Proposed Tool to Compare and Assess the Applicability of Taste Assessment Techniques for Pharmaceuticals

David Clapham , Joanne Bennett , Anne Cram ,
Angela Discihnger , Sabine Inghelbrecht ,
Anne-Marie Pense-Lheriter , Fabrice Ruiz , Smita Salunke ,
Julia Schiele , Jessica Soto , Jennifer Walsh , Catherine Tuleu

PII: S0022-3549(21)00471-8
DOI: https://doi.org/10.1016/j.xphs.2021.09.002
Reference: XPHS 2511

To appear in: *Journal of Pharmaceutical Sciences*

Received date: 8 July 2021
Revised date: 2 September 2021
Accepted date: 2 September 2021

Please cite this article as: David Clapham , Joanne Bennett , Anne Cram , Angela Discihnger , Sabine Inghelbrecht , Anne-Marie Pense-Lheriter , Fabrice Ruiz , Smita Salunke , Julia Schiele , Jessica Soto , Jennifer Walsh , Catherine Tuleu , Proposed Tool to Compare and Assess the Applicability of Taste Assessment Techniques for Pharmaceuticals, *Journal of Pharmaceutical Sciences* (2021), doi: https://doi.org/10.1016/j.xphs.2021.09.002

Proposed Tool to Compare and Assess the Applicability of Taste Assessment Techniques for Pharmaceuticals

Authors: -

David CLAPHAM [1*], Joanne BENNETT[2] , Anne CRAM[2], Angela DISCIHNGER[3], Sabine INGHELBRECHT[4],Anne-Marie PENSE-LHERITER[5], Fabrice RUIZ[6], Smita SALUNKE[7], Julia SCHIELE [8], Jessica SOTO[9], Jennifer WALSH [10] , Catherine TULEU[7]

Affiliations

1. Independent pharmaceutical consultant, Bishops Stortford, UK (* Corresponding Author)

2 Pfizer R&D UK LTD, Sandwich, UK

3 Pharmaceutical Research and Development, F. Hoffmann-La Roche Ltd., CH-4070 Basel, Switzerland

4 Janssen Research & Development, Johnson& Johnson, Beerse, Belgium

5 Ecole de Biologie Industrielle, Cergy, France

6 ClinSearch, Malakoff, France

7 UCL School of Pharmacy, London, UK

8 Abbvie, Ludwigshafen, Germany

9 Novartis, Basel, Switzerland

10 Jenny Walsh Consulting, Nottingham, UK

All authors are current or previous members of the EuPFI Taste Assessment and Taste Masking work stream.

Graphical Abstract

Abstract

Palatability is amongst the most important determinants of whether or not a child will take a medicine. In order to increase concordance with treatment regimens it is often necessary to utilise a range of formulation techniques to improve the palatability of medicines. These can include selecting a different molecule or version of a molecule (such as a different polymorph or salt form), various taste masking techniques and/or the inclusion of flavours and sweeteners.

In order to be able to understand the taste of the Active Pharmaceutical Ingredient (API) and to validate the formulation approach used, it is necessary to be able to use the most reliable taste evaluation method possible. Multiple in vivo and in vitro methods exist nowadays or are proposed

in the literature but are often little understood by the pharmaceutical product development community. In particular, different methods may be more relevant at different stages of product development. The aim of this article is to propose a tool to guide the selection of the most appropriate method for the desired evaluation. A spreadsheet-based tool is proposed that is designed to allow the systematic assessment of the applicability of any taste assessment technique existing or new to the users proposed application. A series of criteria are defined that will allow the user to assess the analytical, usability and availability factors for the technique that is being considered. Such a systematic review will help the user to understand the benefits and risks of using each methodology for that application. The use of the tool is illustrated based on currently available data and literature. As new/existing methods are developed/improved, the outcomes of the tool may change.

Introduction

The assessment of taste is a vital aspect of pharmaceutical product development [1,2, 3] of a formulation that may be detected in the mouth. It is well known that aversive taste is likely to inhibit the acceptability of medication to the patient. It is the single most cited reason for lack of concordance with prescribed oral medication. This is particularly true in children [4, 5, 6].

