Dear Ms. PD Dr. Wild, dear Mr. PD Dr. Wernly, dear Mr. Goetz,

Thank you very much for translating our Open letter for review of interested colleagues. It is of course necessary that all relevant documents are available in a common language, in this case, English. We allowed us to improve the “deepl.com” translation in few places, when our message was not clear. We appreciate very much the opportunity to discuss our points of critic openly.

As you put your comments in three different documents, a letter in German, a not identical letter in English and a comprehensive table addressing our specific points of critic, we structure our response the same. The quoted literature of this manuscript as well as of the table can be found at the end of this document in alphabetical order. We apologize if we may have cited not clear enough in our Open letter.

Let us first state that your response and explanations do not at all clarify the diverse inconsistencies of your HTA. You several times mix up your and our different interpretation of the evidence (which is normal in a scientific discourse) with your inconsistent conduction of the assessments. Those are clearly two different issues. You often discredit evidence or documents coming from industry side just with the argument that they come from industry side. The aihta itself is a GmbH.

We think it is not important from which side people or studies come, but how the quality of the people and their work is. Often, just the money comes from industry side for the preparation, not for the content. You censure this, as well, thereby discrediting physicians who stand for their opinion without being payed. In this sense, you extensively prefer the Meta-analysis of Masri et al. (Masri, JACC 2019), because there is no industry sponsoring declared, not considering that their good results and their negative discussion hardly fit together. By discrediting and ignoring publications, which transparently declare conflicts of interests, one gives reason to hide such conflicts. This cannot be in the interest of any party.

We recommend the interested reader to build his/her own opinion by going through our diverse points of critic and ask you, the HTA authors, again to retract this inappropriate HTA Update from the internet and databases.

Dr. Michael Nürnberg

Dr. Frank Semrau

Reply to the AIHTA-response, German letter

The authors of the HTA-update, Goetz, Wernly, and Wild (in the following called, The Authors) state, that it was not their intention to question narrow indications for the WCD. We appreciate that.

They furthermore say, in their current reply to our critic, they completely refuted specifically accusations of not understanding the specific medical topic and incorrect conduction of the HTA. We see this differently and we encourage the interested reader to dive into the written discussion for every single point, even if this is a lot of material and might be a bit strenuous.

We would like to excuse our sometimes straight-forward wording. We would not be so direct and impolite, if it was not that low scientific quality, discrediting people and institutions and threat of patient´s lives came together.

We do not doubt that the systematic literature search was conducted according to international standards. We do not doubt that RCTs are the current gold standard for comparative studies and that the ITT Analysis is the generally preferred analysis type.

However, as David L. Sackett put it, "Evidence based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions. […] …some questions about therapy do not require randomized trials (successful interventions for otherwise fatal conditions) (…)." (Sackett, BMJ 1996)

This exactly applies to the WCD. The WCD is "a successful intervention for an otherwise fatal condition." Furthermore, the WCD uses defibrillation, which has notoriously a so-called "dramatic effect". For verification, see a publication by one of the members of the GRADE group. (Glasziou, BMJ 2007)

We furthermore like to point to the famous Bradford Hill Criteria (Hill, Proceedings of the royal society of medicine 1965), which also clearly indicate that the causation between a WCD, the termination of SCA and the survival of "an otherwise fatal condition" is true.

When we are clear about the truth of an effect, measures to protect us from drawing erroneously positive conclusions about the confidence in this effect, such as looking on the primary outcome only, or even corrections for multiple testing, are superfluous.

Interestingly, The Authors imply also indirect/historical comparisons to proof a dramatic effect in certain indications. Unfortunately, they excluded, though prospective, the independent case series of Auricchio (Auricchio, 1998) and Reek (Reek, 2003). Both studies tested the WCD on patients, artificially induced with ventricular fibrillation (VF). The WCD detected and terminated VF reliably. Had those studies carried along a control group, all patients of the control groups were dead. This surely stands the test of the IQWiG, The Authors mention. It is important to acknowledge that the WCD does not treat indications or etiologies but **life threatening events**, which are similar in different etiologies. The efficacy to do so is close to 100%. This is also true in real life with a successful termination Rate of about 95% (Nguyen 2018). The chance of dying by VF is close to 100%, as well.

We appreciate that the conduction of this HTA was transparently reported, because otherwise, there was not a chance to discover the inconsistencies.

The Authors report about the publication of their first HTA on the WCD (Ettinger, 2017) and the discourse after a critical appraisal by a group of physicians and scientists. The citation of The Authors includes the original critical appraisal by Sperzel et al. with a first reply of Ettinger et al. (Sperzel, 2018a), as well as the second reply of Ettinger et al. (Ettinger, 2019). However, the mentioning of the

second critical reply by Sperzel et al. between the two Ettinger replies (Sperzel, 2018b) was omitted. This is selective reporting. The reader is not provided with all relevant information for building his own opinion. In the context of this discourse, it might be of relevance that some points, strongly defended by Ettinger et al. in their replies to the Sperzel group, were subject to change during their collaboration with the AGENAS colleagues. In fact, some issues, such as the initial comparator use, were corrected. (Chiarolla, 2018)

In the following, The Authors address again their judgement of low confidence in the outcome total mortality in VEST and argue with the problem of multiple testing. We already mentioned above, that such a test is not necessary in the context of a true dramatic effect. Interestingly, as support for their opinion The Authors quote a publication, which critically illuminates several studies, including the VEST trial, from a statistical point of view, including the issue of multiple testing (Pocock, 2018). Of note, Pocock and Collier do not bring up the point of multiple testing in the context of the VEST trial. Therefore, this is a misleading quotation by The Authors. Even more important, the statisticians themselves seemingly do not consider multiple testing as appropriate in this context.

The Authors state seriously, that VEST showed, randomization is (still) possible in the context of a WCD without ethical concerns. We strongly disagree. Not only is it in our view impossible to expose patients knowingly to a deadly threat, when at the same time a therapy with dramatic effect is available, but furthermore, as pointed out above, such an RCT is absolutely unnecessary, let alone in several indications as implied by The Authors. It is unclear, which specific question The Authors want to address with an(other) RCT.

Reply to the AIHTA-response, English letter
("Response to the open letter from Dr Semrau (dating February 09th 2023) on the second update health technology assessment (HTA) of the wearable cardioverter defibrillator (WCD)")

The Authors emphasize, that verification of treatments can ("definitely") only be accomplished in comparative trials. They state, *"This is crucial to avoid exposing patients to unnecessary risks (including from the omission of other, potentially more effective alternatives) and optimize the use of resources."*
In fact, The Authors named several main outcomes in their HTA, which do not at all benefit from a control group. They specifically looked for arrhythmic and total mortality, **compliance**, **patient satisfaction**, QoL, hospitalization rate, **appropriate therapies**, **shock success rate**, **delayed shocks** (response button use), as well as **serious adverse events** and adverse events. We see that about six main outcomes of this HTA are device related, meaning they **can only occur in patients with the device**. For example, the inappropriate shock rate in a control group without defibrillator will foreseeable be zero. It is unreasonable to demand comparative trials for such parameters. Rather are large populations, such as registries, qualified for reliable figures.
It is indeed crucial to avoid exposing patients to unnecessary risks. Therefore, it would have been beneficial to compare the risk of inappropriate shocks to the risk of dying, or the risk of temporary skin irritation to the risk of wearing life long an implanted device. We are curious to know, which "potentially more effective alternatives" to the WCD The Authors have in mind, because the WCD currently serves an **unmet medical need**. In terms of optimizing use of resources, it can be stated, that WCD use is surely cost saving compared to a hospital stay of the same time-period (Cortesi, 2021, Boriani, 2021).

The Authors say, one needs biostatistical methods for causal inference, and this is correct. However, in case of the WCD we speak about a true cause-effect relationship, not just inference.

The Authors say furthermore, there can only be two scientific standpoints in the interpretation of the VEST results, either inconclusive or absence of an effect. This is again interesting, because the term inconclusive contains two meanings in itself, yes/true or no/untrue, both are possible. Omitting the yes-option is, in our view, selective thinking.

The Authors state we (Semrau/Nürnberg) suggested, they misquoted and/or misinterpreted clinical practice guidelines on the WCD. This is correct and provable.
To be specific, The Authors reported wrongly, "ESC guidelines [4, 14] recommend that the WCD may be used in the following narrow indications (**all recommendations: IIb**/ Grade B or C)".
They did not mention a substantially better **IIa indication for inflammatory heart diseases in the exact guideline they quoted** (Priori, ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death, EuHeartJ 2015, see figures below).

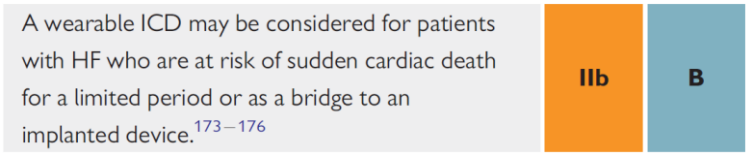
Wearable cardioverter defibrillator

Recommendation	Class[a]	Level[b]	Ref.[c]
The WCD may be considered for adult patients with poor LV systolic function who are at risk of sudden arrhythmic death for a limited period, but are not candidates for an implantable defibrillator (e.g. bridge to transplant, bridge to transvenous implant, peripartum cardiomyopathy, active myocarditis and arrhythmias in the early post-myocardial infarction phase).	IIb	C	167, 168
A wearable defibrillator should be considered for bridging until full recovery or ICD implantation in patients after inflammatory heart diseases with residual severe LV dysfunction and/or ventricular electrical instability.	IIa	C	598, 599

Both figures taken from the cited ESC Guideline. (Priori, 2015, pages 19, 52)

Additionally, The Authors **altered the wording of the indications**. Instead of the original text above, they wrote:
"- Temporary explantation of an ICD (e.g., due to infection)
- those with **accepted indicators for ICD implantation but also other contraindications** (e.g., infection)
- Patients on the waiting list for heart transplantation (without an ICD)
- Patients with an active myocarditis
- Peripartum cardiomyopathy (PPCM)
- Patients in the early postinfarction phase with "arrhythmias"
- Patients with post-coronary intervention (90 days) and impaired LV function"

Furthermore, they completely **ignored the recommendation of the second guideline** they quoted (McDonagh, ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure, EuHeartJ 2021).



In their current letter (reply to our Open letter), The Authors quote another guideline, which again gives one IIa recommendation (Zeppenfeld, 2022). Thus, saying there were only IIb recommendations is wrong for both of these guidelines (Priori 2015, Zeppenfeld 2022).

The Authors see I a high SCD risk always in combination with the need for an ICD. They do not take into consideration that in about 40-50% of patients with a diagnosed high SCD-risk, the risk is of temporary nature, and those patients do not need an ICD after some months treatment with appropriate medication (Exner, 2007, Sjöblom, 2014, Merlo, 2011, Kutyifa, 2015).

The Authors say, “However, if it were true that the evidence be that clear and strong for all indications (as claimed by Dr Semrau), …”.
We want to stress the point that the interpretation of the available evidence and the quality of the conduction of an HTA are two different things. I (Semrau) did not say, the evidence is clear, because it is indeed the current general view that the concept of an RCT with ITT-analysis fits to every subject. Therefore, we do not blame The Authors for thinking this way, as well. We criticize that The Authors do not work accurately in the first place, and furthermore do not the least consider their view could be inappropriate in some way or the other. On the contrary, they blame physicians for caring for their patient´s lives.

The Authors state, “The conclusions of our report are based on the available evidence and **appear to be aligned** with both the interpretation of clinical practice guidelines (above) and independently conducted systematic reviews.”
We already showed substantial discrepancies above. We would furthermore like to clarify the differences between the clinical guidelines and The Authors view. While the guidelines appreciate that the SCD-risk is defined by risk factors, such as a low LVEF, independently from the use of ICDs, The Authors try to promote the importance of ICDs by modifying the original text to, “patients with **accepted indicators for ICD implantation but also other contraindications**” instead of the original “patients with **poor LV systolic function who are at risk of sudden arrhythmic death for a limited period, but are not candidates for an implantable defibrillator…”** There is **apparently no alignment**, when The Authors feel the need to manipulate the original statement.

In terms of independently conducted systematic reviews, we already mentioned above that The Authors abuse a transparent declaration of potential conflicts of interests as a reason to disregard such publications. We think, transparency should be acknowledged. How do The Authors know whether conflicts of interests are always transparently declared? I would like to emphasize again, that from a scientific point of view the quality of the content should be assessed in the first place, not the potential conflicts of interests.

The Authors state, “It appears that numerous of the points of critique raised in the letter by Dr Semrau regarding the discussion section of our HTA report concern cited information from, among others, Ass. Prof. Masri and colleagues (instead of “personal opinions” of the AIHTA-authors,…”

We have to set this in the right light. The Authors start the chapter (p. 39), “Embedding our evidence into existing knowledge”, with extensive use of the publication of Masri et al.. They then report shortly on another systematic review. Then, they start a new paragraph with, **“The WCD is a historical example in which evidence-based decision-making falls short [16], with the WCD use increasing all over the world.”** However, *without* quotation marks. They later on quote once correctly and transparently *with* quotation marks, “Cardiologists who wrote one of the aforementioned systematic reviews [16] see this treatment pattern as likely to be “driven by the finality of SCD and partly by fear of litigation, despite the absence of data to support it”.”
The Authors go on like in the beginning (***without* any quotation marks**), “However, fear may only be one factor that led to **the assumption-based adoption of this medical device in clinical practice**. That is to say; there is no need for proof of clinical effectiveness or long-term safety to receive a CE mark in the European Union. Instead, evidence on the performance (purposes defined by the manufacturer) and safety are sufficient to receive a CE mark from notified bodies [48]. […]”
Needless to say, such statements are out of place in an HTA assessment.
The inappropriate, accusatory tone goes on (*without* quotation marks), “The VEST trial results may further be an example of what is sometimes referred to as spin bias [51]: […] In contrast, the **results of the secondary endpoint (all-cause mortality) were well promoted in scientific meetings and the media**, without mentioning that the RCT did not meet its primary endpoint [18, 52].”

