



Commentary

Safety vs. efficacy assessment of pharmaceuticals: Epistemological rationales and methods

Barbara Osimani

University of Camerino, School of Pharmacology, P.zza dei Costanti, 62032 Camerino, Italy

ARTICLE INFO

Available online 26 August 2014

Keywords:

Safety assessment
Adverse drug reactions
Evidence-based medicine
Evidence hierarchies
Causal assessment
Hypothesis testing
Inductive methods
Bayesian epistemology

ABSTRACT

In their comparative analysis of Randomised Clinical Trials and observational studies, Papanikolaou et al. (2006) assert that “it may be unfair to invoke bias and confounding to discredit observational studies as a source of evidence on harms”. There are two kinds of answers to the question why this is so. One is based on metaphysical assumptions, such as the problem of causal sufficiency, modularity and other statistical assumptions. The other is epistemological and relates to foundational issues and how they determine the constraints we put on evidence. I will address here the latter dimension and present recent proposals to amend evidence hierarchies for the purpose of safety assessment of pharmaceuticals; I then relate these suggestions to a case study: the recent debate on the causal association between paracetamol and asthma. The upshot of this analysis is that different epistemologies impose different constraints on the methods we adopt to collect and evaluate evidence; thus they grant “lower level” evidence on distinct grounds and at different conditions. Appreciating this state of affairs illuminates the debate on the epistemic asymmetry concerning benefits and harms and sets the basis for a foundational, as opposed to heuristic, justification of safety assessment based on heterogeneous evidence.

© 2014 The Author. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

Introduction

Evidence standards are supposed to provide “quick and dirty rules” for assessing the quality of evidence, as a function of the greatest possible reduction of confounding and bias. Hence, randomized controlled studies are followed by comparative studies which are not randomized (e.g. cohort or case–control studies), and these are followed by reasoning about pathophysiologic mechanisms underlying the observed outcome. Expert judgment is regarded as the weakest form of evidence and put at the bottom of the hierarchy (see [Howick, 2011](#), for a philosophical overview).

At present, no distinction is explicitly made concerning the role of such standards for assessing safety instead of efficacy. However, recent contributions by philosophers and health scientists have acknowledged the role of so called “lower level” evidence as a valid source of information contributory to assessing the risk profile of medications. Some of them are based on empirical surveys attesting that both randomized clinical studies and observational ones deliver the same incidence estimates for a series of risks associated with medical interventions, both pharmaceutical and surgical ([Benson and Hartz, 2000](#); [Golder et al., 2011](#); [Papanikolaou et al., 2006](#)). Others draw on various methodological considerations ([Aronson and Hauben, 2006](#); [Howick et al.,](#)

[2009](#); [Vandenbroucke, 2008](#)). These suggestions have noteworthy implications when considering current emphasis on evidence hierarchies, since they imply an asymmetry in the way evidence of benefits and risks of health technologies should be evaluated. However such suggestions fail to be grounded on a sound epistemic basis and seem rather ad hoc, although intuitively correct. Thus my aim here is to present their epistemological underpinnings by relating them to the common partition into (statistical) hypothetico-deductive, abductive, and inductive(-Bayesian) approaches to scientific inference. Furthermore, I will point out a series of pragmatic constraints for which inductive rather than deductive approaches to scientific inference should be considered as better suited to the purpose of risk assessment.

The rationales underpinning evidence hierarchies and alternative approaches: necessary vs. uncertain inference

Scientific inference may be categorized in two main typologies: on one side, inferences whose conclusion follows necessarily from the set of premises and “rules” involved (e.g. laws and initial conditions). On the other side, cases falling under the heading of “uncertain inference”, i.e. situations in which the conclusion is highly probable or plausible, but fallacious from a strictly logical point of view. The former generally fall into the category of deductive inference, whereas the latter are generally labeled with the umbrella category of “induction”. An additional form of uncertain inference falls under the heading of inverse induction

E-mail addresses: barbaraosimani@gmail.com, barbara.osimani@unicam.it.

