Check for updates

RESEARCH ARTICLE

REVISED Associations between chlorophyll *a* and various microcystin health advisory concentrations [version 2; referees: 1 approved, 2 approved with reservations]

Jeffrey W. Hollister, Betty J. Kreakie

Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, US Environmental Protection Agency, Narragansett, RI, USA

v2 **First published:** 09 Feb 2016, **5**:151 (doi: 10.12688/f1000research.7955.1)
Latest published: 13 Jun 2016, **5**:151 (doi: 10.12688/f1000research.7955.2)

Abstract

Cyanobacteria harmful algal blooms (cHABs) are associated with a wide range of adverse health effects that stem mostly from the presence of cyanotoxins. To help protect against these impacts, several health advisory levels have been set for some toxins. In particular, one of the more common toxins, microcystin, has several advisory levels set for drinking water and recreational use. However, compared to other water quality measures, field measurements of microcystin are not commonly available due to cost and advanced understanding required to interpret results. Addressing these issues will take time and resources. Thus, there is utility in finding indicators of microcystin that are already widely available, can be estimated quickly and *in situ*, and used as a first defense against high levels of microcystin. Chlorophyll *a* is commonly measured, can be estimated *in situ*, and has been shown to be positively associated with microcystin. In this paper, we use this association to provide estimates of chlorophyll *a* concentrations that are indicative of a higher probability of exceeding select health advisory concentrations for microcystin. Using the 2007 National Lakes Assessment and a conditional probability approach, we identify chlorophyll *a* concentrations that are more likely than not to be associated with an exceedance of a microcystin health advisory level. We look at the recent US EPA health advisories for drinking water as well as the World Health Organization levels for drinking water and recreational use and identify a range of chlorophyll *a* thresholds. A 50% chance of exceeding one of the specific advisory microcystin concentrations of 0.3, 1, 1.6, and 2 µg/L is associated with chlorophyll *a* concentration thresholds of 23, 68, 84, and 104 µg/L, respectively. When managing for these various microcystin levels, exceeding these reported chlorophyll *a* concentrations should be a trigger for further testing and possible management action.

Open Peer Review

Referee Status: ? ? ✓

	Invited Referees 1	2	3
REVISED **version 2** published 13 Jun 2016			✓
version 1 published 09 Feb 2016	? report	? report	✓ report

1 **Alan E. Wilson**, Auburn University USA

2 **Jason W. Marion**, Eastern Kentucky University USA

3 **Zofia E Taranu**, Université de Montréal Canada

Discuss this article

Comments (0)

Corresponding author: Jeffrey W. Hollister (hollister.jeff@epa.gov)

How to cite this article: Hollister JW and Kreakie BJ. **Associations between chlorophyll *a* and various microcystin health advisory concentrations [version 2; referees: 1 approved, 2 approved with reservations]** *F1000Research* 2016, **5**:151 (doi: 10.12688/f1000research.7955.2)

Copyright: © 2016 Hollister JW and Kreakie BJ. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: No competing interests were disclosed.

First published: 09 Feb 2016, **5**:151 (doi: 10.12688/f1000research.7955.1)

REVISED **Amendments from Version 1**

This second version was edited as suggested by the three referees and details are outlined in our response. A summary of these changes are:

- Bootstrapped conditional probabilities were re-run with 10,000 iterations. This resulted in slightly different final values, although conclusions are unchanged
- Several references were added about related studies
- More details about the National Lakes Assessment sampling protocol were added.
- A static copy of the dataset is now included with the repository at https://github.com/USEPA/Microcystinchla/raw/master/inst/extdata/nla_dat.csv
- Corrected typos and unclear text
- Added discussion about possible regional relationships
- Clarified how measurements at the detection limit were handled

See referee reports

Introduction

Over the last decade, numerous events and legislative activities have raised the public awareness of harmful algal blooms[1–3]. In response the US Environmental Protection Agency (USEPA) has recently released suggested microcystin (one of the more common toxins) concentrations that would trigger health advisories[4–6]. Additionally, the World Health Organization (WHO) has microcystin advisory levels for drinking water and for a range of recreational risk levels[7,8]. While these levels and associated advisories are likely to help mitigate the impacts from harmful algal blooms, they are not without complications.

One of these complications is that they rely on available measurements of microcystin. While laboratory testing (e.g., chromatography) remains the gold standard for quantifying microcystin concentrations in water samples, several field test kits have been developed. Even though field tests provide a much needed means for rapid assessment, they are not yet widely used and are moderately expensive (approximately $150–$200 depending on specific kit) with a limited shelf life (typically one year)[9,10]. Additionally, each technique requires nuanced understanding of the detection method (e.g., limit of detection, specific microcystin variants being measured, and sampling protocol).

Fortunately, cyanobacteria and microcystin-LR has been shown to be associated with several other, more commonly measured and well understood components of water quality that are readily assessed in the field[11]. For instance, there are small or hand held fluorometers that measure chlorohpyll *a*. Additionally, chlorophyll *a* is a very commonly measured component of water quality that is also known to be positively associated with microsystin-LR concentrations[12,13]. Recently, Yuan *et al.*[13] explored these associations in detail and controlled for other related variables. In their analysis they find that total nitrogen and chlorophyll *a* show the strongest association with microcystin. Furthermore, they identify chlorophyll *a* and total nitrogen concentrations that are associated with exceeding 1 μg/L of microcystin. These findings suggest that chlorophyll *a* concentrations could also track the new USEPA microcystin health advisory levels for drinking water. Identifying this association would provide an important tool for water resource managers to help manage the threat to public health posed by cHABs and would be especially useful in the absence of measured microcystin concentrations.

In fact, this is a similar tact to the World Health Organization who, in addition to advisory levels for microcystin, have also proposed related advisory levels for cyanobacteria abundance and chlorophyll *a*[7,8]. The chlorophyll *a* concentrations proposed by the WHO are for low (< 10 μg/L), moderate (between 10 and 50 μg/L, high (between 50 and 5000 μg/L), and very high risk (>5000 μg/L)[8]. While these advisories have proven to be useful tools they do suffer from being coarse, broad, and have been found to overestimate actual risk[14].

In this paper we build on these past efforts and utilize the National Lakes Assessment (NLA) data and identify chlorophyll *a* concentrations that are associated with higher probabilities of exceeding several microcystin health advisory concentrations[6,8,15]. We build on past studies by exploring associations with the newly announced advisory levels and by also applying a different method, conditional probability analysis. Utilizing different methods strengthens the evidence for suggested chlorophyll *a* levels that are associated with increased risk of exceeding the health advisory levels as those levels are not predicated on a single analytical method. So that others may repeat or adjust this analysis, the data, code, and this manuscript are freely available via https://github.com/USEPA/microcystinchla.

Methods

Data

We used the 2007 NLA chlorophyll *a* and microcystin-LR concentration data[15]. These data represent a snapshot of water quality from the summer of 2007 for the conterminous United States and were collected as part of an ongoing probabilistic monitoring program[15]. Water quality data, including chlorophyll *a* and microcystin-LR were obtained via an integrated sample taken from the surface of the lake down to 2 meters. Samples were taken at the same time from the index site (e.g. near the centroid of the lake) and these provide the source for both chlorophyll *a* and microcystin-LR[15].

For our analysis we only used samples that were part of the probability sampling design (i.e. no reference samples) and from the first visit to the lake (e.g. some lakes were sampled multiple times). The detection limit for microcystin-LR was 0.05 μg/L. Approximately 67% of lakes reported microcystin-LR at the detection limit. For this analysis we retained these values as removing them would erroneously reduce the confidence intervals around the conditional probabilities. Data on chlorophyll *a* and microcystin-LR concentrations are available for 1028 lakes.

Analytical methods

We used a conditional probability analysis (CPA) approach to explore associations between chlorophyll *a* concentrations and World Health Organization (WHO) and USEPA microcystin health advisory levels[17]. Many health advisory levels have been suggested (Table 1), but lakes with higher microcystin-LR concentrations in the NLA were rare. Only 1.16% of lakes sampled had a concentration greater than 10 μg/L. Thus, for this analysis we focused on the microcystin concentrations that are better represented in the NLA data. These were the USEPA children's (i.e. bottle fed infants to pre-school age children) drinking water advisory level of 0.3 μg/L (USEPA Child), the WHO drinking water advisory level of 1 μg/L (WHO Drinking), the USEPA adult (i.e. beyond pre-school aged individuals) drinking water advisory level of 1.6 μg/L (USEPA Adult), and the WHO recreational, low probability of effect advisory level of 2 μg/L (WHO Recreational).