Another example can be found on page 42, “Suffice it to say that evidence generation, following evidence-based medicine principles, needs to be prioritised **in light of intensive marketing** [19] and the increased risk of bias present in available observational studies.”
We think marketing should not have an influence on the conduction and results of an HTA, because HTAs are supposed to concentrate on evidence. Therefore, this is not an appropriate subject for an HTA discussion.

Page 19, “However, **intensive marketing hampered the scientific debate** regarding the appropriateness of the WCD [18, 19].

The non-scientific, non-factual discussion goes much further. One highlight for us is, “As per-protocol analysis is strongly inferior to the intention-to-treat (ITT) analysis, **the results of the VEST study should, therefore, not be re-interpreted.**” (page 41) We think, when scientific discussions are muted, science is at its end.

To be absolutely clear about the inappropriateness of the discussion of the HTA, we cite again from page 41, “**Decision-makers** in the health sector must also be particularly careful that, once a niche indication has been approved, it is not deliberately expanded in the context of everyday clinical practice. Hence, **the use of WCD should still be restricted to cardiological centres in Austria.**”
If the reader allows me this inappropriate transfer, this is a sort of *IIa recommendation:* “should be made” from a (supposedly) neutral scientific institution addressing decision makers. Level of evidence C.

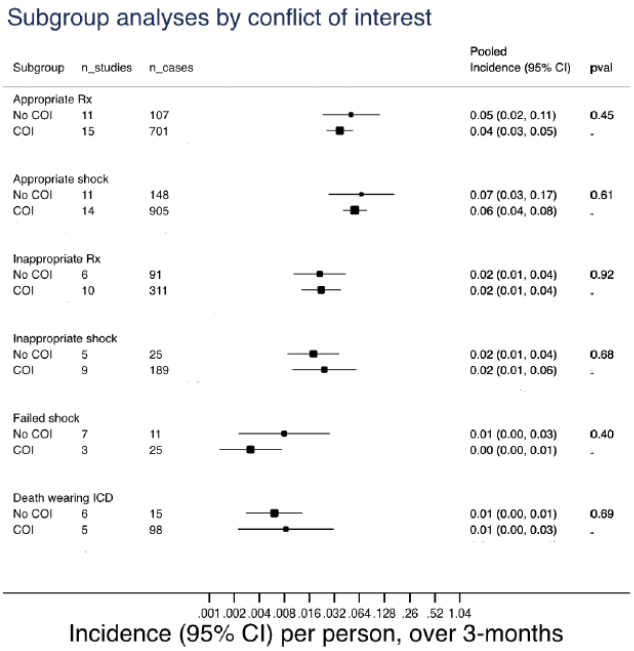
The Authors report in their current Reply to our Open letter that their manuscript was accepted by the Journal IJC Heart & Vascular. We congratulate the Authors for this success, even though we are indeed surprised. We encourage the editors and reviewers to read our critics in detail and are gladly available for discussion.

In the HTA and the reply letter, The Authors repeat again and again opinions from the seemingly independend systematic review of Masri et al.. Therefore, **we would like to reveal some of the results of Masri et al., which The Authors did not report.**

Interestingly enough, Masri et al. did a lot more research in their meta analysis. They **conducted several tests to measure bias**. They found **no different pooled incidence rates** (e.g. for appropriate

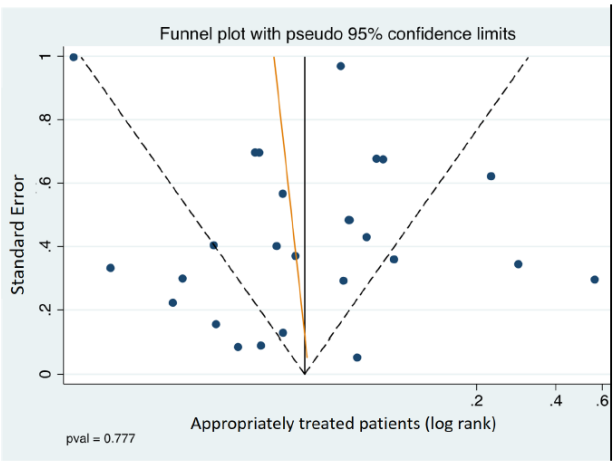
or inappropriate shocks) **for primary preventive vs. secondary preventive indications**. They found no differences between **prospective vs. retrospective studies**. They found no differences **between studies, which used ZOLL data bases vs. independent data bases**. They found no differences between **studies sponsored by ZOLL vs. independent studies**. They did not find a single study influencing results significantly in the meta analyses and, finally, Masri et al. **did not find publication bias**. (see Online Figures 6 and 10 below)

Online Figure 6. Pooled incidence rate per 1 person over 3 months of appropriately and inappropriately treated patients, appropriate and inappropriate shock, failed shock, and death while wearing the wearable cardioverter-defibrillator, stratified by authors having financial disclosures involving Zoll. Multiple rate by a 100 to get incidence rate per 100 persons over 3 months. Rx: treatment; COI: conflict of interest; WCD: wearable cardioverter-defibrillator; CI: confidence interval.



From Masri et al., JACC 2019. ZOLL sponsoring had no influence on results.

Online Figure 10. Funnel plot of studies reporting the rate of appropriately treated patient. Funnel plot is constructed by plotting the study specific incidence estimates against the standard errors of the estimates. Each dot represents a single study. In the absence of bias, the dots are expected to spread symmetrically in a funnel shape. P-value obtained from Egger test (p=0.777) suggests no evidence of publication bias.



From Masri et al., JACC 2019. Test suggests no evidence of publication bias.

Furthermore, Masri et al. differentiated their main results more than The Authors revealed. Masri et al. conducted their meta-analysis separately for **patients with non-ischemic (NICM) vs. patients with ischemic cardiomyopathies (ICM)**. The pooled incidence rate for appropriately shocked patients overall per 100 patients over 3 months was 5 (as reported by The Authors). The respective results for appropriately shocked patients with **ICM was 8** and for patients with **NICM 6** (not reported by The Authors).
Masri et al. indeed stated, "Selection bias and **including mixed indications in observational studies** was likely the **major determinant of the higher rate of appropriate treatment** in patients prescribed a WCD as compared with the WCD arm of the VEST trial."
Taking their own results serious (no publication bias), and acknowledging that ICM (including myocardial infarction!) has the highest appropriate treatment rates, **how can a mix of such highest risk patients (VEST population) with lower risk patients give even higher rates?** This try of an explanation is obviously not reasonable. Reasonable is, however, that the notoriously better compliance in real life studies (so stated by The Authors, as well) leads to a higher coverage of SCA events and therefore higher appropriate shock rates. We know that in VEST 16 of 25 patients who died from adjudicated arrhythmic death did not wear the WCD at the time of death (Olgin, 2018). With better compliance, such as in real life, the appropriate shock rate in VEST had been considerably higher. This is the reason, why the appropriate shock rate of respective patients in registries is higher than in the RCT, as reported by Masri.
In conclusion, the meta-analysis of Masri et al. was seemingly well conducted. However, the discussion seems not to reflect their results. The Authors omitted reporting those results, which did not support their story.

The Authors did a similar distortion of facts related to the publication of Weiss et al. by extracting (and altering) a seemingly negative phrase instead of the results. **Weiss et al. said**,
"Thus, **WCD is clearly not associated with increased anxiety and depression, but may have also positive impact on depressive symptoms.** […]. **In contrast to ICD treatment** [28, 29], […], **the WCD might enable patients to feel more secure**." (Weiss, 2021)
Instead, The Authors concluded, "**One registry study found a statistical (positive) association between WCD and baseline anxiety** when comparing the anxiety score and rate of anxiety between WCD therapy (n=85) to standard care (n=38), ..."
Please see more details in the table addressing the specific points of critic (3.2, right box).

In conclusion, The Authors of this HTA were biased. In several places, they omitted publications and results, which contradicted their opinion. They also manipulated quotes, results and tools of evidence-based medicine. The discussion of the HTA is inappropriate and touches one-sided opinions and non-therapy related subjects more than evidence. The Authors denounce cardiologists and scientists, health administrations, health regulations as well as health industry and even try to impinge politician´s decisions, instead of giving neutral data for decision-making. The Authors overstep the boundaries of a neutral HTA in many ways. The lack of insight in their wrong-doing suggests that their way of working in this HTA may not be restricted to this HTA.

Therefore, we request withdrawing this HTA-Update from the aihta website as well as from every HTA list. We furthermore invite the *IJC Heart & Vasculature* to reconsider their acceptance of the respective manuscript.

Literature (for letter and table)

Reply to aihta response to the Open letter of Nürnberg/Semrau:

Auriccio, A et al. (1998) Clinical efficacy of the Wearable Cardioverter Defibrillator in acutely terminating episodes of ventricular fibrillation. Am J Cardiol 81:1253-56

Balshem H et al. (2011) GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology 64, 401-406

Boriani, G et al. (2021) Cost-minimization analysis of a wearable cardioverter defibrillator in adult patients undergoing ICD explant procedures: Clinical and economic implications. Clin Cardiol 2021;1–9.

Chiarolla, E., M. Orso, G. Götz, M. Stanak, C. Wild and T. Jefferson. Wearable cardioverter-defibrillator (WCD) therapy in primary and secondary prevention of sudden cardiac arrest in patients at risk. Update 2018. Decision Support Document 103/ Update 2018. Rome and Vienna: AGENAS & LBI HTA, 2019 Report No. 15.04.2022. Available from: https://eprints.aihta.at/1186/.

Cortesi, Health Technology Assessment on the use of the Wearable Cardioverter Defibrillator in Patients with Myocardial Infarction and with ICD Explant. Farmeconomia, Health economics and therapeutic pathways 2021; 22(Suppl 1): 3-54; https://doi.org/10.7175/fe.v21i1S.1486

Ettinger S, Stanak M, Szymański P, Wild, C et al. (2017) Wearable cardioverter defibrillators for the prevention of sudden cardiac arrest: a health technology assessment and patient focus group study. Med Devices: Evidence and Research 14;10:257–271.

Ettinger, S., M. Stanak, P. Szymański, C. Wild, R. T. Haček, D. Erčević, R. Grenkovic, G. Goetz, M. Huic (2019) Reply to: Comments on the authors' reply to the critical appraisal concerning "Wearable cardioverter defibrillators for the prevention of sudden cardiac arrest: a health technology assessment and patient focus group study". Med Devices (Auckl). 2019;12:129-131. Epub 2019/04/10. DOI: 10.2147/mder.S202581.

Exner, DV et al. (2007) Noninvasive Risk Assessment Early After a Myocardial Infarction, The REFINE Study. *J Am Coll Cardiol* 50(24):2275-2284

Garcia, R et al. (2021) Wearable cardioverter-defibrillator in patients with a transient risk of Sudden cardiac death: the WEARIT-France cohort study. Europace 23, 73–81

Glasziou, P et al. (2007) When are randomised trials unnecessary? Picking signal from noise. BMJ Analysis 334:349-351

Goetz, G and Wernly, B (2023) Wearable cardioverter-defibrillator (WCD) therapy for primary and secondary prevention of sudden cardiac arrest. Update 2022. Decision Support Documents No.: 103/2. Update 2022. Vienna: Austrian Institute for Health Technology Assessment GmbH.

Guyatt GH et al. (2011) GRADE guidelines: 9. Rating up the quality of evidence. Journal of Clinical Epidemiology 64, 1311e1316

Hill, AB (1965) The Environment and Disease: Association or Causation? Proceedings of the royal society of medicine, 295-300

Hohnloser, SH et al. (2004) Prophylactic Use of an Implantable Cardioverter–Defibrillator after Acute Myocardial Infarction. N Engl J Med 351:2481-2488

Kutyifa, V et al. (2015) Use of the Wearable Cardioverter Defibrillator in High-Risk Cardiac Patients: Data from the Prospective Registry of Patients Using the Wearable Cardioverter Defibrillator (WEARIT-II Registry). Circulation 132(17):1613-9

Lange S et al. (2018) Klinische Studien und Equipoise, Dt. Ärzteblatt 115, 3, A70-A73

Masri, A et al. (2019) Wearable Cardioverter-Defibrillator Therapy for the Prevention of Sudden Cardiac Death. A Systematic Review and Meta-Analysis. JACC Clin Electrophysiol. 2019 Feb;5(2):152-161

McDonagh T. A., M. Metra, M. Adamo, R. S. Gardner, A. Baumbach, M. Böhm, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599-3726. Epub 2021/08/28. DOI: 10.1093/eurheartj/ehab368.

Merlo, M et al. (2011) Prevalence and Prognostic Significance of Left Ventricular Reverse Remodeling in Dilated Cardiomyopathy Receiving Tailored Medical Treatment. (J Am Coll Cardiol 2011;57:1468–76

Nguyen, E et al. (2018) Wearable Cardioverter-defibrillators for the Prevention of Sudden Cardiac Death: A Meta-analysis. Journal of Innovations in Cardiac Rhythm Management, 9:3151–3162

Olgin JE, Pletcher MJ, Vittinghoff E, Wranicz J, Malik R, Morin DP, et al. (2018) Wearable Cardioverter-Defibrillator after Myocardial Infarction. The New England journal of medicine. 2018;379(13):1205-15.