(or abduction, as proposed by C.S. Peirce).¹ Both induction and abduction are rationally compelling but logically unwarranted methods of hypothesis confirmation. The distinction between them mainly consists in the former relying on probabilistic evidence, whereas the latter focuses on explanatory considerations. Current theories of scientific justification have coined an alternative term for abduction; this is “Inference to the Best Explanation” (also known as IBE, see Lipton, 2004); although not completely overlapping, the two concepts heavily rely on theoretical virtues such as simplicity/parsimoniousness (Ockham’s razor) and informativeness/explanatory power as (imperfect indicators) of reliability: explanatory power of a theory is a mark of its truth. Deductive and inductive/abductive rationales underpin diverse methods of evidence evaluation for causal assessment as illustrated below.

Hypothesis testing as a deductive approach to scientific inference

In classical hypothesis-testing, the result is expressed as the probability of observing the experimental result or more “extreme” results in the sample space (p-value), if the treatment makes no difference (so called null Hypothesis: H_0). The underlying epistemology is hypothetico-deductive (Popper, 1992): one assumes an entailment relationship between lack of efficacy and lack of difference between treated and untreated group ($H_0 \rightarrow E$). If the outcome shows a statistically significant difference ($\neg E$), then the hypothesis of lack of treatment can be rejected ($\neg H_0$), following classical modus tollens:

$$\frac{H_0 \rightarrow E}{\neg E} ; \quad \frac{}{\neg H_0}$$

In order to be able to draw a causal inference from the observed result, one must be confident that the difference between the two comparison groups is due to the contribution of the investigated factor, and only to it, otherwise $\neg E$ might be due (also) to some alternative cause. Blinding, intervention and randomization are essential instruments in warranting this causal claim (see also Papineau, 1993; Worrall, 2007; Osimani, 2013a,b,c) and evidence hierarchies are based on such warrants of internal validity. The EBM paradigm has been developed as a way to meet the desideratum that efficacy should be evaluated on the basis of the “best evidence” available, where “best” refers to quality criteria mainly informed by the requirement of internal validity.

The focus on internal validity is evident also in allowed deviations from evidence hierarchies in specific cases, i.e. where “lower level evidence” such as case reports and observational data are considered sufficient evidence for causal claims *to the extent that other conditions warrant for lack of bias and confounding* as alternative to randomization, blinding and intervention. Glasziou et al. (2007) for instance, consider cases where the relation between treatment and effect is so dramatic that bias and confounding can be safely excluded even if studies are based on just observational evidence: these are represented as phenomena of sudden and drastic changes in the clinical/epidemiological pattern and are formalized in terms of signal to noise ratio. Howick et al. (2009) relax the requirement of dramatic effect and reduce it to the desideratum that the effect size be greater than the combined effect of plausible confounders. Vandembroucke (2008) considers that unintended effects, qua unintended, are not known in advance, and thus also not known by the drug prescriber, who cannot calculate on them and thereby possibly bias treatment allocation. It follows that observational studies concerning adverse reactions do not suffer from confounding in the same way as observational studies for intended effects do.

¹ Peirce introduced the term “abduction” first in “Deduction, Induction and Hypothesis” (1934, Collected Papers 2.623), then in the Cambridge Conferences (1898) and in the 1903 Harvard Lectures (with different semantic nuances: see also Thagard, 1988, Section 4.2.1).

Uncertain inference: probabilistic and explanatory approaches

Non-deductive methods abandon the goal of outright hypothesis acceptance or rejection and track uncertainty while updating the degree of confidence in a given hypothesis upon new evidence by also taking into account background knowledge. This allows them to be more flexible with regard to the kind of evidence which is allowed to inform hypothesis confirmation and the methods for amalgamating it.

Within this framework, the two somewhat contending paradigms are constituted by probabilistic approaches to hypothesis confirmation (e.g. Bayesian epistemology) and abductive reasoning (also fleshed out as “inference to the best explanation”, Lipton, 2004).