Conditional probability analysis provides information about the probability of observing one event given another event has also occurred. For this analysis, we used CPA to examine how the conditional probability of exceeding one of the health advisories changes as chlorophyll *a* increases in a lake. We expect to find higher chlorohpyll *a* concentrations to be associated with higher probabilities of exceeding the microcystin health advisory levels. We also calculated bootstrapped 95% confidence intervals (CI) using 10,000 bootstrapped samples. Thus, to identify chlorophyll *a* concentrations of concern we identified the value of the upper 95% CI across a range of conditional probabilities of exceeding each health advisory level. Using the upper confidence limit to identify a threshold is justified as it ensures that a given threshold is unlikely to miss a microcystin exceedance.

As both microcystin-LR and chlorophyll *a* values were highly right skewed, a log base 10 transformation was used. Additional details of the specific implementation are available at https://github.com/USEPA/microcystinchla. A more detailed discussion of CPA is beyond the scope of this paper, but see Paul *et al.*[18] and Hollister *et al.*[19] for greater detail. All analyses were conducted using R version 3.2.3 and code and data from this analysis are freely available as an R package at https://github.com/USEPA/microcystinchla.

Lastly, we assessed the ability of these chlorophyll *a* thresholds to predict microcystin exceedance. We used error matrices and calculate total accuracy as well as the proportion of false negatives. Total accuracy is the total number of correct predictions divided by total observations. The proportion of false negatives is the total number of lakes that were predicted to not exceed the microcystin guidelines but actually did, divided by the total number of observations.

Results

In the 2007 NLA, microcystin-LR concentrations ranged from 0.05 to 225 μg/L. Microcystin-LR concentrations of 0.05 μg/L represent the detection limits. Any value greater than that indicates the presence of microcystin-LR. Of those lakes with microcystin-LR, the median concentration was 0.51 μg/L and the mean was 3.17 μg/L. Of all lakes sampled, 21% of lakes exceeded the USEPA Child level, 8.8% of lakes exceeded the USEPA Adult level, 11.7% of lakes exceeded the WHO Drinking level, and 7.3% of lakes exceeded the WHO Recreational level. Chlorophyll *a*, ranged from 0.07 to 936 μg/L and this captures the range of trophic states from oligotrophic to hypereutrophic. All lakes had detectable levels of chlorophyll *a*. The median concentration was 7.79 μg/L and the mean was 29.63 μg/L. The association between chlorophyll *a* and the upper confidence interval across a range of conditional probability values are shown in Table 2. Specific chlorophyll *a* concentrations that are associated with greater than even odds of exceeding the advisory levels were 23, 68, 84, and 104 μg/L for 0.3, 1.0, 1.6, and 2.0 μg/L advisory levels, respectively (Table 2 & Figure 1).

The chlorophyll *a* cutoffs may be used to predict whether or not a lake exceeds the microcystin health advisories. Doing so allows us to compare the accuracy of the prediction as well as evaluate false negatives. Total accuracy of the four cutoffs predicting microcystin exceedances were 74% for the USEPA children's drinking water advisory, 86% for the WHO drinking water advisory, 89% for the

Table 1. Various microcystin health advisory concentrations from the USEPA and World Health Organization.

Source	Type	Concentration
USEPA	Child Drinking Water Advisory	0.3 μg/L
WHO	Drinking Water	1 μg/L
USEPA	Adult Drinking Water Advisory	1.6 μg/L
WHO	Recreational: Low Prob. of Effect	2–4 μg/L
WHO	Recreational: Moderate Prob. of Effect	10–20 μg/L
WHO	Recreational: High Prob. of Effect	20–2000 μg/L
WHO	Recreational: Very High Prob. of Effect	>2000 μg/L

Table 2. Chlorophyll *a* concentrations that are associated with a 50% probability of exceeding a microcystin health advisory concentration.

Cond. Probability	USEPA Child (0.3 μg/L)	WHO Drink (1 μg/L)	USEPA Adult (1.6 μg/L)	WHO Recreational (2 μg/L)
0.1	0.07	0.07	0.07	1
0.2	0.07	4	12	17
0.3	3	17	32	45
0.4	11	37	68	77
0.5	23	68	84	104
0.6	39	97	115	185
0.7	66	126	871	871
0.8	116	271	871	871
0.9	170	516	871	871

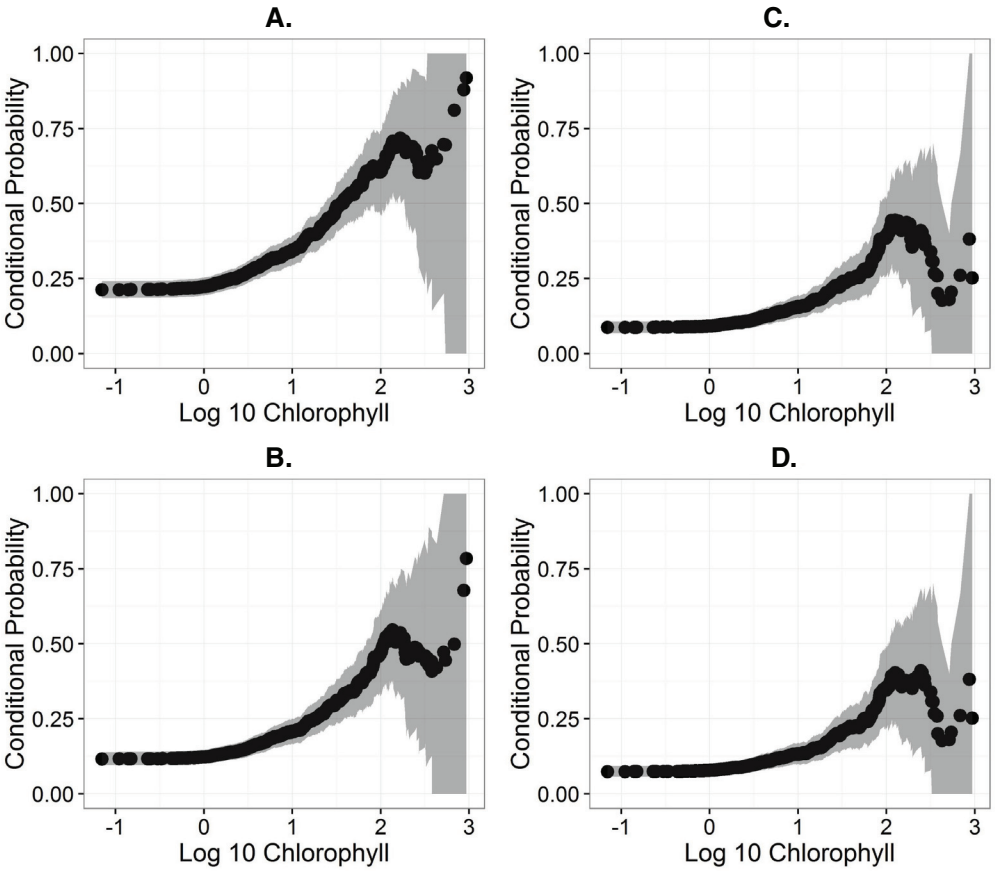

Figure 1. Conditional probability plots showing association between the probability of exceeding various microcystin-LR (MLR) health advisory Levels. A.) Plot for USEPA Child (0.3 μg/L). **B**.) Plot for WHO Drinking (1 μg/L). **C**.) Plot for USEPA Adult (1.6 μg/L). **D**.) Plot for WHO Recreational (2 μg/L).

USEPA adult drinking water advisory, and 91% for the WHO recreational advisory (Table 3, Table 4, Table 5, & Table 6). However, total accuracy is only one part of the prediction performance with which we are concerned.

When using the chlorophyll *a* cutoffs as an indicator of microcystin exceedances, the error that should be avoided is predicting that no exceedance has occurred when in fact it has. In other words, we would like to avoid Type II errors and minimize the proportion of false negatives. For the four chlorophyll *a* cut-offs we had a proportion of false negatives of 9%, 8%, 6%, and 5% for the USEPA children's, the WHO drinking water, the USEPA adult, and the WHO recreational advisories, respectively. In each case we missed less than 10% of the lakes that in fact exceeded the microcystin advisory. While this method performs well with regard to the false negative percentage, it is possible that is a relic of the NLA dataset and testing with additional data would allow us to confirm this result.

Discussion

The log-log association between microcystin-LR and chlorophyll *a* indicates that, in general, higher concentrations of microcystin-LR almost always co-occur with higher concentrations of chlorophyll *a* yet the inverse is not true (Figure 2). Higher chlorophyll *a* is not necessarily predictive of higher microcystin-LR concentrations; however, chlorophyll *a* may be predictive of the probability of exceeding a certain threshold.