Current regulatory guidance requires that applicants demonstrate the acceptability of new formulations of 'innovator' products for paediatric use which must be developed concurrently with those for adults unless a waiver is obtained [7]. Regulators also encourage patient centric formulation in other age groups and particularly the elderly [8].

In response to the requirement to formulate acceptable products and to demonstrate acceptability of formulations, a range of possible methods for assessing the taste of the Active Pharmaceutical Ingredient (API) used in the formulation, the formulation itself and the effectiveness of any taste masking methodologies applied, have been developed, proposed or are in development. A review of each potential method is outside the scope of this current paper.

An ideal taste assessment technique would be one that could be applied to; help facilitate the selection/design of an API with acceptable organoleptic characteristics; assess the efficacy of taste masking approaches where these are required; choose formulations from a range of prototypes for further taste testing; and assess the stability of the organoleptic aspects of formulations over the shelf life of the product.

No single technique is currently optimal in all these applications. Therefore, a systematic assessment tool is proposed to aid users in selecting the appropriate taste assessment methodology for use in any specific application. Previous literature [9] has reported a similar evaluation of various taste assessment methodologies. However the tool we propose is more extensive, is not limited to non-human techniques and is designed to be applicable to the user's specific application. Furthermore the proposed tool is designed to be applicable to both existing technologies and any that will be developed in the future. Application of the tool will allow a clearer understanding of where a particular assessment method is applicable and where its use may not be suitable.

The criteria proposed can also be viewed as attributes of an ideal taste assessment methodology. As such we hope that these attributes will assist developers of existing and novel taste assessment methodologies.

The Tool

This tool has been developed by a workstream of the European Paediatric Formulation Initiative (EuPFI) consisting of members from industry, academia and independent consultants with experience of pharmaceutical product development and some familiarity with various taste assessment methodologies.

To illustrate the use of the tool we have employed the proposed criteria to assess the application of human panel testing; two types of e-tongue; the rat brief access taste aversion method (rat BATA); a zebrafish model; an amoeba model and cell-based models to the determination of the bitterness of a moderately bitter, highly water soluble API and its formulation into a liquid dosage form for the paediatric market. In principle the tool is applicable to any patient group (paediatric, adult, geriatric). Given the potential differences in taste responses of different age groups and/or the amount of information available it is possible that the outcome of the use of the tool may differ. Our evaluation is paediatric focused due to the expertise of the authors. The results of this evaluation are summarised below. A detailed exposition of how they are arrived at can be gained by consulting the Tool and Comments tables in the Supplementary Data file.

The tool is a spreadsheet designed to allow the assessment of the applicability of any new or existing taste assessment technique to the user's proposed application. The criteria have been designed to be capable of being scored numerically or at least categorically (1 [red], 2 [amber], 3 [green]). Score guidance is provided for each criterion in the Score guidance table in the Supplementary Data file. The criteria have all been ascribed a descriptor beginning with the letter R in order to help make them memorable. Techniques can be compared on the basis of a 'heat map' or an overall numerical score. Where a 'heat map' approach is used the technique with the highest number of green scores will be the best, whilst for the numerical approach the one with the highest overall score will be optimum. Clearly users should select the taste assessment approach (or approaches) that are best suited to their application.

Scores should be assigned based on the best available evidence which can be found either in the literature, from the suppliers of any equipment or taste assessment services that the user wishes to evaluate or from in house knowledge.

It is acknowledged that in some cases information will not be available. In these cases a score of 0 will be assigned. Where some data exists but is not considered fully reliable a score of 0.5 can be used. If colours are being employed, then a score of zero will be represented by a white cell whilst a score of 0.5 will be represented by a grey cell. This will prompt the user that this aspect is yet to be fully studied for that technique.

There may also be cases where supporting evidence is available but there may be some difficulty in deciding which of two categories to assign. In this case a pale colour code should be used or a numerical value of, for example 1.5 (if the data is somewhere between red and amber). An example would be where there are quality studies addressing the particular aspect, but they are inconsistent.

