Pesticide risk indices: A review of practical implications for policy and interpretation

James Rainford

Marc Kennedy

Glyn Jones

FERA Science Limited

1. Summary

Plant Protection Products (PPPs) play a key role in agricultural practice and the maintenance of yield but are also associated with potential risks to non-target systems of human health and the wider environment. In general, 'risk' is a measure that combines both impact and the likelihood of the impact. A key challenge for policy makers and other stakeholders is the development of tools to adequately characterise these risks and to support decision making, particularly around when and what to spray and the interpretation of trends in relative usage and impact. As it is rarely possible to characterise the 'true' impacts of PPPs in a cost effective and consistent manner many of these decisions rely on the development of proxies, collectively known as Pesticide Risk Indicators (PRIs), which combine information on product usage, usually with the results of laboratory testing or known chemical properties. We review 5 categories of PRI with differing relationships in how they represent potential risk. For each category we highlight several examples, focusing on those with the greatest relevance for policy decisions in the UK or EU Member States and particularly those used in National Action Plans to promote the sustainable use of PPPs. Categories and indices reviewed include:

- **'Quantity only' measures**, which describe the amount of PPP applied based on various underlying measures, without reference to relative toxicity or other chemical data. Key examples include *economic indicators of sales, the quantity of active substance applied (QA),), the number of unit doses (NUD), the treatment frequency index (TFI)* and the *standardised treatment index (STI).*
- **Qualitative indicators**, which rely on expert opinion or crude chemical designation such as 'Risk phrases' or chemical class as are used in *the EUs harmonised risk indicators (HR1 and HR2)*. The fuzzy expert index *I-Phy* is discussed here due to its semi-qualitative 'grading' of expert opinion.
- **Weighted multi-component PRIs** are a large category of quantitative indicators, which take the toxicity or chemical properties of a substance and combine them into a weighted mathematical formula (and may include splitting underlying continuous variables to represent different levels of perceived risk) in order to generate an overall active substance 'score', which is the basis for risk assessment. Examples detailed here include the *Environmental Impact Quotient (EIQ), the Norwegian Environmental Risk Indicator (NERI), the Danish Pesticide Load Indicator (PLI)* and *PestScreen.*
- **Exposure toxicity ratio (ETR) methods** differ from the above in that they explicitly include within their calculation an estimation of the concentration ('exposure') of substance within a specified environmental 'compartment' (such as nearby freshwater) based on local conditions around application (e.g. recent rainfall). Risk in these measures is expressed based on the ratio of exposure to the toxicity, with toxicity based on testing of relevant laboratory organisms. Examples discussed include the *Synoptic Evaluation Model for Plant Protection Agents (SYNOPS), the Environmental Yardstick for Pesticides (EYP), p-EMA, the Pesticide Occupational and Environmental Risk Indicator (POCER)* and various 'harmonised' indicators developed through EU or OECD research efforts.

- **Mechanistic and complex models** is a catch all term used to discuss the various modelling frameworks used to assess the movement of PPPs through various environmental compartments, most notably soil groundwater and runoff. They may also be referred to as process-based or fate models and aim to capture the physical processes involved in the transport or movement of PPPs. The outputs of these models are not generally used as PRIs in their own right, but often underpin other approaches, particularly the exposure toxicity ratio family, and are included to provide scientific context.

We briefly review the evidence base, validation, and core criticisms of the various methods discussed. We note the increasing evidence for inadequacy in the 'quantity only' methods for the prediction of risk as well as highlighting the general trade-off between the complexity, and specifically the data requirements, of an index relative to its performance under field conditions. This trade-off is important for decision making about PRIs and the suitability of different approaches for a) supporting farmers and their advisors at the micro-scale, b) supporting policy makers and other stakeholders in understanding the trends in PPP impact at a landscape scale, c) the administration of policy instruments relative to PPP usage, d) supporting regulators in the decisions around authorisation and monitoring of novel substances, and e) the characterisation of the wider environmental life cycle surrounding PPP application. We briefly review the suitability of different approaches to these different challenges and outline how PRIs are used in practice across various EU Member States.

Finally, we review the data and policy environment around PPP use in the UK and the potential future role of PRIs within this system. We outline the challenges that the existing system of pesticide monitoring present for UK PRIs and briefly discuss potential ways forward in the context of the ongoing revision of the UK National Action Plan.