Indeed, the probability of exceeding each of the four tested health advisory levels increased as a function of chlorophyll *a* concentration (Figure 1). We used this association to identify chlorophyll *a* concentrations that were associated with a range of probabilities of exceeding a given health advisory level (Table 2). For the purposes of this discussion we focus on a conditional probability of 50% or greater (i.e., greater than even odds to exceed a health advisory level). The 50% conditional probability chlorophyll *a* thresholds represents 28.6%, 11%, 8.9%, and 7.2% of sample lakes for the USEPA Child, the WHO Drinking, the USEPA Adult, and the WHO recreational levels, respectively.

There are numerous possible uses for the chlorophyll *a* and microcystin advisory cut-off values. First, in the absence of microcystin-LR measurements, exceedence of the chlorophyll *a* concentrations could be a trigger for further actions. Given that there is uncertainity around these chlorophyll *a* cutoffs the best case scenario would be to monitor for chlorophyll *a* and in the event of exceeding a target concentration take water samples and have those samples tested for microcystin-LR.

A second potential use is to identify past bloom events from historical data. As harmful algal blooms are made up of many species and have various mechanisms responsible for adverse impacts (e.g., toxins, hypoxia, odors), there is no single definition of a bloom. For cHABs, one approach has been to utilize phycocyanin to screen for or identify bloom events[20]. This is a useful approach, but phycocyanin is not always available, thus limiting its utility especially for examining historical data. Using our chlorophyll *a* cutoffs provides a value that is also associated with microcystin-LR and can be used to classify lakes, from past surveys, as having bloomed.

Table 3. Confusion matrix comparing chlorophyll *a* predicted exceedences (rows) versus real exceedances (columns) for the USEPA childrens drinking water advisory.

	Not Exceed	Exceed	Row Totals
Not Exceed	643	95	738
Exceed	168	122	290
Column Totals	811	217	1028

Table 4. Confusion matrix comparing chlorophyll *a* predicted exceedences (rows) versus real exceedances (columns) for the WHO drinking water advisory.

	Not Exceed	Exceed	Row Totals
Not Exceed	841	78	919
Exceed	66	43	109
Column Totals	907	121	1028

Table 5. Confusion matrix comparing chlorophyll *a* predicted exceedences (rows) versus real exceedances (columns) for the USEPA adult drinking water advisory.

	Not Exceed	Exceed	Row Totals
Not Exceed	884	57	941
Exceed	53	34	87
Column Totals	937	91	1028

Table 6. Confusion matrix comparing chlorophyll *a* predicted exceedences (rows) versus real exceedances (columns) for the WHO recreational water advisory.

	Not Exceed	Exceed	Row Totals
Not Exceed	908	51	959
Exceed	45	24	69
Column Totals	953	75	1028

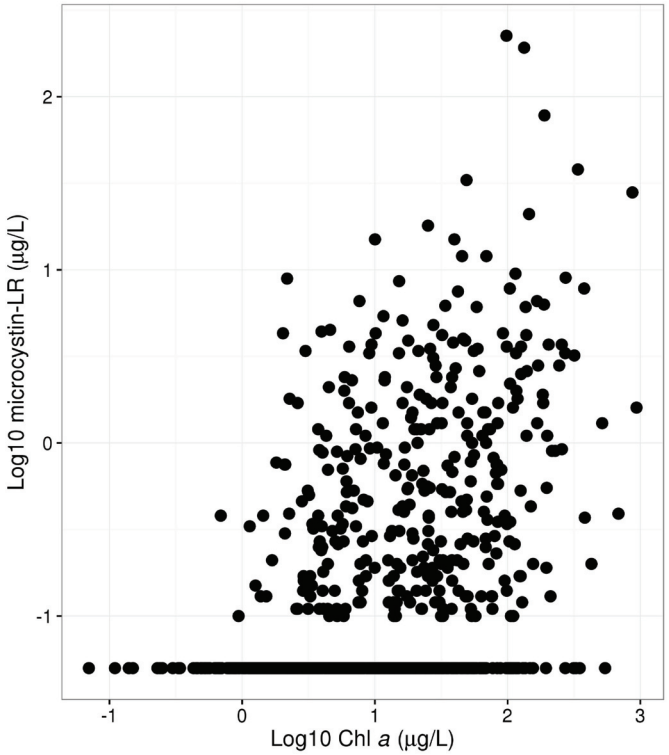

Figure 2. Scatterplot showing association between chlorophyll *a* and microcystin-LR.

The values we propose are national and may miss regional variation in water quality, including, chlorophyll *a* and microcystin-LR[22]. A set of regional conditional probabilities would be interesting; however, limiting the analysis to the data available per region would make interpretation difficult. The sample size for each of the regional conditional probabilities would be reduced and the number of lakes in each region that exceed the microcystin values would also be reduced. Thus, our confidence in the conditional probabilities would be less (i.e. greatly increased confidence intervals) and the relationships less pronounced as we have fewer lakes on which to base the probabilities. Thus, this dataset is best for making national scale recommendations.

There are two other limitations with the 2007 National Lakes Assessment dataset. First, it represents a single sample from a lake and does not capture temporal dynamics. Second, validation of the predictions with the 2007 data alone would be challenging as the data would need to be subset and this would only sever to increase the uncertainty of our conditional probabilities, reducing our ability to validate the presence of microcystin-LR. The 2012 National Lakes Assessment would be ideal for this task. However, as of this writing, the 2012 National Lakes Assessment data are not public. When these data are released, a validation of this approach can be completed then.

Lastly, using chlorophyll *a* is not meant as a replacement for testing of microcystin-LR or other toxins. It should be used when other, direct measurements of cyanotoxins are not available. In those cases, which are likely to be common at least in the near future, using a more ubiquitous measurement such as chlorophyll *a* will provide a reasonable proxy for the probability of exceeding a microcystin health advisory level and provide better protection against adverse effects in both drinking and recreational use cases.

Data and software availability

Data and latest source code

https://github.com/USEPA/microcystinchla

Archived data and source code at time of publication

http://dx.doi.org/10.5281/zenodo.55273[24]

License

Creative Commons Zero 1.0: http://creativecommons.org/publicdomain/zero/1.0/

Author contributions

JH and BK conceived of and conducted the analysis. Both authors reviewed the results and contributed to writing the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

We would like to thank Anne Kuhn, Bryan Milstead, John Kiddon, Joe LiVolsi, Tim Gleason, Wayne Munns, and Leslie D'Anglada for constructive reviews of this paper. Special thanks to Jason Marion, Alan Wilson, and Zofia Taranu for reviews of the submitted manuscript. This paper has not been subjected to Agency review. Therefore, it does not necessary reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This contribution is identified by the tracking number ORD-015143 of the Atlantic Ecology Division, Office of Research and Development, National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency.