Pocock, J. Stuart and J. Collier Tim. (2018) Critical Appraisal of the 2018 ACC Scientific Sessions Late-Breaking Trials From a Statistician's Perspective. J Am Coll Cardiol. 2018;71(25):2957-2969. DOI: 10.1016/j.jacc.2018.04.015.

Priori, SG et al. (2015) 2015 ESC Guidelines for the management of patients with ventricular ar-rhythmias and the prevention of SCD. The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC). Eur Heart J 36(41):2793-867

Reek, S et al. (2003) Clinical efficacy of a wearable defibrillator in acutely terminating episodes of ventricular fibrillation using biphasic shocks. Pacing Clin Electrophysiol 26:2016-22

Reek, S et al. (2017) The wearable cardioverter-defibrillator: current technology and evolving indications. EHRA Scientific Documents Committee. Europace 19, 335–345

Sackett DL et al. (1996) Evidence based medicine: what it is and what it isn't. BMJ 312, 71-72

Sjöblom, J et al. (2014) Evolution of Left Ventricular Ejection Fraction After Acute Myocardial Infarction, Implications for Implantable Cardioverter-Defibrillator Eligibility. Circulation 26;130(9):743-8

Sperzel, J et al. (2018a) Critical appraisal concerning “Wearable cardioverter defibrillators for the prevention of sudden cardiac arrest: a health technology assessment and patient focus group study”. Medical Devices: Evidence and Research 2018:11 201–204

Sperzel, J et al. (2018b) Comments on the authors’ reply to the critical appraisal concerning “Wearable cardioverter defibrillators for the prevention of sudden cardiac arrest: a health technology assessment and patient focus group study”. Medical Devices: Evidence and Research 2018:11 377–378

Steinbeck, G et al. (2009) Defibrillator Implantation Early after Myocardial Infarction. N Engl J Med 361:1427-1436

Weiss, M et al. (2019) Anxiety, depression and quality of life in acute high risk cardiac disease patients eligible for wearable cardioverter defibrillator: Results from the prospective multicenter CREDregistry. PLoS ONE 14(3): e0213261.

Zeppenfeld, K et al. (2022) 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. Eur Heart J. 2022;43(40):3997-4126. Epub 2022/08/27. DOI: 10.1093/eurheartj/ehac262.

Zishiri, ET et al. (2013) Early risk of mortality after coronary artery revascularization in patients with left ventricular dysfunction and potential role of the wearable cardioverter defibrillator. Circ Arrhythm Electrophysiol 6(1): 117–128

Reply to aihta response to open letter Nürnberg/Semrau, specific criticism

No.	Point of critique (translated)	AIHTA response	Nürnberg/Semrau reply
1.	**Selective Reporting** In the HTA of Goetz, Wernly and Wild, so-called "selective reporting" is regularly applied. Publications, text passages or data and results that contradict the own opinion are not mentioned	This was an update systematic review: our selection criteria were, hence, based on the EUnetHTA report 2017 and AGENAS/LBI-HTA update report 2019 [8, 9]. A systematic review is based on a search and multiple databases and on a selection of studies (with pre-defined in-/exclusion criteria) that represent the best available [29].	The more it is incomprehensible that selective reporting is so frequently used in the HTA update by The Authors. The same is true for excessive use of personal opinion, especially of only one side of the medal, in the discussion.
1.1	**Exclusion of retrospective studies** - Goetz et al. categorically exclude retrospective studies, but at the same time criticize that there are not enough studies on the WCD and also resort to "personal opinion" in the discussion to support their views. ("Personal opinion" i.e. an opinion of any person, is the lowest level of evidence and does not require verifiable data).	It is correct that a personal opinion is the lowest level of evidence. We have excluded any kind of opinions – regardless whether these were held by scientific/ clinical experts, ZOLL or both – from the analysis and narrative synthesis of evidence (see results section of our report). Our inclusion criteria were pre-defined and were based on previous EUnetHTA 2017 and AGENAS/LBI-HTA 2019 reports [8, 9]. The rationale for exclusion of retrospective studies was explained in-depth in numerous Zoll-sponsored letter to the editors of the EUnetHTA report 2017: inclusion of retrospective studies would have not changed the comparative evidence [10].	Still, we do not understand the necessity to report The Author´s personal opinion in the discussion, let alone from one side, only; especially in combination with not reporting evidence in their hands. Furthermore, there is another retrospective comparative study (Zishiri et al. 2013) which showed similar results as the RCT (total mortality reduction). Though this was an indirect comparison, it might have been worth to consider. The exclusion of retrospective studies was *criticized*, not explained in two letters to the editor by Sperzel et al. 2018a, 2018b. The Authors try to make their point in comparative evidence, and simultaneously included several main outcomes, where a control group does not add any benefit. At least for device-specific adverse events and device specific parameters (such as appropriate and inappropriate shocks, and compliance), there is **no reason why a comparator group would produce more reliable results, when in such a group the measured value can only be zero.** **The Authors cite almost exclusively from only one meta analysis (Masri 2019), though there are more recent, more**

		The only other independently conducted systematic review [26] on the topic included all available studies (incl. retrospective studies) and did not derive at a different conclusion and critiqued the inability to draw firm conclusions on indication-specific utility of the device.	**comprehensive and more consistent systematic reviews in their scope.** Generally, the most recent and most comprehensive HTA should be preferred**.** The Authors themselves say that even a ZOLL sponsored HTA made rather careful conclusions. "Although industry-sponsored, the study authors [46] did not conclude on the comparative effectiveness of the WCD more broadly. Instead, it was noted that large registries confirm the device's safety and that the WCD detects and terminates VT/VF reliably." **It seems, the "industry sponsored" HTA took a neutral and objective position.**
1.2	**Shortness of breath** - The RCT VEST reports several side effects that could potentially occur (clustered) with WCD. Goetz et al. correctly mention that skin irritation was significantly more common in the WCD group. Furthermore, it is mentioned that all other parameters were not different. The equally relevant side effect of shortness of breath, which occurred significantly less frequently in the WCD group, is not addressed. This can be considered critical, since shortness of breath is a parameter that can be caused by anxiety or other discomfort and thus is actually given special attention by the authors as a "Patient Reported Outcome".	Methodologically, this was a second HTA update of the EUnetHTA report and AGENAS/LBI-HTA report [8, 9]. The evidence synthesis from the last report conducted by AGENAS/LBI-HTA was not changed, but updated. Dr Semrau has already critiqued this point within the past assessment and the rationale was provided within the last letter we received on 18th of April 2019: [11].	The Authors knew that no more RCTs were published since their last HTA (2018/19). If one only accepts RCTs and knows that no further RCTs were conducted, and one does not change anything compared to the previous HTA (apart from incorporation of some low level studies), why should one conduct another HTA? It would have been helpful, if The Authors stated their rationale here, because otherwise, the reader cannot come to a conclusion whether their rationale is appropriate. **For us, the question remains, why did The Authors not correct their mistakes and mention the complete results of the RCT VEST (PRO, shortness of breath) (Olgin 2018), including positive results, especially in light of their emphasizing on patient reported outcomes (PRO)?** **This has also to be seen in the light that The Authors also did not appreciate the QoL results of the RCT.**

1.3	**No overall appraisal of the results of the various analyses on the RCT** - Detailed presentation or discussion of the results of the ITT, as-treated, and per-protocol analyses is largely omitted on the part of the authors, Goetz et al. The opinion is expressed that an ITT analysis is basically the best form of analysis and that other results are negligible. This is in contrast to a scientific approach that basically considers all available analyses with an open mind. In the few sentences, which are to be found to the accomplished analysis forms, characteristics of As-treated analysis with those of the Per-Protocol analysis are mixed up and assigned to a large extent wrongly.	*We quote directly from a New England Journal of Medicine (NEJM) methods paper on primary outcomes in clinical research written By medical statistician Prof. Stuart Pocock and Prof. Gregg W. Stone (MD) [4]: "Analysis conducted according to the intention-to-treat principle is the main method used to make a valid comparison between two treatment strategies according to the treatments that were actually delivered to all patients who underwent randomization. When an intention-to-treat analysis fails to reach statistical significance, arguments are advanced that nonadherence and treatment crossovers may have masked real treatment effects and that as-treated or per-protocol analyses may get closer to the truth. Unfortunately,* ***the use of as-treated or per-protocol populations introduces selection bias, because patients who do not adhere to the treatment regimen and those who cross over to the other treatment strategy may have a different prognosis that is unrelated to actual treatment.*** *Hence, such analyses rarely influence conclusions regarding treatment efficacy that are based on the intention-to-treat principle. However, on-treatment analyses may be considered appropriate when safety issues are examined"* As seen within the ESC guidelines above [16, 25], these post-hoc analyses did not affect the interpretation of the VEST results from guideline groups either. And the Cochrane handbook [30] writes the following: *"An* ***ITT analysis maintains the benefit of randomization****: that, on average, the intervention groups do not differ at baseline with respect to measured or unmeasured prognostic factors. Note that the term 'intention-to-treat' does not have a consistent definition and is used inconsistently in study reports"*	The Authors hunker down in the safety of the general average interpretation of the gold standard for evidence generation. However, **we would like to invite the reader back to the fundamental intention of evidence based medicine.** The preference of an ITT analysis over as-treated or per protocol analyses is reasonable to protect patients from therapies without true effect and to protect patients from adverse effects of therapies without a true (positive) effect - **as long as an effect of the assessed therapy is questionable.** **RCTs were developed to calculate probabilities for cause-effect relationships** between a therapy and an effect in certain populations. For most (accepted) therapies, we have only probabilities (p-values and confidence intervals) of population effects after years. Without statistics, you had no chance to explore, whether there might be a probable cause-effect relationship or not. **This is a completely different situation compared to defibrillation and the WCD.** One can witness every day a thousand times, that **patients with ventricular fibrillation (VF), which are determined to die with a chance of about 100%, survive by defibrillation with a chance of close to 100%.** A larger effect is hardly imaginable. **This effect is real and works in less than a second.** This is also true for the WCD. There were two independent case series with patients who were induced with VF. The WCD detected and terminated all of them. The limitation of a successful defibrillation is only the availability of the events for the device. (Auricchio 1998, Reek 2003) **Had those studies been conducted as RCTs, all the patients in the control groups were dead after the study. This is indeed a dramatic effect!**

		Patients and other stakeholders are often interested in the effect of adhering to the intervention as described in the trial protocol (the 'per-protocol effect'), because it relates most closely to the implications of their choice between the interventions. However, two approaches to estimation of per- protocol effects that are commonly used in randomized trials may be seriously biased. These are: 'as-treated' analyses in which participants are analysed according to the intervention they actually received, even if their randomized allocation was to a different treatment group; and naïve 'per-protocol' analyses restricted to individuals who adhered to their assigned interventions. ***Each of these analyses is problematic because prognostic factors may influence whether individuals adhere to their assigned intervention****. If deviations are present, it is still possible to use data from a randomised trial to derive an unbiased estimate of the effect of adhering to intervention (…). However, appropriate methods require strong assumptions and published applications of such methods are relatively rare to date. When authors wish to assess the risk of bias in the estimated effect of adhering to intervention, use of results based on modern statistical methods may be at lower risk of bias than results based on 'as-treated' or naïve per-protocol analyses."*	When one now looks at registry data, there is a **successful termination of events of about 95% by the WCD in clinical practice (Meta analysis by Nguyen 2018)**. One does not need a control group for this assessment, because there would be no VF terminations in the control group. This dramatic effect is also true in a preventive setting, when not all patients have an event. The population effect is, of course, much lower, depending on the risk in the population. However, the **efficacy for every single patient with VF is always about 100%**. One can verify this in every patient, because the **WCD records all events**. There is no space for doubts. Furthermore, total mortality was significantly reduced in all analyses of the RCT VEST. Additionally, in as-treated and per protocol analyses, all (!) sub-mortalities were significantly reduced, meaning, **definitely no threat by the WCD**. The last point is **substantially different for implantable defibrillators** (see also Hohnloser 2004, Steinbeck 2009).
1.4	**Failure to find important literature** - Goetz et al. also criticise the fact that quality of life data was collected in the RCT but not published. However, corresponding results were already published in 2020. These were not addressed by the authors.	Conference abstracts do not fulfil our inclusion criteria that were defined a priori. Abstract 14913 was, hence, excluded from the synthesis. We explicitly state that conference abstracts are not subjected to peer review.	This is correct. However, The Authors also excluded personal opinion and still use it extensively in the discussion. It should be worth a communication to the authors before blaming them for selective reporting. It would have been very helpful to mention the QoL/POR results of the RCT (VEST) at least in the discussion, even if it was published only as a conference abstract.

