THERAPY OF ENDOCRINE DISEASE

Response and toxicity of small-molecule tyrosine kinase inhibitors in patients with thyroid carcinoma: a systematic review and meta-analysis

E N Klein Hesselink1,2, D Steenvoorden3, E Kapiteijn4, E P Corssmit3, A N A van der Horst-Schrivers1, J D Lefrandt2, T P Links1 and O M Dekkers3,5

1Department of Endocrinology and 2Department of Internal Medicine, Division of Vascular Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, Departments of 3Endocrinology, 4Medical Oncology and 5Clinical Epidemiology, C7-99, Leiden University Medical Center, University of Leiden, PO Box 9600, 2300 RC Leiden, The Netherlands

Abstract

Context: Many tyrosine kinase inhibitors (TKIs) have been studied in patients with thyroid carcinoma (TC). However, the effect and toxicity of various TKIs in differentiated TC (DTC) and medullary TC (MTC) patients have not been directly compared. The aim of the present systematic review and meta-analysis was to systematically summarize response and toxicity of TKIs in TC patients.

Methods: All major databases were systematically searched for publications on TKIs in TC. Primary endpoint was objective response; secondary endpoints were clinical benefit, percentage TKI dose reduction/discontinuation, hand–foot syndrome, diarrhea, and nausea/vomiting. Meta-analysis was performed using an exact likelihood approach and a logistic regression. Pooled percentages and 95% CIs were reported.

Results: In total, 22 publications were included. For DTC patients, gefitinib induced no objective responses. Pooled percentage was highest for pazopanib, 49 (95% CI 33–64)%, and was 17 (95% CI 12–24)% for sorafenib. For MTC, gefitinib and imatinib induced no objective responses, whereas sunitinib induced objective response in 43 (95% CI 14–77)%. For vandetanib and cabozantinib, these numbers were 40 (95% CI 34–46)% and 27 (95% CI 22–32)% respectively. Clinical benefit was found in 53 (95% CI 48–59)% of DTC patients on sorafenib, and in 84 (95% CI 79–88)% and 55 (95% CI 49–61)% of MTC patients on vandetanib and cabozantinib respectively. All TKIs were associated with considerable toxicity.

Conclusion: The currently studied TKIs show a modest response, while side effects are not negligible. Therefore, we suggest to solely consider TKIs in TC patients with rapid progressive disease, for whom the benefits of treatment outweigh toxicity.

Introduction

Thyroid carcinoma (TC) is the most common endocrine malignancy, with an increasing incidence over the past decades. In 2013, the estimated number of new TC cases was ~60,000 in the USA (1), compared with 22,000 incident TC cases 10 years earlier (2). Differentiated TC (DTC, which consists of papillary TC (PTC) and follicular TC (FTC)) is the most common TC type (3), whereas medullary TC (MTC) accounts for only 5%. Anaplastic TC (ATC) is rare, as it constitutes 1% of all TCs (3). Surgery is the main treatment for TC patients, with an important
additional role for radioiodine treatment in DTC (3). Ten-year overall survival rates vary according to the type of TC, being 98% for PTC, 92% for FTC, 80% for MTC, and 13% for ATC (4). However, uncontrolled local disease (including radioiodine refractory DTC) or metastatic spread is associated with a poorer survival with, for example, a 10-year overall survival of 10% in DTC patients with radioiodine refractory disease (5). Surgical debulking, radiofrequency ablation, and external beam radiotherapy can induce local control and palliation (6, 7, 8), but a well-established effective treatment is still lacking for these patients.

Small-molecule tyrosine kinase inhibitors (TKIs) are a promising new class of systemic therapy for TC patients with progressive disease (PD). These agents target the molecular TC signaling pathway at single or multiple sites (9). Many different TKIs have been studied (9), and several combinations are currently under investigation. Nevertheless, the effectiveness and toxicity of various TKIs in TC patients have not been directly compared, although this would be of great importance for clinical decision-making. The aim of the present systematic review and meta-analysis was to systematically summarize the response and toxicity of treatment with small-molecule TKIs in patients with TC.

Methods

Search strategy and study selection

PubMed, Embase, Web of Science, Cochrane, Academic Search Premier, and CINAHL were systematically searched on 9th January 2014 for relevant full-text articles, using the keywords TKI and TC. Supplementary information can be referred (Supplementary information 1, see section on supplementary data given at the end of this article) for the full search string.

