Abstract
Discussions regarding resident duty-hour restrictions have been ongoing and heated. One influential argument for restrictions has been patient safety. Two trials, FIRST and iCOMPARE, were performed to investigate this relationship with surgical and medicine training, respectively. As the authors are approaching this discussion from a medicine-based perspective, iCOMPARE will serve as the primary basis of our discussion. Results from the iCOMPARE trial comparing flexible (28-hour shifts allowed) to the original 2011 ACGME shift requirements (maximum 16 hours) were recently published in the *New England Journal of Medicine*. This non-inferiority trial used 30-day post-hospitalization mortality as its primary endpoint. Results met qualifications for non-inferiority, and ACGME policy was changed to allow for 28-hour shifts for medicine residents. iCOMPARE results were highly lauded and used as primary justification for extending resident duty hours. Despite this sweeping impact, few have critically evaluated what this study actually adds to the literature. Herein, we argue that serious questions regarding trial design are apparent. Most importantly, the non-inferiority margins chosen were large, and represent an ambiguous marker of resident performance. Additionally, we question the lack of both patient consenting and direct patient-reported or patient-centered outcomes within the hospital stay. As more discussion arises in the medical literature surrounding patient-reported outcomes and shared decision making, we argue that the results of iCOMPARE disregarded the patient perspective or meaningful patient outcomes in an attempt to maintain status quo. Lastly, we discuss how iCOMPARE missed the broader question of actual duty-hour restrictions, and some practical methods already in practice at some programs, which may more directly balance resident work hours with patient care and resident learning.
Keywords
General Medical Education, Resident Work Hours, Resident Duty Hours, Resident Wellness, Medical Education, iCOMPARE

Corresponding author: Zachary H. Hopkins (zhopkinsderm@gmail.com)
Author roles: Hopkins ZH: Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Secrest AM: Supervision, Writing – Review & Editing
Competing interests: No competing interests were disclosed.
Grant information: The author(s) declared that no grants were involved in supporting this work.
Copyright: © 2019 Hopkins ZH and Secrest AM. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hopkins ZH and Secrest AM. iCOMPARE, what value does it add to resident duty-hour discussions? [version 1; peer review: 1 approved with reservations] F1000Research 2019, 8:1856 https://doi.org/10.12688/f1000research.21130.1
First published: 05 Nov 2019, 8:1856 https://doi.org/10.12688/f1000research.21130.1
Mr. C was friendly and rapport came quickly. He ultimately needed thoracentesis and during the ensuing discussion he perfunctorily asked, “How long have you been on your shift?” After replying, “12 hours,” he responded, “I understand the necessity, but I want someone performing their best for this procedure. Can you wait until morning?” Fortunately, his case could be safely deferred until morning. This was neither the first, nor the last, time during my training a patient asked this question.

Recently, the New England Journal of Medicine published the final primary endpoint analysis for iCOMPARE with an accompanying editorial. This trial sought to evaluate the impact of flexible (program directed shift lengths up to 28 hours) versus 16-hour capped shifts on 30-day post-hospitalization patient mortality. As the only trial investigating duty-hour effects on internal medicine patients, its impact is large. Indeed, in the ACGME’s new duty-hour guidelines, results from the iCOMPARE and FIRST (similar study from surgical programs), are cited as a major justification for this change. However, as with any practice changing study, critical appraisal of its results and application is required.

First, iCOMPARE lacked informed consent from either patients or residents. This is concerning with a primary endpoint of patient mortality. One participant institution’s IRB application stated that they were not involved in human research because the residency program itself was the research subject—this claim seems implausible. Patients are both interested in, and concerned by, the length of their doctors’ shifts. Indeed, in one study, a significant number of patients stated they would like to be informed of any clinician who had worked over 12 hours. Patients and residents deserve to be informed of factors affecting their care, particularly in a trial with the primary endpoint involving their 30-day mortality. For residents, this is especially concerning given their vulnerable workforce position. A 2004 lawsuit against the National Resident Matching Program asserted that residents were “forced to participate in a system that ensures they work long hours for low wages.” However, after a strong lobbying effort, national legislation exempted the residency match process from anti-trust laws. In this setting the potential for exploitation is high, and policies should err on the side of fair work conditions.

Third, the 1% non-inferiority margin utilized suggests that 2,646 additional deaths is an acceptable justification for longer shifts. Extrapolating this margin to the U.S. population, nearly 80,000 deaths annually would be considered “non-inferior.” Understandably, limitations in recruiting and study design apply, especially for non-inferiority. However, we also question the appropriateness of a non-inferiority design for this question. A proposed ethical prerequisite for using non-inferiority analysis is that an experimental therapy “must have known advantages such as reduced cost, greater convenience, or fewer side effects to justify the randomization of patients to a therapy with unknown efficacy.” Long-duty hours have been linked to several resident, patient, and community safety concerns. Likewise, patients do not appear to feel comfortable with physicians who worked long hours. With these concerns, we argue that iCOMPARE should have demonstrated superiority of unrestricted shift lengths. Indeed, one oft-used argument for extended shift hours is fewer handoffs and improved continuity. Inherent in the design appears to be a belief of superiority otherwise why put people through it? Furthermore, the hours worked between the two groups were the same, thus it is wholly expected that no difference in endpoint be observed. Why not choose a restricted hour work week, as is done in Europe, for the comparator? Or, if extended lengths of shift are believed to be better for patients, why not use this shift structure as the comparator? Lastly, we suspect superiority in patient clinical endpoints would be the only results acceptable to patients for continuing a training system of which they do not approve.