	(Cheung CC, Olgin J, Pletcher MJ, Hue T, Vittinghoff E, Lin F, Lai M, Lee BK. (2020) Abstract 14913: The Impact of Wearable Cardioverter-defibrillators on Quality of Life: Insights from the Vest Trial. Circulation 142, Issue suppl 317).		Furthermore, It would have been helpful to mention the results of a patient questionnaire (Garcia 2021) in the discussion, even if it was not generated by a validated tool. Both would have been much more appropriate than citing extensively personal opinion from the meta analysis of Masri 2019. Because, as claimed by The Authors, "It is correct that a personal opinion is the lowest level of evidence."
1.5	**Non-reporting of the patient survey of an included study** - Goetz et al. attach particular importance to patient-reported outcomes. Thus, two of the five defined endpoints of the HTA refer to corresponding outcomes (health-related quality of life/QoL and patient satisfaction). The prospective study by Garcia et al. included by the authors contains data on precisely this question (Europace 2021). However, the results of this publication were not discussed.	After carefully reading this paper again, we still could not find data on quality of life or satisfaction **measured with a validated instrument**. It appears that Garcia et al. [31] used a questionnaire with regard to user satisfaction using a **tool that was not scientifically validated** but – rather – created by Zoll[12]. Please see footnote 49 and our PICO question – a rationale was provided why these data did not meet our inclusion criteria.	I agree that this was not a validated tool. Still, the scientific value is certainly higher than the extensively used personal opinion of Masri 2019 and The Authors. Furthermore, the answers in the study of Garcia et al. came from a substantial number of patients who all wore the WCD themselves, instead of the **answers of five males who did not know the WCD at all.** (See focus group analysis of Ettinger et al. in their first HTA, Ettinger et al. 2017) For additional information, the "focus group analysis" was one main point of critic in two letters to the editor (Sperzel et al. 2018a, 2018b), following this first HTA publication. The Authors even put this questionable assessment in the title of their publication.
1.6	**Incomplete/incorrect citation from European guidelines** - When listing WCD indications from guidelines, certain indications were not reported. One of the indications not mentioned received a "IIa" recommendation in the cited guideline, (meaning: "should" be done). Goetz et al., on the other hand, incorrectly report	Although Dr Semrau did not provide us with a reference of the specific guideline, we assume that he means the ESC guideline for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death [25]. It appears that the guideline was published/ available on Pubmed on the 21st of October 2022. At this time, the AIHTA update report was already published. Also, neither did we conduct a guideline synopsis nor was there room for a systematic search for guidelines	A specific reference was not necessary because **we referred to the exact guideline mentioned in the HTA update by The Authors themselves** (Priori et al. 2015). On page 52 of the quoted guideline **one can find the recommendation that was omitted by The Authors** (see our current reply, "Reply to the AIHTA-response, English letter", page 4) However, now that The Authors lay their finger in the wound, they once more did not quote correctly. While they said, "ESC guidelines [4, 14] recommend that the WCD may be used in

	that only "IIb" recommendations (meaning: "can" be made) were made in the guidelines.	within this update systematic review. After careful inspection of the mentioned guideline, it appears that there is a misunderstanding when it comes to the correct interpretation of evidence-based guideline recommendations: The mentioned **IIa** recommendation is applicable for a niche indication (adult patients with a secondary prevention ICD indication, who are temporarily not candidates for ICD implantation) and has **a level of evidence of C** (based on consensus of opinion of experts and/or small studies, retrospective studies, registries). There is no single class I (=is recommended) recommendation on the WCD. All available recommendations are class II recommendations meaning more broadly that (direct quotation from ESC terminology): “conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of the given treatment or procedure” is present. While a IIa recommendation means that weight of evidence/opinion is in favour of usefulness/efficacy, a class IIb recommendation means that the usefulness/efficacy is less well established by evidence/opinion. Usually, guideline recommendations are only sporadically mentioned within these systematic update reviews of the clinical evidence. A guideline synopsis [17] would be the method of choice if one wanted to synthesise (and potentially contrast) available guideline recommendations with regard to the WCD. We urge the importance, however, to consider not only the recommendation, but also the level of evidence which was omitted by Dr Semrau.	the following narrow indications (all recommendations: IIb/ Grade B or C): ...", **The Authors did not highlight that they altered the text of the original guidelines**. (Please see again our letter page 4.) Furthermore, they did not mention the recommendation of their second quoted guideline (McDonagh 2021). This is biased selection and interpretation instead of correct reporting.
2.	**Inadequate use of Risk of Bias (ROB) assessment tools GRADE**		

2.1	**Assessment** The GRADE group is a highly esteemed group of international scientists who have made a special contribution to evidence-based medicine. One of the basic approaches is to independently determine the confidence in effects for each parameter analyzed. The GRADE tool used by Goetz et al. also serves this purpose. The evaluations derived by Goetz et al. from this tool, in relation to the WCD, show serious errors in places. However, the reason for this can only be understood to a limited extent, as it is not discussed in detail in the course of the study. For example: The "total mortality" is assessed together with the "arrhythmic mortality. This cannot be considered adequate for various reasons. While all-cause mortality is the most reliable of all conceivable parameters, since there are no two opinions on death or non-death, the classification "arrhythmic mortality" is dependent on regularly incomplete data in the case of a (usually unobserved) sudden cardiac death, which, moreover, must be interpreted by people from a distance. So there are at least two relevant uncertainty factors here. One would therefore most likely place high confidence in all-cause mortality	We have separately GRADED these endpoints (please see previous AGENAS/LBI report [9] for a nuanced description). In order to be more concise for the update report (and given that this RCT evidence was not newly identified), we have shortened the GRADE evidence profile, by combining arrhythmic and all-cause mortality. It is still separately GRADED (all explanations are inserted in footnotes). Main reasons for downgrading arrhythmic mortality: deviation from intended intervention (especially due to poor compliance) and statistical imprecision. Main reasons for downgrading all-cause mortality: Besides the aforementioned reasons for downgrading, the endpoint "death from any-cause" was set as a secondary outcome in the included RCT. In addition, the study did not statistically correct the analysis for multiple testing (please see footnote c in GRADE table)" If Bonferroni-correction was applied, the endpoint all-cause mortality is not statistically significant anymore. We quote the NEJM methods paper again [4]: "If the primary outcome is negative, positive findings for secondary outcomes are usually considered to be hypothesis-generating." It appears that the opinion of Dr Semrau (regarding considering unadjusted stat. difference in all-cause mortality as moderate or high certainty evidence for strong additive benefit of WCD in post-MI) is not only in stark contrast to the AIHTA report using GRADE, the NEJM methods paper, but also to other scientific interpretation of how the stat. difference in all-cause mortality (note: unadjusted for multiple testing; not significant after adjustment for multiple testing) can be interpreted. Dr C. Israel – a cardiologist and advisory board member of Zoll – and colleagues [32], for instance, discussed, among others, VEST results within a	We do not think that repeating a mistake makes it correct. **It is highly questionable, whether an analysis for multiple testing according to Bonferroni is appropriate in this setting.** We are not talking about several but only two outcomes and those **outcomes are certainly not independent**, because one is included in the other. In fact, in the beginning of this trial, total mortality **was** the primary outcome. Therefore, it was not the case that Olgin et al. were randomly looking for any significant outcome. Furthermore, **multiple testing shall prevent from erroneously taking any observation by chance for a real effect.** In this case, we know for certain that defibrillation has a dramatic effect on mortality. Therefore, we expected an effect on mortality. We are curious to know on how many trials The Authors (or somebody else) applied Bonferroni correction. An alternative approach for multiple testing, already published in the original paper by Olgin et al., still shows a significant result for total mortality in the ITT analysis: "...Using an alternative approach 8-10 that takes into account correlations between endpoints, the corrected p value for total mortality is 0.046 with adjustment for two comparisons

	and possibly moderate confidence in arrhythmic mortality. Goetz et al., on the other hand, assign equal low confidence to both mortality parameters in this large randomized trial. This is incomprehensible because all three, the ITT, and the as-treated, and per-protocol analysis of the RCT consistently show significantly reduced all-cause mortality. Thus, while the most reliable parameter is given a low confidence rating, Goetz et al. give a high confidence rating to compliance, which in their own assessment was a major shortcoming of the study. This assessment is not consistent and not comprehensible.	narrative review: the difference within all-cause mortality was hereby addressed (although without GRADE assessment): "(…) Perhaps, the most unexpected finding of the VEST trial was the observation that the secondary endpoint all-cause mortality was apparently reduced by the WCD. This finding has to be considered hypothesis-generating and not confirmatory as the primary end point was not significantly different. Even though it can only be speculated what the underlying mechanism for this observation might be, it is reasonable to assume that the WCD affected some patients' compliance."	(viewing total mortality, the previous primary outcome, as uniquely important among the secondary outcomes)." (Olgin et al. 2018, referring to 8. Sankoh AJ, Huque Mf Fau - Dubey SD, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Stat Med 1997;16:2529-42. 9. Armitage P, Parmar M. Some approaches to the problem of multiplicity in clinical trials. Proceedings of the XIIth International Biometrics Conference; 1986; Seattle. 10. Dubey S. Adjustment of p-values for multiplicities of intercorrelating symptoms,. Proceedings of the VIth International Society for Clinical Biostatisticians; 1985; Germany.) The Authors state the NEJM paper again [4]: "If the primary outcome is negative, positive findings for secondary outcomes are usually considered to be hypothesis-generating." We agree and point to the word "usually". We do not think that a significant total mortality reduction in a large (larger than DINAMIT and IRIS combined) correctly randomized trial, assessing a device well known to dramatically reduce mortality, can easily be downgraded (for explanations on "dramatic effects", see Glasziou 2007). The Authors are clearly misinterpreting the GRADE guidelines. It is notorious that it is the most difficult task in itself to demonstrate total mortality reduction as a single outcome. DINAMIT and IRIS (independent ICD-studies) both failed to show a total mortality benefit in populations similar to VEST. Evidence based medicine is about confidence in effects and not about RCTs.

		We agree with Dr Semrau that generation of scientific hypotheses, speculations and assumptions are generally important. C. Israel and colleagues hereby further formulated very interesting hypotheses: The WCD could serve as a tool to increase patient awareness of a heart disease and improve compliance (e.g., within live-style modification). "Potential reductions in non-arrhythmic mortality related to better compliance, combined with a significant reduction of arrhythmic mortality, could result in reduced total mortality". Although interesting, these are defined by authors themselves – and need to be regarded as – hypotheses. [13] The authors correctly concluded that "(…) these hypotheses derived from the VEST trial merit validation in future, prospective studies".	This is a viable interpretation of Israel et al.. Another viable interpretation is, that the data for arrhythmic mortality was (is always) scarce and the remote adjudication of the subtype of mortality in VEST (not only in VEST) was therefore flawed (already suggested by Olgin in the original publication). Importantly, both interpretations take the total mortality reduction for real. One can also think that both effects contributed to the mortality results of VEST.
		From a statistical standpoint, we agree that VEST was statistically underpowered (which was reflected within the domain "imprecision" within our GRADE assessment) and agree with the appraisal of Stuart J. Pocock and Tim J. Collier, when stating the following [12]: *"The hypothesis posed for VEST (Vest Prevention of Early Sudden Death Trial) (…) is: can a wearable cardioverter-defibrillator (WCD) reduce the risk of sudden death in the immediate post-MI period (up to 90 days) in patients with reduced ejection fraction (EF)? The trial recruited 2,309 patients within 7 days of*	**Pocock and Collier wrote this statement after the initial congress presentation of the VEST trial, before the publication was out.** They did not know then, that an as-treated and a PPA analysis would show univocally significant reduction of every kind of mortality. They did not know that 16 of the 24 patients, which were adjudicated to SCD in the WCD group, did not wear the WCD at the time of death. They did not know that the success rate of a WCD for terminating a SCA is about 95% (Nguyen 2018). They did not know that compliance with a WCD is much better in real life, and they did not know, that another real life study - although

		hospital discharge after acute MI who had EF ≤35%. They were randomized in a 2:1 ratio to WCD + guideline treatment (n = 1,524) versus guideline treatment only (n = 778) and were then followed for 90 days. *Results for the primary outcome (sudden death) and several pre-defined fatal and nonfatal secondary outcomes are shown in Table 2[note: within the publication of [12]]:. There is not a significant reduction in sudden death (p = 0.18), and hence, some have called this a "negative" trial. This we find too dismissive, because the observed difference in incidence of sudden death (1.6% vs. 2.4%) is in favor of WCD: a 32.8% relative reduction, but with a wide 95% CI ranging from a 21.2% increase to a 62.8% decrease.* ***A better term is to call the trial "inconclusive."*** *The problem is that the trial only has good statistical power to detect very marked treatment differences. For instance, had the total of 44 sudden deaths split 22 (1.4%) on WCD and 22 (2.8%) on control, then this hypothetical 50% risk reduction would have been significant with p = 0.02. Even if the trial had been twice as big (n = 4,604) the observed 32.8% reduction would still only have p = 0.06. It would require 3 times as many patients (n = 6,906) for such a risk reduction to achieve p = 0.02. This is the dilemma we face when undertaking trials of an intervention strategy (9), such as wearing a WCD in the VEST trial. Patient recruitment is much harder than in drug trials (in VEST it took almost 10 years to recruit 2,302 patients), so that definitive evidence of efficacy is much harder to achieve. A further issue is patients' compliance with wearing the WCD; this averaged around 18 h/day initially and declined to around 12 h/day by 90 days (including nonusers). Such reduced compliance over time must inevitably compromise the ability to prevent sudden deaths.*	retrospective with an indirect control - came to similar results as the RCT (significant total mortality reduction, Zishiri, 2013). **Pocock and Colliers thoughts are neutral and completely correct from a statistical point of view, neglecting only the fact, that defibrillation is one of the very rare therapies with a dramatic effect**, which one can see working within less than a second. There is no doubt about the truth of the effect. Interestingly enough, **the statistical experts <u>did not</u> mention Bonferroni correction related to the VEST trial.**