References

1. Jetoo S, Grover VI, Krantzberg G: **The Toledo drinking water advisory: Suggested application of the water safety planning approach.** *Sustainability.* 2015; **7**(8): 9787–9808.
Publisher Full Text
2. Rinta-Kanto JM, Konopko EA, DeBruyn JM, *et al.*: **Lake Erie *Microcystis*: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake.** *Harmful Algae.* 2009; **8**(5): 665–673.
Publisher Full Text
3. Harmful Algal Bloom and Hypoxia Research and Control Amendments Act: **Harmful Algal Bloom and Hypoxia Research and Control Amendments Act of 2014.** 2014; S. 1254.
Reference Source
4. McElhiney J, Lawton LA: **Detection of the cyanobacterial hepatotoxins microcystins.** *Toxicol Appl Pharmacol.* 2005; **203**(3): 219–230.
PubMed Abstract | Publisher Full Text
5. Zurawell RW, Chen H, Burke JM, *et al.*: **Hepatotoxic cyanobacteria: a review of the biological importance of microcystins in freshwater environments.** *J Toxicol Environ Health B Crit Rev.* 2005; **8**(1): 1–37.
PubMed Abstract | Publisher Full Text
6. US EPA: **Drinking water health advisory for the cyanobacterial microcystin toxins.** EPA-820-R-15100. 2015.
Reference Source
7. World Health Organization: **Cyanobacterial toxins: Microcystin-LR in drinking-water.** Background document for development of WHO *guidelines for drinking-water quality*. Geneva, Switzerland. World Health Organization, 2nd ed. Geneva. 2003.
Reference Source
8. Chorus I, Bartram J: **Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management.** World Health Organization. 1999.
Reference Source
9. James R, Gregg A, Dindal A, *et al.*: **Environmental technology verification report: Abraxis microcystin test kits.** Online document. Accessed online: June 22, 2011.
Reference Source
10. Aranda-Rodriguez R, Jin Z, Harvie J, *et al.*: **Evaluation of three field test kits to detect microcystins from a public health perspective.** *Harmful Algae.* 2015; **42**: 34–42.
Publisher Full Text
11. Yuan LL, Pollard AI: **Deriving nutrient targets to prevent excessive cyanobacterial densities in U.S. lakes and reservoirs.** *Freshwater Biol.* 2015; **60**(9): 1901–1916.
Publisher Full Text
12. Pip E, Bowman L: **Microcystin and algal chlorophyll in relation to nearshore nutrient concentrations in Lake Winnipeg, Canada.** *Environ Pollut.* 2014; **3**(2): p36.
Publisher Full Text
13. Yuan LL, Pollard AI, Pather S, *et al.*: **Managing microcystin: Identifying national-scale thresholds for total nitrogen and chlorophyll *a*.** *Freshwater Biol.* 2014; **59**(9): 1970–1981.
Publisher Full Text
14. Loftin KA, Graham JL, Hilborn ED, *et al.*: **Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the ePA national lakes assessment 2007.** *Harmful Algae.* 2016; **56**: 77–90.
Publisher Full Text
15. US EPA: **National lakes assessment: A collaborative survey of the nation's lakes.** EPA 841-R-09-001. 2009.
Reference Source
16. USEPA: **Survey of the nation's lakes field operations manual.** EPA841-B-07-004. 2007.
Reference Source
17. Paul JF, Munns WR Jr: **Probability surveys, conditional probability, and ecological risk assessment.** *Environ Toxicol Chem.* 2011; **30**(6): 1488–1495.
PubMed Abstract | Publisher Full Text
18. Paul JF, McDonald ME: **Development of empirical, geographically specific water quality criteria: A conditional probability analysis approach.** *J Am Water Resour Assoc.* 2005; **41**(5): 1211–1223.
Publisher Full Text
19. Hollister JW, Walker HA, Paul JF: **CProb: a computational tool for conducting conditional probability analysis.** *J Environ Qual.* 2008; **37**(6): 2392–2396.
PubMed Abstract | Publisher Full Text
20. Miller TR, Beversdorf L, Chaston SD, *et al.*: **Spatiotemporal molecular analysis of cyanobacteria blooms reveals *Microcystis*--*Aphanizomenon* interactions.** *PLoS One.* 2013; **8**(9): e74933.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Marion JW, Lee J, Wilkins JR 3rd, *et al.*: ***In vivo* phycocyanin flourometry as a potential rapid screening tool for predicting elevated microcystin concentrations at eutrophic lakes.** *Environ Sci Technol.* 2012; **46**(8): 4523–4531.
PubMed Abstract | Publisher Full Text
22. Beaver JR, Manis EE, Loftin KA, *et al.*: **Land use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation.** *Harmful Algae.* 2014; **36**: 57–62.
Publisher Full Text
23. Wagner T, Soranno PA, Webster KE, *et al.*: **Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes.** *Freshwater Biol.* 2011; **56**(9): 1811–1824.
Publisher Full Text
24. Hollister JW, Kreakie BJ: **Associations between chlorophyll a and various microcystin health advisory concentrations.** *Zenodo.* 2016.
Data Source

Open Peer Review

Current Referee Status: ? ? ✔

Version 2

Referee Report 28 June 2016

doi:10.5256/f1000research.9675.r14343

✔ **Zofia E Taranu**
Department of Biological Sciences, Université de Montréal, Montreal, QC, Canada

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 11 April 2016

doi:10.5256/f1000research.8562.r12719

✔ **Zofia E Taranu**
Department of Biological Sciences, Université de Montréal, Montreal, QC, Canada

General comments
This study provides an elegant framework to predict the severe impairment of U.S. lakes and reservoirs by cyanobacterial harmful algal blooms. I especially appreciated the clever use of conditional probability analysis to identify chlorophyll *a* threshold above which MC concentrations exceed WHO drinking water and recreational provisional guidelines. Chlorophyll *a* is a regularly measured water quality variable, and this study indeed offers a promising approach that can me used in many lakes around the world to identify problematic lakes or regions. In that regard, it would have been interesting to account for the spatial heterogeneity across this landscape -- as indicated by Beaver *et al.* (2014), some regions of the continental U.S. are more likely to be MC hotspots. Accounting for this heterogeneity will likely help explain some of the noise in the biplot shown in Fig. 2. I have also analyzed the MC data from the same dataset and found that accounting for different ecoregions in my model (as presented in Beaver *et al.,* 2014) helped further explain the probability of detecting versus failing to detect MC in lakes. I am curious to know how this would play out with your modeling approach (conditional probability analysis).

Minor comments
I recommend the following minor typographical corrections:

P3

Change: "Yuan *et al.* (2014) explore these associations in detail and control for other related variables. In their analysis they find that total [...]"
To: "Recently, Yuan *et al.* (2014) explored these associations in detail and controlled for other related variables and found that total [...]"

Change: "Given these facts, it should be possible to identify chlorophyll a concentrations that would be associated with the [...]"
To: "These findings suggest that chlorophyll a concentrations could also track the [...]"

Change: "Identifying these associations would provide another tool for [...]"
To: "Identifying this association would provide an important tool for [...]"

Change: "We add to past studies by exploring associations with newly announced advisory [...]"
To: "We build on past studies by exploring associations with the newly announced advisory [...]"

P4

Change: "Thus, to identify chlorophyll a concentrations of concern we identify the value [...]"
To: "Thus, to identify chlorophyll a concentrations of concern we identified the value [...]"

Change: "were highly skewed right,"
To: "were highly right skewed,"

Change: "Lastly, we assess the ability of"
To: "Lastly, we assessed the ability of"

Change: "We use error matrices and calculate total accuracy"
To: "We used error matrices and calculate total accuracy"

Change: "For chlorophyll a, the range was"
To: "Chlorophyll a ranged from"
Please specify that this chlorophyll a range corresponds to a range from oligotrophic to hypereutrophic lakes.

Change: "The associations between chlorophyll a and the upper confidence interval"
To: "The association between chlorophyll a and the upper confidence interval"

Figure 2 should first be presented in the Results section.

Change: "This is the case as the probability of exceeding each of the four tested health advisory levels increases as a"
To: "Indeed, the probability of exceeding each of the four tested health advisory levels increased as a"

Change: "We used this association to identify chlorophyll a concentrations that are associated"
To: "We used this association to identify chlorophyll a concentrations that were associated"

References

1. Beaver J, Manis E, Loftin K, Graham J, Pollard A, Mitchell R: Land use patterns, ecoregion, and

microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation. *Harmful Algae*. 2014; **36**: 57-62 Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 01 Jun 2016

Jeffrey Hollister, US EPA, USA

Thank you for your review. We have just submitted our revisions and expect the new version to be available in the next few days. Below are our responses to the specific issues you raise. We feel the paper is stronger after this revision. Any questions, feel free to contact us or comment here. Thanks again!

Spatial/Regional Heterogeneity in Chlorophyll/Microcystin realtionship
We agree that there are likely regional differences and would like to account for this; however, sample sizes for each region vary (67 to 155) and are relatively small. The resulting conditional probability analysis would have very wide confidence intervals. Thus, comparison between regions would be difficult and inferring a pattern would not be possible. We have added additional text in the discussion (second to last paragraph) that raises this issue. Additionally, we have added the Beaver et al. reference in this discussion.

Specific Edits

P3
- Change: “Yuan et al. (2014) explore these associations in detail and control for other related variables. In their analysis they find that total [...]”, To: “Recently, Yuan et al. (2014) explored these associations in detail and controlled for other related variables and found that total [...]”
  - **Response:** Changed

- Change: “Given these facts, it should be possible to identify chlorophyll a concentrations that would be associated with the [...]”, To: “These findings suggest that chlorophyll a concentrations could also track the [...]”
  - **Response:** Changed

- Change: “Identifying these associations would provide another tool for [...]”, To: “Identifying this association would provide an important tool for [...]”
  - **Response:** Changed

- Change: “We add to past studies by exploring associations with newly announced advisory [...]”, To: “We build on past studies by exploring associations with the newly announced advisory [...]”
  - **Response:** Changed

P4
- Change: “Thus, to identify chlorophyll a concentrations of concern we identify the value [...]”, To: “Thus, to identify chlorophyll a concentrations of concern we identified the value [...]”
  - **Response:** Changed

- Change: “were highly skewed right,”, To: “were highly right skewed,”
  - **Response:** Changed

- Change: “Lastly, we assess the ability of”, To: “Lastly, we assessed the ability of”
  - **Response:** Changed

- Change: “We use error matrices and calculate total accuracy”, To: “We used error matrices and calculate total accuracy”
  - **Response:** Changed

- Change: “For chlorophyll a, the range was”, To: “Chlorophyll a ranged from”
  - **Response:** Changed

- Please specify that this chlorophyll a range corresponds to a range from oligotrophic to hypereutrophic lakes.
  - **Response:** Changed

- Change: “The associations between chlorophyll a and the upper confidence interval”, To: “The association between chlorophyll a and the upper confidence interval”
  - **Response:** Changed

- Figure 2 should first be (Figure \ref{fig:chla_micro_scatter})presented in the Results section.
  - **Response:** We did not changes this as we fill Figure 2 is not presenting results of our analysis but justifying the approach as is a better fit to be introduced in the Discussion.

