Received: 12 July 2019 | Revised: 5 September 2019 | Accepted: 14 September 2019

DOI: 10.1002/rth2.12267

COMMENTARY

rpth research & practice in thrombosis & haemostasis

A method to determine if more than surrogate outcomes were improved: The EMR glitch experiment

Jennifer Gill MS[1] | Vinay Prasad MD, MPH[1,2]

[1]Division of Hematology Oncology/Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon

[2]Department of Public Health and Preventive Medicine, Oregon Health & Science University, Portland, Oregon

Correspondence
Vinay Prasad, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239.
Email: prasad@ohsu.edu

1 | INTRODUCTION

Surrogate endpoints offer both benefits and limitations in medicine.[1–7] To strengthen inferences from surrogates, interventions are often additionally assessed for effects on clinical endpoints. Yet clinical outcomes may not always offer concrete proof of efficacy. For instance, in thrombotic thrombocytopenic purpura (TTP), a drug that increases the platelet level (surrogate) may also result in fewer days of plasmapheresis (clinical endpoint). In this case, improvement in the clinical endpoint of plasmapheresis may be indirectly influenced by knowledge of changes in the surrogate endpoint, that is, platelet levels. Providers may cease plasmapheresis simply upon seeing normalized platelet levels, and thus plasmapheresis may be shorter, but this might occur in the absence of any true change in disease biology. When data supporting a treatment are based on surrogate endpoints and clinical endpoints that are contingent on knowledge of those surrogates, we cannot definitively determine whether the intervention is effective. In this commentary, we wish to separate clinical endpoints that may be mediated by surrogate outcomes from those that are independent of it. We propose that clinical endpoints potentially effected by provider interpretation of surrogate endpoints do not offer stronger inference about interventions compared to surrogate endpoints alone.

How do we know which clinical endpoints are mediated by provider interpretation of surrogate outcomes and which are not? A simple thought experiment can clarify. When randomized controlled studies reach positive results, imagine what would occur if we substitute the experimental drug with an electronic medical record (EMR) glitch that gives the appearance of improved surrogate outcomes in patient charts (with no treatment differences compared to control). Imagine what the effects of spurious EMR records (the glitch) might be on the clinical outcomes investigated in the original study. If outcomes could be similar in the randomized trial and the EMR glitch thought experiment, it is likely that clinical endpoints would not offer independent evidence of benefit. We offer a few examples.

2 | CAPLACIZUMAB

The HERCULES trial compared the effects of caplacizumab to placebo on normalization of platelet count in patients with TTP.[8] TTP is a blood disorder in which autoantibodies block the von Willebrand factor–cleaving protease ADAMTS-13, causing platelets to adhere to uncleaved von Willebrand factor multimers in excess. This leads to thrombocytopenia, hemolytic anemia, and eventually multiorgan dysfunction, thromboembolic events, and death.[9] Plasmapheresis is used to remove autoantibodies from the blood and replenish ADAMTS-13 enzymes. Caplacizumab was developed to block platelets from adhering to von Willebrand factor multimers and thereby prevent microvascular thrombosis. In the HERCULES trial, the caplacizumab group had significantly shorter median time to platelet count normalization (the surrogate endpoint) compared to placebo (2.69 days vs. 2.88 days). Days of plasmapheresis and days of hospitalization (the clinical endpoints) were also shorter in the caplacizumab group, and TTP recurrence during treatment was lower.

Normalization of platelet count (>150 000 platelets/mm^3) is a common response criterion for patients with TTP, and one may equate quicker normalized platelet count to less thromboembolic events and death. Yet there are some issues with this assumption. First, caplacizumab does not change the underlying biology of the disease (eg, it does not remove or attack autoantibodies); it prevents platelet adherence. Second, platelet count has not been verified as

an effective surrogate for death or thromboembolic events in this setting. Instead of powering their study to determine caplacizumab's effect on outcomes such as death, the HERCULES investigators powered the study for effects on platelets.