		Among the pre-defined secondary outcomes (…), the one that really matters is all-cause death, with a 90-day incidence of 3.1% on WCD versus 4.9% on control. This is a 35.5% relative risk reduction with 95% CI: 2.2% to 57.5% reduction; p = 0.04. It is a natural instinct to now label VEST as a "positive" trial. After all, surely a significant result for all-cause death justifies such a claim? But a more cautious interpretation is warranted. First, the result is statistically fragile: if there had been just 1 less death in the control arm, the p value becomes >0.05. Second, all-cause death is not the primary outcome. Third, it seems illogical that the WCD is equally effective in preventing both sudden and nonsudden deaths. Thus, ***although it is plausible that a WCD really does reduce mortality, the VEST trial's evidence is not sufficiently convincing*** *by itself."*	As Pocock and Collier state, "The one outcome that really matters is all-cause death [total mortality]." We could not agree more. That is why initially total mortality had been chosen as the primary outcome in VEST. Correct, the statistical evidence seems to be fragile, however, **a p-value of 0.05** does not decide between truth or fantasy. It is an artificially chosen limit to prevent us from wrongly assuming effects. It **does not suspend natural laws.** Human beings cannot live without oxygen. Patients with a SCA do not get oxygen. A WCD restarts oxygen supply.
2.2	**IHE tool Assessment** When using the RoB assessment tool for single-arm observational studies, Goetz et al. make two methodological errors, each of which leads to a systematically worse rating of all studies. In this tool, 20 questions are given on the quality of the studies to be assessed. Each positively assessed question results in one point. While the developers of the tool (IHE) remove certain questions that do not fit the context (and thus would always have to be answered with NO) before the analysis, Goetz et al. leave such questions in their analysis, so that - regardless of the quality of the study - all studies are already assigned negative points in advance.	Assessing the validity of single-arm observational studies needs harmonisation throughout the health technology assessment world. As no clear guidance is available to reach overall bias by using the IHE-20 tool, we used the point system clearly and transparently that can be seen on page 25 within our report. In future HTA assessments throughout Europe, the risk of bias of single arm clinical trials will not be assessed anymore by following EUnetHTA guidance. Instead a high RoB will be assumed for these studies by default. Direct quotation of EUnetHTA new practical guideline 2022 [14]: *"****Uncontrolled trials per se are of very limited value for performing relative effectiveness assessment****. Although the (partial) use of some tools for RoB assessment is possible, the overall conclusion on the (very limited) internal validity of uncontrolled studies is very unlikely to be changed by RoB assessment. Therefore, RoB assessment is not required"*	**The Authors manipulated the IHE tool assessment in order to get the (negative) results they wanted to show.** See left box (IHE tool Assessment) for details. There are two important points to make: 1. **There are outcomes, The Authors said were of specific importance in their HTA, which are completely independent from a control group** (and consequently an RCT, as well) and therefore can be taken at least with similar evidence from large single arm studies (- which they in majority excluded). For example, appropriate or inappropriate shock rates are zero in a control group without a defibrillator. **It is not reasonable to demand a control group (specifically RCTs) to get a result, which is already known before the trial starts.** 2. **The Authors make an interesting statement: In their view, bending the rules in a (should be) neutral assessment is ok,**

	The IHE states in its guidance, for example, on the question: "Were outcome assessors blinded to the intervention that patients received": "Answer YES, when blinding is not applicable or is unnecessary ". In a single-arm study, blinding is not appropriate because all patients receive the same intervention. Goetz et al. nevertheless answer the question with NO throughout. If only this one question were taken out of the evaluation, according to the evaluation scale of Goetz et al. there would already be 7 studies with moderate and only 3 with high bias risk (instead of 7 with high and 3 with moderate risk). Another factor to be questioned is how Goetz et al. chose the scaling for assessing very high, high, moderate or low bias risk. (It would also have been possible to add very low risk as a 5th grade). The easiest to understand would be 25% fulfilled conditions each to move from one risk class to the next (0-5 points very high risk, 6-10 points high risk, 11-15 points moderate risk, 16-20 points low risk). With this classification, all included studies fell at least into the moderate bias risk, three even into the low risk. In contrast, the classification chosen by Goetz et al., in which 57.5% are necessary to be considered a high	So we see this point of critique obsolete, as new HTA guidelines [14] will consider these studies to be of high risk of bias by default in future (without RoB assessment using a tool such as the IHE-20 checklist).	**when sometime afterwards a commission says, one need not look at that point so much**. This is confusing. **We think every assessment that is done in a serious HTA must be as correct as possible.** There is no way in washing one´s hands clean, afterwards.

	bias risk (<57.5% a very high risk), in which 80% of the points are necessary to be considered moderate and 92.5% to be considered a low risk, does not seem comprehensible. This classification is not intuitive. Only because of the combination of these non-intuitive modulations, together with the retention of inadequate questions/questions not applicable to the underlying studies, the majority of studies appear to be at high risk of bias.		
3.	**Inaccurate allegations** Furthermore, Goetz et al. make several claims which, on closer examination, are not technically or factually correct. For example, they suggest that the authors of the VEST study themselves "suspected" that the overall mortality in VEST was only significantly reduced by chance. In fact, as is usual and correct according to scientific standards, Olgin et al. merely noted that, if interpreted conservatively, one could just as well say that the result was due to chance. In reality Olgin et al. stated several times, including in the original publication, that a misclassification of arrhythmic mortality was the most likely explanation for the non-significant primary endpoint. It can be assumed that Goetz et al. were aware of this fact, since they cite a	As systematic review authors, we summarise evidence from randomised trials. We do not speculate. Clear evidence for superiority based on randomised studies is needed [4]. We wrote: "*The statistically significant reduction in all-cause mortality could also occurred due to chance, as suggested by the authors of the VEST study in the course of the first publication*" (see p.41). Olgin et al. [13] write that *"(…) the uncorrected P value for comparison was 0.04 in favor of the wearable cardioverter–defibrillator. However, this result was not corrected for multiple testing, and given the use of most such corrections, the difference between the device and control groups would not be significant.* ***Thus, the conservative interpretation is that this result was a chance finding****."* By writing that it "could also" have occurred due to chance, we meant that the authors discussed that the conservative interpretation is that this result was a chance finding. We apologise if this was not clear enough to the reader of our report.	The sentence of The Authors could have interpreted so that Olgin et al. *suggested* that the significant all-cause mortality reduction occured due to chance. I would like to believe that The Authors meant differently. However, in their German version (their mother tongue) it reads, "Die statistisch signifikante Senkung der Gesamtmortalität kann, **wie von den Autor*innen** der VEST-Studie im Zuge der ersten Publikation **vermutet**, zufällig sein." In German, The Authors state, **Olgin et al. *suppose*** that the significant reduction of total mortality was by chance. This is obviously wrong. It is therefore clear that **The Authors** intended to manipulate the reader´s opinion. **The Authors put Olgin et al. something in their mouths, they never said.**

	"personal opinion" in which Olgin holds this opinion. Furthermore, it must be noted that the results obtained in this HTA, which are already quite negative in themselves, are presented even more negatively in the course text. Some statements are therefore not covered by the reported results. For example, Goetz et al. state: "The trustworthiness of the evidence from observational studies was very low". This statement cannot be derived from the results presented.	Regarding misclassification, the following is written by Olgin et al. [13]: *"The trial may have been underpowered to detect a beneficial effect of the wearable cardioverter–defibrillator on the primary outcome. Our power calculation anticipated a 58% lower rate of arrhythmic death with the device than without it. The power was, in part, reduced because 5% of the deaths were adjudicated as being of indeterminate cause and were thus removed from the primary analysis.* **Misclassification of the adjudicated cause of death may have further reduced the power for the primary outcome**. *It is difficult to determine an arrhythmic cause of death accurately for unwitnessed deaths or deaths with limited documentation."*	Misclassification of the cause of death is, in our view, the central point in this trial. I am happy to discuss this in detail, if interest be. The Authors did not respond to this last point. We therefore assume, that they acknowledge their mistake.
3.1	**There is an additional and not comprehensible negation of the actual results** At best, the evidence of the studies was low (as shown above, it was rather moderate from an objective point of view), but by no means very low. Similar negative overstatements can be found in	Thank you for highlighting this import issue within GRADE methods. When following GRADE guidelines [33], the lowest certainty of evidence on the outcome level reaches the overall certainty of evidence for the whole body of evidence. We agree that the "overall certainty of evidence" to be derived from the endpoint with the lowest certainty of evidence does not perfectly reflect the actual findings of a systematic review more broadly. We believe that a	Unfortunately, the answer of The Authors does not fit to the question. This is not a GRADE problem. In fact, **The Authors did not evaluate single arm studies with ROBINS tools but with the IHE tool.** **Therefore, as we stated, they used overly negative terms.** (Analysis 3.3, page 24) The Authors state, "Two independent researchers (GG, BW) systematically assessed the risk of bias (RoB) of the included studies using the Cochrane RoB tool v.2

	other parts of the HTA. For example, a study with "critical" ROB is mentioned as a result of the assessment. This study and/or the assessment of "critical risk of bias" does not exist within this HTA. The statement made, therefore does not correspond to the documented results and therefore appears to have no relation to reality.	range (certainty of evidence ranged from … to …) would be a good methodological improvement within GRADE. In so doing, the GRADE methods may be enhanced. Thanks for this valid point of critique that we may use in future HTA reports for deriving overall certainty of a body of evidence. For the credibility to consider results of a secondary endpoint as high certainty of evidence, please see answer to 2.1 For the term "critical RoB", please see explanation to the utilised ROBINS-I tool (answer to point 4.3)	[24] and the ROBINS-I tool [25] for RCTs and studies with a control group, respectively. **The quality of uncontrolled observational studies was appraised using the Institute of Health Economics (IHE-20) checklist [26]."**
3.2	**Another study is said to have shown an association between WCD and anxiety (Weiss 2019**). In fact, anxiety was assessed before assignment to the groups and issuance of the WCD. Thus, there can be no association at baseline between a patient's level of anxiety and wearing the WCD. Goetz et al., on the other hand, incorrectly suggest that the WCD was causal for an increased level of anxiety and repeat this view in various places.	We appreciate our reports are read in-depth, but it appears that there is a significant misinterpretation of our text. Errare humanum est: We are happy to clarify what we have written/ not written. It is correct that we included the comparative registry-analysis of Weiss 2019 for our evidence synthesis**. It is incorrect that we suggested the WCD to be causal for increased level of anxiety based on this registry.** The only thing that a registry can provide are data for associations (not for causation!). We recommend the article of Altman and Krywinski for a detailed description of distinctions of these two terms [34]. Mingling association and causation is scientifically incorrect. As systematic review authors, we generally try to minimise interpretation of data within the results section and use the description of data used from the respective included studies instead. In the context of the registry, we have written the following: *"One registry study found **a statistical (positive) association between WCD and baseline anxiety** when comparing the anxiety score and rate of anxiety between WCD therapy (n=85) to standard care (n=38), with 41 ± 11 vs 39 ± 13 (p = 0.22) and 58.9% vs 29.2% (p = 0.02), respectively (State- Trait Anxiety Inventory). Further, there was a non-statistical trend*	**We could not disagree more. The Authors cited wrongly, and interpreted wrongly.** They turned the original results of the study upside down. For a fact, Weiss et al. mentioned this observation but did not include it in their abstract or conclusions, because it is not a result of the comparative assessment. The Authors wrote, *"One registry study found a statistical (positive) association between WCD and baseline anxiety when comparing the anxiety score and rate of anxiety **between WCD therapy** (n=85**) to standard care** (n=38), ... ".* Crucial is the phrase,"Comparing anxiety score between **WCD therapy** to standard care." However, there cannot even be an association between WCD therapy and higher baseline anxiety, because anxiety was measured before patients got a WCD. There was an association between anxiety and the patient group, which got (subsequently) a WCD, but in no way between baseline anxiety and WCD wearing. Weiss et al. state quite clearly, "Patients with **subsequent** WCD prescription showed a higher baseline state anxiety score..."

		toward better improvement of depression scores in patients with WCD, with a mean change in score points of -4.1 ±6.1 and -1.8 ±3.9 (p =0.09) in patients receiving WCD and patients receiving no WCD, respectively. The change in anxiousness score was not statistically significantly different between patients enrolled in the WCD registry compared to patients receiving no WCD." The authors of the study themselves stated (direct quotation!) [35]: "Patients with subsequent WCD prescription showed a higher baseline state anxiety score (41 ± 11) compared to those without WCD (39 ± 13, p = 0.22), and had a significantly higher rate of anxiety (58.9% versus 29.2%, p = 0.02). **The association between WCD and baseline anxiety was still significant when adjusting for significant differences in baseline characteristics** between patients with and without WCD such as age, education level and history of malignancy (p = 0.02)."	Let us quote from the original paper of Weiss et al. to make clear, that misinterpretation by The Authors is hard to believe, "Thus, **WCD is clearly not associated with increased anxiety and depression, but may have also positive impact on depressive symptoms**. This does not contribute to the hypothesis that, in analogy to patients with ICD treatment, WCD might remember the patient of his life-threatening cardiac disease and the anticipation of shocks, thus triggering phobic anxiety and depressed mood [27]. **In contrast to ICD treatment [28, 29], due to the exposed wearing compared to ICDs, the WCD might enable patients to feel more secure.**" In the German part of their HTA, The Authors say on page 13: „Hinsichtlich der Lebensqualität wurde eine statistische Assoziation zwischen Defibrillator-Weste und Angst zu Beginn der Therapie innerhalb einer komparativen Beobachtungsstudie (Defibrillator-Weste: n=38, keine Defibrillator-Weste: n=38) gefunden." Our translation: „**Related to QoL a statistical association between WCD and anxiety was found at the start of the therapy** within a comparative observational study (…)." This statement is, additionally highlighted in bold letters on the right side of the same page: "**stat. association between WCD and anxiety**." **The Authors highlight repeatedly the *opposite of the study results* as conclusion of the study.**
3.3	**In addition, reference should be made again to an issue already briefly mentioned above**. Goetz et al. are of the opinion that intention to treat is the superior form of analysis to all others. This leads them to the assumption and statement that no attention should	In the results section, we have both reported the ITT results and the results of the PPA and as-treated analysis. Please have a look at page 34 and the data-extraction table (Table A-1). We agree that all these analyses have their own purpose. This is exactly why we have both extracted and synthesised the evidence with regard to ITT and other analyses.	We appreciate that The Authors reported the results from ITT, PPA and As-treated Analyses in the results section. However, they nowhere considered the implications.

