

abstract

Purpose Despite the plethora of randomized controlled trial (RCT) data, most cancer treatment recommendations are formulated by experts. Alternatively, network meta-analysis (NMA) is one method of analyzing multiple indirect treatment comparisons. However, NMA does not account for mixed end points or temporality. Previously, we described a prototype information theoretical approach for the construction of ranked chemotherapy treatment regimen networks. Here, we propose modifications to overcome an apparent straw man effect, where the most studied regimens were the most negatively valued.

Methods RCTs from two scenarios—upfront treatment of chronic myelogenous leukemia and relapsed/refractory multiple myeloma—were assembled into ranked networks using an automated algorithm based on effect sizes, statistical significance, surrogacy of end points, and time since RCT publication. Vertex and edge color, transparency, and size were used to visually analyze the network. This analysis led to the additional incorporation of value propagation.

Results A total of 18 regimens with 42 connections (chronic myelogenous leukemia) and 28 regimens with 25 connections (relapsed/refractory multiple myeloma) were analyzed. An initial negative correlation between vertex value and size was ameliorated after value propagation, although not eliminated. Updated rankings were in close agreement with published guidelines and NMAs.

Conclusion Straw man effects can distort the comparative efficacy of newer regimens at the expense of older regimens, which are often cheaper or less toxic. Using an automated method, we ameliorated this effect and produced rankings consistent with common practice and published guidelines in two distinct cancer settings. These findings are likely to be generalizable and suggest a new means of ranking efficacy in cancer trials.

Clin Cancer Inform. © 2017 by American Society of Clinical Oncology

INTRODUCTION

Health care data can be highly convoluted, given the significant dimensionality, nonlinearity, and temporality present in most clinical contexts. In oncology, knowledge has been painstakingly built over decades, primarily through carefully designed randomized controlled trials (RCTs). RCT data, which evolve longitudinally over years and usually involve many indirect comparisons, are known to be subject to many potential biases, ones that can be difficult to discern.[1] As a likely result of this complexity, the conventional approach to the ranking and recommendation of cancer treatments studied in RCTs has been expert consensus–driven guidelines (eg, the National Comprehensive Cancer Network [NCCN] guidelines). Alternatively, work by others has

shown that network meta-analysis applied to RCTs can yield powerful insights[2-7]; however, the networks in these studies have been relatively simple, do not allow for mixed end points (eg, overall survival and response rate), and do not account for temporal factors. In complex networks, layout, animation, and visual parameters such as size and color take on increasing importance.[8] For example, visual analytics have been successfully applied to temporal associations of laboratory results, phenotype relationship networks, and patterns of publication by biomedical specialty and primary degree.[9-11] Visual analysis of networked RCT data may help uncover previously underappreciated biases.

In previous work, we described a prototype approach for the automated construction of a ranked

Jeremy L. Warner

Peter C. Yang

Gil Alterovitz

Author affiliations and support information (if applicable) appear at the end of this article.

Corresponding author: Jeremy L. Warner, MD, MS, Vanderbilt University, 2220 Pierce Ave Preston Research Building 777, Nashville, TN 37232; e-mail: jeremy.warner@vanderbilt.edu.

chemotherapy treatment regimen network using information-theoretical techniques, which were applied to the first-line treatment of chronic myelogenous leukemia (CML-1).[12] Here, we demonstrate how extension of the approach through additional information theoretical measures help overcome the apparent presence of a straw man phenomenon. The straw man effect is a bias that causes new studies to appear more promising because they are compared with regimens that are comparatively ineffective.[1] Although this bias has been described, the degree to which it pervades clinical trial design is unknown. The objective of this paper is to present a new algorithm, built on prior foundations,[12] as well as to visually analyze this putative straw man phenomenon in the CML-1 scenario and a second scenario, the treatment of relapsed/refractory multiple myeloma (RRMM).

METHODS

Context-Specific Regimen Identification

The RCTs that were previously identified in the context of CML-1 were also used in this study, along with several newly published RCTs. Briefly, RCTs were identified through a PubMed query and by hand searches of the literature and published guidelines. There were 27 RCTs identified between 1968 and 2016, with 18 distinct regimens, representing 10,282 patients studied (Data Supplement). To identify RCTs for the context of RRMM, we used a combination of an established knowledge base of chemotherapy regimens, HemOnc.org,[13] along with RCTs identified by two recent network meta-analyses in this setting.[6,7] This yielded a total of 25 RCTs published between 2004 and 2016 containing outcome information for 28 distinct regimens, representing 9,737 patients studied (Data Supplement).

Algorithm Modifications

The previous valuation algorithm,[12] which was used for ranking as well as for coloration of vertices, was revised to include strength of evidence, effect sizes, aging effects, propagation, and refresh as explained in the following paragraphs.

Efficacy measure. For all trials, we selected the trial-defined primary end point, as described in the published manuscript, as the main efficacy measure for the valuation algorithm. For trials with more than one predefined primary end point, we used the least-surrogate end point. If the primary end point was met, we used less-surrogate secondary end points in the algorithm if they had marginal or better statistical significance (ie, $P \leq .10$).[14]

Conversely, if the primary end point was not met but secondary end points were met, we still used the primary end point for the valuation algorithm. We assigned a relative value (RV) as follows: 1.0 for strong, 0.8 for intermediate, and 0.7 for weak end points (Table 1; Equation 1). To determine the stability of the rankings, we varied RV by ±5%, ±10%, and ±20% in a sensitivity analysis.