Let's apply our thought experiment. If we were to use standard of care as the control (like HERCULES) and replace caplacizumab with an EMR glitch that adds 40 000 platelets/mm^3 to each patient chart, what would the outcomes look like? Providers would likely report improved time to normalization in the experimental arm, as every patient's records mistakenly received a boost in platelet counts. Due to this misinformation, providers would also likely prescribe less plasmapheresis and discharge these patients sooner. Endpoints such as thromboembolic events or mortality would likely not differ significantly between groups, as there was no biological intervention given to EMR glitch group and these outcomes are not influenced by (incorrect) numbers on a chart. Moreover, the EMR glitch trial, paralleling the HERCULES trial, is not powered to measure those outcomes. Instead, the chosen clinical endpoints were days of plasmapheresis and hospitalization, endpoints that can be influenced by a glitch in the EMR system without any biological changes present. This means that the clinical endpoints investigated in the trial could be mediated by provider interpretation of the surrogate, and provide no independent evidence of benefit. As a contrast, thromboembolic events and mortality are clinical endpoints that would have been uninfluenced by surrogate outcomes or provider interpretation of surrogates in this setting. In the case of HERCULES, the trial was not designed to find a difference in these endpoints.

3 | AVATROMBOPAG AND LUSUTROMBOPAG

Avatrombopag[10] and lusutrombopag[11] are 2 thrombopoeitin receptor agonists whose US Food and Drug Administration approvals were based on trials with platelet-related surrogate outcomes. Both drugs are indicated for patients with thrombocytopenia and chronic liver disease scheduled to undergo procedures. The ADAPT trials[12] and L-PLUS trials[11] investigated avatombopag and lusutrombopag, respectively, with the primary outcome of proportion of platelet transfusions. The trials found that the drugs lowered number of prophylactic platelet transfusions compared to placebo.

To prevent excess bleeding, guidelines recommend platelet transfusions prior to surgery for patients with counts <50 × 10^9/L.[13] Avatrombopag and lusutrombopag work to increase platelets, and it is therefore unsurprising that administering these drugs led to fewer transfusions. Yet patients are given transfusions to prevent major bleeding events, and the aforementioned trials were not powered to investigate this clinically important endpoint. The ADAPT trials did not demonstrate lower incidence of bleeding in the avatrombopag group, a limitation acknowledged in their discussion, and the L-PLUS trials have not yet published full results. The surrogate endpoint of platelet transfusion count is not validated from past trials, and therefore investigating platelet transfusion as a primary outcome may be insufficient in determining efficacy.

Consider our thought experiment: If there were an EMR glitch that increased platelet count by 20 × 10^9/L in records of patients randomized to the experimental group, providers would likely read their charts and prescribe fewer platelet transfusions compared to placebo. Platelet counts would appear to increase (if only on paper) and more patients would head to surgery without prophylactic transfusions. To parallel the ADAPT and L-PLUS trials, we would not power the EMR glitch study to measure bleeding complications, and therefore could not determine if our EMR glitch improved these clinical outcomes, similar to the above studies. Again, all clinical differences in the trial are mediated through the surrogate's influence on provider decision making. Whether or not there was a mistake in patient EMRs, bleeding complications would remain the same, and therefore could act as an independent clinical outcome. On the other hand, rates of platelet transfusions are influenced by chart numbers, even if there is no biological intervention, making it susceptible and dependent on surrogate outcomes.

4 | G-CSF

Granulocyte colony-stimulating factor (G-CSF) was approved as a prophylactic treatment after chemotherapy to decrease the incidence of infection due to febrile neutropenia in patients with nonmyeloid malignancies.[14] Febrile neutropenia puts patients at risk of serious infections and death if conditions persist beyond 10 to 14 days.[15] G-CSF is indicated for chemotherapy regimens associated with a >20% incidence of febrile neutropenia, based on a trial that found that G-CSF significantly decreased febrile neutropenia (temperature ≥38.2°C and absolute neutrophil count [ANC] <1.0 × 10^9/L) in patients with small-cell lung cancer compared to placebo.[16] Overall use of antibiotics and days of hospitalization, secondary endpoints in the study, were also found to be less in the G-CSF group.