As discussed above, this practice will prompt the user to exercise due caution in interpreting data regarding that aspect of that technique.

Where a criterion in the chart is a summary of a number of several factors, the score/colour should be assigned on the basis of either an overall average of the individual factors or on the basis of the most important factor as judged by the user for their individual application. Where this would be misleading it may be desirable to include up to two colours/numbers in the appropriate cell. The reason for decisions made should be captured and justified.

Depending on the application in mind a white, grey/pale or even red coding or their numerical equivalents need not necessarily mean that the technique has no utility. However, it does mean that results need to be interpreted cautiously in the light of these findings and reported with appropriate caution/caveats.

Unless it is self-evident why a score/value/colour has been assigned it is desirable to capture the reason. We suggest that this is done by including comment in the spreadsheet or a linked document.

The criteria we have suggested may also be regarded as an aspirational list of the features of an ideal taste assessment technique. We hope that they will stimulate the development, or further development, of methodologies to aid taste assessment in pharmaceutical product development. They are presented in an approximate order of importance from 'need to have' to 'nice to have'.

As an example of how to apply the tool, the authors have scored some of the more common taste assessment methodologies (and a few more experimental ones) based on their understanding of the available data and a typical application in pharmaceutical product development. Other users may arrive at slightly different scoring based on their own application and knowledge.

Since research into taste assessment techniques and methodologies in on-going, with the emergence of new technologies, and enhanced understanding of existing ones, the scoring presented in this paper is of necessity a 'snapshot' of our current understanding of the taste assessment methodologies that we have assessed. Users should update their assessment in the light of emerging data.

The choice of criteria is based around a number of themes. These include analytical factors (how well the method performs as an analytical technique), usability factors (how easy it is to perform an assessment and understand the results obtained) and availability factors (how readily accessible the method is). Where appropriate we suggest that techniques are reviewed under three headings.

API - The utility for the assessment of a single component such as the API

Formulation - The utility for the assessment of a mixture of components including soluble and insoluble ingredients. This classification will also consider the ability of the technique to measure the extent of taste masking required and/or achieved

Stability - The utility for the assessment of changes in taste over time

Scope

The tool as currently envisaged covers only the utility of the various taste assessment methodologies to provide useful information about any aversive taste of the molecule or formulation. For many medicines it is the taste, and in particular the bitterness of the molecule/formulation that is the major determinant of the overall organoleptic experience.

Other factors that are also important for the assessment of palatability and/or overall patient acceptability (e.g. smell, mouthfeel, grittiness, trigeminal effects etc.) are generally outside the scope of these techniques. However similar questions may be relevant to methodologies designed to measure these aspects also.

The tool also concentrates only on the technical aspects of the methodologies assessed. Ethical considerations are also clearly important but are outside the scope of this review.

Criteria

For the purposes of this tool we have identified the criteria as shown below. In several cases understanding the definition may be aided by referring to the scoring guidance in the supplementary data file. Given the specialist and often multifactorial nature of sensory assessment, the definitions we have used may differ somewhat from similar terms used for other analytical techniques (e.g. HPLC) and their validation.

The outcome of the case studies we have evaluated are summarised after each individual criterion.

1) Rationale

The main aspect to consider is whether there is a clear understanding of what the technique is actually measuring and the mechanism via which the technique identifies both the nature and intensity of the taste. Clearly, ideally the rational will be understood in which case it will be scored 3 [green]. If the mechanism of action is understood, it is easier to be able to deduce *a priori* whether the technique is likely to be applicable a sample being measured for the first time, or not. In some cases there may well be a clear demonstrated correlation between the output from the technique and the human response to the sample even where the mechanism of that correlation is not known. This is not ideal, particularly for samples which are being measured for the first time since it is hard to be sure that the correlation will still apply for those samples. Depending on how much is known, a score of 2[amber] or even 1[red] may be appropriate. This knowledge, if available, could help the design of molecules with reduced aversive taste and/or guide taste masking activities.

Although not essential, a secondary but desirable outcome would be information about the mechanism of action of the generation of the taste response in the human (e.g. which taste receptors are being stimulated/blocked).

