Cancer Chemotherapy and Pharmacology (2019) 84:213–216
https://doi.org/10.1007/s00280-019-03821-6

SHORT COMMUNICATION

Qualitative and quantitative variations in liver function thresholds among clinical trials in cancer: a need for harmonization

Edwin Wang[1] · Fei Song[2] · Jessica K. Paulus[3] · Douglas Hackenyos[4] · Paul Mathew[5]

Received: 13 January 2019 / Accepted: 8 February 2019 / Published online: 22 April 2019

Abstract

Purpose The liver is critically involved in drug metabolism pathways and the potential for hepatic toxicity is significant with specific cancer therapeutics. Variations in the definition of liver function thresholds that may generate heterogeneity of toxicity and efficacy outcomes across therapeutics trials in cancer require assessment.

Methods A random sample of therapeutic trials in cancer ($n=500$, general category), trials using hepatotoxic drugs (abiraterone, pazopanib: $n=181$), trials using drugs metabolized by the liver (doxorubicin, vincristine: $n=606$), and therapeutic trials in hepatic dysfunction ($n=49$) were each identified on clinicaltrials.gov. Definitions of liver function thresholds and their distribution were collated and categorized in each group.

Results A third of all trials listed on clinicaltrials.gov across the four categories failed to provide an explicit definition of liver function. Among trials with an explicit definition, a combination of bilirubin and transaminase levels was used in 33–64%, whereas a miscellaneous combination of definitions (in the general category consisting of 11 unique liver function parameters creating 17 unique combinations) was used 29–58% of the time across the four categories of studies. The Child–Pugh or National Cancer Institute Organ Dysfunction Working Group (NCI-ODWG) criteria were rarely employed (0–12% studies). Allowance for Gilbert's disease in bilirubin thresholds was identified in only 6–23% studies and for liver metastases in 2–15% of studies.

Conclusions There is a marked heterogeneity in the liver function definitions used across cancer clinical trials even when the potential for drug toxicity and altered drug metabolism is significant. Harmonization of criteria will streamline eligibility and mitigate variations in key outcomes across trials.

Keywords Liver function thresholds · Clinical trials · Cancer

Introduction

Unlike the kidney, the liver lacks evidence-based guidelines for the definition of hepatic function suitable for application to the large range of drugs that are either hepatotoxic or whose toxicity is defined by liver metabolism. Two very commonly used liver tests, bilirubin or transaminase levels, are not direct estimates of hepatic synthetic or excretory function [1]. There have been attempts to consolidate different liver function tests to assess hepatic impairment. The Child–Pugh classification consists of five clinical features, encephalopathy, ascites, bilirubin, albumin, and international normalized ratio. Each feature is rated one to three points, with three being the most severe, which contributes to an overall score. While FDA guidance in 2003 recommends the use of Child–Pugh to grade hepatic function in phase 1 clinical trials [2], there are limitations when applying this

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00280-019-03821-6) contains supplementary material, which is available to authorized users.

✉ Paul Mathew
pmathew@tuftsmedicalcenter.org

1 Tufts University School of Medicine, Boston, MA 02111, USA

2 Division of Internal Medicine, Tufts Medical Center, Boston, MA 02111, USA

3 Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA 02111, USA

4 Pharmacy Department, Tufts Medical Center, Boston, MA 02111, USA

5 Division of Hematology-Oncology, Tufts Medical Center, Boston, MA 02111, USA

metric to cancer therapeutics. Child–Pugh partly relies on two subjective measurements (the presence of ascites and encephalopathy), which could introduce misclassification. In addition, the Child–Pugh classification has not been validated with respect to elimination of drugs metabolized by the liver in the setting of cancer. In 2004, the National Cancer Institute Organ Dysfunction Group (NCI-ODWG) developed criteria to characterize liver dysfunction using four categories of severity based on degrees of elevation of total bilirubin and serum aspartate aminotransferase [3]. The CTEP (Cancer Therapy Evaluation Program) has recommended adoption of the NCI-ODWG criteria in organ dysfunction studies in cancer [4].

With the premise that variations in liver function thresholds carry the risk of generating imbalances in the observed toxicities and efficacy across clinical studies of individual therapeutic agents, we assessed the current practice of defining liver function thresholds in a sample of contemporary cancer clinical trials. We additionally characterized how liver function is defined in subgroups of trials featuring agents that were potentially hepatotoxic, agents whose toxicity is critically influenced by hepatic clearance, or therapeutic studies in the setting of hepatic dysfunction.