Strength of evidence. In our pilot work,[12] we used a simple win-lose-draw framework with win and lose defined as a superior or inferior finding with a P value $\leq .05$, and draw defined as statistical nonsignificance or formal noninferiority.[15] Here, we introduce a weighted entropy measure: the negative logarithm of the P value. Because very small P values are difficult to interpret, this coefficient is allowed to take a maximum value of 3 (ie, P values $<.001$ were truncated to .001).

Effect size. We replaced the win-lose-draw framework with a coefficient representing the effect size reported in the trial. For time-based outcomes (eg, overall survival), we ideally used the hazard ratio (HR) as the effect size.[16] When HR was not reported, we defined the effect size either as the ratio of the median survival times or as the point estimate reported as significant in the publication (eg, the 3-year event-free survival). For nontemporal measures (eg, response rate), we used the calculated odds ratio as the effect size. In all cases, the effect size > 1 was transformed into a coefficient E, which is positive for the winning side and negative for the losing side (eg, if a publication reports HR = 0.5, $E = 2$ for the winning side and $E = -2$ for the losing side).

Aging effects. To incorporate outdating of scientific evidence, we introduced an exponential decay coefficient as a function of the time since publication of trial results; additional details are in the Data Supplement.

Vertex valuation algorithm. After incorporation of strength of evidence, effect size, and aging effects, the empirical vertex valuation formula is as follows:

Table 1. End Points Used in the Examined CML-1 or RRMM Trials, With Relative Value

End Point	Relative Value
Overall survival	1.0
Progression-free survival	0.8
Time to progression	0.8
Overall response rate	0.7
Response rate	0.7

Equation 1

$$\hat{v}_n = \sum\nolimits_{y=1}^{m} -\log_{10}(P_y) \times RV_y \\ \times E_y \times \log_{10}(N_y) \times f(\alpha_y)$$

where for the nth vertex v, there are m incident edges, E is the effect size coefficient of the yth edge, N is the total number of patients in each pairwise comparison, P is the P value of the yth outcome, and $f(\alpha_y)$ is the aging coefficient described in the previous paragraph. A positively valued vertex is considered recommendable, and a negatively valued vertex is considered contraindicated. A vertex with value near zero is considered to have lacking evidence, conflicting results, or poor study quality such that there is insufficient evidence on whether to recommend. Although the valuation coefficient is unitless, the magnitude informs the power of the valuation and, thus, it is not normalized.

Propagation and refresh. To overcome the apparent straw man effect (discussed in Results), we investigated the introduction of indirect evidence propagation. In our pilot work, we did not assign any node valuation on the basis of indirect evidence, such that the calculated network is akin to a single-layer perceptron (aka, pairwise network analysis). We augmented this model with information propagation, which has been studied in the context of social networks.[17-19] Specifically, we allow nodes that were calculated to lose value as a result of newly introduced evidence to pass some of their value loss to regimens to which they had previously been superior (ie, single-generation value propagation). Conversely, nodes that were calculated to gain value as a result of newly introduced evidence pass some of their value gain to regimens to which they had previously been inferior. For example, in CML-1, dasatinib was demonstrated to be superior to imatinib[20]; however, imatinib had been shown 7 years earlier to be superior to interferon α and low-dose cytarabine (IFNA/LoDAC).[21] Therefore, a portion of the value loss assigned to the imatinib node is propagated to the IFNA/LoDAC node. This has the effect of restoring some value to imatinib. When value is propagated under these constraints, we also refresh the age-related devaluing coefficient by one half-life. This has the simultaneous effect of allowing more value to be propagated while also restoring some relevance to the older regimen; this is analogous to decreasing impedance in an electrical circuit, and is one possible solution to the problem of hindsight bias.[22] See the Data Supplement for a more detailed, graphical description.

Treatment Network Visualization

We used multiple visual variables to display the treatment regimen networks: size, color, transparency, and position.[23] See the Data Supplement for details.

Statistical and General Methods

R version 3.4.0 and RStudio version 1.0.143 (https://www.r-project.org/) were used for the calculations. Graphs were created and displayed using the igraph package version 1.0.1[24] (http://igraph.org/r/); coloration was by the RColorBrewer package.[25] Correlations between vertex value and size before and after value propagation were calculated using the Pearson product-moment correlation; unadjusted P values $< .05$ were considered statistically significant. Animations of all networks and the R code used to develop them are available upon request.

RESULTS

Visualization of the Treatment Regimen Networks

The resultant networks for CML-1 and RRMM in the most recent year of analysis (2016), after incorporation of evidence strength, effect size, and aging into the valuation algorithm, are shown in Figures 1A and 2A; a complete list of regimens and the number of patients studied for each are shown in the Data Supplement. On initial visualization of the CML-1 regimen network, a few things are immediately evident: (1) there are severe aging effects on regimens 1 through 8, with most of these being valued somewhere near zero; (2) the quality of the outcome measure degrades over time, with the newer regimens almost exclusively evaluated with weak surrogate end points (blue edges); and (3) the largest vertex, regimen 9 (imatinib), is also the lowest ranked. Visualization of the RRMM network reveals that (1) aging effects only seem prominent for regimens 1 through 5; (2) outcome measures are mostly intermediate surrogates (eg, progression-free survival); and (3) the largest and most connected regimens are the lowest ranked. The visually apparent link among connectedness, size, and low valuation on the initial visual inspection led us to suspect that straw man effects were present in both networks, but potentially overstated.