Managing the potential impacts of PPPs on non-target systems relies on having consistent and interpretable proxies that can be used to inform relevant decisions makers. Here we review various approaches that could be used in the calculation of relevant metrics and their applicability to the UK, and existing data collections. One of the key challenges when discussing PRIs is the tendency for tools to be developed for very specific decision making and data requirements. As a result, there is a tendency for the proliferation of similar tools in different national and academic contexts, which makes it challenging to directly compare models or draw conclusions about best practice. If PRIs are to play a larger role, particularly in national or transnational policy, a greater emphasis on harmonisation and common frameworks is likely to be required, particularly around transparency and stakeholder engagement. PRIs have the potential to be key instruments in the management of non-target impacts of PPPs, and they have been applied for this purpose in several other European countries. However, targeted development is needed to adapt any PRIs to the specific requirements of UK policy and PPP use.

Contents

2. Introduction: Metrics for describing the non-target impacts of Plant Protection Products

The importance of PPPs in agriculture

Plant protection products (PPPs) play a key role in the maintenance of yield and food security (Cooper and Dobson 2007; Popp, Pető, and Nagy 2013). A recent global review estimated that the range of potential losses of wheat to pests and pathogens worldwide was between 10 and 28% (Savary et al. 2019), with older estimates suggesting that across a wide range of crops, losses due to weeds alone may run as high as 34% (Oerke 2006). PPPs are the principal and favoured tool available to farmers and growers to help mitigate crop loss from harmful organisms and so increase the predictability and stability of crop yield (Wilson and Tisdell 2001; Carpentier and Reboud 2018). Between 1960 and 2003, world average yields of rice, wheat and maize more than doubled, in large part due to a 15 to 20 fold increase in global PPP use (Oerke 2006; Silva et al. 2019). Additionally, Cooper and Dobson (2007) identified 26 primary and 31 secondary additional benefits of the use of PPPs including among others, improvements in food safety, human disease suppression and improvements in the efficacy of farm labour (Lamichhane et al. 2015; Sud 2020). Part of the reason for PPP use arises from the structural nature of modern farming practice, with the use of high-yielding monocultures and short crop rotation cycles necessitating high degrees of protection from potentially devastating pest and disease outbreaks (Savary et al. 2019). Chemical control is often favoured due to its economic efficiency and ease of implementation within a conventional farming setting, and is further reinforced by the ready availability of cost-effective compounds as well as the strong aversion exhibited by farmers towards crop health risks (Carpentier and Reboud 2018; Möhring, Wuepper, et al. 2020). Increasingly PPP use occurs in the context of a wider programme of pest management incorporating variety choice, the timings of application and use of other tools to help manage pest occurrence under the framework of 'Integrated Pest Management' (Peshin et al. 2009; Lee, den Uyl, and Runhaar 2019).

Despite their importance in the delivery of consistent yields, the predominant focus of scientific and policy debate around PPPs has been on the non-target impacts on human health and the wider environment. Since the mid 60's with the publication of works like Silent Spring (Carson 1964) considerable policy attention has been placed on the potential environmental and human health impacts of PPP use, with a number of recent reviews focusing on various aspects and policy instruments that could be used to minimised negative consequences (Bourguet and Guillemaud 2016; Kim, Kabir, and Jahan 2017; Lee, den Uyl, and Runhaar 2019; Dereumeaux et al. 2020; Sud 2020). One of the key challenges in characterising and controlling the impact of PPPs lies in the issue of measurement and the choice of indicators used to inform policy and application decisions.