Finally, the primary endpoint of iCOMPARE is inadequately justified. Residency training should have sufficiently robust supervision that shift length has minimal effect on 30-day mortality. Orders made by an intern pass through supervising residents, attendings, pharmacists, and nurses. While errors can and do get through, post-hospitalization mortality is buffered from the source of errors in question. This is supported by one study showing an increased length of hospital stay and increased number of ICU admissions with long resident work hours, but a non-significant change in hospital readmissions and within-hospital mortality. Likewise, we suspect 30-day mortality is not the only issue for which patients are concerned. Quantifying errors and near-misses made within the hospital as well as the patient experience could more directly evaluate causal links with resident performance. Additionally, while markers of resident performance (testing performance, wakefulness, etc.) were evaluated, we argue that with work hours being equal and minimal implementation of 28 hour shifts occurring between the two groups, no meaningful comparison between these groups can be derived. With no difference in hours worked and a large non-inferiority margin of an ill-defined primary endpoint, we argue that little can be gained in regards to the preceding data.

While the idea of flexible shift structure is sensible as policy, it severely diluted the applicability and generalizability of the results. Currently, policy extends further than data. A program could, in theory, make a majority of its rotations utilize a 28-hour shift structure and this would be within policy. However, the safety of this structure is untested, as this was not directly evaluated in iCOMPARE. Further, even if the results were to be taken at face value, the reciprocity of the results is seemingly ignored. This trial, as a non-inferiority trial, demonstrated that capped shifts, with the changes in care they bring, do not worsen patient mortality. However, what if even these are too long? Perhaps both capped and flexible structures as they currently stand with 80-hour work week restrictions are inferior to a more restricted system? We posit that this is a more pressing and sensible question.

We believe that the results of iCOMPARE are critically lacking—we can and must do better (Table 1). We do not argue in favor of 2011 ACGME requirements nor the new adjustments. Rather, we suggest that iCOMPARE itself was constructed to test an artificial argument, and did so poorly. Thoughtfully tested data is necessary to further duty hour discussions and guide
| Issues | Concerns | Possible Changes |
|--------|----------|------------------|
| Lack of informed consent | Residents were considered the study "subjects." Patients and residents were not required to give consent. This reasoning is flawed. | - In trials investigating patient safety, patients should be informed. Residents should likewise be informed and consented as participants in the trial. Medical students should be made aware of programs participation preMATCH and current residents should be consented and involved in decisions to participate. |
| Culture | "I survived residency living at the hospital, so should everyone else." | - Cultivate a culture supportive of resident concerns and needs. |
| Exploitation | Residents are a cheap labor force. | - These data suggest that restricted duty-hours are likewise non-inferior to flexible/extended hours. |
| Obstructed outcome design | 30-day patient mortality is not ideal for studying resident duty-hours. | - This bias in favor of extended duty-hours. If one wants to argue that extended duty-hours are preferred, with associated physician and patient risks, superiority should be demonstrated. |
| Lack of concern for resident well-being | Previous results from COMPARE resident survey demonstrate resident dissatisfaction with the flexible group. | - Conclusion in favor of flexible group despite no superiority in results and increased concern for resident harm. |

---

Table 1. Summary of concerns and proposed changes.
further policy refinement as well as an openness to consider new paradigms. Let us work to cultivate a group of compassionate, empathetic, and well-trained physicians; armed with not only with technical prowess, but the emotional reserve to respond to and alleviate patient suffering. In contrast to this goal, the iCOMPARE study is sadly a self-fulfilling prophecy—a superficial justification to continue the status quo.

Data availability
No data are associated with this article.

Acknowledgements
We would like to thank Dr. Vinay Prasad for his helpful editorial comments on this manuscript.