	be paid to other types of analysis, as already noted above. This is not a scientifically defensible view. The aim of science is to ask questions openly and to consider them logically, neutrally and objectively. For example, ITT and Per-Protocol Analysis (PPA) have different questions: ITT asks what the outcome is after allocation of a therapy, while PPA asks what the outcome is after application of a therapy (similar to as-treated analysis). Both analysis types and questions have their scientific justification. In case of ambiguity, it is always useful to look at different sides - in this case - analyses. This is omitted by Goetz et al. or at least cannot be understood on the basis of the published HTA.	But as correctly stated by Dr Semrau, each form of analyses has their own strength and limitations. While forms of on-treatment analyses (for instance) may be considered appropriate when safety issues are examined, the ITT is still the main method used to make a valid comparison between treatments. We quote the NEJM methods [4] paper again: *"**Analysis conducted according to the intention-to-treat principle is the main method used to make a valid comparison between two treatment strategies** according to the treatments that were actually delivered to all patients who underwent randomization. When an intention-to-treat analysis fails to reach statistical significance, arguments are advanced that nonadherence and treatment crossovers may have masked real treatment effects and that as-treated or per-protocol analyses may get closer to the truth**. Unfortunately, the use of as-treated or per-protocol populations introduces selection bias, because patients who do not adhere to the treatment regimen and those who cross over to the other treatment strategy may have a different prognosis that is unrelated to actual treatment**. Hence, such analyses rarely influence conclusions regarding treatment efficacy that are based on the intention-to-treat principle. However, on-treatment analyses may be considered appropriate when safety issues are examined."*	In the cited NEJM methods paper, the authors suggest taking as-treated analyses for safety issues. **The Authors** did not consider as treated or per-protocol analyses for safety implications in VEST, though they quoted this NEJM publication several times. **Neglecting all evidence and implications but primary endpoints of ITT analyses of RCTs makes evidence based medicine assessments quite easy.** **The GRADE Guideline 3** (Balshem, 2011) Takes it not so easy, "The optimal application of GRADE requires systematic reviews of the **impact of alternative management approaches on all patient-important outcomes** [1]. In the context of a systematic review, the ratings of the quality of evidence reflect the extent of our confidence that the estimates of the effect are correct." In fact, there are far more than 20 GRADE Guidelines on very different aspects of evidence assessments. It may rather not be as easy as The Authors think. We should not forget that **fundamental for science** (and so for evidence-based medicine, ebm) **is logic** not dogmatism.
3.4	Instead, attempts are made to discredit forms of analysis other than ITT. Thereby confusion occurs between the properties of As-treated and Per-Protocol	See above. ITT is the gold standard confirmatory [4].	**Evidence based medicine is about the confidence in effects** – not about the results of RCTs, in the first place. From a logical standpoint, confidence in effects (or in measured parameters or rates) have at least two sides. One is the character of the parameter, how it is measured, how many steps are needed to gain the results, how much

	analysis. When looking objectively at the results for all-cause mortality - the most objective parameter imaginable - it is significantly reduced in the WCD group in ITT, as-treated and per-protocol analysis. (This parameter was assessed by Goetz et al. in the GRADE assessment as the only parameter with a high risk of bias or low confidence).		interpretation is planted in the results (patient side), how much interpretation is necessary (physician side/event assessment), which confounders are known or possible. Second is, under which circumstances are the results obtained. Here is the first question, are we talking about parameters or effects. If it is parameters, we do not need a control group. If we are looking for effects, we probably need a comparative trial (at best an RCT) - except, when we are dealing with a true dramatic effect. It is by far too easy to just look for comparative effectiveness with always the same tool (RCT), no matter which parameter or effect. If one wants to get from A to B. Is it legitimate to discredit a bike when you prefer driving Mercedes? We recommend reading (among others) GRADE Guideline 3 and 9 (Balshem, 2011, Guyatt 2011).
4.	**<u>Inconsistency of assessment and reporting</u>**		
4.1	The risk of bias in the RCT is assessed inconsistently and logically incomprehensible. Thus, parameters with high confidence are assessed as low confidence (total mortality), parameters with different confidence are assessed as the same (total mortality, arrhythmic mortality) and, for various reasons, less confidence is assessed as the highest confidence (compliance). In addition, parameters that are always automatically recorded in the same way and are independent	We do not fully understand this point of critique. But we believe answer to 2.1. gives justification to our reasoning that can also be found in our report. It is aligned with the interpretations of VEST by both renowned guidelines such as the ESC (see above) and the other independently conducted SR by Masri and colleagues [26].	This is a different issue, because in 2.1, The Author´s operationalization of the assessments is not addressed. Therefore, the answer to our point of inconsistency in the assessments cannot be found there. We think, the basis for an assessment should be **logic** in the first place. Let us give the reader one example of the inconsistency of The Authors assessments. **When the daily wear-time of a WCD is always automatically documented by the device, then the confidence in this parameter is the same, whether the study is an RCT or a registry**. Furthermore, a control group has no benefit at all in this assessment, because the wear time in the control group will of course be zero (no WCD!).

	of the study type or of a control group (e.g. compliance) are assessed with two levels of higher trustworthiness in the RCT than in the observational studies.		In Tab-A-7, GRADE Evidence Profile, **The Authors judge the compliance in the RCT, as well as in the observational trials absolutely the same**, risk of bias, inconsistency, indirectness, and imprecision are "*not serious*" in both lines. **However, their conclusions in the last cell is "certainty high" for the RCT and "certainty low" for the observational trials.** This is two levels difference for the same outcome with the same assessment for RCT vs. registry. **This is very inconsistent to say the least.**
4.2	In the RCT, the Risk of Bias tool states "some concerns" with regard to the measurement of the outcome parameters, with the comment that the assessors of the Adjudication Committee may have known which intervention which patient had received (Table A-4). This statement is a serious, incomprehensible accusation, as the RCT was apparently conducted correctly and the adjudication committee was also fully blinded. The original paper by Olgin et al. explicitly states: "The cause of death was adjudicated by an independent panel of experts Who were unaware of the group assignments (and therefore did not have any data from the wearable cardioverter-defibrillator)". The assumption on which the authors' statement is based is not explained further in the text. Because of this, it is also not comprehensible how Goetz et al. arrive at the overall assessment "Overall risk of bias: high".	Please see Table A-4 in our report. The risk of bias for the ITT effect of VEST is high due to bias due to deviations from intended intervention (low compliance). The risk of bias for the PP effect in VEST was high due high bias due to missing data and the fact that PP/as-treated analyses represent inadequate methods to fully estimate the effect of assignment to intervention (= the effect of interest for the systematic review at hand).	With their writing on the left side of this box (middle box), The Authors try to lead away from their baseless accusation of Olgin et al.. The Authors state in their HTA, page 65, table A-4, Risk of bias, Bias in measurement of the outcome, *"Some concerns"*. The explanation can be found in footnote 53: **"53 Outcome assessors may have been aware of the intervention received".** **This is a serious accusation that the assessors of the Adjudication Committee may have deceived.** **However, there is not the slightest reason for this baseless accusation of Olgin et al. and the adjudication committee of the VEST trial.** **The Authors** use their fantasy to discredit not only the therapy but also renowned investigators.

4.3	In general, it should be noted that Goetz et al. make several unspecific, general criticisms without specifying them further or clarifying the underlying problem and its potential impact (quote: "Some concerns were additionally found with bias in the measurement of outcome and selection of reported results."). In addition, the reporting of own results seems inconsistent in parts. For example, the result of the ROB assessment of the observational studies (tool of the IHE) is reported differently in two places in the HTA.	We used the terminology of the respective risk of bias tools. These consider the following options for domain specific and overall bias of a study: RoB v.2 for RCTs [36]: low, some concerns, high [36] ROBINS-I for comparative studies [37]: low risk, moderate risk, serious risk, critical risk IHE-20 for single-arm studies [38]: Low, moderate, high, very high using a self-defined scoring system.	The different reporting of the (IHE) tool results is not explained. On page 12 The Authors say in German: “The Confidence in the evidence of observational trials was very low.” Even so, the results of the (inappropriately applied) IHE tool said at least “low”. The Authors supplied a table with assessments according to ROB-2. However, the specific answers to the specific questions of this tool (Stern, BMJ 2019) are not disclosed. Therefore, it is not comprehensible, how those judgements were derived. The direction of bias or confounding is not even mentioned. In fact, **low compliance leads to lower effectiveness, meaning, without this bias, results were even more in favour of the WCD with even greater mortality reduction.** See GRADE Guideline 9, Rating up the quality of evidence (Guyatt, 2011).
4.4	Elsewhere, the term "critical risk of bias" was used for a study in the text, although this term can be found neither in the document itself nor in the corresponding table.	See above (ROBINS-I checklist) and table A-5	**This is correct, our apologies.**
5.	**Omission of scientific discussion of results**		
5.1	A decisive factor for a scientifically sound, neutral evaluation of a study is an open discussion of the results at the end. Thus, in the case of the VEST study, it is of course possible that, as Goetz et al. assume, the reduction in all-cause mortality may have been random. However, it is	Basic biomedical research tests hypotheses [39, 40]. Scientifically, we can either a) reject H0 and accept H1 as there is sufficient evidence (based on primary endpoint and ITT) or b) not reject H0 because of insufficient evidence [6, 39, 40] As an HTA institute, we assess the evidence – from randomised studies – that are able to sufficiently show	Sometimes it is better to take the screw driver instead of always using the hammer. The Authors do not the least consider GRADE Guidelines, e.g. GRADE Guidelines 3 and 9 (Balshem 2011, Guyatt 2011). **They do not consider, why the statistical concepts they report correctly, were developed. The goal is always to**

	just as likely that this can be assumed to be real. This is supported by various facts listed above. An open discussion of the results of all points speaking for and against was omitted or at least not documented in the HTA. Furthermore, it must be noted that Goetz et al. do not consider or discuss effects (such as arrhythmic mortality or total mortality) and their influence (direction of influence) in the sense of a sensitivity analysis. Goetz et al. rightly criticize the relatively poor compliance in the RCT. However, the consequence of a possible better compliance on the mortality outcome is not considered or included and discussed. It can be assumed that with better compliance, the reduction in mortality would have been even greater. Sixteen patients with adjudicated sudden cardiac death were not wearing the WCD at the time of death. With an effectiveness of around 95% (Nguyen 2018), it is legitimate to postulate that a large proportion of these patients would have remained alive if they had been wearing their WCD. This would correspond to up to 1% absolute further mortality reduction in an already significantly reduced overall mortality (at best WCD 2.1% vs.	that H0 can be rejected and write a report based on this evidence. We do not speculate and postulate what would have happened if… The facts are: VEST [13] failed its primary endpoint (h0 cannot be rejected = insufficient evidence) with poor compliance being a major limitation. Secondary post-hoc analyses on VEST [41] can be used to formulate hypotheses, as done by numerous authors. These need to be confirmed in future trials. Based on real-world evidence [31, 35, 42-54], we know that the compliance with WCD is good (which is in fact stated within our conclusion of the report), being a major limitation of VEST [13]. We acknowledge, however, that there are different standpoints with regard to what VEST results indicate. After VEST failed its primary endpoint, there were essentially two scientific standpoints/ opinions what VEST implies: 1. Although VEST failed its primary endpoint, evidence is still inconclusive and VEST does not provide evidence against the WCD as highlighted by, for instance, Lee et al. [20]. In this context, results of the secondary endpoints and per-protocol analyses could be used as data for the hypotheses that some selected patients within the post-MI patient group would still benefit from the WCD, although VEST failed to show an arrhythmic mortality benefit. Fauchier et al. [55] hereby also state the need to optimise patient selection and highlight that the risk for sudden cardiac death could be considered. OR 2. VEST results are similar to DINAMIT and IRIS and would "(…) provide robust evidence showing no role for defibrillators, whether implantable or wearable, within 40 days after myocardial infarction in the absence of sustained ventricular tachycardia or fibrillation (…)" as Stecker and colleagues [21] point out.	**explore, whether there is a cause-effect relationship and therefore a true effect or not.** We cite from GRADE Guideline 9 (Guyatt 2011) to remind of the goal of evidence based medicine. "If methodologically rigorous observational studies (…) show a sufficiently large effect, one can reasonably deduce that effect is real (that is, nonzero, and causally attributable to the intervention)." This is the underlying question: **are we talking about a real effect**, not whether or not we have positive RCT results. RCTs are developed to help answer that question. If the question can be answered right away, no RCT is needed to answer that question. Interestingly enough, there are indeed certain therapies, which do not need statistical tests. **Dramatic, and even more, immediate effects may be experienced directly.** We expect from ebm/HTA experts a differentiation of the assessed subject, being able to deal with more than the average, where causal inferences can only be drawn from RCTs. **SCA (i.e. ventricular fibrillation, VF) leads to death with an almost 100% certainty in few minutes.** (Goetz et al., page 18, left side (translated) "SCA occure without warning and lead untreated to death.") **Defibrillation is an accepted therapy with a dramatic effect** (Glasziou 2007). **Defibrillation terminates such SCA events with almost 100% certainty (WCD 95%, Nguyen 2018) in less than a second.** We would again like to remember that **RCTs were developed, to get probabilities for cause-effect relationships.** („Randomisierte kontrollierte Studien stellen unverändert den Goldstandard für den kausalen Wirksamkeitsnachweis medizinischer Interventionen und deren Nutzenbewertung dar." (Lange et al.(IQWiG), Dt. Ärzteblatt 2018)) **If such a relationship is confirmed, efficacy is confirmed.**