Two investigators (E N Klein Hesselink and D Steenvoorden) independently assessed all studies obtained from the database search. First, all titles and abstracts were screened. Clinical trials or observational studies performed in patients with any type of TC studying the effect of any TKI were eligible. If considered relevant, full-text articles were retrieved and thoroughly assessed. Only full-text articles published in English, Dutch, German, or French, reporting on efficacy of a single TKI (defined as any -nib) in patients with DTC or MTC, with a follow-up of at least 3 months and at least ten participants of at least 18 years old were included. In case two or more publications reported on the same cohort, the publication with the highest number of patients or the longest follow-up was included for analysis.

The current meta-analysis complied with the Meta-Analysis Of Observational Studies in Epidemiology (MOOSE) guidelines (10).

Data extraction and study endpoints

Data extraction was independently performed by two investigators (E N Klein Hesselink and D Steenvoorden), using a prespecified data extraction form. In case of disagreement, a third reviewer (O M Dekkers) decided. Eligible publications and accompanying online supplements as well as data from the clinicaltrials.gov results section were considered during data extraction. Parameters extracted included sample size, median age, sex, tumor histology, prior TC therapy, tumor responses, median overall and progression-free survival (PFS), and adverse events (AEs). Of clinical trials with a placebo arm, data of both the intervention and control groups were extracted separately. The Response Evaluation Criteria In Solid Tumors (RECIST) version 1.0 (11) or 1.1 (12) were used to assess tumor response; complete response (CR) was defined as complete regression of target lesions, partial response (PR) as a decrease of at least 30% in the sum of the longest diameter of target lesions, PD as an increase of at least 20%, and stable disease (SD) as any response between a 30% decrease and a 20% increase in the size of target lesions (11). To assess the percentage of patients with a tumor response, the numbers of patients with CR, PR, and durable SD of at least 24 weeks after start of treatment were extracted. As recommended (11), these numbers were divided by the total number of patients at the start of the study rather than the number of patients evaluable at the end of the study to not overestimate tumor response. AEs were reported in accordance with the Common Terminology Criteria for Adverse Events (CTCAE) version 3.0. Grade III/IV AEs as well as AEs independent of grade were obtained. AEs of any grade were used in the analyses as all grades of AEs can affect quality of life (13, 14) and are important to consider when systemic therapy is initiated and monitored.

The primary endpoint of this systematic review and meta-analysis was objective response (PR+CR) of TKIs in patients with DTC and MTC. Clinical benefit (PR+CR+ durable SD of at least 24 weeks), dose reduction and/or permanent TKI discontinuation due to AEs, and hand-foot syndrome (HFS, or palmar–plantar erythrodysesthesi, diarrhea, and nausea/vomiting were secondary endpoints, as these AEs were most frequently reported.
We present results stratified according to the histological subtypes DTC and MTC for the endpoints objective response and clinical benefit. For toxicity endpoints, results are presented for all TC histological subtypes combined as well as stratified by subtype.

Risk of bias assessment

The following potential sources of bias were assessed for each study: i) adequacy of endpoint reporting. If RECIST (tumor response) and CTCAE (side effects) were used for reporting, an independent or central review board assessed responses, and when the duration of SD was properly stated, a study was considered at low risk of bias with respect to endpoint reporting. ii) Duration of follow-up. A median follow-up of at least 12 months was considered a low risk of bias. As TC (with exception of the anaplastic variant) is a relatively slow-growing tumor, treatment effect can hardly be distinguished from the natural course of the disease after shorter time intervals. iii) Loss to follow-up. No or limited (defined as <10%) loss to follow-up within 1 year was defined as low risk of bias. iv) Selection of patients. We considered inclusion of consecutive patients or a random sample of the inception cohort to represent a low risk of bias. Furthermore, establishment of PD according to RECIST no longer than 12 months before the start of study as inclusion criterion was considered to cause a low risk of bias. The reason is that patients with indolent disease may be included when PD is established more than 12 months before the start of the study. These patients are more likely to have durable SD as a result of the slow-growing tumor rather than as a result of TKI treatment. As the large majority of included studies were single-arm non-randomized studies, randomization procedure and concealment of allocation were not an issue. Results of the risk of bias assessment were used to explore potential between-study heterogeneity.