Uncovering and Countering the Straw Man Effect

When the vertices are plotted by vertex value versus size (ie, the total number of patients studied under the regimen), the apparent tendency for large vertices to be negatively valued becomes more evident, as shown in Figure 3. In both
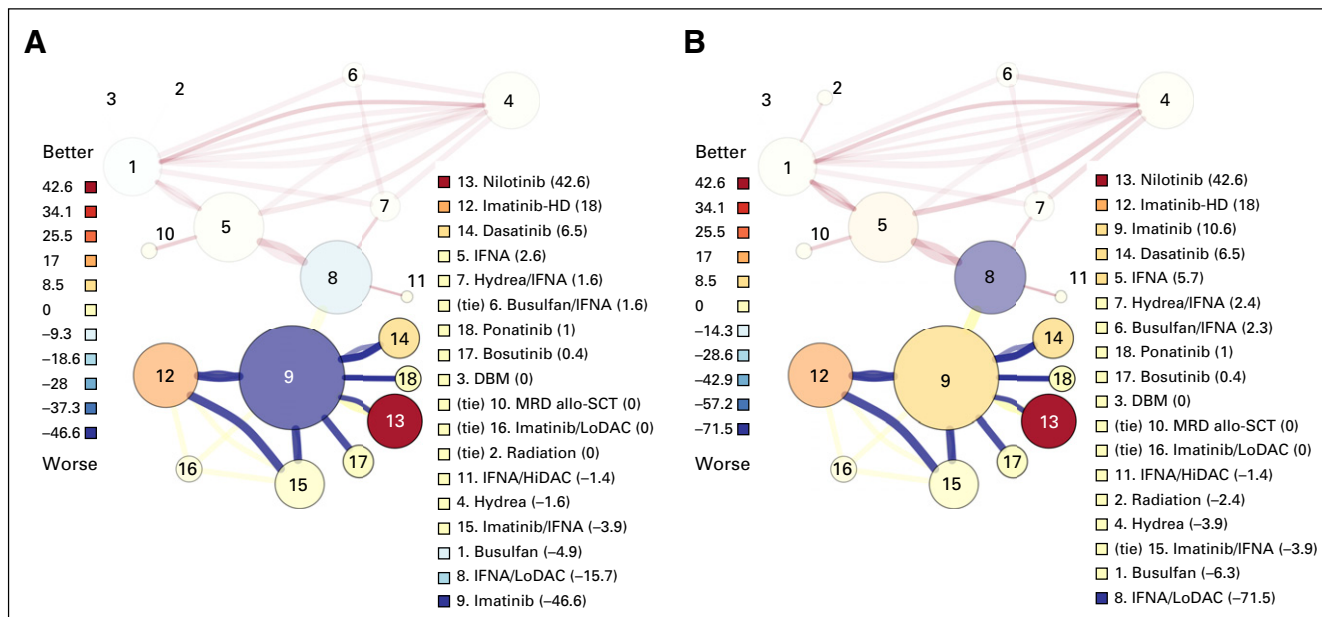
Fig 1. Chemotherapy regimen network for first-line treatment of chronic myelogenous leukemia-1 (CML-1), through 2016. (A) Initial valuations, before application of propagation and refresh. The current standard of care for CML is the use of tyrosine kinase inhibitors (TKIs) in the upfront setting. Consistent with this, TKIs are highly ranked, with the exception of imatinib, which is the lowest-ranked regimen. (B) Applying propagation and refresh to the network changes several valuations, most notably imatinib. DBM, dibromomannitol; HD, high dose; HiDAC, high-dose cytarabine; IFNA, interferon α; LoDAC, low-dose cytarabine; MRD allo-SCT, matched related-donor allogeneic stem-cell transplant.

contexts, this negative correlation was initially statistically significant: For CML-1, the value of r for the correlation of value and size was -0.52 (95% CI, -0.80 to -0.07; $P = .026$). For RRMM, the value of r for the correlation of value and size was -0.65 (95% CI, -0.82 to -0.37; $P = .0002$).