- Change: “This is the case as the probability of exceeding each of the four tested health advisory levels increases as a”, To: “Indeed, the probability of exceeding each of the four tested health advisory levels increased as a”
  - **Response:** Changed

- Change: “We used this association to identify chlorophyll a concentrations that are associated”, To: “We used this association to identify chlorophyll a concentrations that were associated”
  - **Response:** Changed

Competing Interests: No competing interests were disclosed.

Referee Report 05 April 2016

doi:10.5256/f1000research.8562.r12688

Jason W. Marion
Department of Environmental Health Science, Eastern Kentucky University, Richmond, KY, USA

Overview: The manuscript/article addresses a critical question applicable to recreational and drinking water managers: Can we rapidly predict potentially harmful cyanobacteria blooms using traditional water quality methods? This question is likely to become more relevant according to the most current literature as blooms are expected to increase in frequency in the midst of a warming global climate facing more

extreme storm and drought events [1-6]. Using the rather large and nationally applicable National Lakes Assessment database for the United States, the authors demonstrate some of the strengths and weaknesses of using chlorophyll *a* as an indicator for at-risk conditions that could warrant management action or follow-up testing for cyanotoxins. The authors are also able to assign action levels or at least share possible action levels for management action using conditional probabilities. The strengths and weaknesses of selected probabilities are described using analyses similar to specificity and sensitivity in the form of accuracy and 'avoiding type II errors'. The authors do not propose using chlorophyll *a* as a proxy to replace toxin measurement, but as a tool to help facilitate targeted monitoring of toxins during at-risk conditions.

Overall Comments: The article is meritorious in that it does provide a meaningful starting point for lakes with no phycocyanin measures and for providing a meaningful starting point for developing some semblance of an action level that could be employed by recreational water and drinking water managers concerned with cyanotoxins. The article does stand to improve significantly in some key areas, which are as follows:

(1) Additional discussion in methods related to the National Lakes Assessment
(2) Additional discussion needed on how the data were organized for data analysis
(3) Improved discussion needed on alternative indicators for cHABs and cyanotoxins not assessed in the NLA
(4) Consideration of region-specific criteria or limitations of national recommendations for chl *a*.
(5) Greater discussion on limitations of NLA and need for model validation/future studies.

<u>(1) Additional discussion needed in methods related to the National Lakes Assessment:</u>
The readership may not be aware of the U.S. NLA performed in 2007. The author(s) should clarify where samples were collected (nearshore or from the surface in the deeper waters). NLA chlorophyll *a* samples were take from the profundal zone rather than the littoral zone. The readership may also be interested in how many chl *a* samples were collected from each lake. Where were the microcystin-LR samples collected?

<u>(2) Additional discussion needed on how the data were organized for data analysis:</u>
Were these samples paired (collected at the same time from the same locale) or are these some type of aggregated value over a lake season? Describing this in the methods will really help for understanding the importance of this work. Paired results (MC-LR and Chl *a* from the same day) are much more impactful for demonstrating the rapid advantage of chl *a* compated to using results that are a seasonal average indicating that the hypereutrophic and eutrophic lakes (ones with the highest chl *a*) are also the ones that are most likely to have a cyanoHAB event sometime during the year.

<u>(3) Improved discussion needed on alternative indicators for cHABs and cyanotoxins not assessed in the NLA:</u> Brief mention is given to phycocyanin (one study), and the additional language (about phycocyanin not always being available for measure and when measured, it is for only measuring pigment and not toxins) is equally relevant for chl *a*. The same *in vivo* handheld fluorometers and continuous monitoring solutions available for chl *a* are now widely available for phycocyanin, often at the same cost as a rapid measure for chl *a*. Phycocyanin, like chl *a,* does not measure toxin either, but phycocyanin in many studies has outperformed chl *a*, and in some studies it has not (especially when toxin concentration is low). Historical records on PC are likely not as great as chlorophyll *a*. Overall, several studies on this topic have been produced in the last two to four years (see Zamyadi and Dorner's work), with one study using phycocyanin to predict non-alcoholic liver disease presuming a relationship with cyanotoxins (Zhang *et al.* 2015)

(4) Consideration desired on region-specific criteria or limitations of national recommendations for chl *a*: With nearly 30% of the lakes in the temperate plains being coded as poor for chlorophyll *a* in the 2007 NLA, what impact would these conditional probabilities have on these lakes? Should the lake managers in this region be monitoring continuously all the time? What are the mean/median chlorophyll *a* levels for this part of the U.S? Regional variability may be really important and did the conditional probability approach take this into consideration or can it take it into consideration? Is there a way to evaluate if there are significant regional effects in the U.S? For nutrient standards in the U.S. and macroinvertebrate assessments, EPA has had to issue region-specific guidelines/criteria, etc. for some parameters.

(5) Greater discussion needed on limitations of NLA and need for model validation/future studies: The paper fails to address the limitations of the NLA – as a reader, I'm not aware of the limitations. I have much respect for the NLA, but I do have questions regarding the number of samples for each lake. Furthermore, a statement or two discussing the need to validate modeled data may be worthwhile. Is there a way to see if the probabilities actually align with the accuracy and type II error rates predicted by the conditional probability approach?

Abstract-Specific Comments: Near the bottom of the abstract, the units seem quite high for microcystins (g/L) rather than micrograms/L. The micro Greek symbol (mu) may have been lost during uploading.

Results Comments: (1) In discussing the lake exceedances of the various recommended levels by EPA, the addition of 'drinking water' is appropriate in my opinion. Although it is mentioned earlier in the methods, further providing the information in the results is helpful to a novice reader or a person just becoming familiar with drinking water regulations and guidelines, as the U.S. EPA child level may be presumed by a reader to be a level for recreation in a lake rather than a level associated with finished drinking water after water treatment. (2) "All lakes had reported chl *a* concentrations that exceeded detection limits" Does this mean that some were over range? Or does this mean that "All lakes had detectable levels of chl a"

Discussion Comments: The wedge pattern in figure 2 is not apparent in figure 2, however, the logic makes sense and is supported visually by the conditional probability plots in fig 1. If figure 2 could have two lines of best fit (similar to the way some researchers do for funnel plots on publication bias papers), it may be easier to see the wedge shape.

Climate Articles Highlighting Current Importance of Topic:

Harvell, C. D.; Kim, K.; Burkholder, J. M.; Colwell, R. R.; Epstein, P. R.; Grimes, D. J.; Hoffman, E. E.; Lipp,E. K.; Osterhaus, A. D. M. E.; Overstreet, R. M.; Porter, J. W.; Smith, G. W.; Vasta, G. R.Emerging marine diseases – climate links and anthropogenic factors *Science* **2000**, 285, 1505– 1510

Peperzak, L.Climate change and harmful algal blooms in the North Sea *Acta Oecol.* **2003**, 24, 139–144

Edwards, M.; Johns, D. G.; Leterme, S. C.; Svendsen, E.; Richardson, A. J.Regional climate change and harmful algal blooms in the northeast Atlantic *Limnol. Oceanogr.* **2006**, 51 (2) 820– 829

Ye, C.; Shen, Z.; Zhang, T.; Feng, M.; Lei, Y.; Zhang, J.Long-term joint effect of nutrients and temperature increase on algal growth in Lake Taihu, China *J Environ. Sci.* **2011**, 23 (2) 222– 227

Paerl, H. W.; Hall, N. S.; Calandrino, E. S.Controlling harmful cyanobacterial blooms in a world

experiencing anthropogenic and climatic-induced change *Sci. Total Environ.* **2011**, 409, 1739– 1745

Davis, T. W.; Berry, D. L.; Boyer, G. L.; Gobler, C. J.The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms *Harmful Algae* **2009**, 8, 715– 725

Articles on Phycocyanin and Toxin Indicators Ought to Be Considered:

Ahn, C.-Y.; Joung, S.-H.; Yoon, S.-K.; Oh, H.-M.Alternative alert system for cyanobacterial bloom, using phycocyanin as a level determinant *J. Microbiol.* **2007**, 45 (2) 98– 104.