Methods

To identify a set of cancer clinical trials for descriptive analysis (Supplementary Fig. 1), we used the search term "cancer" among the 19,826 active, recruiting, and therapeutic clinical trials registered in clinicaltrials.gov (accessed 6/5/2018). In a pilot phase, we randomly sampled 100 studies to create a classification scheme for how liver function is most commonly described. The seven-category classification scheme was: transaminase alone, bilirubin alone, a composite of transaminase and bilirubin, Child–Pugh, "adequate organ function," Other definitions and no definition reported. For the primary analysis of describing liver function classification in cancer clinical trials in the general category, a sample of 500 was drawn from the 19,826 trials meeting search criteria. To identify studies with hepatotoxic drugs, the search strategy of "pazopanib" or "abiraterone" and "cancer" was applied (181 trials identified). For studies of agents critically metabolized by the liver, the search terms "vincristine" or "doxorubicin" were employed (606 trials identified). Therapeutic studies in cancer and hepatic dysfunction were similarly identified (49 trials). For each category of trial, the number and proportion of studies that provided an explicit liver function threshold were tabulated.

Results

Among the random sample of 500 clinical trials in cancer listed on clinicaltrials.gov, 38% ($n=191$) either did not offer a liver function definition or listed a non-specific definition of "adequate organ function" as an eligibility criterion. The number of studies that listed either "adequate organ function" or no definition of liver function in the protocol was 36% ($n=65$) for trials of hepatotoxic drugs, 38% ($n=233$) for trials of drugs metabolized by the liver, and 33% ($n=16$) for trials of patients with hepatic dysfunction.

As displayed in Table 1, among the cancer clinical trials that provided a liver function definition, 64% ($N=199$) used a combination of bilirubin and transaminase levels to define liver function, and 29% ($n=89$) used other definitions. The other definition group consisted of as many as 17 unique liver function parameters creating 11 unique combinations of definitions. (Supplementary Fig. 2). Forty-five percent ($n=52$) of clinical trials with hepatotoxic drugs that reported definitions used bilirubin and transaminase levels, and 54% ($n=63$) used other definitions. For the clinical trials that

Table 1 Table showing the proportion of liver function definitions among cancer clinical trials that provided definitions

Criteria	Bilirubin alone, % (*n*)	Transaminase alone, % (*n*)	Bilirubin and transaminase, % (*n*)	Other definitions, % (*n*)	Child–Pugh, % (*n*)	NCI-ODWG, % (*n*)
Trial category						
General category ($n=309$)	3% (9)	3% (8)	64% (199)	29% (89)	1% (4)	0% (0)
Hepatotoxic agents ($n=116$)	0% (0)	1% (1)	45% (52)	54% (63)	0% (0)	0% (0)
Abiraterone ($n=64$)	0% (0)	2% (1)	40% (26)	58% (37)	0% (0)	0% (0)
Pazopanib ($n=52$)	0% (0)	0% (0)	50% (26)	50% (26)	0% (0)	0% (0)
Liver metabolized ($n=373$)	14% (54)	2% (9)	56% (207)	26% (97)	2% (6)	0% (0)
Vincristine ($n=189$)	17% (33)	2% (3)	65% (123)	15% (30)	0% (0)	0% (0)
Doxorubicin ($n=184$)	11% (21)	3% (6)	46% (84)	36% (67)	4% (6)	0% (0)
Hepatic dysfunction ($n=33$)	3% (1)	0% (0)	33% (11)	45% (15)	12% (4)	6% (2)

Clinical trials failed to provide an explicit definition were excluded

involved drugs metabolized by the liver and listed a definition, the transaminase and bilirubin composite definition was used in 56% of studies ($n=207$), whereas other definitions were employed in 26% ($n=97$). In the setting of hepatic dysfunction, 33% of trials ($n=11$) used transaminase and bilirubin thresholds, whereas other definitions were utilized in as many as 52% of studies ($n=17$).

We found that the numerical thresholds for total bilirubin across all four categories of clinical trials had a wide distribution, with the most common threshold 1.5 times the upper limit of normal (ULN) (Fig. 1a). Among the trials that specified a 1.5 × ULN numerical threshold for total bilirubin, there was also a wide distribution of AST numerical thresholds (Fig. 1b).