Upon inclusion of propagation and refresh, the valuation of some vertices changes dramatically, as shown in Figures 1B and 2B. In the CML-1 network, imatinib moves from the lowest-ranked regimen to the regimen ranked third highest; IFNA/LoDAC inherits most of the negative value from imatinib and becomes the lowest-ranked regimen. In the RRMM network, almost all aging effects disappear due to refresh, bortezomib and lenalidomide -dexamethasone become more positively valued, dexamethasone (Dex) becomes even more negatively valued, and pomalidomide-dexamethasone (Pom-Dex) moves from the fourth-highest ranked regimen to the second-highest ranked. With this adjustment, the correlation between value and size changes and is no longer significant; for CML-1, the value of r for the correlation between value and size becomes -0.07 (95% CI, -0.52 to 0.41; $P = .78$). For RRMM, the value of r for the correlation of value and size becomes -0.33 (95% CI, -0.62 to 0.05; $P = .09$).
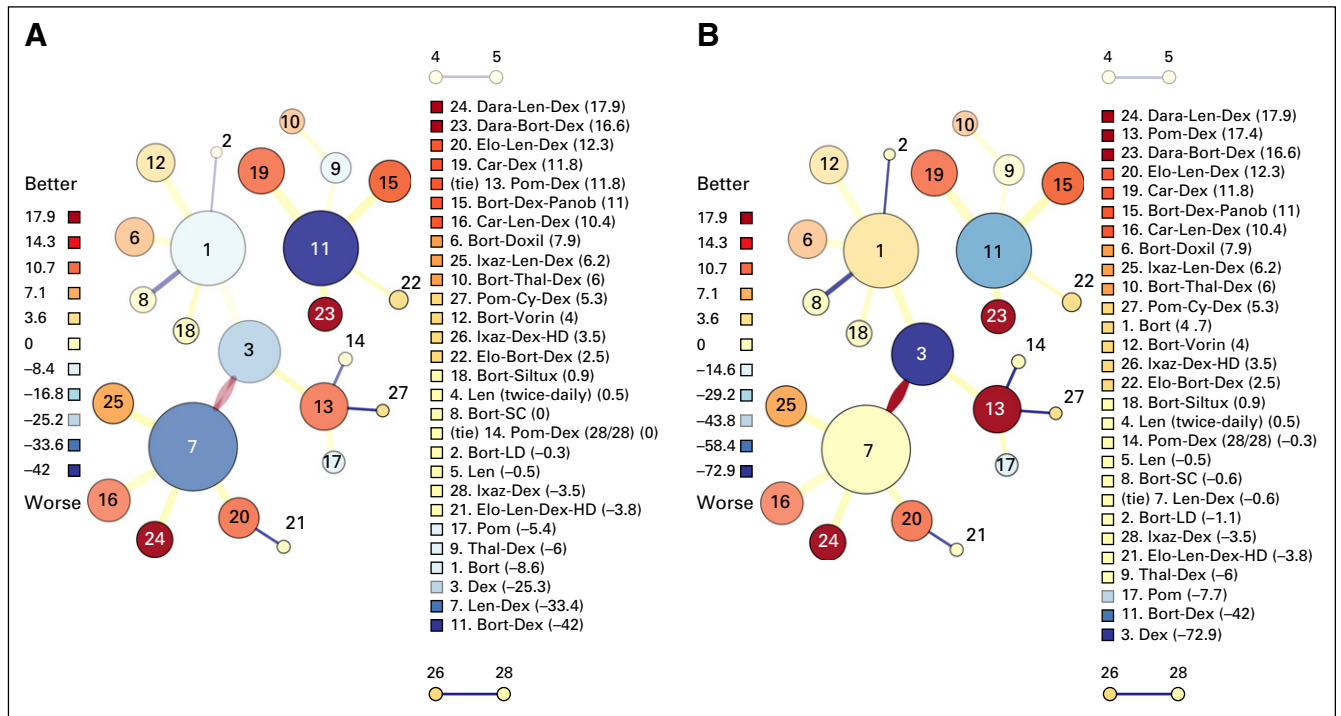
Sensitivity Analysis

With systematic variation in RV, the magnitudes of the vertex values changed slightly, but the rank order did not change for CML-1 or RRMM. Positively valued regimens remained positively valued and vice versa. See the Data Supplement.

DISCUSSION

The interpretation of complex networked data benefits from computational approaches and visualization of the results. In the examples discussed here, multiple visual channels (ie, color, transparency, size, position) provided an integrated picture of context-specific treatment scenarios that evolved over many years (CML-1, 49 years; RRMM, 13 years). We were able to leverage human color perception through the use of a divergent color scale,[26] as compared with the rainbow color map often used in scientific visualizations.[27] The human visual system is particularly well adapted for anomaly detection, owing to enhanced perception of color, edges, and outliers.[28,29] As such, we were able to immediately recognize a potential anomaly in that the largest nodes in the CML-1 and RRMM networks seemed to be both highly connected (ie, compared with many other regimens) and negatively valued. This evidence from visual inspection led to further investigation into a possible straw man effect, which was initially supported by the existence of a statistically significant negative correlation between vertex value and size for both contexts. Through the introduction of propagation and refresh into our algorithm, we were able to ameliorate the straw man effect, although it was not eliminated entirely.

Generally, the straw man effect is most evident when new interventions are compared with clearly inferior regimens.[1,30] A subtler version is the tendency to compare a new regimen with a comparatively effective regimen using a weaker surrogate end point, such as progression-free survival.[31-33] It

A

4 5

24. Dara-Len-Dex (17.9)
23. Dara-Bort-Dex (16.6)
20. Elo-Len-Dex (12.3)
19. Car-Dex (11.8)
(tie) 13. Pom-Dex (11.8)
15. Bort-Dex-Panob (11)
16. Car-Len-Dex (10.4)
6. Bort-Doxil (7.9)
25. Ixaz-Len-Dex (6.2)
10. Bort-Thal-Dex (6)
27. Pom-Cy-Dex (5.3)
12. Bort-Vorin (4)
26. Ixaz-Dex-HD (3.5)
22. Elo-Bort-Dex (2.5)
18. Bort-Siltux (0.9)
4. Len (twice-daily) (0.5)
8. Bort-SC (0)
(tie) 14. Pom-Dex (28/28) (0)
2. Bort-LD (−0.3)
5. Len (−0.5)
28. Ixaz-Dex (−3.5)
21. Elo-Len-Dex-HD (−3.8)
17. Pom (−5.4)
9. Thal-Dex (−6)
1. Bort (−8.6)
3. Dex (−25.3)
7. Len-Dex (−33.4)
11. Bort-Dex (−42)