All the techniques that we evaluated have a clear rational or it is clear where they do and do not apply and hence scores of 3[green] or 2[amber] (for the amoeba and Zebrafish models) have been assigned.

2) Repeatability

This is the ability to repeat a result within acceptable limits on multiple occasions. The term is defined in several guidance documents relating to standard analytical method validation. Given that standards with known quantified taste are yet to be agreed we are using it in a slightly different way.

A sample (or samples) measured on different occasions provides the same numerical value or at least the same rank order with the same magnitude of interval between samples. For example sample A is always twice as bitter as sample B. Ideally this should be because samples A and B give the same numerical value on each occasion the samples are measured. If this is not the case, then they should move proportionately. At the least the technique should be applicable to samples measured side by side on the same day. The utility improves if results are repeatable over several days. For assessment of taste in stability studies it will either be necessary to compare the aged sample to a standard which is known not to change in terms of taste or for results to be repeatable for the full period of the stability protocol. Refer to the Score Guidance in the supplementary data for more detail.

For the example we investigated, repeatability was uniformly good (3[green] within 1 day ; varied from unknown (Zebrafish, Cell models) to poor (some e-tongues) inter day and tended to be poorer the longer the gap between measurements. The rat BATA model was assessed as best overall in terms of repeatability.

3) Reproducibility

This criterion is designed to capture the ability to get the same results, preferably numerically but at least proportionally, for nominally the same experimental set up on different devices of the same type, in different laboratories or using replacement sensor sets/assessors. It measures the ability to get the same results at different times and in different places. Reproducibility is important if results are to be transferred between users and if results generated at different times, such as is the case for assessment of the stability of taste is being evaluated or when the taste of API's discovered at different times are being compared based on historical data. A methodology that is not reproducible will require a careful design of the evaluation protocol in order to achieve reliable data. For example, extensive calibration may be required.

In our example, all methods assessed gave some concern in terms of reproducibility where it was possible to assess this, but in most cases strategies to mitigate the issue have been identified. For the zebra fish and cell based models there was insufficient data available to fully evaluate this and a score of 0.5 was given.

4) Robustness

This factor covers two aspects of robustness. Firstly, there is the physical robustness of the equipment/assessors/sensors i.e. their resistance to breakage, physical damage or fatigue and the speed of recovery from any factors that affect their performance. Secondly, it assesses the tolerance to a wide and realistic (or at least well characterised) range of factors such as

- Temperature
- pH
- Ionic strength

- Buffer capacity
- Co-solvents
- Viscosity
- Presence of insoluble components
- Dosage form (e.g. solution, suspension, paste, powder)
- Sample volume
- Sample concentration

The more robust the methodology the higher will be the score.

All the methods we evaluated were either quite robust or their limitations are well known. The least robust methodologies were human taste testing due mostly to the phenomenon of assessor fatigue; e-tongues where sensors could be saturated/poisoned and cell based methods where cells damage/death could occur.

5) Range

This criterion aims to capture three aspects namely, the applicability to a wide concentration range (dynamic range); to a wide range of pharmacophores (and/or it is clear where it does not apply) and; to a range of taste modalities. In terms of the concentration range it is desirable to be able to produce a full dose response curve whether the sample has a strong taste (e.g. intensely bitter) or a weak taste (e.g. mild bitterness). The more widely applicable the better the score will be.

In our examples all methods were able to cope with a reasonable concentration range, albeit some more readily than others. In this regard the ratBATA model was best adapted to providing a full taste/concentration profile in a reasonable experimental time frame. Most were also sensitive to a range of pharmocophores and different taste modalities. One exception is the amoeba model which is only sensitive to bitterness and has only been applied to a limited number of pharmacophores to date and on at least one occasion did not correlate with the human response.

6) Reliable Quantitative measurement

A quantitative measurement in readily understood units that allow a clear understanding of the likely human response is preferable to a qualitative or rank order output. It is also necessary to understand the parameters within which the results are likely to be reliable and those where the data should be discarded or at least treated with caution. This factor also covers external validation versus a widely accepted taste assessment, preferably that in humans. If there is an offset between the output of the method being evaluated and the reference method this should be consistent and well characterised.