Bayesian epistemologies (Howson and Urbach, 2006) insist on hypothesis confirmation rather than testing, and allow statistics to measure the degree of confirmation provided by evidence E to a given set of hypotheses $H = \{h_1, \dots, h_n\}$, by relying both on the likelihood of the evidence in relation to each hypothesis $P(E / h_i)$, as well as on the probability measure associated to each hypothesis prior to collecting the evidence, $P(h_i)$, and by updating it through conditionalization (or other means, depending on the specific Bayesian approach adopted). This distinguishes them sharply from frequentist statistics where the p-value measures instead the probability of observing the evidence obtained in the experiment (or “more extreme results”) if the hypothesis under investigation is false.

In the Bayesian paradigm the main requirement is that all available evidence is used (Carnap, 1947; Carnap, 1950): this is because all non-deductive logics are non-monotonic. Non-monotonicity is a phenomenon which characterizes defeasible reasoning, i.e. contexts where the addition of further data to the initial premises may invalidate some previous conclusion (Kyburg and Teng, 2001): formally, you may for instance have that the probability of hypothesis H is greater than its negation given evidence E : $P(H/E) > P(\neg H/E)$; but by adding another datum to the previous body of evidence, the opposite inequality may hold: $P(H/E,F) < P(\neg H/E,F)$. This may be illustrated by a diagnosis of celiac disease (H) with evidence of immune reactions to certain kinds of food (E), and then weakening of this hypothesis after a laboratory test (F).

In the IBE paradigm, hypotheses are justified by their explanatory power: the greater the amount of data the hypothesis is able to explain, the greater its plausibility. Thus, explanatory power is considered to be truth-conducive (Lipton, 2004). This paradigm is seldom explicitly adopted in causal assessment for health technologies; however it often underlies systematic reviews and qualitative reports, where heterogeneous evidence is combined in a narrative fashion.

The first important advocate of alternative approaches to statistical hypothesis testing was Sir Austin Bradford Hill with his most cited President’s Address (Hill, 1965) inaugurating the Section of Occupational Medicine of the Royal Society of Medicine; that is, a discipline mostly concerned with exposure to hazards. After presenting his nine guidelines for detecting and assessing causal relationships he claims: “None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us make up our minds in the fundamental question — *is there any other way of explaining the set of facts before us, is there any other equally, or more, likely than cause and effect?*” (emphasis added). Thus, Hill both refers to explanatory power and hypothesis likelihood as reliable grounds to justify causal judgments.

In recent times other authors have endorsed similar proposals. Aronson and Hauben (2006) put forward that “In some cases other types of evidence may be more useful than a randomized controlled trial. Combining randomized trials with observational studies and case series can sometimes yield information that is not available from randomized trials alone”. This idea is also at the basis of the recent proposal by Howick et al. (2009) to integrate evidence hierarchies with Bradford Hill criteria for causal inference (see also Stegenga, 2011). Vandembroucke (2008) proposes to invert hierarchies for

the purpose of evaluating unintended outcomes on the basis of Bayesian as well as abductive considerations (see Osimani, 2013a,b,c). In all these cases the implicit epistemic framework is non-deductive and requires all available evidence (Bayesian epistemology) — or as much as possible of it (explanatory paradigm) — to be accommodated in the assessment. Table 1 illustrates the epistemological rationales for the adoption of lower level evidence following the deductive vs. non-deductive categorisation of scientific inference.

The next section presents a case study which illustrates the different epistemologies underlying, often implicitly, the distinctive stances in evaluating evidence for causal assessment. Such debates may be clarified by making the respective epistemic framework explicit. Furthermore, with the help of this case study, four points will be illustrated as to show why non-deductive approaches fare better in safety assessment.