Recombinant G-CSF was developed to stimulate proliferation and maturation of neutrophils,[15] thereby lowering rates of neutropenia. While G-CSF drugs are successful in reducing febrile neutropenia caused by chemotherapy, febrile neutropenia is a surrogate endpoint. Without verifying ANC as an appropriate proxy for hard clinical endpoints such as serious infections, overall survival, and quality of life (QOL), the surrogate may be insufficient in determining the clinical efficacy of a drug. Furthermore, low neutrophil count is one component of the multifaceted pathophysiology leading to infections in patients receiving chemotherapy. Chemotherapy affects gut microflora, mucosal integrity, and other aspects of the immune system that defend against infections.[15]

If we were to conduct a thought experiment using the G-CSF trial design, comparing placebo to an EMR glitch that increased records of ANC by 0.5 × 10^9/L, results would likely be comparable. The primary endpoint of febrile neutropenia would be less frequent in the EMR glitch group, as ANC (a surrogate endpoint) would automatically be increased. Due to reported healthy neutrophil levels in the experimental group's charts, practitioners would likely be influenced by the numbers

to reduce days on antibiotics and discharge patients earlier (clinical endpoints). The thought experiment would not be designed to make any claims about serious infection, overall survival, or QOL, similar to the G-CSF study, but because there is no difference in actual treatments, outcomes would likely be similar across groups.

As it turns out, EMR glitches and G-CSF have similar effects on independent clinical endpoints. Various meta-analyses found that G-CSF and granulocyte-macrophage colony-stimulating factor did not improve infection-related death, QOL, or overall survival in patients with nonmyeloid cancers undergoing chemotherapy.[17–19] ANC counts and reported infections were found to be lower, but this did not translate to reduced days of hospitalization or duration of parental antibiotic treatment.[17] Endpoints that may be influenced by ANC count were improved by treatment, but patient-centered, independent outcomes were largely the same compared to placebo or no treatment.

5 | CONCLUSION

This discussion does not question these drugs' abilities to interact upon some part of human biology or imply that a computer glitch is mimicking their effect. Our thought experiment instead suggests that the effect of these agents on the chosen clinical endpoints (days receiving plasmapheresis, transfusions, days on antibiotics) are mediated through the provider's knowledge of the surrogate and do not, in and of themselves, offer independent evidence favoring clinical benefit. We still lack evidence that these drugs improve patient health and well-being.

One must therefore be cautious interpreting provider-controlled clinical endpoints when a surrogate endpoint may affect their outcome. Novel drugs boast tremendous price tags. For instance, caplacizumab is priced at $270 000 for treating a typical TTP episode.[20] To be certain of clinical benefit, we must study the effect of agents on clinical outcomes uninfluenced by provider interpretation of surrogate outcomes. It is wrong to say these drugs improved the surrogate *and* accomplished another goal, but rather they improved the surrogate *and, as a result, could have* yielded the same conclusions.

RELATIONSHIP DISCLOSURE

VP reports receiving royalties from his book *Ending Medical Reversal*; that his work is funded by the Laura and John Arnold Foundation; that he has received honoraria for Grand Rounds/lectures from several universities, medical centers, nonprofit groups, and professional societies; and that he is a writer for Medscape. VP is host of the Plenary Session podcast, which is partially supported by contributions from the Patreon platform. JG has nothing to disclose.

AUTHOR CONTRIBUTIONS

VP contributed to conception of the article. JG and VP both contributed to the design, acquisition, and analysis of content and drafting and revision of the manuscript.