Since correlation with the human taste response is the target all data from human evaluation scores 3- green for this criterion. A growing range of studies using the rat BATA model shows an excellent correlation with human data. The rat response is driven and thus the concentration response is offset from that of the human, but this offset is consistent and so this also scores 3-green. Published data for e-tongues suggests quantitative data can be obtained but relies on correlation with prior knowledge of the human taste response to the API under study. It is not always possible to deduce the human response from the sensor response alone although in some cases broad categories (very

bitter, bitter, sweet etc) can be deduced if the sensors are properly calibrated, hence the score is lower in our example. To date information concerning how to quantify and correlate information from the cell models is lacking in the literature.

7) Readily Interpretable

A competent trained scientist can readily provide information from the data set. The technique clearly identifies both the type (aversive, bitter, sweet, sour etc.) and the magnitude of the taste of the sample in a manner that has a clear correlation/relationship to the human response. If this is not the case it must be clear where the technique applies and does not apply.

In our evaluation even where reliable quantitative data were not readily available useful information could be gleaned from all the techniques though the human and ratBATA were most universally applicable in this regard.

8) Rapid Results

For maximum utility in pharmaceutical product development an ideal technique would be easy and quick to set up, have high to moderate throughput and provide results for samples within a few (preferably one) working day(s).

Aspects to be considered are

- System set up – including calibration, conditioning (of sensors etc.), training (e.g. of rats), equilibration time and other similar analytical factors before test samples can be measured.
- Duration of the test - including the time for measurement of a single sample including washout/re equilibration time and the number of replicate measurements required for a single sample and hence the overall time required to obtain a data set for a single sample
- Data manipulation/ analysis and correlation with human data. If complex manipulation is required this will tend to be long - but IT solutions may facilitate this.

Some judgement will be required here. For example a particular technique may require considerable set up and calibration but if it is repeatable and reliable then this may only need to be undertaken infrequently and hence the time required may be mitigated in the light of the number of samples that can be assessed before a recalibration is required. Rather than give this criterion a colour/score, state a typical sample turnaround time.

We did not score this aspect.

9) Readily Available

Availability and the related aspect of affordability encompasses acquisition costs and maintenance costs as well as ongoing sample costs (such as costs of sample preparation and disposable components). This may be summed up in a low fee for service cost from an external supplier. Costs will typically be higher if the technique is only available from a single supplier and lower if it is readily available from multiple suppliers.

For our example the availability of several sensory analysis companies that can undertake such studies and the availability at reasonable cost of e-tongue assessments either in house or via

contractors led to a score of 3- green. The rat BATA model is not yet as readily available, but the model is stable for many months once calibrated and so the score given was 2 – amber. The other methods are not widely available and so a score of 1- red was given.

10) Regulatory awareness

Although data that is scientifically reliable may guide molecule choice and/or formulation development it would be clearly beneficial if regulators were willing to accept the output from the technique either alone or as supporting data, alongside studies in humans, to demonstrate the acceptability of a formulation to the potential user. When used as supporting data the results may perhaps aid the design of the definitive human study and reduce the number of volunteers that are required to produce a reliable result.

Clearly properly conducted human sensory studies are acceptable to regulators. Regulators are aware of e-tongue and in some instances will accept e-tongue data as supporting data. The other techniques are in their infancy and so are unlikely to be acceptable with the possible exception of cell data since many other cell-based assays used in drug discovery and toxicology do provide data that regulators accept.

11) Research Activity

This aspect is designed to capture the level of use and understanding of the technique within the research community, both within academia and industry. One way of measuring this factor is the number of published papers on the use and understanding of the technique. Patents may also provide a useful information source.

This aspect is likely to change rapidly.

12) Response Kinetics

This is the final scored aspect of the tool. Human taste response includes not only the type and maximum intensity of the taste but also how quickly the taste develops (taste latency) and how quickly it clears (aftertaste and persistence of aftertaste). It would be very beneficial to be able to obtain all this data from a single sample measurement. The authors acknowledge however that this could be very difficult, and the minimum requirement is that the experimental output does not mislead the user in this respect.