Case study: acetaminophen and asthma

The case of acetaminophen and asthma emerged with a study by Varner et al. (1998) who detected a precise correspondence between the increase of asthma incidence and increased acetaminophen use as a substitute for aspirin (which had been recognized to be associated with Reye's syndrome). The trend leveled off in the 1990s, i.e. at a time when acetaminophen had already become one of the most widespread analgesics. The suspicion raised by this study was further investigated by Newson et al. (2000) who conducted an ecological study showing a significant relationship between per-capita sales of acetaminophen and asthma morbidity across countries. Subsequent investigations explicitly aimed to examine the hypothesis of causal connection between paracetamol and asthma (Newson et al., 2000; Lesko et al., 2002; Barr et al., 2004; McKeever et al., 2005; Karimi et al., 2006; Beasley et al., 2008; Beasley et al., 2011; Shaheen et al., 2008; Lowe et al., 2010; Amberbir et al., 2011).

Possible biological pathways underlying the causal hypothesis have also been investigated. Eneli et al. (2005) summarize these findings and present five possible (non-exclusive) causal pathways accounting for the role of acetaminophen in asthma exacerbation. An additional immunologic pathway has been hypothesized by Nassini et al. (2010), namely the production of neurogenic airway inflammation caused by the transient receptor potential ankyrin-1 (TRPA1). Farquhar et al. (2010) have proposed another possible acetaminophen–asthma mediating mechanism based on its antipyretic effect.

The evidence gathered so far in support of the hypothesis of causal association between acetaminophen and asthma has generated two

opposite stances. On one side, a series of authors show some reluctance in accepting such evidence as a sufficient basis for practice change and for establishing a causal relationship between acetaminophen and asthma, on grounds that it does not result from randomized clinical trials (Eneli et al., 2005; Allmers et al., 2009; Johnson and Ownby, 2011; Karimi et al., 2006; Wickens et al., 2011; Chang et al., 2011). Particularly, these authors express the concern that the acetaminophen–asthma relationship may be explained by reverse causation, or confounding by indication. Other authors are less imperative on the matter but equally require or recommend the performance of adequately powered placebo-controlled trials to establish causation (Holgate, 2011; Henderson and Shaheen, 2013). On the other side, Martinez-Gimeno and García-Marcos (2013) recommend against a too liberal use of acetaminophen in children, while waiting for regulatory agencies to do their part and reconsider the safety profile of acetaminophen on grounds that the acetaminophen–asthma association appears to be stronger and more robust than any other endogenous or exogenous candidate. Beasley et al. (2011) assert that “when the study findings are considered together with other available data, there is substantive evidence that acetaminophen use in childhood may be an important risk factor for the development and/or maintenance of asthma” (p. 1570, emphasis added). An even stronger commitment to the hypothesis of causal association is expressed by McBride (2011) who, considering all the evidence available claims that evidence of causal association can by now be regarded as strong enough to warrant a change in prescription practice.

McBride justifies his claim by appealing to the consistency of interdisciplinary evidence: 1) strength of the association displayed in comparative studies; 2) robustness of association across geography, culture and age; 3) dose–response relationship; 4) coincidence of time trends in acetaminophen use and asthma increase; 5) lack of other equally strong causal explanations; 6) relationship between asthma epidemic and per-capita sales of acetaminophen across countries; and 7) plausible mechanism. McBride explicitly warns against the use of acetaminophen in children with asthma or at risk for asthma and claims that if further evidence is required, then this is for documenting product safety rather than the contrary.

Methodological dissent on causal assessment and safety issues

The dissent concerning the best course of action among scholars is ultimately caused by differing epistemological views which are left implicit. Those recommending the performance of placebo-controlled RCTs are in line with the rationales underlying evidence hierarchies.

Table 1

Rationales for justifying “lower level” evidence, as proposed in recent contributions, in relation to the corresponding epistemological paradigms.