Statistical analysis

Meta-analysis was performed using an exact likelihood approach, as this method has been shown to give less biased estimates when compared with the approximate method, especially in studies with a smaller within-study sample size and larger between-study variance (15). A logistic regression was performed; given the expected between-study heterogeneity, random-effects models (that allow for between-study heterogeneity (15)) were used, unless the number of studies for a certain endpoint was below 5, in which case a fixed-effect analysis was used.

Pooled percentages and 95% CIs per TKI were reported for each endpoint. Stata version 11.0 (Stata Corp., College Station, TX, USA) was used for all analyses.

Results

Search results

A total of 1535 unique publications were identified by the systematic search (see Fig. 1). Of these, 1384 were excluded based on assessment of title and abstract. The remaining 151 publications were assessed full-text. A total of 130 publications were excluded, for following reasons: study population consisted of less than ten patients ($n=105$), TKI response was no endpoint ($n=13$), several TKIs or TKI combination therapy was studied ($n=5$), the study reported on an already described cohort ($n=2$), patients under 18 years old were studied ($n=1$), only intermediate results were available ($n=1$), or no patients with DTC or MTC were included ($n=3$). A phase III randomized controlled trial published in July 2014 was added (16). Finally, 22 publications were included in the present meta-analysis (16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37).

Figure 1

Flow chart of search results and study selection.
Study characteristics

Included studies were published between 2007 and 2014 (details are given in Supplementary Table 1, see section on supplementary data given at the end of this article). A total of 20 studies were trials (mainly phase II), and two had a cohort design. Of 20 trials, four were double-blind randomized controlled trials with a placebo arm (16, 33, 36, 37). The 22 studies reported on 1435 patients. Of these patients, 599 (42%) were females. The median age of the patients per study ranged from 45 to 65 years, and if described, the median follow-up per study ranged from 11 to 25 months. Furthermore, 494 controls were analyzed. Eight studies reported on sorafenib, four on vandetanib, and two each on cabozantinib and motesanib. Sunitinib, pazopanib, axitinib, gefitinib, imatinib, and selumetinib were studied in a single study only.

Responses and adverse events

An overview of primary and secondary outcome parameters at the levels of individual studies is shown in Supplementary Tables 2 and 3, see section on supplementary data given at the end of this article. One CR was described (24); this concerned a patient with DTC on sunitinib. A total of 111 DTC patients (16%) and 200 patients with MTC (28%) had an objective response. For clinical benefit, these numbers were 269 (51%) for patients with DTC and 452 (66%) for MTC patients. AEs and dose reductions or TKI discontinuation due to AEs were common. The latter was described in 7–100% of patients treated with TKIs. HFS was reported most frequently in patients on sorafenib and cabozantinib, affecting 30–93% of patients treated with these TKIs. Diarrhea and nausea and/or vomiting of any grade were frequently described in all studied TKIs, affecting 7–81 and 7–73% of treated patients respectively.

Risk of bias assessment

RECIST and CTCAE were uniformly used (Supplementary Table 4, see section on supplementary data given at the end of this article). In eight studies (36%), an independent or central review board confirmed tumor responses, and a total of 19 studies (86%) adequately reported the duration of SD. A median follow-up of at least 12 months was applicable to 17 studies (77%). Loss to follow-up was not well described in the majority of studies; only two (9%) adequately reported numbers of patients lost to follow-up. Whether included patients were consecutive patients or a random sample was not well described either; only three (14%) reported inclusion of consecutive patients or a random selection of patients. Five studies (23%) used PD according to RECIST within a maximum of 1 year before the start of the study as an inclusion criterion.

Results of meta-analysis

Figure 2 shows the results of the meta-analysis for objective response (CR+PR). Results are shown by TKI and stratified by DTC and MTC. Results from four placebo arms were pooled as well, with complete or partial tumor regression in 0.4% of DTC patients (95% CI 0–2)% and in 6% of MTC patients (95% CI 4–10%). For patients with DTC, gefitinib induced no objective responses, whereas pazopanib induced responses in 49 (95% CI 33–64)% of patients. Gefitinib and imatinib induced no objective responses in patients with MTC; sunitinib induced most

![Graph showing objective response rate](http://dx.doi.org/10.1530/EJE-14-0788)
responses, in 43 (95% CI 14–77)% of patients. However, these agents were studied in only one or two studies. Sorafenib was the best studied TKI among DTC patients (seven studies) with objective response in 17 (95% CI 12–24)%. For MTC, vandetanib and cabozantinib were the best studied, with objective responses in 40 (95% CI 34–46)% and 27 (95% CI 22–32)% respectively.