Better
17.9
14.3
10.7
7.1
3.6
0
−8.4
−16.8
−25.2
−33.6
−42
Worse

26 28

B

4 5

24. Dara-Len-Dex (17.9)
13. Pom-Dex (17.4)
23. Dara-Bort-Dex (16.6)
20. Elo-Len-Dex (12.3)
19. Car-Dex (11.8)
15. Bort-Dex-Panob (11)
16. Car-Len-Dex (10.4)
6. Bort-Doxil (7.9)
25. Ixaz-Len-Dex (6.2)
10. Bort-Thal-Dex (6)
27. Pom-Cy-Dex (5.3)
1. Bort (4 .7)
12. Bort-Vorin (4)
26. Ixaz-Dex-HD (3.5)
22. Elo-Bort-Dex (2.5)
18. Bort-Siltux (0.9)
4. Len (twice-daily) (0.5)
14. Pom-Dex (28/28) (−0.3)
5. Len (−0.5)
8. Bort-SC (−0.6)
(tie) 7. Len-Dex (−0.6)
2. Bort-LD (−1.1)
28. Ixaz-Dex (−3.5)
21. Elo-Len-Dex-HD (−3.8)
9. Thal-Dex (−6)
17. Pom (−7.7)
11. Bort-Dex (−42)
3. Dex (−72.9)

Better
17.9
14.3
10.7
7.1
3.6
0
−14.6
−29.2
−43.8
−58.4
−72.9
Worse

26 28

Fig 2. Chemotherapy regimen network for treatment of relapsed/refractory multiple myeloma, through 2016. (A) Initial valuations, before application of propagation and refresh. The current standard of care is doublet or triplet therapy using an immunomodulatory drug or proteasome inhibitor backbone. This is reflected in the ranking, although Bort-Dex and Len-Dex are the lowest-ranked regimens. (B) Applying propagation and refresh to the network changes several valuations; most notably, Dex is much more negatively valued. Bort, bortezomib; Car, carfilzomib; Cy, cyclophosphamide; Dara, daratumumab; Dex, dexamethasone; Elo, elotuzumab; HD, high dose; Ixaz, ixazomib; LD, low dose; Len, lenalidomide; Panob, panobinostat; Pom, pomalidomide; SC, subcutaneous; Siltux, siltuximab; Thal, thalidomide; Vorin, vorinostat.

has been suggested that pharmaceutical industry support, along with reluctance to sponsor head-to-head comparisons of drugs manufactured by perceived competitors, may exacerbate such effects.[34,35] Two examples where the straw man effect may be operational in cancer domains other than those examined here are dacarbazine in melanoma, where at least nine RCTs demonstrating inferiority have been published between 2000 and 2017[36]; and docetaxel in non–small-cell lung cancer, where at least eight RCTs demonstrating inferiority have been published between 2014 and 2017.[37] The straw man effect is particularly hard to identify directly from the medical literature because the design and execution of RCTs may precede their publication by years. Also, many contexts have the fortunate situation in which prognosis is improving as a result of treatment, including both CML-1[38,39] and RRMM[40]; this will naturally lead to the need to substitute surrogate end points so that new RCTs can be completed within a reasonable time. An intriguing possible way to mitigate the biases of RCT design is the use of a treatment of physician's choice control arm, which was used in the recent (negative) CheckMate 026 trial.[41,42]

Any algorithmic ranking algorithm must be judged on face validity. In the final CML-1 network (Fig 1B), the algorithm ranks nilotinib, imatinib (standard and high dose), and dasatinib the highest, in close concordance with NCCN guidelines.[43] Imatinib, in particular, is the subject of ongoing contention for

many reasons, including (1) all regimens prospectively evaluated against imatinib subsequent to the IRIS (International Randomized Study of Interferon and STI571) trial[21] have either been neutral or superior to imatinib; (2) surrogate end points have been substituted extensively; and (3) imatinib is now generic and cost may be lower than treatments still under patent protection.[44] Given these challenges, it is notable that our updated algorithm still ranks imatinib highly. In the final RRMM network (Fig 2B), the algorithm ranks DRd as the highest and Dex as the lowest ranked regimen, in agreement with the analyses by Botta et al[6] and van Beurden-Tan et al.[7] Interestingly, our algorithm ranks a doublet, Pom-Dex, as second-highest, just behind Dara-Len-Dex. This is primarily due to the refresh of Dex, which allows for substantial negative value to be propagated from Pom-Dex to Dex. The validity of this finding is supported by the fact that Pom-Dex has a category 1 recommendation by the NCCN, although it is noted by the NCCN that triplet regimens are preferred to doublet regimens except in frail or elderly patients.[45] Notably, the only triplet compared with Pom-Dex, pomalidomide-cyclophosphamide-dexamethasone, was in a randomized phase II trial with a weak surrogate primary end point (overall response rate) and borderline significance ($P = .035$).[46] Given that Pom-Dex is relatively well tolerated,[47,48] and that toxicities are a serious concern for modern myeloma drugs,[49] our findings may have real-world significance.