Recent proposals for the adoption of “lower level” evidence in causal assessment	Related methodological assumptions	Underlying epistemology	
Howick (2011) and Howick et al. (2009): effect size greater than the combined effect of plausible confounders. Glasziou et al. (2007): dramatic relation between treatment and effect (sudden and drastic change in the clinical/epidemiological pattern). Vandenbroucke (2008): unintended effects cannot bias allocation in that they are not known in advance, hence non-randomized studies may be equally reliable.	Hypothesis testing: likelihood of evidence if H_0 is true (p-value) Internal validity (homogeneous populations and conditions with regard to all possible relevant causal factors): “best evidence” (EBM)	Hypothetico-deductive (statistical mode)	Deductive approaches
Vandenbroucke (2008): probability of hypothesis given (“lower level”) evidence and prior (theoretical) knowledge.	Bayes theorem, conditionalization Requirement of total evidence and coherence: all available evidence must be considered for inference to be valid	Bayesian epistemology	Uncertain inference: probabilistic and explanatory approaches
Stegenga (2011) and Howick et al. (2009): integration of BradfordHill criteria with standard EBM criteria. Aronson and Hauben (2006): evidence amalgamation.	1. Several heterogeneous indicators of causality; 2. Causal hypothesis is confirmed by the amount of data it is able to accommodate. The more evidence the hypothesis is able to explain the more reliable it is: <i>explanatory power</i> is truth conducive.	Evidence amalgamation (possibly based on explanatory considerations)	

Thus they insist on the elimination of any suspicion of confounding, especially confounding by indication (Henderson and Shaheen, 2013; Chang et al., 2011). On the other side there are those who point to the joint support of different and independent sources of evidence as a valid basis for dropping any need for RCTs, thus implicitly relying on alternative ways to justify causal claims. Indeed, there are several reasons for preferring the latter to the former approach when evaluating evidence for safety assessment (McBride, 2011; Martínez-Gimeno and García-Marcos, 2013):

- 1) The threshold for practice change is determined by the risk–benefit balance: the higher the risk in comparison to the therapeutic effect, the weaker may be the hypothesis of the causal link in order to change the marketing status of the pharmaceutical product (Osimani, 2007, 2013a,b,c). For instance, if an analgesic is found to be associated with carcinogenic risk, then, given its modest therapeutic importance in comparison to the suspected risk, the strength of the hypothesis of the causal association is not required to be high in order to take countermeasures (such as drug withdrawal) since it is counterbalanced by an unfavorable risk–benefit relationship; even a very low probability of causal association may suffice to retire the product from the market. Hypotheses of causal relationships need not be rejected or accepted: it is sufficient that they are strong enough with respect to the risk with which the treatment is suspected to be associated. In the case at hand, by claiming that if further evidence is required, then this is for documenting product safety rather than the contrary, McBride implies that such threshold has been trespassed. By shifting the burden of proof, McBride assumes that, given the expected harm and benefit, the probability of causal connection between acetaminophen and asthma is high enough as to shift the balance against its use.
- 2) Following Holland (1986), hypothesis testing should be seen as a method developed in order to assess *effects* of causes, rather than *causes* of effects, with the latter rather than the former being the focus of risk detection. Indeed when testing efficacy one is interested in ascertaining what kinds of effects a given treatment produces and in proving that no other causes is at play. Instead the main point of risk detection is to explain observed adverse events and to identify their causes in the first place. As Rudén and Hansson (2008) analogously point out, the focus of research in risk detection is on false negatives, rather than on false positives; which means that the problem is failure to see causation, rather than discerning spurious from authentic causation. Hence, standards for assessing causal claims concerning side-effects should not be parasitic on those developed for assessing treatment efficacy/effectiveness;
- 3) Concerning side-effects, most information comes gradually through “lower level” kinds of evidence, and there comes a point where the signal strongly suggests causation without demonstrating it. In such cases, one cannot pretend to know nothing about the side-effects of intervention until one doesn't have an RCT proving causality; yet this is exactly what the epistemology underpinning evidence hierarchies asks one to do.
- 4) RCTs deliver limited and purposely decontextualized information; they have been developed in order to test fertilizers on plant growth (Fisher, 1935). The causal structure here is much closer to physical causality (Thompson, 2011) — plots of lands do not react to fertilizers in the same way as human beings absorb and metabolize drugs (pharmacokinetics, pharmacodynamics) — and is not as rich in feedback loops, threshold effects, interactive causality as complex biological systems are, where such phenomena are much more frequent and entrenched. Hence another important reason for preferring a more flexible view on causality and justification of causal claims is also related to the ontological complexity of biological mechanisms upon which one should intervene.