Results for clinical benefit (CR+PR+durable SD) are shown in Fig. 3. Pooled percentages for clinical benefit were 33 (95% CI 28–39)% for controls with DTC and 39 (95% CI 33–46)% for MTC patients on placebo. For DTC patients on sorafenib, these numbers were 53 (95% CI 48–59)%. Patients with MTC on vandetanib and cabozantinib had clinical benefit in 84 (95% CI 79–88)% and 55 (95% CI 49–61)% respectively.

Figure 4A shows the meta-analysis for toxicities in TC patients independent of histological subtype. Control patients had pooled percentages for dose reductions and/or drug discontinuation, HFS, diarrhea, and nausea/vomiting of eight (95% CI 6–11)%, seven (95% CI 5–10)%, 21 (95% CI 18–25)%, and 15 (95% CI 12–18)% respectively. Sorafenib and cabozantinib were associated with the highest percentage of dose reductions and/or TKI discontinuation; 70 (95% CI 54–82)% and 77 (95% CI 71–82)% respectively. The pooled percentage of TC patients experiencing HFS was highest for sorafenib with 78 (95% CI 68–85)%. Diarrhea was most frequent in TC patients treated with sorafenib: 70 (95% CI 65–74)%. For nausea/vomiting, these numbers were highest for pazopanib: 73 (95% CI 57–85)%. Toxicity subanalyses for DTC and MTC patients are shown in Fig. 4B and C.

Discussion

The current systematic review and meta-analysis aimed to systematically summarize response and toxicity of treatment with various TKIs in patients with TC. The currently studied TKIs were found to have a modest response, with pooled percentages for objective response of 17% in DTC patients on sorafenib and 40 and 27% for MTC patients on vandetanib and cabozantinib respectively. Importantly, many TKIs were studied in one or two publications only, leading to considerable uncertainty in estimated treatment effects.

To the best of our knowledge, the present review is the first to provide an overview on response and toxicity of all studied TKIs for treatment of both DTC and MTC. Previously, only one meta-analysis focusing on sorafenib in patients with DTC has been published (38). Our study provides an overview of the response and toxicity to be expected in TC patients treated with TKIs. Such prognostic information is of clear value for clinical practice as without such information shared decision-making is hampered. Our review, however, cannot provide an answer to the question as to what the best TKI is, both in terms of effect and in terms of side effects. This has two reasons. First, head-to-head comparisons are lacking, which deprives from direct comparisons between TKIs. Secondly, the number of placebo-controlled trials is rather low, which precludes the possibility of a network meta-analysis.

It should be acknowledged that included studies display considerable heterogeneity. First, inclusion criteria for eligible TC patients were not consistent as different definitions for progressive TC disease were applied. Second, the allowed time span between established PD and the start of TKI treatment differed from a maximum of 6 months to no limitation at all, which allows for inclusion of patients with less aggressive disease.
Figure 4 (legend continued)
Figure 4 (legend continued)
Figure 4
Toxicities, (A) all patients, (B) DTC patients, and (C) MTC patients. Pooled percentages and 95% CIs are shown. DTC, differentiated thyroid carcinoma; MTC, medullary thyroid carcinoma; ND, no data; HFS, hand–foot syndrome; % dose R/D, percentage of dose reductions or TKI discontinuation due to adverse events. A full colour version of this figure is available at http://dx.doi.org/10.1530/EJE-14-0788.
Furthermore, some studies only reported SD as best response within a few months after TKI start rather than durable SD. The former is an inferior endpoint to evaluate TKI efficacy, given the generally relative slow progression of TC. Another limitation is that quality of life was not assessed in any of the included studies. Toxicity due to TKI treatment occurs frequently, but the important question that matters is to what extent the quality of life is affected, and this is yet to be answered. A blinded review board for confirmation of RECIST responses was absent in the majority of studies, which may have led to overestimation of results. In the trial by Leboulleux et al. (33), for example, six PRs were observed by the researchers; however, only one could be confirmed by independent review. Finally, follow-up duration was limited. In the majority of trials, median overall survival could therefore not be estimated.