This is one aspect where the reported capabilities of the two e-tongues assessed markedly differ. Due to the sensor design and the way the data is collected the INSENT e-tongue claims the capability to capture the kinetic response. Clearly humans can do the same, though the experimental protocol needs to be designed to capture such data. Some studies of the amoeba model suggest that this might be possible. All other methods are not (at least currently) able to provide such data.

13) Key limitations

This is a free text field allowing the user to highlight any known limitations that may affect the use of the technique for their application. Examples include but are not limited to: -

- Prerequisites for use that are likely to be problematic to meet

- Non applicability to certain sample types (e.g. sweet samples, suspensions)
- Lack of correlation with human data
- Restricted availability
- Regulatory acceptance

Many of these aspects will have been captured elsewhere in the scorings assigned but adding a note here will alert users of the document in a more readily accessible form.

14) Number of compounds

Again this is an (optional) free text field allowing the user to record the number of discrete compounds that have been assessed using the technique. This information may be hard to identify but may be available from literature and/or in house experience. Such data could help inform the scoring of the Research Activity criterion, however recording the number of compounds separately here will help users of the document to see at a glance the level of use.

Results

The authors have applied this tool to taste assessment methodologies with which they are familiar. The results of this review are provided below as an example of how this tool may be used in practice. It provides the overall evaluation showing that no method, not even human evaluation is fully ideal but, as expected this is the 'gold standard' against which other methods may be assessed. The next most appropriate method in all applications is the ratBATA method. E-tongues can provide useful data in some applications but even then, samples usually need to be assessed side by side to provide reliable data and correlation with human data needs to be established on a case-by-case basis.

The other methods assessed have more limited utility or do not yet have sufficient data published to be sure where they are useful and where they are not.

Include Overall results here

Consult the supplementary data file for details of how these overall scores were derived.

Conclusions

A spreadsheet-based tool is proposed that permits a thorough assessment of any taste assessment tool that a user wishes to evaluate for a specific application. Use of the tool will help the user identify the best assessment methodology to utilise in any specific application and will highlight the benefits and risks of using that methodology. The authors also hope that the identification of these criteria will assist those developing new applications of existing techniques and/or development of novel taste assessment technologies.

References

[1] European Medicine Agency. (2013) Guideline on pharmaceutical development of medicines for paediatric use. EMA/CHMP/QWP/805880/2012 Rev. 2.

[2] Bryson, S.P. (2014) Patient-centred, administration friendly medicines for children - an evaluation of children's preferences and how they impact medication adherence. *Int J Pharm* 469 (2), 257-259

[3] Ranmal, S.R. et al. (2018) Methodologies for assessing the acceptability of oral formulations among children and older adults: a systematic review. *Drug Discov Today* 22 (17), 30477-30474

[4] Nordenmalm, S. et al. (2019) Children's views on taking medicines and participating in clinical trials. *Archives of disease in childhood*, archdischild-2018-316511

[5] Venables, R. et al (2015) Determination of Formulation Factors That Affect Oral Medicines Acceptability in a Domiciliary Paediatric Population. Int J Pharm 480(1-2):55-62

[6] Venables R, et al (2015) Problems with oral formulations prescribed to children: a focus group study of healthcare professionals. Int J Clin Pharm. 37(6):1057-1067.

[7] European Medicine Agency. (2013) Guideline on pharmaceutical development of medicines for paediatric use. EMA/CHMP/QWP/805880/2012 Rev. 2.

[8] European Medicines Agency, Concept paper on the need for a reflection paper on quality aspects of medicines for older people, 44 (2013). https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-need-reflection-paper-quality-aspects-medicines-older-people-first-version_en.pdf.

[9] Mohamed-Ahmed AH, Soto J, Ernest T, Tuleu C (2016) Non-human tools for evaluation of taste in the design and development of medicines: a systematic review'., *Drug Discovery Today* 21(7):1170-80.

Overall Results