In sum, the complexity of the causal structure in biological domains, the need for flexible decision tools which require the possibility to act

upon uncertainty, and the focus on risk/benefit balance, strongly suggest a paradigm change in the evaluation of evidence for harm.

Conclusion

Different epistemologies grant different methodological actions and grant “lower level” evidence on distinct grounds and at different conditions, thus it is worthy and also crucial to bear in mind the criteria underlying the evidence constraints one imposes on oneself. Appreciating this state of affairs illuminates the debate on the epistemic asymmetry concerning benefit and harm assessment and sets the basis for a foundational, as opposed to heuristic, justification of evidential support for pharmaceutical harm. Particularly, granted that knowledge about the drug risks comes from different sources and grows cumulatively over the course of time, probabilistic methods of causal assessment which rely on evidence amalgamation are better suited to the purpose. Instead, hypothesis testing is ill suited for risk management and pharmaco-surveillance in that it has been developed in order to address the issue of false positives, rather than false negatives, which are the main concern in safety assessment. In sum, categorical approaches are too rigid for safety assessment on the basis of probabilistic hypotheses and of heterogeneous evidence which possibly fails to meet their strict desiderata.

Conflict of interest statement

The authors declare that there are no conflicts of interests.

Acknowledgments

Funding for the paper has been provided by an intramural project conducted at the University of Camerino (# IM 250300): “Epistemic Asymmetries in Benefit vs. Risk Assessment of Pharmaceuticals”.

References

- Allmers, H., Skudlik, C., John, S.M., 2009. Acetaminophen use: a risk for asthma? *Curr. Allergy Asthma Rep.* 9 (2), 164–167.
- Amberbir, A., Medhin, G., Alem, A., Britton, J., Davey, G., Venn, A., 2011. The role of acetaminophen and geohelminth infection on the incidence of wheeze and eczema: a longitudinal birth-cohort study. *Am. J. Respir. Crit. Care Med.* 183 (2), 165–170.
- Aronson, J.K., Hauben, M., 2006. Anecdotes that provide definitive evidence. *BMJ* 333 (16), 1267–1269.
- Barr, R.G., Wentowski, C.C., Curhan, G.C., et al., 2004. Prospective study of acetaminophen use and newly diagnosed asthma among women. *Am. J. Respir. Crit. Care Med.* 169 (7), 836–841.
- Beasley, R., Clayton, T., Crane, J., et al., 2008. Association between acetaminophen use in infancy and childhood, and risk of asthma, rhinoconjunctivitis, and eczema in children aged 6–7 years: analysis from phase three of the ISAAC programme. *Lancet* 372 (9643), 1039–1048.
- Beasley, R.W., Clayton TO, Crane, J., et al., 2011. Acetaminophen use and risk of asthma, rhinoconjunctivitis, and eczema in adolescents: International Study of Asthma and Allergies in Childhood Phase Three. *Am. J. Respir. Crit. Care Med.* 183 (2), 171–178.
- Benson, K., Hartz, A.J., 2000. A comparison of observational studies and randomised, controlled trials. *NEJM* 342 (25), 1878–1886.
- Carnap, R., 1947. On the application of inductive logic. *Philos. Phenomenol. Res.* 8 (1), 133–148.
- Carnap, R., 1950. *Logical foundations of probability*. The University of Chicago Press, Chicago, (Routledge and Kegan Paul).
- Chang, K.C., Leung, C.C., Tam, C.M., Kong, F.Y., 2011. Acetaminophen and asthma: spurious association? *Am. J. Respir. Crit. Care Med.* 183 (11), 1570 (1).
- Eneli, I., Sadri, K., Camargo, C., Barr, R.G., 2005. Acetaminophen and the risk of asthma: the epidemiologic and pathophysiologic evidence. *Chest* 127 (2), 604–612.
- Farquhar, H., Stewart, A., Mitchell, E., Crane, J., Eyers, S., Weatherall, M., Beasley, R., 2010. The role of paracetamol in the pathogenesis of asthma. *Clin. Exp. Allergy.* 40 (1), 32–41.
- Fisher, R.A., 1935. *The Design of Experiments*. Oliver & Boyd, Oxford, England.
- Glasziou, P., Chalmers, I., Rawlins, P., McCulloch, P., 2007. When are randomized trials unnecessary? Picking signal from noise. *BMJ* 334, 349–351.
- Golder, S., Loke, Y.K., Bland, M., 2011. Meta-analyses of adverse effects data derived from randomized controlled trials as compared to observational studies: methodological overview. *PLoS Med.* 8 (5), 1–13.
- Henderson, A.J., Shaheen, S.O., 2013. Acetaminophen and asthma. *Paediatr. Respir. Rev.* 14 (1), 9–15.
- Hill, A., 1965. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58 (5), 295–300.