Objective response rate, which includes PR and CR according to RECIST, was chosen as the primary endpoint in the present analysis. It might, however, be argued that RECIST are not optimal to determine tumor response in patients with TC. First, not all tumor lesions are assessable with RECIST as small lung metastases, for example, are not accounted for (11). Moreover, TKIs may inhibit cell growth rather than induce cell death and a decrease in tumor size, which is measured by RECIST (and induced by cytotoxic drugs, for which RECIST were originally developed) (39). It has been proposed that PFS is a more accurate response measure; however, there is no standardized way other than RECIST to determine PFS (39), despite the proposed Choi criteria, which incorporate tumor density as well as tumor size but have not been validated for TC (40). Moreover, TC patients with progressive or metastasized disease can survive for many years (5) and can have PFS and SD for prolonged intervals without any treatment. It would therefore be opportune to determine PFS in uncontrolled studies especially when follow-up is reasonably short. Therefore, we only included durable SD of at least 24 weeks as a secondary endpoint.

Overall survival is probably the most reliable outcome measure unless validated TC-specific response criteria are available, as it accounts not only for TKI efficacy but also for drug-induced mortality. However, overall survival benefit of TKIs in TC patients has not been demonstrated yet. In randomized trials performed thus far, a PFS benefit of 5–11 months was found though (16, 33, 36, 37), while the majority of patients also experienced TKI toxicity. Future studies should therefore demonstrate whether PFS benefit of several months outweighs the impact of side effects on quality of life. Although toxicity of TKIs is considerable in both the earliest and most recent studies, tolerability of adverse events probably improved over time due to increased experience with toxicity management. Data regarding potentially improved tolerability in terms of quality of life during TKI treatment are unfortunately lacking. Importantly, long-term adverse effects of TKIs remain largely unknown for patients with TC, while these are clinically relevant as patients with TC may live for many years despite PD. The optimal treatment strategy including the timing of start of TKI for patients with progressive TC is therefore still unclear.

A real effective TKI has unfortunately not been identified yet. It has been postulated that this may be a result of resistance for certain TKIs, such as BRAF inhibitors (41). In BRAF-mutant thyroid cancer cell lines, for example, it has been demonstrated that inhibition of RAF can induce enhancement of the signaling pathway of the upstream receptor tyrosine kinases HER2 and HER3 by lifting the negative feedback. This may result in subsequent RAS activation and can therefore induce resistance to RAF inhibitors (42). Owing to the lack of clearly effective therapies, new kinase inhibitors such as vemurafenib and fostamatinib have already been studied in small numbers of TC patients (43, 44). Furthermore, several combinations of TKIs or a TKI combined with other targeted therapies such as sorafenib and everolimus have been examined. The latter has shown promising results, with 53% of TC patients experiencing a PR (45). However, the higher efficacy of combination therapy is probably associated with higher toxicity levels. In addition, selumetinib seems to be promising as it has been demonstrated that this kinase inhibitor can reverse radioiodine refractoriness in a subset of patients with metastatic TC (46). The results of a recently presented phase III randomized controlled trial of lenvatinib in DTC patients is favorable as well, as 2% of patients experienced a CR and 63% had a PR (47).

We can conclude that TKI treatment for DTC patients is thus far disappointing with respect to objective responses. The recently presented TKI lenvatinib is promising though. For MTC patients, TKIs seem more favorable with a reasonable number of PRs, especially for vandetanib. Unfortunately, side effects are not negligible. For TC patients with PD without any treatment options left, initiation of TKI therapy may be a serious last option. TKI treatment is, however, still initiated in patients without rapid progression, for whom the toxicity may not outweigh the benefits of TKI treatment. We therefore suggest to consider TKIs in case of rapid PD, until data become available demonstrating that treatment at an earlier stage might be beneficial.
Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/EJE-14-0788.

Declaration of interest
All of the authors can assure that the manuscript represents honest work, and that no actual or potential financial interest is capable of influencing judgment.

Funding
T P Links received research support from Genzyme and was part of an advisory board for Astra Zeneca and Bayer. E Kapiteijn was part of an advisory board for Astra Zeneca.

Acknowledgements
The authors are grateful for the important suggestions for the manuscript advisory board for Astra Zeneca.

References
15 Hamza TH, van Houwelingen HC & Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. Journal of Clinical Epidemiology 2008 61 41–51. (doi:10.1016/j.jclinepi.2007.03.016)


Received 15 September 2014
Revised version received 20 November 2014
Accepted 7 January 2015