Very few trials accounted for patient exclusion for elevated bilirubin from Gilbert's disease or elevated transaminases from liver metastases. Gilbert's disease was accounted for in 13% ($n=65$) of studies in the general category, 23% ($n=42$) of the trials using hepatotoxic drugs, 13% ($n=83$) of the trials using drugs metabolized by the liver, and 6% ($n=3$) of the trials in the setting of hepatic dysfunction. The presence of liver metastases was integrated into eligibility of clinical trials in the general category in 15% ($n=74$) of 500 cancer clinical trials, 13% ($n=24$) of the trials using hepatotoxic drugs, 13% ($n=81$) of the trials using drugs metabolized by the liver, and 2% ($n=1$) of the trials in the setting of hepatic dysfunction.

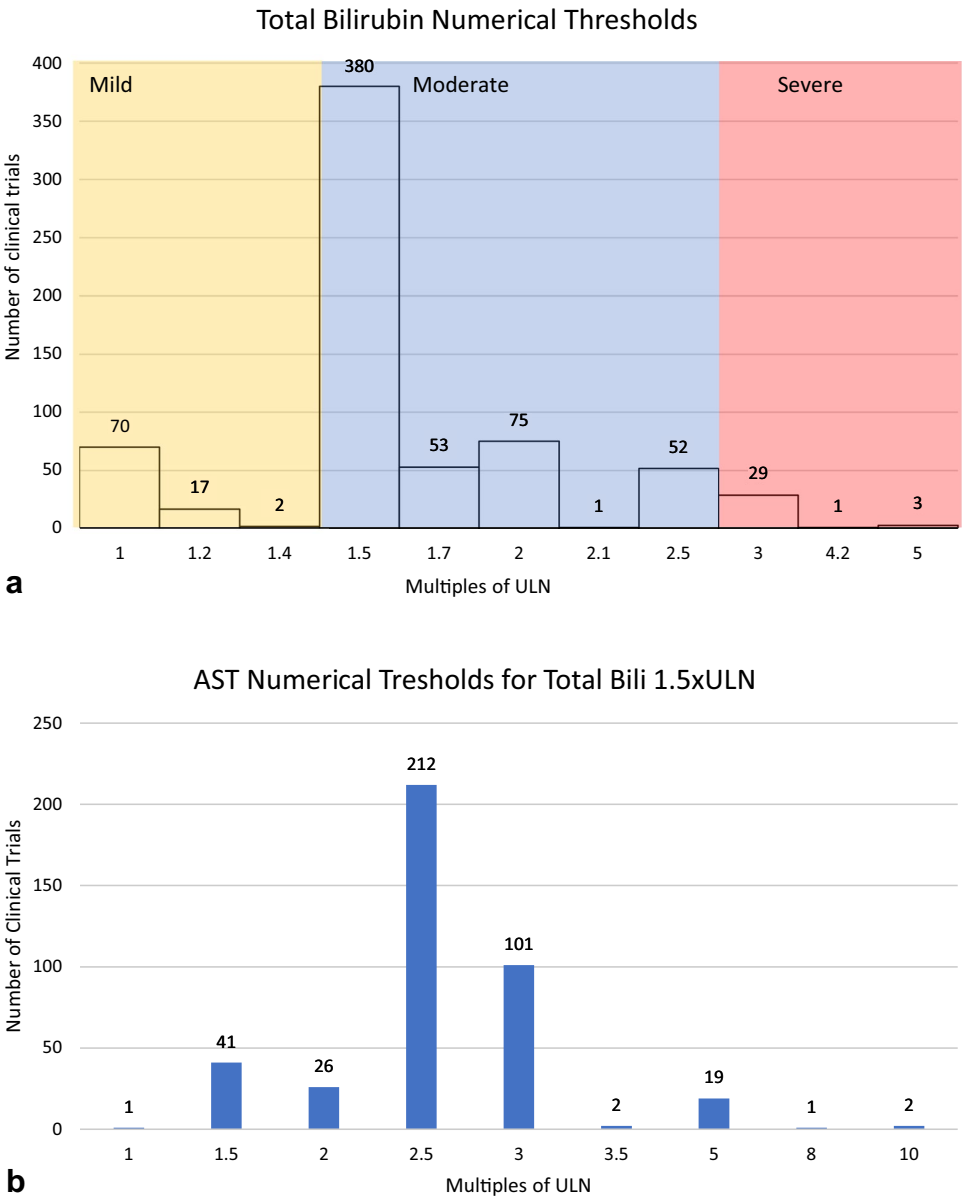

Fig. 1 **a** Distribution of the exclusion criteria for the numerical thresholds of total bilirubin across three levels of hepatic dysfunction as denoted by the NCI-ODWG and in all four categories of clinical trials. **b** Distribution of exclusion criteria for AST numerical thresholds for clinical trials that specified 1.5 × ULN for total bilirubin as an exclusion criteria