Outside of highly controlled, and thus highly artificial settings, the direct impacts of pesticides on non-target systems can rarely be assessed at the level of individual farms or applications (although see Woodcock et al. (2016); Larsen, Gaines, and Deschênes (2017) for partial examples). Hence policy makers concerned with understanding the change in impact on the landscape and/or farmers interested in stewardship are usually forced to rely on indirect measures that attempt to extrapolate the potential 'risk' from the volumes applied, the conditions around spraying, the chemical properties of the compounds involved, and standardised laboratory assessment against well-known indicator organisms (see Milner and Boyd (2017), for commentary). These indirect measures are collectively termed Pesticide Risk Indicators (PRIs), e.g. Reus et al. (2002); Labite, Butler, and Cummins (2011); Feola, Rahn, and Binder (2011); Pierlot et al. (2017). Broadly PRIs can be defined as any measure, quantitative score or systematic classification which attempt to characterise the potential or realised impact of PPPs on non-target systems, and are reviewed here in the context supporting agricultural decision making and the development of policy instruments intended to moderate non-target impact (Lee, den Uyl, and Runhaar 2019). The choice of PRI can have major consequences for the perception of risk as differing trends both spatially and through time may be observed when alternative PRIs are used over the same dataset (Möhring, Gaba, and Finger 2019; Uthes et al. 2019). It is important therefore that prospective users understand the scope and usefulness of different approaches to generating PRIs and how differing measures have been used in practice in various countries.

Legislative context around management of PPPs

Before moving on to an in-depth review of approaches to the development of PRIs (Section 3) it is worth pausing to review the legislative and policy context in which many PRIs were developed and deployed. Our focus here is primarily on the Member States of the European Union, as well as the UK. This is in part a reflection of the significance that PPP management has traditionally held within the EU (Skevas, Oude Lansink, and Stefanou 2013), and also the fact that the majority of described approaches to PRIs have been developed in a European context (Feola, Rahn, and Binder 2011). In Section 5 we briefly review the linkage between PRI development and policy instruments across Europe and in Section 6 we specifically relate the discussed methodologies back to the existing infrastructure and policy development for the UK. Our aim is to provide a discussion of the role of PRIs and how the various forms and classes have been used in practice to support decision making around PRIs and as measures or proxies for monitoring environmental change. For discussion of the wider context of policy development around PPPs in Europe we recommend the reviews of Skevas, Oude Lansink, and Stefanou (2013); Pedersen and Nielsen (2017); Lee, den Uyl, and Runhaar (2019); and Sud (2020).

At the centre of European discussion of PPP use is the 2009 Directive 2009/128/EC of the European Parliament and of the Council, which established the framework for Community action to achieve the sustainable use of PPPs by reducing the risks and impacts of PPP use on human health and the environment and promoting integrated pest management (European Court of Auditors 2020). This regulation was introduced to address perceived inadequacies in then applicable PPP regulatory frameworks, as well as concerns over trends of increasing PPP sales in major EU Member States (Skevas, Oude Lansink, and Stefanou 2013). This framework introduced the requirement for Member States (at the time including the UK) to draw up National Action Plans (NAPs) promoting the directives goals. These NAPs were scheduled for introduction by November 2012, with the last being introduced in 2014 (European Court of Auditors 2020), and were scheduled for revision on a five year cycle, although many Member States including the UK have delayed revision of NAP (European Court of Auditors 2020), such that the current UK plan (Defra 2013) is expected to be released for public consultation at some point during 2020 (Claydon 2020) (see Section 6). The resulting NAPs provide the core mechanism for monitoring the potential impact of PPPs in different Member States (and the UK) and, where applicable for informing impact reduction targets (Barzman and Dachbrodt-Saaydeh 2011; Sud 2020). The current implementation of the UK NAP, as in the majority of Member States, does not include explicit impact or reduction targets for PPP use and there are no clear indicators of their implementation in the context of the revised National Action Plan[1]. In the near future, following the UK exit from the EU it is expected that measures relating to the usage and authorisation of PPP with be revised into national legislation the structure and contest of which, remain unclear at the time of writing..

Absent from the original EU directive was guidance on how 'risk' to non-target systems was to be measured and analysed in a comparable manner across Europe. The directive calls on Member States to calculate 'harmonised risk indicators, identify trends in the use of certain active substances, and identify priority items that require particular attention', but was adopted with an empty Annex, which does not specify the nature of these indicators[2] (European Court of Auditors 2020). While there has been some progress on adoption of harmonised indicators since 2018 (see Section 5), several Member States (including Norway, Demark, France, Germany, Belgium and the Netherlands (Sud 2020)) have developed their own PRIs, focused around their specific national objectives, with little consideration for harmonisation. This proliferation, combined with a large quantity of academic work on developing increasingly sophisticated tools for characterising agricultural impacts (Bockstaller et al. 2008), and/or