Fig 3. Scatterplots of vertex value versus size for CML-1 and RRMM. Linear regression lines using the least squares approach are overlaid. For all panels, the results show that value is inversely correlated with size, although the correlation becomes nonsignificant once propagation and refresh are instituted. Proportionate size and transparency are preserved from the treatment network visualization; numbering of regimens is omitted to improve clarity. CML-1, chronic myelogenous leukemia; RRMM, relapsed/refractory multiple myeloma.
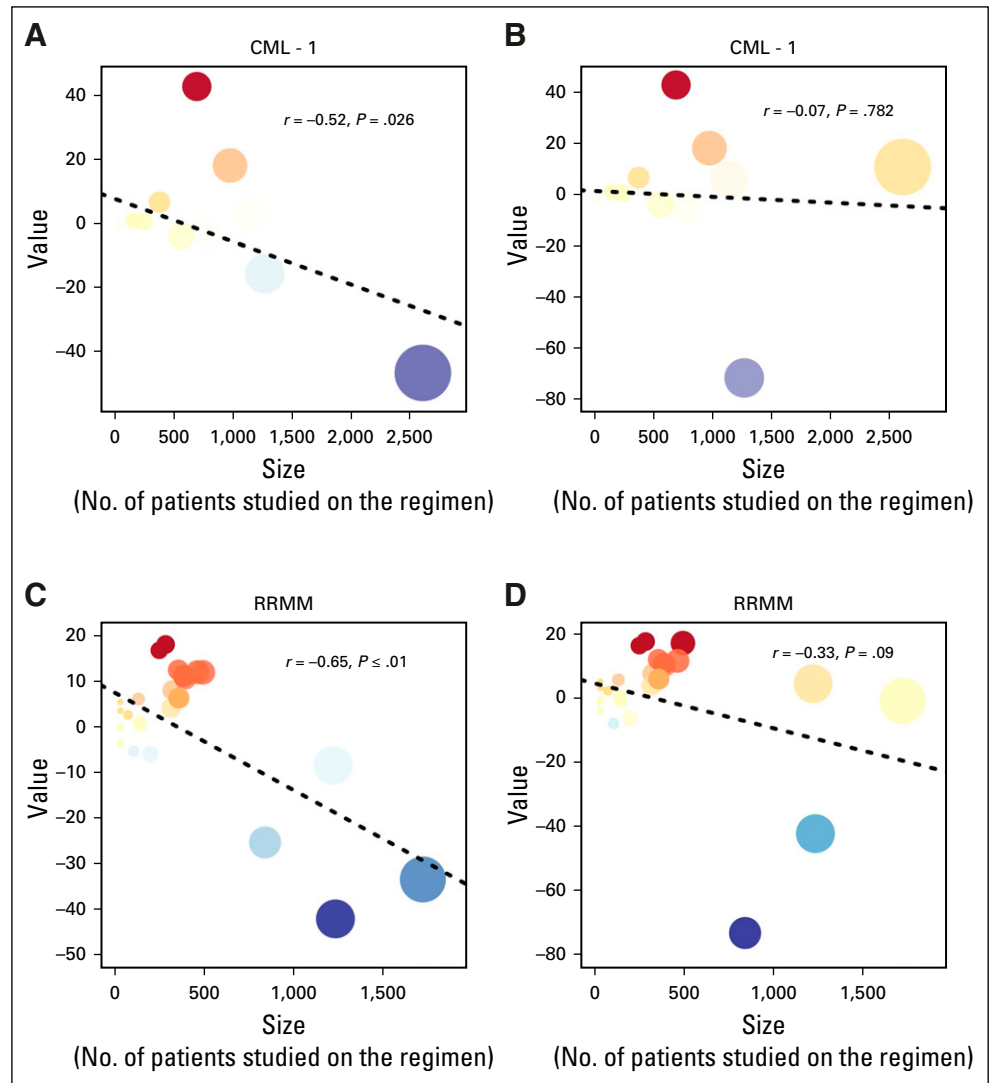
There are several limitations to our current approach. First and foremost, significant methodological challenges remain in the field of dynamic network analysis, especially with respect to in-study and between-study effect modifiers.[50-52] The vertex valuation algorithm contains several empirically derived coefficients and, therefore, could be subject to unmeasured bias. However, it is notable that the ASCO Value Framework has adopted similar weighting metrics to those we use for the surrogacy of end points.[53,54] The ASCO Value Framework and other approaches to comparative valuation, such at the NCCN Evidence Blocks,[55] are also empirically derived. We have used several definitions for effect sizes, including using point estimates when HRs are not reported. Older publications are more likely to report point estimates rather than HRs, so this could introduce a time-based systematic bias.[56] In recognition of this limitation, we support efforts such as SAMPL to encourage uniform reporting of HRs.[57] Current work is focused (as future work will be) on the use of ensemble methods to evaluate how rankings change with perturbation of the weighting coefficients, with a goal of choosing a consensus model that represents best fit.[58] Beyond this methodological limitation, the measured valuations may also be subject to positive publication bias (ie, RCTs that demonstrate a statistically significant result are more likely to be published).[59] We have tried to ameliorate the known tendency for positive publication bias by including so-called gray literature when possible. Future work will investigate incorporating information from clinical trial registries (eg, ClinicalTrials.gov) so that unpublished trials might be incorporated into the model. It is also possible that supplementation with comparative effectiveness data may change the valuations and subsequent rankings. We will also investigate more granular definitions

of time of publication and incorporate evidence updates into the valuation algorithm, as opposed to the current method of using interim updates only to modify the aging coefficient.