Makarewicz, J. C.; Boyer, G. L.; Lewis, T. .; Guenther, W.; Atkinson, J.; Arnold, M.Spatial and temporal distribution of the cyanotoxin microcystin-LR in the Lake Ontario ecosystem: Coastal embayments, rivers, nearshore and offshore, and upland lakes *J. Great Lakes Res.* **2009**, 35, 83– 89.

Murby, A. L.Assessing spatial distributions of cyanobacteria and microcystins in N.H. lakes with implications for lake monitoring. Master's Thesis. University of New Hampshire, **2009**; p 89.

Lehman, E. M.Seasonal occurrence and toxicity of *Microcystis* in impoundments of the Huron River, Michigan, USA *Water Res.* **2007**, 41, 795– 802

Lee C, Marion JW, Cheung M, Lee CS, Lee J. Associations among Human-Associated Fecal Contamination, Microcystis aeruginosa, and Microcystin at Lake Erie Beaches. International journal of environmental research and public health. 2015 Sep 11;12(9):11466-85.

Zamyadi A, Dorner S, Ndong M, Ellis D, Bolduc A, Bastien C, Prévost M. Application of in vivo measurements for the management of cyanobacteria breakthrough into drinking water treatment plants. Environmental Science: Processes & Impacts. 2014;16(2):313-23.

Zhang F, Lee J, Liang S, Shum CK. Cyanobacteria blooms and non-alcoholic liver disease: evidence from a county level ecological study in the United States. Environmental Health. 2015 May 7;14(1):41.

References

1. Harvell CD, Kim K, Burkholder JM, Colwell RR, Epstein PR, Grimes DJ, Hofmann EE, Lipp EK, Osterhaus AD, Overstreet RM, Porter JW, Smith GW, Vasta GR: Emerging marine diseases--climate links and anthropogenic factors.*Science*. 1999; **285** (5433): 1505-10 PubMed Abstract
2. Peperzak L: Climate change and harmful algal blooms in the North Sea. *Acta Oecologica*. 2003; **24**: S139-S144 Publisher Full Text
3. Edwards M, Johns D, Leterme S, Svendsen E, Richardson A: Regional climate change and harmful algal blooms in the northeast Atlantic. *Limnology and Oceanography*. 2006; **51** (2): 820-829 Publisher Full Text
4. Ye C, Shen Z, Zhang T, Fan M, Lei Y, Zhang J: Long-term joint effect of nutrients and temperature increase on algal growth in Lake Taihu, China.*J Environ Sci (China)*. 2011; **23** (2): 222-7 PubMed Abstract
5. Paerl HW, Hall NS, Calandrino ES: Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change.*Sci Total Environ*. 2011; **409** (10): 1739-45 PubMed Abstract | Publisher Full Text
6. Davis T, Berry D, Boyer G, Gobler C: The effects of temperature and nutrients on the growth and

dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. *Harmful Algae*. 2009; **8** (5): 715-725 Publisher Full Text
7. Ahn CY, Joung SH, Yoon SK, Oh HM: Alternative alert system for cyanobacterial bloom, using phycocyanin as a level determinant.*J Microbiol*. 2007; **45** (2): 98-104 PubMed Abstract
8. Makarewicz J, Boyer G, Lewis T, Guenther W, Atkinson J, Arnold M: Spatial and temporal distribution of the cyanotoxin microcystin-LR in the Lake Ontario ecosystem: Coastal embayments, rivers, nearshore and offshore, and upland lakes. *Journal of Great Lakes Research*. 2009; **35**: 83-89 Publisher Full Text
9. Murlby AL: Assessing spatial distributions of cyanobacteria and microcystins in N.H. lakes with implications for lake monitoring. 2009; **Master's Thesis** (University of New Hampshire): 89
10. Lehman EM: Seasonal occurrence and toxicity of Microcystis in impoundments of the Huron River, Michigan, USA.*Water Res*. 2007; **41** (4): 795-802 PubMed Abstract | Publisher Full Text
11. Lee C, Marion JW, Cheung M, Lee CS, Lee J: Associations among Human-Associated Fecal Contamination, Microcystis aeruginosa, and Microcystin at Lake Erie Beaches.*Int J Environ Res Public Health*. 2015; **12** (9): 11466-85 PubMed Abstract | Publisher Full Text
12. Zamyadi A, Dorner S, Ndong M, Ellis D, Bolduc A, Bastien C, Prévost M: Application of in vivo measurements for the management of cyanobacteria breakthrough into drinking water treatment plants. *Environ Sci Process Impacts*. 2014; **16** (2): 313-23 PubMed Abstract | Publisher Full Text
13. Zhang F, Lee J, Liang S, Shum CK: Cyanobacteria blooms and non-alcoholic liver disease: evidence from a county level ecological study in the United States.*Environ Health*. 2015; **14**: 41 PubMed Abstract | Publisher Full Text

I have read this submission. I 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.

Competing Interests: Several of the recommended papers suggested for the author to consider in any subsequent revision of this manuscript involve authors for whom I have collaborated with on research.

Author Response 01 Jun 2016

Jeffrey Hollister, US EPA, USA

Thank you for your review. We have just submitted our revisions and expect the new version to be available in the next few days. Below are our responses to the specific issues you raise. We feel the paper is stronger after this revision. Any questions, feel free to contact us or comment here. Thanks again!

Additional discussion needed in methods related to the National Lakes Assessment:
The readership may not be aware of the U.S. NLA performed in 2007. The author(s) should clarify where samples were collected (nearshore or from the surface in the deeper waters). NLA chlorophyll a samples were take from the profundal zone rather than the littoral zone. The readership may also be interested in how many chl a samples were collected from each lake. Where were the microcystin-LR samples collected?
- **Response:** We agree that additional information was needed describing the NLA. We have added this to the first paragraph on the Data section.

Additional discussion needed on how the data were organized for data analysis:
Were these samples paired (collected at the same time from the same locale) or are these some type of aggregated value over a lake season? Describing this in the methods will really help for understanding the importance of this work. Paired results (MC-LR and Chl a from the same day)

are much more impactful for demonstrating the rapid advantage of chl a compated to using results that are a seasonal average indicating that the hypereutrophic and eutrophic lakes (ones with the highest chl a) are also the ones that are most likely to have a cyanoHAB event sometime during the year.

- **Response:** We agree and have added some additional wording to the Data section indicating that the samples are taken at the same time.

Improved discussion needed on alternative indicators for cHABs and cyanotoxins not assessed in the NLA:
Brief mention is given to phycocyanin (one study), and the additional language (about phycocyanin not always beiThese blooms are expected to increase in frequecy and severity due to the impacts of climate change ng available for measure and when measured, it is for only measuring pigment and not toxins) is equally relevant for chl a. The same in vivo handheld fluorometers and continuous monitoring solutions available for chl a are now widely available for phycocyanin, often at the same cost as a rapid measure for chl a. Phycocyanin, like chl a, does not measure toxin either, but phycocyanin in many studies has outperformed chl a, and in some studies it has not (especially when toxin concentration is low). Historical records on PC are likely not as great as chlorophyll a. Overall, several studies on this topic have been produced in the last two to four years (see Zamyadi and Dorner's work), with one study using phycocyanin to predict non-alcoholic liver disease presuming a relationship with cyanotoxins (Zhang et al. 2015)

- **Response:** We agree that phycocyanin is more closely linked to microcystin than is chl *a*. Our paragraph mentioning phycocyanin was confusing and did suggest that chl *a* had a stronger association. Wording of that paragraph has been changed and the Ahn et al paper was added as reference. We feel that further discussion of phycocyanin, while important, is beyond the scope of our paper with its focus on chlorophyll.

Consideration desired on region-specific criteria or limitations of national recommendations for chl a:
With nearly 30% of the lakes in the temperate plains being coded as poor for chlorophyll a in the 2007 NLA, what impact would these conditional probabilities have on these lakes? Should the lake managers in this region be monitoring continuously all the time? What are the mean/median chlorophyll a levels for this part of the U.S? Regional variability may be really important and did the conditional probability approach take this into consideration or can it take it into consideration? Is there a way to evaluate if there are significant regional effects in the U.S? For nutrient standards in the U.S. and macroinvertebrate assessments, EPA has had to issue region-specific guidelines/criteria, etc. for some parameters.

- **Response:** We agree that there are likely regional differences and would like to account for this; however, sample sizes for each region vary (67 to 155) and are relatively small. The resulting conditional probability analysis would have very wide confidence intervals. Thus, comparison between regions would be difficult and inferring a pattern would not be possible. We have added additional text in the discussion (second to last paragraph) that raises this issue. Additionally, we have added the Beaver et al. reference in this discussion.