[1] https://publications.parliament.uk/pa/cm201719/cmselect/cmeuleg/301-ix/30112.htm

[2] Indicators subsequently developed or adopted across the Union are described in Section 3.

to support the process of substance authorisation (Labite, Butler, and Cummins 2011), has resulted a vast array of different tools and approaches which attempt to characterise or provide proxies for the impact of PPPs on non-target systems. In addition, separate regulatory silos such as human dietary exposure versus operator or bystander exposure have led to individual models being developed to predict exposure and risk in support of the relevant legislation. This is particularly true for the more complex models, although simpler indicators such as those described here often encompass multiple risk criteria (particularly those belonging to the class of weighted multi-component PRIs; see Section 3). There is a general desire for more harmonised treatment of PPP risk across different areas of regulation (More et al. 2019) but this has not been achieved yet, in part due to the complexity of data collection and harmonisation.

3. What defines a 'Pesticide Risk Indicator'?

When considering the *potential* effects PPPs may have on non-target systems, regulators and policy developers tend to adopt a 'risk' based framework to help support decision making (Levitan, Merwin, and Kovach 1995; Levitan 2000; Schäfer et al. 2019). 'Risk' in this context, refers to the combined function of the potential aggregate impact on non-target systems associated with exposure to some suite of PPP compounds, and the likelihood that these impacts occur under the assumptions of a specific PRI (Meek et al. 2011; OECD 2018; More et al. 2019). It is worth noting that, as a rule, PRI tend to focus on the 'hazard' posed by PPP to non -target systems, i.e. the potential for negative impact, based on information from laboratory or other standardised settings. Less consideration is often given to the probability of these impacts being expressed in reality, usually due to lack of supporting data. Hence some sources refer to PRI as measures of "consequences of hazard" as opposed to risk per se, e.g. Maud, Edwards-Jones, and Quin (2001). For example, an indicator might consider the health risk posed to operators of PPP application, due to the nature of the compounds used, without necessarily considering whether the operator in question in wearing appropriate PPE that might mitigate some of the health impacts (this is typically because the latter is often not recorded in datasets used to construct the indicator). Likewise impacts on non-target organisms, will generally not consider the distribution of susceptible individuals within the population, beyond broad partitioning of model organism appropriate to various environmental 'compartments' such as local freshwater. A useful way to think about PRI is to consider them as proxies for "standardised", or 'typical' risks to non-target systems, based on our knowledge of what is being applied, and the circumstances around application, as opposed to being tools for measuring impact on non-target systems. This standardisation of risk necessitates assumptions regarding how applied PPP will act under field conditions which are implemented differently in different approaches. It is the nature and structure of these assumptions, and their associated data requirements, that provides the classification for different PRIs used in this review.

The review presented here focuses on the use of PRIs to compare aggregate impacts across a suite of multiple PPPs applied over some defined geographic area, as this is what has driven the development of most of the described approaches (although there are important examples of related methods used for comparing rates of application for a single substance, e.g. in recent studies of global glyphosate use (Maggi et al. 2020)). Issues of scale, both in terms of spatial scope for the indicator (i.e. applicability of a PRI at the field, farm, national or international level) and/or the time period the indicator would be assessed over (days, seasons years etc.), will be touched on only briefly in our discussion, as this is driven largely by available local data sources and the use to which the PRI is being put, as opposed to being intrinsic elements of the calculation, (see Section 5/6 for discussion). Where possible we will note the original location and usage for each described PRI and how these might influence its future use. It is important to bear in mind throughout that PRIs are developed primarily as decision support tools (DST) and that this can influence both their suitability for use outside of their original context and the ease of drawing direct comparisons between different indicators, see Feola, Rahn, and Binder(2011); Labite, Butler, and Cummins (2011); Pierlot et al. (2017) for discussion.