In conclusion, we have described how the creation and visual inspection of chemotherapy treatment regimen networks can rapidly lead to new insights. In both of the described use cases, the visual variables of color, transparency, size, and position led to an almost immediate recognition of an anomaly that appeared to be the manifestation of the straw man effect. Additional modifications to the algorithm led to hierarchical rankings with face validity for both scenarios. These findings are likely to be generalizable to any cancer setting with multiple indirect comparisons, and they suggest a new means of ranking efficacy in cancer trials.

AUTHOR CONTRIBUTIONS
Conception and design: Jeremy L. Warner, Gil Alterovitz
Collection and assembly of data: Jeremy L. Warner, Peter C. Yang
Data analysis and interpretation: Jeremy L. Warner, Peter C. Yang
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

Affiliations
Jeremy L. Warner, Vanderbilt University, Nashville, TN; **Peter C. Yang**, Massachusetts General Hospital; and **Gil Alterovitz**, Harvard Medical School and Harvard-Massachusetts Institute of Technology Division of Health Science, Boston; and Massachusetts Institute of Technology, Cambridge, MA.

REFERENCES

1. Ioannidis JP: Perfect study, poor evidence: Interpretation of biases preceding study design. Semin Hematol 45:160-166, 2008

2. Baldwin D, Woods R, Lawson R, et al: Efficacy of drug treatments for generalised anxiety disorder: Systematic review and meta-analysis. BMJ 342:d1199, 2011

3. Liao WC, Chien KL, Lin YL, et al: Adjuvant treatments for resected pancreatic adenocarcinoma: A systematic review and network meta-analysis. Lancet Oncol 14:1095-1103, 2013

4. Elliott WJ, Meyer PM: Incident diabetes in clinical trials of antihypertensive drugs: A network meta-analysis. Lancet 369:201-207, 2007

5. Stettler C, Wandel S, Allemann S, et al: Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis. Lancet 370:937-948, 2007

6. Botta C, Ciliberto D, Rossi M, et al: Network meta-analysis of randomized trials in multiple myeloma: Efficacy and safety in relapsed/refractory patients. Blood Adv 1:455-466, 2017

7. van Beurden-Tan CHY, Franken MG, Blommestein HM, et al: Systematic literature review and network meta-analysis of treatment outcomes in relapsed and/or refractory multiple myeloma. J Clin Oncol 35:1312-1319, 2017

8. Chen C: An information-theoretic view of visual analytics. IEEE Comput Graph Appl 28:18-23, 2008

9. Klimov D, Shahar Y, Taieb-Maimon M: Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med 49:11-31, 2010

10. Hidalgo CA, Blumm N, Barabási AL, et al: A dynamic network approach for the study of human phenotypes. PLOS Comput Biol 5:e1000353, 2009

11. Weber GM: Identifying translational science within the triangle of biomedicine. J Transl Med 11:126, 2013

12. Warner J, Yang P, Alterovitz G: Automated synthesis and visualization of a chemotherapy treatment regimen network. Stud Health Technol Inform 192:62-66, 2013

13. Warner JL, Cowan AJ, Hall AC, et al: HemOnc.org: A collaborative online knowledge platform for oncology professionals. J Oncol Pract 11:e336-e350, 2015

14. Ellis LM, Bernstein DS, Voest EE, et al: American Society of Clinical Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 32:1277-1280, 2014

15. Wangge G, Klungel OH, Roes KC, et al: Interpretation and inference in noninferiority randomized controlled trials in drug research. Clin Pharmacol Ther 88:420-423, 2010

16. Spruance SL, Reid JE, Grace M, et al: Hazard ratio in clinical trials. Antimicrob Agents Chemother 48:2787-2792, 2004

17. Bakshy E, Rosenn I, Marlow C, et al: The role of social networks in information diffusion. Presented at the 21st International Conference on World Wide Web, Lyon, France, April 16-20, 2012.

18. Cha M: Propagation phenomena in large social networks. Presented at the 23rd International Conference on World Wide Web, Seoul, Korea, April 7-11, 2014.

19. Han S, Zhuang FZ, He Q, et al: Energy model for rumor propagation on social networks. PHYSICA A 394:99-109, 2014

20. Kantarjian H, Shah NP, Hochhaus A, et al: Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 362:2260-2270, 2010

21. O'Brien SG, Guilhot F, Larson RA, et al: Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 348:994-1004, 2003

22. Hoffrage U, Hertwig R, Gigerenzer G: Hindsight bias: A by-product of knowledge updating? J Exp Psychol Learn Mem Cogn 26:566-581, 2000

23. Bertin J: La graphique. Communications 15:169-185, 1970

24. Csardi G, Nepusz T: The igraph software package for complex network research. InterJournal Complex Systems 1695, 2006

25. Neuwirth E: RColorBrewer: ColorBrewer palettes, 2011. R Package Version 1.0-5. https://cran.r-project.org/web/packages/RColorBrewer/index.html

26. Moreland K: Diverging color maps for scientific visualization. Presented at the 5th International Symposium on Advances in Visual Computing: Part II, Las Vegas, Nevada, November 30-December 2, 2009.