- Holland, P.W., 1986. Statistics and causal inference. *JASA* 396, 945–960.
- Holgate, S.T., 2011. The acetaminophen enigma in asthma. *Am. J. Respir. Crit. Care Med.* 183, 147–148.
- Howick, J., 2011. *The Philosophy of Evidence-based Medicine*. Wiley-Blackwell.
- Howick, J., Glasziou, P., Aronson, J.K., 2009. The evolution of evidence hierarchies: what can Hill's 'guidelines for causation' contribute? *J. R. Soc. Med.* 102, 186–194.
- Howson, C., Urbach, P., 2006. *Scientific Reasoning: The Bayesian Approach*. Open Court, Illinois.
- Johnson, C.C., Ownby, D.R., 2011. Have the efforts to prevent aspirin-related Reye's syndrome fuelled an increase in asthma? *Clin. Exp. Allergy* 42 (3), 296–298.
- Karimi, M., Mirzaei, M., Ahmadi, M.H., 2006. Acetaminophen use and the symptoms of asthma, allergic rhinitis and eczema in children. *Iran. J. Allergy Asthma Immunol.* 5 (2), 63–67.
- Kyburg, H.E., Teng, C.M., 2001. *Uncertain Inference*. Cambridge University Press, Cambridge.
- Lesko, S.M., Louik, C., Vezina, R.M., Mitchell, A.A., 2002. Asthma morbidity after the short-term use of ibuprofen in children. *Pediatrics* 109 (2) (Available at <http://pediatrics.aappublications.org/content/109/2/e20.full.pdf+html>).
- Lipton, P., 2004. *Inference to the Best Explanation*, 2nd edition. Routledge.
- Lowe, A.J., Carlin, B., Bennett, C.M., et al., 2010. Paracetamol use in early life and asthma: prospective birth cohort study. *BMJ* 341, c4616.
- Martinez-Gimeno, A., Garcia-Marcos, L., 2013. The association between acetaminophen and asthma: should its pediatric use be banned? *Expert Rev. Respir. Med.* 7 (2), 113–122.
- McBride, J.T., 2011. The association of acetaminophen and asthma prevalence and severity. *Pediatrics* 128 (Available at: <http://pediatrics.aappublications.org/content/128/6/1181.full.pdf+html?sid=29b845ce-0023-48dd-9835-3c87a6b45b69>).
- McKeever, T.M., Lewis, S.A., Smit, H.A., Burney, P., Britton, J.R., Cassano, P.A., 2005. The association of acetaminophen, aspirin, and ibuprofen with respiratory disease and lung function. *Am. J. Respir. Crit. Care Med.* 171 (9), 966–967 (1).
- Nassini, R., MAterazzi, S., Andre, E., Sartiani, L., Aldini, G., Trevisani, M., et al., 2010. Acetaminophen, via its reactive metabolite N-acetyl-p-benzo-quinonemine and transient receptor potential ankyrin-1 stimulation, causes neurogenic inflammation in the airways and other tissues in rodents. *FASEB J.* 24, 4904–4916.
- Newson, R.B., Shaheen, S.O., Chinn, S., Burney, P.G., 2000. Acetaminophen sales and atopic diseases in children and adults: an ecological analysis. *Eur. Respir. J.* 16 (5), 817–823.