There are well over one hundred published PRI methodologies, the vast majority of which have seen limited use outside of the original context for which they were developed (Bockstaller et al. 2008). Rather than attempt to review all methods individually we have followed previous reviews (Reus et al. 2002; Labite, Butler, and Cummins 2011; Pierlot et al. 2017) in defining broad classes of methodologies that share similar properties in terms of the description of risk, which often have related data requirements for calculation. For our discussion we identify five principal groups according to the information required and how this is represented within the calculation. The first group contains those PRIs which are derived only from measures of the amount of PPP that has been applied in a given context (there is some debate as to whether the term PRI is truly appropriate for these measures but we have retained this terminology in light of common practice in the wider literature e.g. (Möhring, Gaba, and Finger 2019)). The second and third groups build information on the chemical properties or toxicity into the value of the indicator, either in a qualitative manner or as a quantitative or pseudo-quantitative measure ("weighted multi-component PRIs"). The fourth group includes the diverse array of indicators built around the so called Exposure-Toxicity Ratio (ETR) framework, originally proposed by Reus et al. (2002) and which has proved highly influential in the development of more recent PRIs. The defining feature of ETR indicators is that they use information on the conditions around application, and/or some suitable worse case assumptions, to explicitly predict the resulting concentration in some specified location (for example nearby fresh water; conventionally termed a 'compartment'). These predicted 'within compartment' concentrations are then described relative to toxicity information for relevant organisms to provide the final value of the index (Reus et al. 2002; Feola, Rahn, and Binder 2011). The final grouping briefly discusses some of the mechanistic or otherwise complex models describing the physical movement of PPPs through various environmental compartments (most commonly soil). These are covered only briefly, as they would rarely be reported as PRIs in their own right but underpin many of the other indicators, most notably those using the ETR approach.

Our selection of indicators to discuss in detail was influenced by considerations of policy influence and the suitability of different PRIs to decision making outside of the academic sphere. Hence, we have prioritised those PRIs which have been integrated into National Action Plans or related surveillance. For a more complete treatment of alternative methods readers are referred to the reviews of Maud, Edwards-Jones, and Quin (2001); Reus et al. (2002); Dubus and Surdyk (2006); Labite, Butler, and Cummins (2011); Feola, Rahn, and Binder (2011); Pierlot et al. (2017) which include a number of other approaches not outlined in detail here. As previously stated, our focus is on indicators used in EU Member States, or which have the potential to be adapted to a UK setting, bearing in mind the data requirements and scope of different methods (see Section 6 for discussion).

An alternative way to classify PRI is to group approaches based on their intended audience, and on the and decisions that the indicator(s) might be used to support, see e.g. Reus et al. (2002). In Section 5, we review the various uses to which PRIs have been put, as well as highlighting their roles in decision making around PPPs with an emphasis on policy making. In general, the simpler 'quantity only' measures are most commonly associated with surveillance and monitoring of change, particularly at national or international scales, while the more complex measures, particularly ETR methods are often used in decision support and for comparing local alternative PPP use or in authorisation decisions (Feola, Rahn, and Binder 2011), although exceptions apply and are discussed in Section 5 .

Quantity only indicators

This simplest class of PRIs aim to characterise the amount of PPP being applied to an area without any consideration of chemical properties or differences in impact/toxicity. These measures, due to their simplicity and ease of calculation, are often politically important (See Section 5) but lack the sophistication of the more advanced tools outlined below. We call these indicators 'quantity only' because they only account for the quantity of PPP applied (under various measures) and make other consideration for example around relative impact (Reus et al. 2002). Other workers have used terms like ‘quantitative’ indicators to describe quantity only indicators (e.g. (Möhring, Gaba, and Finger 2019).

Economic indicators

The first group of indicators treat quantity of PPPs in economic terms and measure the total spend (at a farm level, e.g. (Uthes et al. 2019)) or the total sale of (at national or regional levels) of PPPs. National sales of PPPs, grouped under the ‘Harmonised classification of substances’ is a key indicator at an EU level under Regulation (EC) No 1107/2009 of the European Parliament (see (European Court of Auditors 2020) for a detailed discussion of how sales records are calculated and aggregated at the European level). There are many problems with describing amounts of PPPs in purely economic terms, e.g. because of local economic conditions such as inflation, consumer choice and sales tactics[3], but these indicators are used partly due to the relative ease of collecting sales records and, partly due to expressing quantity in monetary terms in order to compare to value to agricultural outputs.