27. Borland D, Taylor MR II: Rainbow color map (still) considered harmful. IEEE Comput Graph Appl 27:14-17, 2007

28. Marr D, Hildreth E: Theory of edge detection. Proc R Soc Lond B Biol Sci 207:187-217, 1980

29. McCollough C: Color adaptation of edge-detectors in the human visual system. Science 149:1115-1116, 1965

30. Prasad V, Berger VW: Hard-wired bias: How even double-blind, randomized controlled trials can be skewed from the start. Mayo Clin Proc 90:1171-1175, 2015

31. Fleming TR, DeMets DL: Surrogate end points in clinical trials: Are we being misled? Ann Intern Med 125:605-613, 1996

32. Booth CM, Eisenhauer EA: Progression-free survival: Meaningful or simply measurable? J Clin Oncol 30:1030-1033, 2012

33. Prasad V, Kim C, Burotto M, et al: The strength of association between surrogate end points and survival in oncology: A systematic review of trial-level meta-analyses. JAMA Intern Med 175:1389-1398, 2015

34. Peppercorn J, Blood E, Winer E, et al: Association between pharmaceutical involvement and outcomes in breast cancer clinical trials. Cancer 109:1239-1246, 2007

35. Lathyris DN, Patsopoulos NA, Salanti G, et al: Industry sponsorship and selection of comparators in randomized clinical trials. Eur J Clin Invest 40:172-182, 2010

36. Dacarbazine monotherapy. HemOnc.org. https://hemonc.org/wiki/Melanoma#Dacarbazine_monotherapy

37. Carrboplatin & paclitaxel. HemOnc.org. https://hemonc.org/wiki/Non-small_cell_lung_cancer#Carboplatin_.26_Paclitaxel_3

38. Kantarjian H, O'Brien S, Jabbour E, et al: Improved survival in chronic myeloid leukemia since the introduction of imatinib therapy: A single-institution historical experience. Blood 119:1981-1987, 2012

39. Bower H, Björkholm M, Dickman PW, et al: Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. J Clin Oncol 34:2851-2857, 2016

40. Rajkumar SV, Kumar S: Multiple myeloma: Diagnosis and treatment. Mayo Clin Proc 91:101-119, 2016

41. Twelves C, Jove M, Cortes J: Recognizing the place of trials with treatment of physician's choice as the control arm. J Clin Oncol 33:1300-1301, 2015

42. Carbone DP, Reck M, Paz-Ares L, et al: First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. N Engl J Med 376:2415-2426, 2017

43. National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology: Chronic Myelogenous Leukemia, Version 1.2017, 2017. https://www.nccn.org/professionals/physician_gls/pdf/cml.pdf

44. Agrawal M, Garg RJ, Kantarjian H, et al: Chronic myeloid leukemia in the tyrosine kinase inhibitor era: What is the "best" therapy? Curr Oncol Rep 12:302-313, 2010

45. National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology: Multiple Myeloma Version 3.2017, 2017. https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf

46. Baz RC, Martin TG III, Lin H-Y, et al: Randomized multicenter phase 2 study of pomalidomide, cyclophosphamide, and dexamethasone in relapsed refractory myeloma. Blood 127:2561-2568, 2016

47. Dimopoulos MA, Palumbo A, Corradini P, et al: Safety and efficacy of pomalidomide plus low-dose dexamethasone in STRATUS (MM-010): A phase 3b study in refractory multiple myeloma. Blood 128:497-503, 2016

48. Conticello C, Parisi M, Romano A, et al: Pomalidomide experience: An effective therapeutic approach with immu-nomodulatory drugs in a patient with relapsed-refractory multiple myeloma. Future Oncol 13(5s):3-6, 2017

49. Delforge M, Ludwig H: How I manage the toxicities of myeloma drugs. Blood 129:2359-2367, 2017

50. Li T, Puhan MA, Vedula SS, et al: Network meta-analysis-highly attractive but more methodological research is needed. BMC Med 9:79, 2011

51. Mills EJ, Ioannidis JP, Thorlund K, et al: How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 308:1246-1253, 2012

52. Jansen JP, Naci H: Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med 11:159, 2013

53. Schnipper LE, Davidson NE, Wollins DS, et al: American Society of Clinical Oncology statement: A conceptual framework to assess the value of cancer treatment options. J Clin Oncol 33:2563-2577, 2015

54. Schnipper LE, Davidson NE, Wollins DS, et al: Updating the American Society of Clinical Oncology Value Framework: Revisions and reflections in response to comments received. J Clin Oncol 34:2925-2934, 2016.

55. National Comprehensive Cancer Network: NCCN clinical practice guidelines in oncology with NCCN evidence blocks. https://www.nccn.org/evidenceblocks/.

56. Tierney JF, Stewart LA, Ghersi D, et al: Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 8:16, 2007

57. Equator Network: SAMPL guidelines for statistical reporting. http://www.equator-network.org/2013/02/11/sampl-guidelines-for-statistical-reporting/

58. Dietterich TG: Ensemble methods in machine learning, in Kittler J, Roli, F (eds.): Multiple Classifier Systems. First International Workshop, MCS 2000 Cagliari, Italy, June 21-23, 2000 Proceedings. Berlin, Germany, Springer 2000: 1-15.

59. Chan AW, Hróbjartsson A, Haahr MT, et al: Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. JAMA 291:2457-2465, 2004