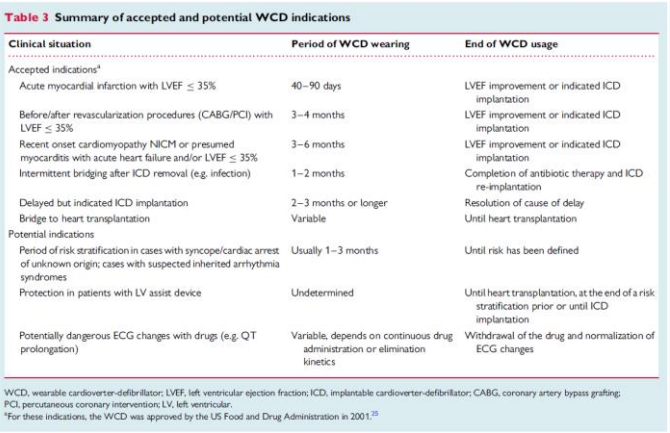
	control 4.9% instead of 3.1% vs. 4.9%).	We fully agree with Dr Semrau when saying that one can postulate and hypothesise about the add-on benefit of WCD in selected post-MI patients. We also agree that it would be an oversimplification to discredit the utility of the WCD fully on the basis of VEST. In fact, we have clearly written this in the discussion of the update report 2022: Absence of evidence should not be confused with evidence for no effect. We even highlighted that there are plausible indications for WCD use despite sufficient RCT evidence (please see p. 41 in our report [22]). It appears that guidelines also interpret the evidence as inconclusive instead of evidence favouring the null hypothesis: The ESC guideline task force, for instance, "(…) does not recommend routine use of the WCD in the early post-MI phase. Nevertheless, the use of the device may be considered in selected post-MI patients deemed to be at high risk for SCD" [25]. So we would agree with the standpoint of Lee et al. [20] and Fauchier et al. [55] that VEST results do not necessarily provide evidence against the role of a WCD (as highlighted also in our report). But in fact, it does not matter which opinions from cardiologists we agree upon. There are no further RCTs that can confirm one or the other assumption. It is beyond our remit to speculate and put weight on numerous assumptions, as suggested by Dr Semrau that we should do. This is not how scientific evaluations are conducted. Sensitivity analysis are important, if decisions are made based on assumptions (e.g., had they worn their WCD, if assuming that, something can be postulated). We acknowledge that modelling approaches do exist. We are not modellers, we are systematic review authors summarising clinical comparative evidence solely. However, hypotheses are important and can encourage further randomised controlled studies – for	As early as 1998 (Auricchio, AmJCardiol) and 2003 (Reek, PacClinEP), case series of patients equipped with a WCD were induced with VF in the EP lab. The WCD detected and terminated the events reliably. **Had those studies been RCTs, all patients of the control groups had died. SCA is a parade example of a deterministic condition. No control needed.** The WCD does not heal an indication. Instead, it turns a deadly condition dramatically into survival in less than a second. The event, SCA, is almost identical, no matter what reason (underlying indication or etiology) lead to the event. Used as a preventive matter, **not all patients in a population with a WCD suffer from an event.** Therefore, **the population effect is much lower than the efficacy.** However, **the efficacy for a patient with an event in such a population is still always about 100%.** This is fact. Facts cannot be compromised by tools (e.g. RCTs), which can only deliver probabilities. With those facts in mind, we may and should interpret the VEST results. **The primary endpoint (SCA) is unfortunately a very unstable outcome with scarce (often no) data plus the need for remote interpretation of those data. Most people with a SCA die unwitnessed.** If we acknowledge this, we see that the adjudication of sub-mortalities (arrhythmic vs. non-arrhythmic mortality) is a challenge and was most likely improper in VEST. The outcome we should rely on is (as always) total mortality. To option 2 on the left (in yellow), "VEST results are similar to DINAMIT and IRIS." **The VEST results are completely different to DINAMIT & IRIS.** While **in those ICD trials**, arrhythmic mortality was significantly reduced, total mortality was not, and unfortunately, **non-arrhythmic mortality was significantly increased**. There were initially two hypothetical options:

		instance in selected post-MI patients that would in theory benefit most from WCD. Given the excellent compliance shown in real-world evidence, it seems unlikely that patient compliance is as poor as it was in VEST again.	1. Patients resuscitated by ICDs died even so a bit later by their severe heart failure. Meaning, no reason to invest in prophylactic systematic defibrillation in this indication. 2. There is an upside (defibrillation) and a downside of ICDs (implantation, etc.). In this early, vulnerable period after a myocardial infarction, the downside of ICDs is about as large as the upside. **VEST showed improved overall survival and thereby proved that patients in this condition are not prone to die anyway after an appropriate defibrillation.** The differences between ICD-study results and VEST are clear and convincing. Accordingly, VEST showed that if a WCD has any effect on non-arrhythmic mortality, this effect is *in favor* of the patient. As mentioned above, **patients with a SCA in a control group of an RCT will die with a close to 100% chance, despite the fact that they could have been equipped with a WCD, turning the threat in a close to 100% chance of survival.** We doubt that an ethical committee would allow such an RCT, again. We would have been grateful for such a statement (yellow) in the HTA. However, The Authors do seemingly not understand, that **the conditions of the RCT are the reason for the under average compliance.** When verum and control group are presented to the patient as equally good care (this is a must in RCTs: called equipoise), the motivation to wear a WCD is substantially lower as in case the WCD was individually prescribed for your specific high risk of SCD. Furthermore, a hospital (study site) has not the capacity and experience for training and close follow up of patients, and to immediately step in, if compliance drops. In real life, a 24/7 hot line is available for every patient to solve potential problems at the earliest stage.

6.	**Criticising data that is missing but actually exists and has been ignored**		
6.1	As already described under "selective reporting", Goetz et al. do not report on the patient-reported outcomes (PROs) of the study by Garcia et al. included in the HTA. This is not comprehensible, especially considering the fact that QoL and PRO are declared to be among their most important five outcomes. Moreover, it does not seem logical to criticize the lack of corresponding data when these are demonstrably available in the included studies. Goetz et al. complain about selective reporting because the QoL data of the RCT were not included in the original paper by Olgin et al. This fact, too, seems hardly understandable, since the QoL data of the RCT had already been published in a separate publication when they prepared their HTA update, but were not identified by them. A request for this to the study leader (Olgin et al.) was apparently not made. Goetz et al. excluded retrospective studies in principle (analogous to the predecessor HTAs of LBl/aihta). At the same time, they criticized the	This first point is repetitive. Answers above. In May 2022, we contacted the study authors of VEST and were informed that, while there was no publication available yet, a manuscript was in preparation. With regard to the rationale for not including retrospective studies, please see the answers to the letter to the editors (four out of nine authors received financial support from Zoll GmbH for the submitted work) [11] published within the Dove press. The authors of the EUnetHTA report [10] clearly provided a sound rationale for their selection criteria. Hence, we did not see a need for including retrospective studies. Again, it makes it really hard to follow the point of critique if no source is adequately referenced. We assume that the following articles are meant by Dr Semrau: Epstein AE, Abraham WT, Bianco NR et al. Wearable cardioverter-defibrillator use in patients perceived to be at high risk early post-myocardial infarction. J Am Coll Cardiol 2013;62:2000–2007.	Please see again above. In their HTA, The Authors did not the least mention the PRO assessment of Garcia 2021, included in their SLR. They did not see this as more relevant than extensively reported personal opinion. It is, however, one step more to demand PRO assessments on one hand without explaining the omission of existing evidence already in their hands. If this is true, why did The Authors not state in the HTA that they contacted the principal investigator and gave the reader the information that a manuscript is in preparation instead of discrediting the authors? **For the interested reader: We recommend reading all the (4) letters/replies related to the first HTA publication (Ettinger 2017).** (Sperzel 2018a with direct reply of The Authors, Sperzel 2018b, Ettinger 2019) **Some issues that The Authors strongly defended even in their last reply (2019) were corrected during their cooperation with the Italian AGENAS.** We apologize. Generally, first authors name, subject, journal and date of publication is sufficient for identifying a publication.

	fact that there were predominantly studies with mixed indications for WCD and that only a few small studies with homogeneous indications were available. In fact, there are many large register studies. The largest of these include the one by Epstein et al. (JACC 2013) and of Ellenbogen et al. (JACC 2017). Both included more than 8,000 patients each and both only include patients from one indication each.	Ellenbogen KA, Koneru JN, Sharma PS, Deshpande S, Wan C, Szymkiewicz SJ. Benefit of the Wearable Cardioverter-Defibrillator in Protecting Patients After Implantable-Cardioverter Defibrillator Explant. JACC: Clinical Electrophysiology These single-arm retrospective studies were included in the only other independently conducted systematic review from Masri and colleagues [26] and there is no notable change with the overall conclusion and interpretation of the evidence from these researchers (in fact, these researchers saw the observational evidence as even more flawed as we have). Selection bias and including mixed populations (diluting the ability to draw firm conclusion on indication-specific utility) was first concluded by Masri et al. Our conclusion may be seen as a confirmation of conclusion of SR conducted by Masri et al. in 2019 [26].	**We do not think that the fact that other authors included certain studies, is an excuse for The Authors for having neglected them.** **Importantly, The Authors criticized such studies would not exist. This is again one step more than just omitting the evidence.** **We would like to highlight this line of argumentation. The Authors first say, such studies are urgently needed and do not exist. When it is clear they exist, and that The Authors just did not include them in their scope, they say such studies are irrelevant.**
7.	**Lack of understanding of the therapy to be evaluated, inadequate questions**		
7.1	Goetz et al. repeatedly criticise that the studies do not explicitly state whether the WCD is used for primary or secondary prevention of sudden cardiac death. However, this deficiency only appears to exist. The indications mentioned in the publications themselves contain the answer to the question of whether it is primary or secondary prevention. For example, patients after explantation are always counted as secondary prevention, whereas the protection of patients	We would never presume lack of understanding out of politeness within the framework of an academic discourse, but: "Homo sum, humani nil a me alienum puto". However, there is a misunderstanding regarding primary and secondary prophylaxis on the part of Dr Semrau. Primary prevention involves the use of ICDs in patients who have not yet experienced a life-threatening arrhythmia but are at high risk of developing one. These patients may have a history of heart disease or other risk factors that increase their likelihood of developing a dangerous arrhythmia. ICD placement is recommended for primary prevention in patients who have a reduced ejection fraction (EF) of	We are happy that now The Authors got a bit deeper into the subject. The Authors state by their own definition of primary vs. secondary prevention **that the underlying disease does not play a major role,** because defibrillation addresses a certain event, not a disease. **The Authors correctly bring it down to reduced LVEF and other factors, which define a high risk for SCD. They do not quote only one indication/etiology (which they strongly demand for the WCD)!**

	with myocarditis or after a myocardial infarction is initially considered primary prevention. It is generally recommended to involve specialists, in this case a cardiologist, in the preparation of an HTA for supportive assessment.	less than or equal to 35% and who meet other criteria as outlined in guidelines. Secondary prevention, on the other hand, involves the use of ICDs in patients who have already experienced a life-threatening arrhythmia, such as ventricular fibrillation or ventricular tachycardia. These patients are at high risk of recurrent episodes and benefit from ICD placement to prevent future events. So, to use the examples given, myocarditis or a previous myocardial infarction do not represent primary prophylactic indications. Rather, these entities are etiologies. Similarly, explanted ICD systems do not necessarily represent a secondary prophylactic setting, as the initial ICD placement may have been for primary prophylaxis. These are indeed delicate clinical situations where it is essential for clinically active	Those risk factors may occur in combination with several "indications" or "etiologies". The resulting event, **a SCA due to ventricular arrhythmia, is practically always the same, no matter, which indication caused the event.** Furthermore, the therapy of choice within few minutes is always the same, as well - defibrillation. If secondary prevention would always involve the use of an ICD, as The Authors suggest, what would be the role of a WCD? It is correct, that secondary prevention means, protection of a patient who already had a life-threatening event, in the first place. However, boundaries are somewhat bend during the times. Treating an explant patient as a typical primary prevention patient is probably inappropriate, considering his/her historical path even without a documented SCA. **It is interesting, why The Authors think primary or secondary prevention would always involve the use of ICDs?** **In fact, about 40-50% of patients eligible for a WCD for primary prevention (e.g. during the high-risk phase after a myocardial infarction with low LVEF) do not need an ICD after up-titration of their medication.** (e.g. Sjöblom 2014, Kutyifa 2015) The crucial point is, when a high risk for SCD is diagnosed (The Authors reported the risk factors correctly), it is hard to tell whether the risk is persistent or only temporary. **If The Authors really do not think that myocarditis or a myocardial infarction with low LVEF (the signs they mentioned) represent indications for primary prevention (e.g. for a WCD), we are indeed confused.** Because, the same "conditions" imply a primary prophylactic ICD implantation - however only after a waiting period of 40d -3 months, recommended by guidelines. This is due to the fact, that patients recover during that time and studies

		physicians to be involved, as was the case in this review.	showed no benefit of early ICD implantation (DINAMIT, Hohnloser 2004, IRIS, Steinbeck 2009). We quote from a publication, authorized by the EHRA Scientific Documents Committee, where (among others) both indications are described: "Indications for use [*of a WCD*] are listed in Table 3 and include primary and secondary prevention of SCD in patients with ischaemic and non-ischaemic heart disease in a variety of situations especially during the mandatory waiting periods after MI and the diagnosis of cardiomyopathy when medical therapy is being optimized to improve LVEF." (Reek et al, Europace 2017)  **Table 3 Summary of accepted and potential WCD indications** **Clinical situation** / **Period of WCD wearing** / **End of WCD usage** Accepted indications[a] Acute myocardial infarction with LVEF ≤ 35% — 40–90 days — LVEF improvement or indicated ICD implantation Before/after revascularization procedures (CABG/PCI) with LVEF ≤ 35% — 3–4 months — LVEF improvement or indicated ICD implantation Recent onset cardiomyopathy NICM or presumed myocarditis with acute heart failure and/or LVEF ≤ 35% — 3–6 months — LVEF improvement or indicated ICD implantation Intermittent bridging after ICD removal (e.g. infection) — 1–2 months — Completion of antibiotic therapy and ICD re-implantation Delayed but indicated ICD implantation — 2–3 months or longer — Resolution of cause of delay Bridge to heart transplantation — Variable — Until heart transplantation Potential indications Period of risk stratification in cases with syncope/cardiac arrest of unknown origin; cases with suspected inherited arrhythmia syndromes — Usually 1–3 months — Until risk has been defined Protection in patients with LV assist device — Undetermined — Until heart transplantation, at the end of a risk stratification prior or until ICD implantation Potentially dangerous ECG changes with drugs (e.g. QT prolongation) — Variable, depends on continuous drug administration or elimination kinetics — Withdrawal of the drug and normalization of ECG changes WCD, wearable cardioverter-defibrillator; LVEF, left ventricular ejection fraction; ICD, implantable cardioverter-defibrillator; CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention; LV, left ventricular. [a]For these indications, the WCD was approved by the US Food and Drug Administration in 2001.[25] **Myocarditis and myocardial infarctions with low LVEF <u>do</u> represent primary prophylactic indications.** It is still not comprehensible why The Authors explicitly demand the categorization in primary and secondary prevention in WCD publications.
7.2	Furthermore, Goetz et al. repeatedly criticize the fact that it is not regularly stated whether the prescription of the WCD is a	The position paper of the DGK [56] on the wearable cardioverter defibrillator states clearly that in certain situations, a WCD can replace hospital observation/ monitoring in the hospital.	The Authors seemingly do not understand the medical background.