Quantity of Active Ingredients

The second major group of quantity only indicators describe the use of PPP in terms of the mass of active substances applied (or, more rarely, the total mass of PPP applied). The most common form, Quantity of Active Ingredients (QA, e.g. (Möhring, Gaba, and Finger 2019)) simply measures the mass of PPP applied multiplied by the concentration of the active ingredients(s), to provide an equivalent kilogram mass of the amount of active applied (often standardised by cropping area). As with many PRIs, QA will typically be calculated separately for each active and reported as aggregated totals. The focus of QA is on the ‘active’ ingredients of PPPs and there have been some criticisms that this fails to account for ‘inert’, often undocumented, components of PPPs, and as such may under-represent overall risk (e.g. Surgan, Condon, and Cox 2010).

Area treated

Another small group of quantity only indicators, generally of lower policy concern, measure the amount of application in terms of area treated. Examples include the UK Pesticide Usage Survey (Garthwaite et al. 2019; also reports Quantity of Active Ingredients as outlined above). As with the economic indicators, area treated has statistical issues for discussing trends through time, particularly when use is not reported relative to cropping area, or when multiple treatments are combined. However, it is again a simple measure that has proved appealing to decision makers and been used as a driver for discussion of landscape scale impacts, particularly by NGO’s and other conservation bodies.

Number of Unit doses, the Treatment Frequency Index and the Standardised treatment index

The final major group of quantity only indicators represents an attempt to incorporate frequency and intensity of treatment into measures of the amount of PPP applied. The Number of Unit Doses (NUD), is an important French indicator, which uses information on total sales, the recommended maximum dose of actives within a cropping context, and the national cropping area to provide an estimate of average total number of applications per unit area (Hossard et al. 2017).

Likewise, the Treatment Frequency Index (TFI) (Coll and Wajnberg 2017; Hossard et al. 2017), developed in Denmark (Gravesen 2003) but now used in a small number of other European countries (e.g. Stenrød et al. (2008); Kudsk, Jørgensen, and Ørum (2018)), is calculated by dividing the observed rate of PPP application by a recommended standard rate for each product within a given cropping context (Möhring, Gaba, and Finger 2019). The TFI can therefore be thought of as the proportion of the recommended standard rate at the relevant spatial scale that is observed over the recorded applications (or as approximated by sales records). The TFI was formerly used in the implementation of pesticide taxation in Denmark prior to the development of the more sophisticated Danish Pesticide Load indicator (see below) and has formed the basis of usage reduction targets in Denmark and France (see Section 5 for discussion).

[3] https://www.eea.europa.eu/airs/2018/environment-and-health/pesticides-sales

The Standardised Treatment Index (STI (Sattler, Kächele, and Verch 2007); also called the Treatment Index (Uthes et al. 2019)) is another indicator that uses the ratio between the observed and the recommended application rates, although in this case it also incorporates information on the number of active substances applied and the treatment area (typically calculated at the level of a treatment rather than an individual product). STI was developed in part to help monitor the multiple ways in which farmers can act to reduce applications of PPPs (Bürger, de Mol, and Gerowitt 2008), as well as to help characterise the dependence of different cropping systems on pesticide inputs, and has been used in farm level comparative analysis across several European countries (Uthes et al. 2019).

The unifying feature of quantity only indicators is that their calculation depends on the spray records that are maintained as a legal requirement in both the UK and EU Member States for a period of three years (Article 67 of the PPP Regulation). This reliance on simple data around PPP use means quantity only measures can be adapted to a range of different scales and situations, and as a result they've had a prominent role in policy making. However, the fact these indicators make no distinction between different compounds, and thus in effect treat risk associated with all PPPs as equal by mass, is clearly inadequate for describing the potential impacts of the changing composition of applied compounds. For example, a misleading impression may be created by the substitution of high impact substances with alternatives, where a greater net quantity of application (necessary to achieve levels of control) may associated with a lower net impact (because the replacement substances are less impactful per kilogram of application) (Möhring, Gaba, and Finger 2019). It is this crudeness that has been the catalyst for the development of many of the more complex PRIs outlined below.