ORCID

Jennifer Gill https://orcid.org/0000-0002-5591-6855
Vinay Prasad https://orcid.org/0000-0002-6110-8221

REFERENCES

1. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989;8:431–40.
2. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. Health Aff (Millwood). 2005;24:67–78.
3. Ball JR, Micheel CM.Evaluation of biomarkers and surrogate endpoints in chronic disease. Washington, DC: National Academies Press; 2010.
4. Yudkin JS, Lipska KJ, Montori V. The idolatry of the surrogate. BMJ. 2011;343:d7995.
5. la Cour JL, Brok J, Gøtzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. BMJ. 2010;341:c3653.
6. Tardif J-C, Heinonen T, Orloff D, Libby P. Vascular biomarkers and surrogates in cardiovascular disease. Circulation. 2006;113:2936–42.
7. Califf RM. Biomarkers, putative surrogates, surrogates, and decision making. Circulation. 2013;6(1):6–7.
8. Scully M, Cataland SR, Peyvandi F, Coppo P, Knöbl P, Kremer Hovinga JA, et al. Caplacizumab treatment for acquired thrombotic thrombocytopenic purpura. N Engl J Med. 2019;380:335–46.
9. Sadler JE. Von Willebrand factor, ADAMTS13, and thrombotic thrombocytopenic purpura. Blood. 2008;112(1):11–8.
10. FDA approves avatrombopag for thrombocytopenia in adults with chronic liver disease www.fda.gov: US Food & Drug Administration; 2018 [updated May 21, 2018]. Available from: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm608323.htm
11. FDA approves lusutrombopag for thrombocytopenia in adults with chronic liver disease www.fda.gov: US Food & Drug Administration; 2018 [updated July 31 2018]. Available from: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm615348.htm
12. Terrault N, Chen Y-C, Izumi N, Kayali Z, Mitrut P, Tak WY, et al. Avatrombopag before procedures reduces need for platelet transfusion in patients with chronic liver disease and thrombocytopenia. Gastroenterology. 2018;155:705–18.
13. Estcourt LJ, Malouf R, Doree C, Trivella M, Hopewell S, Birchall J. Prophylactic platelet transfusions prior to surgery for people with a low platelet count. Cochrane Database Syst Rev. 2017;2017:CD012779.
14. NEUPOGEN® (filgrastim) [package insert]. Thousand Oaks, CA: Amgen Inc.;2015.
15. Bennett CL, Djulbegovic B, Norris LB, Armitage JO. Colony-stimulating factors for febrile neutropenia during cancer therapy. N Engl J Med. 2013;368:1131–9.
16. Crawford J, Ozer H, Stoller R, Johnson D, Lyman G, Tabbara I, et al. Reduction by granulocyte colony-stimulating factor of fever and neutropenia induced by chemotherapy in patients with small-cell lung cancer. N Engl J Med. 1991;325:164–70.
17. Bohlius J, Herbst C, Reiser M, Schwarzer G, Engert A. Granulopoiesis-stimulating factors to prevent adverse effects in the treatment of malignant lymphoma. Cochrane Database Syst Rev. 2008:CD003189.
18. Sung L, Nathan PC, Alibhai S, Tomlinson GA, Beyene J. Meta-analysis: effect of prophylactic hematopoietic colony-stimulating factors on mortality and outcomes of infection. Ann Intern Med. 2007;147:400–11.
19. Clark O, Lyman GH, Castro AA, Clark L, Djulbegovic B. Colony-stimulating factors for chemotherapy-induced febrile neutropenia:

a meta-analysis of randomized controlled trials. J Clin Oncol. 2005;23:4198–214.

20. Sanofi: FDA approves Cablivi® (caplacizumab-yhdp), the first Nanobody®-based medicine, for adults with acquired thrombotic thrombocytopenic purpura (aTTP) globenewswire.com: Globe News Wire; 2019. Available from: https://globenewswire.com/news-release/2019/02/06/1711610/0/en/Sanofi-FDA-approves-Cablivi-caplacizumab-yhdp-the-first-Nanobody-based-medicine-for-adults-with-acquired-thrombotic-thrombocytopenic-purpura-aTTP.html