	supplement to pharmacological therapy or a substitute for hospitalization. However, the question is neither technically nor factually correct, because the WCD does not treat the underlying disease, which is done by the individually prescribed drugs that are necessary in every case, but it protects against death in the event of sudden cardiac arrest (VT/VF). The question of whether a hospital stay is replaced, which can be asked completely independently of this, depends solely on how high a patient's individual risk of sudden cardiac death is perceived by the attending physician.	If the argument of Dr Semrau is to be true (that WCD monitoring cannot replace hospital monitoring), it would mean that the use of the WCD is always used in addition to standard care. It appears that this is in stark contrast to both the DGK position paper [56] and common clinical experience. We urge the need for more clarity with regard to specific use cases of the WCD. Evidence-requirements for strong utility are highly dependent on the question whether something is given as an add-on treatment or as a replacement.	**We would like to make very clear that we did not at all say, a WCD cannot replace hospital monitoring.** It is not comprehensible, how The Authors come to such an interpretation. The opposite is the case. As already explained in the box on the left hand side, the two questions of The Authors are independent from one another. 1. Standard care always includes appropriate medication, no matter whether with or without a WCD. 2. If a physician neglects the SCD-risk, standard care from his/her point of view does not include protective measures outside the hospital (after the initial hospital phase is finished). If a physician is aware of a substantial SCD-risk, standard care would be (after the initial hospital phase is finished) in-hospital 24/7 monitoring, or a WCD for 40d - 3 months. **Both answers (1. and 2.) are always true.** No reason for explicitly mentioning it in publications. **Therefore, The Author´s demand for those answers in every WCD publication is not reasonable.**
7.3	Goetz et al. repeatedly point out that most studies only report on mixed populations. They deduce from this that no indication-specific conclusions can be drawn. Against the background of the purpose of WCD therapy, the question does not seem reasonable. The WCD does not treat indications, but arrhythmic events that are absolutely comparable for different indications. It is possible that different indications have a different risk. However, this is irrelevant to the question of how high the success rate is after adequate WCD	This point is essentially a repetition of the second part of point 6.1. Please see answer above.	We disagree. While point 6 is about the demand of The Authors for missing studies, which in fact exist, point 7.3 is about the interpretation of The Authors that one needs studies (RCTs?) for every indication. **In The Authors comment to 7.1 they give their selves the explanation, why one does not need studies for every single indication.** The Authors define the indications for ICDs without mentioning only one "etiology". This is because **the risk factors, defining a high SCD risk are almost identical between different indications or etiologies.** **The event (VT/VF) is the same and the treatment (appropriate shock) is the same, as well.** There may only be a difference in the risk rate, however, this effects the

	therapy. Goetz et al. could answer this question based on their own collected data or by referring to other HTAs and meta-analyses. However, there is no reference to this in the document.		effectiveness in a population, only, but not the efficacy for a patient with an event.
8.	**Own emotional expression of opinion, anticipation of political decisions (exertion of influence)**		
8.1	By definition, an HTA should collect, document, and qualitatively evaluate the available evidence on a specific therapy in a neutral, open-ended, and factually correct manner, e.g., to provide political decision makers with a neutral basis for decisions. Goetz et al. overstep their boundaries as HTA authors and thus move even more obviously away from an independent, scientific claim. It can be implied that with their formulations they take decisions out of the hands of policy makers and give their personal understanding and thus the line of march. In addition, they denounce in their opinion inadequate behavior on the part of governments, health systems, physicians, and industry. Thus, their HTA manifestly becomes indicative of pre-established particular interests. Goetz et al. generally attack medical technology assessment and	A discussion is structured in the following way: Summary of findings, Embedding new knowledge into existing knowledge, contextualisation, limitations of report, conclusions For contextualisation only, we have also cited opinions from clinical experts/ cardiologists with an evidence-based medicine approach (note: these were neither our opinions/ "emotions" nor are these "attacks"): "WCD as an example in which evidence-based practice falls short" (incl. FDA-approval): A. Masri, A. M. Altibi, S. Erqou, M. A. Zmaili, A. Saleh, R. Al-Adham, et al. Wearable Cardioverter-Defibrillator Therapy for the Prevention of Sudden Cardiac Death: A Systematic Review and Meta-Analysis. JACC Clin Electrophysiol. 2019;5(2):152-161. Epub 2019/02/21. DOI:10.1016/j.jacep.2018.11.011. Discussion of new medical device regulation replacing the old CE mark regulation: A. G. Fraser, R. A. Byrne, J. Kautzner, E. G. Butchart, P. Szymański, I. Leggeri, et al. Implementing the new European Regulations on medical devices—clinical responsibilities for evidence-based practice: a report from the Regulatory Affairs Committee of the European Society of Cardiology. Eur Heart J. 2020;41(27):2589-2596. DOI: 10.1093/eurheartj/ehaa382 Spin bias in VEST:	We think The Authors and the reader will still agree that what The Authors present in their discussion is to a great extend opinion. If The Authors state those opinions (many times from only one publication (Masri 2019), presenting a great discrepancy between results and discussion), which are all from the negative side, the impression arises, that The Authors share those opinions. Why else would someone extensively cite one-sided opinion when his task is gathering evidence? **Very important: The Authors mostly do not mark their "cited opinions" with quotation marks. However, they do it at least once correctly, giving the impression that in all other cases they state their own opinion.** **Of note, The Authors take their time to blame the health system including almost all stakeholders, while they do not take their time for an analysis and discussion of the wider medical subject, such as the dramatic effect of defibrillation, the effect of low compliance on the outcomes, transient vs. persistent risk, or the potential of ICDs for complications.** We do not think that a serious assessment of the evidence of a therapy and the accusation from The Authors of whom ever, e.g. for "promotion of endpoints in scientific meetings", should be mixed in a neutral HTA.

	approval in Europe (CE mark) and present the WCD as a prime example of a product that has entered the European market virtually untested. Interestingly, however, the WCD does not serve as an example, because the WCD was initially approved in the USA by the U.S. Food and Drug Administration (FDA) on the basis of the results of clinical studies. The argumentation of Goetz et al. reveals throughout that not data and facts, but emotions and misconceptions underlie and shape their work.	J. Mandrola. The VEST Trial Failed, and So Did the Press Release. 2018 [cited 15.04.2022]. Available from: https://www.medscape.com/viewarticle/89375614	Apart from potential wrong-doings of marketing or anybody else, A therapy *per se* is innocent and a patient who could benefit from a therapy is innocent, as well.
Conclusion 1	This HTA was conducted contrary to the applicable standards and the requirement for an HTA to present a comprehensive, neutral picture of the data on a therapy. It is neither open-ended nor neutral or objective. Rather, it can be classified as tendentious. This inadequate approach deprives various parties of the opportunity to form an objective, neutral picture of a product, such as the WCD in this case. Politicians in particular trust that an appropriate institute such as AIHTA will produce careful, neutral, scientifically correct analyses that do not have to be subjected to further scrutiny. Due to the manuscript here, which does not meet scientific standards, it is not possible for politicians to make a	It is incorrect to say that the HTA does not meet applicable standards, and we kindly request such statements to be avoided. Our HTA reports follow the principles of evidence-based medicine - currently the highest international standard for summarizing clinical evidence. In this specific case, a systematic literature search was conducted in international research databases and the clinical studies were evaluated according to scientific evidence-based medicine criteria. The randomized controlled trial is thereby considered the gold standard for the evaluation of medical procedures. In this context, we would like to remind the reader that we cannot emphasise on unproven assumptions based on single-arm studies, expert opinions or opinions from Zoll, but apply strict evidence based medicine principles. Our job is to synthesise the available evidence with strict and neutral evidence-based medicine criteria. Such an assessment does not change if assumptions	We agree that The Authors know how an average HTA should be conducted. The frame of this HTA is according to recommendations. However, **The Authors let themselves get carried away by their opinions, which were seemingly fixed before they started their work.** **Inexcusable is their behavior of manipulating the results of their work, such as quotations, and RoB tools.** **It is not enough to know what must be done, it must be done correctly and unbiased.** If evidence based medicine was as simple as reading results from primary endpoints of RCTs only, no experts were needed. We suggest reading the more than 20 GRADE

	scientifically sound, neutral decision for or against a life-saving therapy.	and low-level single-arm evidence is promising, which is essentially what Dr Semrau is suggesting. Scientific and independent scrutiny is highly important and should take place within an editorial of an academic journal and independent researchers (ideally without potential conflict of interests) job is to scrutinise the rigour of a scientific article. As we uphold a high scientific standard and engage within an independent peer review process, we have submitted a scientific manuscript to IJC Heart & Vasculature. The referees (independent cardiologists) critiqued the systematic reviews minimally and provided us with valuable constructive scientific feedback before suggesting to accept the article for publication [27].	Guidelines and recommend starting with Guidelines 3 (Balshem 2011) and 9 (Guyatt 2011). We congratulate The Authors to the acceptance of their work in the IJC Heart & Vasculature. However, we confess, we are surprised of the acceptance. We guess, the reviewers acknowledged The Author´s status as an independent institute and did not seriously challenge how they worked. Furthermore, I guess, they did not consider dramatic effects and what RCTs were developed for, originally.
Conclusion 2	A direction is being set that, in the worst case scenario, will hinder or even prevent adequate care for patients at high risk of sudden cardiac death. Flawed or knowingly influenced policy decisions may also limit doctors' treatment options. This happens, for example, when a product is withdrawn from the market due to a refusal to reimburse it, thus depriving treating physicians of a treatment option. Therefore, a trustworthy and absolutely neutral evaluation of the evidence, and thus of the safety and effectiveness of a product, is essential, as is that of the economic components.	Adequate care is to be defined by evidence-based guidelines (indication-specific as highlighted by guidelines above): We acknowledge that WCD might and should be indicated in some selected patients [25]. We assessed the comparative evidence. We did not judge in which scenarios the WCD is adequate and necessary when randomised evidence confirming superiority of add on use of WCD in post-MI patients is missing (= job of guidelines for clinicians)! Our evidence synthesis is based on available research and our interpretation of evidence is aligned with evidence based guidelines and the only other available recent systematic review without industry-sponsoring [26]. We agree that neutral and objective communication about the facts of the WCD is key for adequate shared decision making. Nobody would disagree that the WCD technically works. Otherwise the device would have never received CE marking. We have also highlighted that the technical performance and safety of WCD are adequate.	We point to our initial wording in the left hand box. We would like to mention again that The Authors omitted many important results from the “only independend systematic review”, which oppose their own line of argumentation (Masri 2019, see page 6). Furthermore, industry sponsoring is certainly not enough as a reason for neglecting the most recent and most comprehensive HTA. If The Authors are convinced that the WCD technically works and the compliance is good, what exactly is their remaining question? **When the WCD technically works, it terminates SCA (as mentioned with 95% success, Nguyen 2018). This is a direct,**

			visible, and provable effect on mortality. When patients wear the WCD well, events will be detected and terminated. This is exactly what we saw in VEST, even at the most reliable outcome total mortality and even with a rather low compliance, compared to real life studies. Masri 2019 calculated, that in real life studies, which reported much better compliance than the RCT VEST (Goetz et al. 2023), the shock rates were much higher than in VEST. This makes perfect sense.
		But evidence confirming superiority based on hard endpoints is a completely other research question. Evidence-based guidelines correctly interpret VEST results hereby (= by concluding only on the statistically conservative interpretation of VEST), as highlighted above.	**There is no harder endpoint than total mortality.** Superiority means in the setting of the WCD, **superiority of a WCD over no protection** of patients from SCD from discharge on, **during the period of the highest SCD-risk**. For the fate of patients with a SCA event in the control group, see our comments at 1.3.