Qualitative indicators and classifications

In addition to the quantity only indicators listed above and the more sophisticated tools given below there are a small number of intermediate indicators which incorporate a small quantity of chemical information as a proxy for risk, but which do not fit in the category of the more complex methods described below. A simple example of a qualitative index would be the PRI's that underly the Norwegian Banded Pesticide Tax Scheme (Norwegian Food Safety Authority 2005), or the similar systems (formerly) in effect in France (Böcker and Finger 2016) and Belgium (ECOTEC et al. 2001; NAPAN 2014). In all three cases the "Risk phrases[4]" (e.g. corrosive, or irritant) are used as indicators of the potential impact of a substance, which in turn is used to define taxation bands applying to pesticide sales. For situations where the goal is rapid classification of novel substances and to minimise any disputes which might arise, this simplistic binning of substances can be effective, although lacking much nuance with respect to the potential impacts of different substances.

Further (typically) older systems of qualitative risk assessment are reviewed in Levitan, Merwin, and Kovach (1995); Foster and Mourato (2000) and Maud, Edwards-Jones, and Quin (2001). Qualitative assessment of risk was for a long time the standard for PRIs, particularly prior to ready availability of computational resources and online datasets (Levitan 2000; Bockstaller et al. 2008; Lewis et al. 2016). It should also be noted that many forms of multi-component PRIs (see following section) contain qualitative elements (Cox, Babayev, and Huber 2005), which we review below under our discussion of the well-known EIQ indicator . The challenges of how to utilise qualitative indicators of risk, particularly outside of the setting for which they were original developed, have been reviewed elsewhere (e.g. Cox, Babayev, and Huber (2005); Levitan (2000)), and these indicators are often associated with issues of subjectivity and ambiguity in how they communicate risk. For example, choice of descriptor ('high', 'moderate' 'extreme' etc) is a key component of communicating risk that is often challenging to agree with relevant stakeholders, who may have widely differing perspectives of what is or is not acceptable risk (Johnson and Chess 2006). Purely qualitative PRIs, based on expert opinion, still exist in the decision-making processes of some policy relevant activities, but in general they've become less significant as more sophisticated tools and software have become more widely available. None of the more recent reviews of PRIs (Feola, Rahn, and Binder 2011; Labite, Butler, and Cummins 2011; Pierlot

[4] Defined in Annex III of European Union Directive 67/548/EEC

et al. 2017) cover measures that could be described as truly qualitative and it appears likely that their importance, at least within the academic sphere, is in decline in favour of more quantitative or model based approaches (Dubus and Surdyk 2006), although see (Jepson et al. 2020) for an alternative perspective.

EU harmonised risk indicators

One of the most significant events of recent years for the policy management of PPP within the EU has been the adoption in 2019 of harmonised risk indicators as the basis for management targets within the Union (European Court of Auditors 2020). Conceptually, harmonised risk indicator 1 (HR1) can be viewed as an example of a qualitative indicator, in that applied substances are categorised into four broad 'groups' defined under Article 22 of the Regulation (EC) No 1107/2009 and the associated Annex of Regulation (EU) No 540/2011 (European Commission and Statistical Office of the European Union 2019).These 'groups' of substances correspond to approved 'low risk' actives, general approved actives, approved actives listed as candidates for substitution, and any actives not approved for use under EC regulations. Each of these groups corresponds to a weighting value, reflecting increasing perceived risk, see European Commission and Statistical Office of the European Union (2019) for details. The final value of HR1 is the sum of the quantities of active substances placed on the market for each group, multiplied by their weightings, and standardised relative to the reported mean of 2011 to 2013 (European Commission and Statistical Office of the European Union 2019). Full details of the calculation of HR1, including group weights, can be found in European Commission and Statistical Office of the European Union (2019).

HR1 is intended to be collated annually by Member States under Amendment C(2019) 3580 of EC; 2019 (European Court of Auditors 2020), and has been viewed more widely as the political basis for reduction targets in PPP usage at an EU level (Foote 2020). The relative recency in which HR1 has been adopted means that to date there has been limited academic discussion of the index, and questions have already be raised as to whether the classification of actives and the associated weighting appropriately captures PPP risks, particularly given the political dimensions of what substances have been authorised at different risk categories, and which are considered 'candidates for substitution' (European Court of Auditors 2020). There are also wider concerns about the quality of datasets collected across the Union on PPP usage (outlined in Section 5 of this report), which may impact on the calculation of HR1, particularly when comparing with Member States with less sophisticated PPP monitoring programmes (European Commission 2019; 2020).