- Osimani, B., 2007. Probabilistic information and decision making in the health context: the package leaflet as basis for informed consent. (Doctoral Thesis) Università della Svizzera italiana. (2007; 2007COM009 oai:doc.rero.ch:20120314100721-NN urn:nbn:ch:rero-006-110691. <http://doc.rero.ch/record/28759?ln=fr>).
- Osimani, B., 2013a. The precautionary principle in the pharmaceutical domain: a philosophical enquiry into probabilistic reasoning and risk aversion. *Health Risk Soc.* 15 (2), 123–143 (special issue: Health Care through the Lens of Risk).
- Osimani, B., 2013b. Until RCT-proven? On the asymmetry of evidence requirements for risk assessment. *J. Eval. Clin. Pract.* 19, 454–462.
- Osimani, B., 2013c. An epistemic analysis of the precautionary principle. *Dilemata* 11, 149–167.
- Papanikolaou, P.N., Christidi, G.D., Ioannidis, J.P.A., 2006. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 174 (5), 635–641.
- Papineau, D., 1993. The virtues of randomization. *Br. J. Philos. Sci.* 45 (2), 437–450.
- Popper, K., 1992. *The Logic of Scientific Discovery*. Routledge, London.
- Rudén, C., Hansson, S.O., 2008. Evidence-based toxicology: "Sound Science" in new disguise. *Int. J. Occup. Environ. Health* 14, 299–306.
- Shaheen, S.O., Potts, J., Gnatiuc, L., et al., 2008. Selenium and Asthma Research Integration Project; GA2LEN. The relation between acetaminophen use and asthma: a GA2LEN European case-control study. *Eur. Respir. J.* 32 (5), 1231–1236.
- Stegenga, J., 2011. Is meta-analysis the platinum standard of evidence? *Stud. Hist. Philos. Biol. Biomed. Sci.* 42, 497–507.
- Thagard, P., 1988. *Computational Philosophy of Science*. MIT Press/Bradford Books, Cambridge, Mass.
- Thompson, R.P., 2011. Causality, theories and medicine. In: Illari, P.M., Russo, F., Williamson, J. (Eds.), *Causality in the Sciences*. Oxford University Press, Oxford, pp. 25–44.
- Vandenbroucke, J.P., 2008. Observational research, randomised trials, and two views of medical science. *PLoS Med.* 5 (3), 339–343 (quotation in the text are from the longer version: 1–28).
- Varner, A.E., Busse, W.W., Lemanske Jr., R.F., 1998. Hypothesis: decreased use of pediatric aspirin has contributed to the increasing prevalence of childhood asthma. *Ann. Allergy Asthma Immunol.* 81 (4), 347–351.
- Wickens, K., Beasley, R., Town, I., et al., 2011. The effects of early and late acetaminophen exposure on asthma and atopy: a birth cohort. *Clin. Exp. Allergy* 41 (3), 399–406. <http://dx.doi.org/10.1111/j.1365-2222.2010.03610.x> (Epub 2010 Sep 29).
- Worrall, J., 2007. Why there's no cause to randomize. *BJPS* 58, 451–488.