Greater discussion needed on limitations of NLA and need for model validation/future studies:
The paper fails to address the limitations of the NLA – as a reader, I'm not aware of the limitations. I have much respect for the NLA, but I do have questions regarding the number of samples for each lake. Furthermore, a statement or two discussing the need to validate modeled data may be worthwhile. Is there a way to see if the probabilities actually align with the accuracy and type II error rates predicted by the conditional probability approach?

- **Response:** We added a paragraph to the discussions about validation and the single sample limitations of the NLA.

Specific Edits

Abstract-Specific Comments:
- Near the bottom of the abstract, the units seem quite high for microcystins (g/L) rather than micrograms/L. The micro Greek symbol (mu) may have been lost during uploading.
  - **Response:** Oops! Looks like it did occur during upload. Have double checked the final for the proper units.

Results Comments:
- In discussing the lake exceedances of the various recommended levels by EPA, the addition of 'drinking water' is appropriate in my opinion. Although it is mentioned earlier in the methods, further providing the information in the results is helpful to a novice reader or a person just becoming familiar with drinking water regulations and guidelines, as the U.S. EPA child level may be presumed by a reader to be a level for recreation in a lake rather than a level associated with finished drinking water after water treatment.
  - **Response:** Added this in.

- "All lakes had reported chl a concentrations that exceeded detection limits" Does this mean that some were over range? Or does this mean that "All lakes had detectable levels of chl a"
  - **Response:** Changed wording to: All lakes had detectable levels of ...

Discussion Comments:
- The wedge pattern in figure 2 is not apparent in figure 2, however, the logic makes sense and is supported visually by the conditional probability plots in fig 1. If figure 2 could have two lines of best fit (similar to the way some researchers do for funnel plots on publication bias papers), it may be easier to see the wedge shape.
  - **Response:** Agreed that the wedge is not pronounced. We have changed the wording to better describe the pattern. Also, we disagree with the lines of best fit as those would then imply some sort of linear (presumably) pattern that we are not actually highlighting. Additionally the added lines would then require discussion and would detract from the focus on conditional probability.

Competing Interests: No competing interests were disclosed.

Referee Report 22 February 2016

doi:10.5256/f1000research.8562.r12330

Alan E. Wilson
School of Fisheries, Aquaculture, and Aquatic Sciences, College of Agriculture, Auburn University, Auburn, AL, USA

Title and Abstract:
For clarity, the authors might consider replacing "various" with "World Health Organization and U.S. Environmental Protection Agency".

Article content:
Using publicly available data produced from the 2007 USEPA National Lakes Assessment, the authors use conditional probability analysis to develop four chlorophyll concentration thresholds associated with two USEPA microcystin drinking water advisory targets, one WHO microcystin drinking water advisory target, and one WHO microcystin recreational use target. For most water resource managers, chlorophyll is much easier and cheaper to measure than microcystin. Given the threat that microcystins pose to human health, the purpose of this study is valuable. With that said, this study significantly overlaps with Yuan *et al.* (2014). The current study adds new advisory microcystin targets recently established by USEPA and uses a different statistical approach than Yuan *et al.* 2014. The authors cite relatively few studies. I think the authors need to more broadly consider the existing literature and describe how their findings relate to and build from past studies. Below, I provide some related studies that the authors might want to consider. I am certain that I have missed other relevant studies.

Ahn *et al.* (2011)
Beaver *et al.* (2014)
Chan *et al.* (2007)
Conti *et al.* (2005)
Dolman *et al.* (2012)
Downing *et al.* (2000)
Giani *et al.* (2005)
Graham *et al.* (2004)
Graham *et al.* (2010)
Jacoby *et al.* (2015)
Kotak *et al.* (2000)
Marion *et al.* (2012)
Orihel *et al.* (2012)
Sarnelle *et al.* (2010)
Scott and Haggard (2015)
Sinang *et al.* (2015)
Stow *et al.* (2015)
Su *et al.* (2015)
Yuan and Pollard *et al.* (2015)

Based on the 2007 National Lakes Assessment report, roughly two-thirds of the waterbodies reported no detectable microcystin (detection limit = 0.05 ug/L) despite covering a huge range of chlorophyll concentrations. And, Fig 2 suggests that a large number of sites had barely detectable concentration of microcystin across a wide range of chlorophyll. It is not clear from the text how the authors dealt with waterbodies with undetectable or barely detectable microcystin concentrations.

Presenting histograms of chlorophyll and microcystin concentrations for the study lakes would be useful.

I am not an expert on conditional probability analysis. Based on the authors' text (second paragraph in Analytical Methods section), it appears that this analysis considers multiple events over time. If their dataset includes single measurements in a waterbody, I don't understand where the temporal component comes into the analysis. Again, I could be totally misunderstanding how this analysis works and should probably read the relevant references the authors provided.

Based on increasing error in the conditional probability plots as chlorophyll increases, the reported

chlorophyll thresholds should not include significant digits (i.e., ± 0.1) but instead be whole numbers.

I would organize the information in table 1 by either concentration (low to high) or advisory type (drinking or recreational) and concentration (low to high). It might also be useful to include the number of lakes represented in each category based on microcystin. In table 2, I would add the specific microcystin concentration target under each advisory type to avoid having to look back at table 1 for these data.

Conclusions:
The purpose of this study is to use a simple measurement (chlorophyll) to determine the threat that microcystins pose to a waterbody relative to existing microcystin concentration targets. Most waterbodies lacked microcystin and Figure 2 clearly shows that there are a huge number of waterbodies across a large chlorophyll range that apparently had microcystin concentrations at the detection limit of 0.05 ug/L. I am concerned about the microcystin data at the detection limit. They appear to be false positives. I agree with the authors who acknowledged that high chlorophyll is not always a good predictor of high microcystin. What should be done for those waterbodies with high concentrations of chlorophyll but that had no or barely detectable microcystin?

Data:
I am confused about the data collected and available for the 2007 National Lakes Assessment. For example, I organized this dataset in July 2010 and found that 1158 lakes were sampled once (1152 of these lakes included data for both chlorophyll and microcystin) and 95 of the 1158 originally sampled lakes were sampled a second time in 2007. Yuan et al. 2014 (Freshwater Biology) used data for 1077 sampled lakes. The current study (as well as the National Lakes Assessment website and report) describes data for 1028 lakes. Clarity about these discrepancies is not necessarily the authors' job, but it would be good to understand why the differences exist across these datasets. Also, for this study, how were data used for lakes sampled twice in 2007?

Although all of the National Lakes Assessment data are publicly available, the authors should provide the dataset that they used for this study.

References
1. Ahn C, Oh H, Park Y: EVALUATION OF ENVIRONMENTAL FACTORS ON CYANOBACTERIAL BLOOM IN EUTROPHIC RESERVOIR USING ARTIFICIAL NEURAL NETWORKS1. *Journal of Phycology*. 2011; **47** (3): 495-504 Publisher Full Text
2. Beaver J, Manis E, Loftin K, Graham J, Pollard A, Mitchell R: Land use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation. *Harmful Algae*. 2014; **36**: 57-62 Publisher Full Text
3. Chan WS, Recknagel F, Cao H, Park HD: Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. *Water Res*. 2007; **41** (10): 2247-55 PubMed Abstract | Publisher Full Text
4. Conti AL, Guerrero JM, Regueira JM: Levels of microcystins in two Argentinean reservoirs used for water supply and recreation: differences in the implementation of safe levels.*Environ Toxicol*. 2005; **20** (3): 263-9 PubMed Abstract | Publisher Full Text
5. Dolman AM, Rücker J, Pick FR, Fastner J, Rohrlack T, Mischke U, Wiedner C: Cyanobacteria and cyanotoxins: the influence of nitrogen versus phosphorus.*PLoS One*. 2012; **7** (6): e38757 PubMed Abstract | Publisher Full Text
6. Downing J, Watson S, McCauley E: Predicting Cyanobacteria dominance in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*. 2001; **58** (10): 1905-1908 Publisher Full Text
7. Giani A, Bird D, Prairie Y, Lawrence J: Empirical study of cyanobacterial toxicity along a trophic