Discussion

Our data indicate that there is marked qualitative and quantitative variation in the definitions of liver function thresholds in cancer clinical trials. Less than 1% of contemporary clinical trials use the NCI-ODWG criteria for defining liver function. The most commonly used liver function threshold definition was a composite of bilirubin and transaminase levels without a consistent numerical threshold, indicating another level of heterogeneity among these definitions. A significant percentage of studies used miscellaneous composites of different liver function tests even with agents with potential for liver toxicity, agents whose metabolic fate could determine liver toxicity, or indeed in the setting of liver dysfunction. The wide variation in how liver function thresholds are currently defined has the potential to generate heterogeneous toxicity and efficacy data within trials of the same agent or in cross-trial comparisons between different agents with hepatotoxic potential. These problems are compounded by the fact that in up to a third of these trials listed on clinicaltrials.gov, no formal definition of liver function thresholds is provided.

There are two criteria endorsed by professional groups to assess hepatic dysfunction: the Child–Pugh, and the NCI-ODWG [5]. The NCI-ODWG uses readily available, objective biochemical data that has clear numerical thresholds, which provides several advantages over the Child–Pugh. There are, however, disadvantages to the NCI-ODWG criteria as well. The NCI-ODWG criteria specify that mild hepatic dysfunction is defined as total bilirubin levels greater than the upper limit of normal, or AST levels greater than the upper limit of normal. The moderate, and severe hepatic dysfunction levels are then defined as greater than 1.5 times the upper limit of normal for total bilirubin, and 3 times the upper limit of normal, respectively, with any concurrent AST level. This shows a heavy reliance on total bilirubin levels to define the metabolic capability of the liver, which is an oversimplification. This is especially concerning when the majority of trials in our study did not consider Gilbert's disease, which is present in as much as 6% of the general population, for elevations of bilirubin. A similar minority of clinical trial designs accounted for the impact of liver metastases on acceptable liver function thresholds.

Our data indicate that the Child–Pugh criteria or NCI-ODWG criteria are rarely employed to define liver function thresholds. Investigators may not see the value of these combinatorial indices given the lack of validated data to establish their use for this purpose. Both recommended guidelines have limitations in their abilities to estimate liver metabolic function. This does not, however, negate the importance of harmonization. While there is a compelling need for an evidence-based, practical, and accurate method to define liver function thresholds, the usage of NCI-ODWG criteria as a starting point across clinical trials, as recommended by the CTEP, may allow for harmonization of practice until better tools are developed. We would additionally advocate for a clear statement on the allowance for Gilbert's disease and the impact of liver metastases on these eligibility thresholds. All clinical trials listed on clinicaltrials.gov should have an explicit definition of liver function thresholds in the interest of investigators and patients who are seeking solutions for their problems.

Compliance with ethical standards

Conflict of interest The authors declare there are no conflicts of interest.

References

1. Superfin D, Iannucci AA, Davies AM. Commentary (2007) Oncologic drugs in patients with organ dysfunction: a summary. Oncologist 12(9):1070–1083. https://doi.org/10.1634/theoncologist.12-9-1070
2. (2003) Guidance for industry pharmacokinetics in patients with impaired hepatic function: study design, data analysis, and impact on dosing and labeling. FDA Guidance. https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm072123.pdf. Accessed 30 Sept 2018
3. Patel H, Egorin MJ, Remick SC et al (2004) Comparison of Child-Pugh (CP) criteria and NCI organ dysfunction working group (NCI-ODWG) criteria for hepatic dysfunction (HD): implications for chemotherapy dosing. J Clin Oncol 22(14_suppl):6051–6051. https://doi.org/10.1200/jco.2004.22.90140.6051
4. Cancer Therapy Evaluation Program (2018) Protocol templates and guidelines | protocol development | CTEP. https://ctep.cancer.gov/protocolDevelopment/templates_applications.htm#policiesAndGuidelines. Accessed 30 Sept 2018
5. Talal AH, Venuto CS, Younis I (2017) Assessment of hepatic impairment and implications for pharmacokinetics of substance use treatment. Clin Pharmacol drug Dev 6(2):206–212. https://doi.org/10.1002/cpdd.336