References

1. Rosenbaum L, Lamas D: Eyes Wide Open — Examining the Data on Duty-Hour Reform. N Engl J Med. 2019; 380(10): 969–970. Published Abstract | Publisher Full Text
2. Silber JH, Bellini LM, Shea JA, et al.: Patient Safety Outcomes under Flexible and Standard Resident Duty-Hour Rules. N Engl J Med. 2019; 380(10): 905–914. Published Abstract | Publisher Full Text | Free Full Text
3. Mallory K, Brigham T, Nasca T: ACGME’s 2017 Revision of Common Program Requirements. Patient Safety Network. 2017; Published August 2017. Accessed August 12: 2019. Reference Source
4. Shepherd L, Macklin R: Erosion of informed consent in U.S. research. Bioethics. 2019; 33(1): 4–12. Published Abstract | Publisher Full Text
5. Blum AB, Raiszadeh F, Shea S, et al.: US public opinion regarding proposed limits on resident physician work hours. BMC Med. 2010; 8: 33. Published Abstract | Publisher Full Text | Free Full Text
6. Drolet BC, Hyman CH, Ghaderi KF, et al.: Hospitalized Patients' Perceptions of Resident Fatigue, Duty Hours, and Continuity of Care. J Grad Med Educ. 2014; 6(4): 658–63. Published Abstract | Publisher Full Text | Free Full Text
7. Robinson S: Antitrust Lawsuit Over Medical Residency System is Dismissed. The New York Times. 2014; A00016. Reference Source
8. Aberegg SK, Hersh AM, Samore MH: Do non-inferiority trials of reduced intensity therapies show reduced effects? A descriptive analysis. BMJ Open. 2018; 8(5): e019484. Published Abstract | Publisher Full Text | Free Full Text
9. Aberegg SK, Hersh AM, Samore MH: Empirical Consequences of Current Recommendations for the Design and Interpretation of Noninferiority Trials. J Gen Intern Med. 2018; 33(1): 88–96. Published Abstract | Publisher Full Text | Free Full Text
10. Garantini S, Bertiet V: Non-inferiority trials are unethical because they disregard patients' interests. The Lancet. 2007; 370(9600): 1875–7. Published Abstract | Publisher Full Text
11. Prasad V: Non-Inferiority Trials in Medicine: Practice Changing or a Self-Fulfilling Prophecy? J Gen Intern Med. 2018; 33(1): 3–5. PubMed Abstract | Publisher Full Text | Free Full Text
12. Sen S, Kranzler HR, Krystal JH, et al.: A prospective cohort study investigating factors associated with depression during medical internship. Arch Gen Psychiatry. 2010; 67(6): 557–565. PubMed Abstract | Publisher Full Text | Free Full Text
13. Ayas NT, Barger LK, Cade BE, et al.: Extended work duration and the risk of self-reported percutaneous injuries in interns. JAMA. 2006; 296(9): 1055–1062. PubMed Abstract | Publisher Full Text
14. Barger LK, Cade BE, Ayas NT, et al.: Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med. 2005; 352(2): 125–134. PubMed Abstract | Publisher Full Text
15. Ware JC, Risser MR, Manser T, et al.: Medical resident driving simulator performance following a night on call. Behav Sleep Med. 2006; 4(1): 1–12. PubMed Abstract | Publisher Full Text
16. Schumi J, Wittes JT: Through the looking glass: understanding non-inferiority. Trials. 2011; 12(1): 106. PubMed Abstract | Publisher Full Text | Free Full Text
17. RodriguezJareno MC, Demou E, Vargas-Prada S, et al.: European Working Time Directive and doctors' health: a systematic review of the available epidemiological evidence. BMJ Open. 2014; 4(7): e004916. PubMed Abstract | Publisher Full Text | Free Full Text
18. Ouyang D, Chen JH, Krishnan G, et al.: Patient Outcomes when Housestaff Exceed 80 Hours per Week. Am J Medicine. 2016; 129(9); 993–999.e1. PubMed Abstract | Publisher Full Text | Free Full Text
19. Desai SV, Asch DA, Bellini LM, et al.: Education Outcomes in a Duty-Hour Flexibility Trial in Internal Medicine. N Engl J Med. 2018; 378(10): 1494–1508. PubMed Abstract | Publisher Full Text | Free Full Text
20. Basner M, Asch DA, Shea JA, et al.: Sleep and Alertness in a Duty-Hour Flexibility Trial in Internal Medicine. N Engl J Med. 2019; 380(10): 915–923. PubMed Abstract | Publisher Full Text | Free Full Text
Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 05 July 2021

https://doi.org/10.5256/f1000research.23259.r88690

© 2021 Ropponen A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Annina Ropponen
Finnish Institute of Occupational Health, Helsinki, Finland

The authors of this commentary have chosen very dramatic words, which I would prefer being more moderate/humble. Studying working hours at hospitals is always tricky and especially if the data is gathered via self-reports. Only until recently, few studies exist utilizing objective, pay-roll based working hours. Furthermore, besides very long working hours among physicians or residents, also combinations of work shifts may cause difficulties for recovery and sleep. Hence the studies should not only focus on the length of working hours, but also on shift intervals between work shifts.

What comes to the informal consents, the procedures may vary. If only register data is utilized, that could be done even without a consent. Hence I would like to suggest to put more emphasis on discussing the design and analyses, than on the permissions for the study.

Is the rationale for commenting on the previous publication clearly described?
Partly

Are any opinions stated well-argued, clear and cogent?
Partly

Are arguments sufficiently supported by evidence from the published literature or by new data and results?
Yes

Is the conclusion balanced and justified on the basis of the presented arguments?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Occupational epidemiology.
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com