gradient of lakes. *Canadian Journal of Fisheries and Aquatic Sciences*. 2005; **62** (9): 2100-2109 Publisher Full Text
8. Graham JL, Jones JR, Jones SB, Downing JA, Clevenger TE: Environmental factors influencing microcystin distribution and concentration in the Midwestern United States.*Water Res*. 2004; **38** (20): 4395-404 PubMed Abstract | Publisher Full Text
9. Graham JL, Loftin KA, Meyer MT, Ziegler AC: Cyanotoxin mixtures and taste-and-odor compounds in cyanobacterial blooms from the Midwestern United States.*Environ Sci Technol*. 2010; **44** (19): 7361-8 PubMed Abstract | Publisher Full Text
10. Jacoby J, Burghdoff M, Williams G, Read L, Hardy J: Dominant factors associated with microcystins in nine midlatitude, maritime lakes. *Inland Waters*. 2015; **5** (2): 187-202 Publisher Full Text
11. Kotak B, Lam A, Prepas E, Hrudey S: Role of chemical and physical variables in regulating microcystin-LR concentration in phytoplankton of eutrophic lakes. *Canadian Journal of Fisheries and Aquatic Sciences*. 2000; **57** (8): 1584-1593 Publisher Full Text
12. Marion JW, Lee J, Wilkins JR, Lemeshow S, Lee C, Waletzko EJ, Buckley TJ: In vivo phycocyanin flourometry as a potential rapid screening tool for predicting elevated microcystin concentrations at eutrophic lakes.*Environ Sci Technol*. 2012; **46** (8): 4523-31 PubMed Abstract | Publisher Full Text
13. Orihel D, Bird D, Brylinsky M, Chen H, Donald D, Huang D, Giani A, Kinniburgh D, Kling H, Kotak B, Leavitt P, Nielsen C, Reedyk S, Rooney R, Watson S, Zurawell R, Vinebrooke R, Smith R: High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences*. 2012; **69** (9): 1457-1462 Publisher Full Text
14. Sarnelle O, Morrison J, Kaul R, Horst G, Wandell H, Bednarz R: Citizen monitoring: Testing hypotheses about the interactive influences of eutrophication and mussel invasion on a cyanobacterial toxin in lakes.*Water Res*. 2010; **44** (1): 141-50 PubMed Abstract | Publisher Full Text
15. Scott JT, Haggard BE: Implementing Effects-Based Water Quality Criteria for Eutrophication in Beaver Lake, Arkansas: Linking Standard Development and Assessment Methodology.*J Environ Qual*. 2015; **44** (5): 1503-12 PubMed Abstract | Publisher Full Text
16. Sinang S, Reichwaldt E, Ghadouani A: Local nutrient regimes determine site-specific environmental triggers of cyanobacterial and microcystin variability in urban lakes. *Hydrology and Earth System Sciences*. 2015; **19** (5): 2179-2195 Publisher Full Text
17. Stow CA, Cha Y, Johnson LT, Confesor R, Richards RP: Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie.*Environ Sci Technol*. 2015; **49** (6): 3392-400 PubMed Abstract | Publisher Full Text
18. Su X, Xue Q, Steinman AD, Zhao Y, Xie L: Spatiotemporal Dynamics of Microcystin Variants and Relationships with Environmental Parameters in Lake Taihu, China.*Toxins (Basel)*. 2015; **7** (8): 3224-44 PubMed Abstract | Publisher Full Text
19. Yuan L, Pollard A: Deriving nutrient targets to prevent excessive cyanobacterial densities in U.S. lakes and reservoirs. *Freshwater Biology*. 2015; **60** (9): 1901-1916 Publisher Full Text
20. Yuan L, Pollard A, Pather S, Oliver J, D'Anglada L: Managing microcystin: identifying national-scale thresholds for total nitrogen and chlorophylla. *Freshwater Biology*. 2014; **59** (9): 1970-1981 Publisher Full Text

I have read this submission. I 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.

Competing Interests: No competing interests were disclosed.

Author Response 01 Jun 2016

Jeffrey Hollister, US EPA, USA

I think the authors need to more broadly consider the existing literature and describe how their findings relate to and build from past studies. Below, I provide some related studies that the authors might want to consider.
- **Response:** First thanks for the fantastic list of refs! Having it linked with this publication is a resource in and of itself. We have looked at those carefully and have added several: including Ahn et al, Beaver et al, Yuan and Pollard, and Marion et al. We have not added signficantly to the background on this paper because our goal was to keep this research communication short and focused on on the chl and microcystin relationship.

Based on the 2007 National Lakes Assessment report, roughly two-thirds of the waterbodies reported no detectable microcystin (detection limit = 0.05 ug/L) despite covering a huge range of chlorophyll concentrations. And, Fig 2 suggests that a large number of sites had barely detectable concentration of microcystin across a wide range of chlorophyll. It is not clear from the text how the authors dealt with waterbodies with undetectable or barely detectable microcystin concentrations.
- **Response:** We have added some text to the Data section indicating how we deal with the detection limit. We feel it is important to keep these values in the analysis as removing them would inflate our confidence around the conditional probabilities. We hope this is clearer in our revision.

Presenting histograms of chlorophyll and microcystin concentrations for the study lakes would be useful.
- **Response:** We have chosen to present the distribution information in text and present for both chlorophyll and microcystin the range, mean, and median. Figure 2 also indicates the distribution of both. Lastly, the data are availble via [code from the GitHub repository](https://github.com/USEPA/Microcystinchla/blob/master/R/get_nla.R).

I am not an expert on conditional probability analysis. Based on the authors' text (second paragraph in Analytical Methods section), it appears that this analysis considers multiple events over time. If their dataset includes single measurements in a waterbody, I don't understand where the temporal component comes into the analysis. Again, I could be totally misunderstanding how this analysis works and should probably read the relevant references the authors provided.
- **Response:** We have added some additional text in the methods about the NLA as well as in the Discussion on NLA limitations. In short, this is not a temporal analysis and is based on a single snap shot.

Based on increasing error in the conditional probability plots as chlorophyll increases, the reported chlorophyll thresholds should not include significant digits (i.e., ± 0.1) but instead be whole numbers.
- **Response:** Done. NEED TO DO on table already in overleaf

I would organize the information in table 1 by either concentration (low to high) or advisory type (drinking or recreational) and concentration (low to high). It might also be useful to include the number of lakes represented in each category based on microcystin.
- **Response:** Table re-orderd based on concentration. Number of lakes (as percentage) included in text. Need to do directly on table in overleaf.

In table 2, I would add the specific microcystin concentration target under each advisory type to avoid having to look back at table 1 for these data.

- **Response:** Done. Need to transfer to overleaf.

Most waterbodies lacked microcystin and Figure 2 clearly shows that there are a huge number of waterbodies across a large chlorophyll range that apparently had microcystin concentrations at the detection limit of 0.05 ug/L. I am concerned about the microcystin data at the detection limit. They appear to be false positives. I agree with the authors who acknowledged that high chlorophyll is not always a good predictor of high microcystin. What should be done for those waterbodies with high concentrations of chlorophyll but that had no or barely detectable microcystin?
- **Response:** We added some discussion about this in the last paragraph of the Data section. We feel that these should be left in as removing them would erroneously inflate our confidence intervals and impact the conditional probabilities. Essentially these are lakes with very low microcystin but widely varying chlorophyll values.

I am confused about the data collected and available for the 2007 National Lakes Assessment. For example, I organized this dataset in July 2010 and found that 1158 lakes were sampled once (1152 of these lakes included data for both chlorophyll and microcystin) and 95 of the 1158 originally sampled lakes were sampled a second time in 2007. Yuan et al. 2014 (Freshwater Biology) used data for 1077 sampled lakes. The current study (as well as the National Lakes Assessment website and report) describes data for 1028 lakes. Clarity about these discrepancies is not necessarily the authors' job, but it would be good to understand why the differences exist across these datasets. Also, for this study, how were data used for lakes sampled twice in 2007?
- **Response:** We share your confusion! There are many "types" of samples included with the raw NLA data. For this analysis, we only used the probability samples (i.e. no reference samples) and only used the first visit to a lake. Additionally, lakes that had no data reported for either chl or microsystin were not included. As noted, this results in 1028 samples

Although all of the National Lakes Assessment data are publicly available, the authors should provide the dataset that they used for this study.
- **Response:** Code to access the data is available from [USEPA/microcysinchla](https://github.com/USEPA/microcystinchla). We have also added in a static .csv file to this repository of the data used for our analysis. This is listed in the "Data and software availability" section.

Thank you for your review. We have just submitted our revisions and expect the new version to be available in the next few days. Below are our responses to the specific issues you raise. We feel the paper is stronger after this revision. Any questions, feel free to contact us or comment here. Thanks again!

Specific Edits:

Title and Abstract:
- For clarity, the authors might consider replacing "various" with "World Health Organization and U.S. Environmental Protection Agency".
- **Response:** As this has already been indexed, we thought it best to limit the edits to the title.

Competing Interests: No competing interests were disclosed.