While HR1 fits naturally within the classification of PRIs used here, the accompanying indicator HR2 is more difficult to interpret. HR2 represents a simple sum of the number of 'emergency authorisations' for PPP across the Union, weighted using the same substance groups as HR1, and expressed relative to the mean of the same baseline period. In context, an emergency authorisation is a power held by Member States to authorise a specific substance for a period not exceeding 120 days, intended to be used to manage specific dangers (such as pest outbreaks or incursions) which cannot be contained by any other reasonable means (European Commission 2017). The interpretation of HR2 thus lies on the question of whether emergency authorisations of substances carry intrinsically higher risk of impact on non-target systems, which is difficult to assess given that such events are often confounded with e.g. delays in the full authorisation of substances for minor uses (European Commission 2017). In the absence of clear guidance on the causes of emergency authorisation, and with inadequate baseline data for comparison, the current authors feel unable to comment on the utility of HR2 and how it should be considered in the wider context of PRIs and their role in policy development.

Fuzzy-expert Indices (I-Phy)

I-Phy (formerly known as IPEST) (van der Werf and Zimmer 1998) is a unique form of indicator, developed specifically to incorporate uncertainty associated with chemical properties or potential impacts, which bridges the gap between a qualitative and fully quantitative approach. The so-called 'fuzzy-expert system' is primarily based on expert opinion but each input variable is graded on a sliding scale between being a favourable or unfavourable condition. These gradings are then the basis for the calculated index value, which is expressed using predefined rules for compartments representing air, surface water and groundwater. The power of this framework lies in the ability to incorporate an extremely wide array of pesticide properties, site specific conditions and application conditions, without necessarily the need for detailed quantitative data.

While having no known policy applications, I-Phy has been increasingly discussed in academic literature, particularly where the aim is to expand on understanding of a poorly or contentiously characterised system (Lindahl and Bockstaller 2012). When compared to other approaches Pierlot et al. (2017), note that, despite their simplicity, both the original I-Phy and the modified version developed by Lindahl and Bockstaller (2012) were relatively strong performers in their comparative assessment (see Section 4; Figure 1). There have also been some attempts to integrate supervised machine learning into the I-Phy framework (I-Phy2v in Pierlot et al. (2017)) although it is not yet clear if this has led to general improvements in performance. Perhaps the greatest challenge with the I-Phy methodology is the issue of repeatability and transparency, given the underlying dependence on subjective expert opinion (Labite, Butler, and Cummins 2011). This may account for why I-Phy has not been more widely adopted in a policy development.

Weighted multi-component PRIs

Weighted multi-component PRIs are by far the most diverse group of indicators and represent one of the major ways in which PRIs have been adapted for local audiences. The defining property of these approaches is that they combine multiple measurements of the chemical properties or experimental findings associated with an active substance, often focusing on the mobility and impacts on non-target organisms. They then calculate a 'score' that describes the relative potential impact based on some predefined formula, varying across indices (Feola, Rahn, and Binder 2011). We term these indices 'weighted multi component' as typically calculation of the score for a given substance incorporates not only underlying measures of impact, but also 'weighting' values, which determine the relative effect that different measures have on the final value for each substance. 'Weightings' are often presented as arbitrary numeric values and can be challenging to objectively interpret, particularly when adapting an index to a novel situation, as the process by which they are defined is often poorly documented and subjective (Dushoff, Caldwell, and Mohler 1994).

The measures contributing to a weighted multi-component PRI are typically derived from measurements collected during the authorisation process, for example the rate of degeneration in soil (known as the DT_{50}) or the toxicity of the substance with respect to standard non target organisms such as birds, bees or *Daphnia* (water fleas) (see Table 1 in Labite, Butler, and Cummins (2011)). In general, a specified set of these measures are combined within the index formula for a specific substance and the field level risk is then represented by the summed value of all applied actives multiplied by their respective quantity of application, typically represented as the quantity of active substance (QA), but occasionally using metrics akin to the treatment frequency index (TFI) as described above.

Before dealing with the true multi-component indicators, it is worth noting there are a small number of PRIs that treat risk as proportional to single experimental derived measures as a proxy for all human health or environmental impacts of a given PPP. In the majority of cases the chosen metric is either the acute or chronic toxicity of the substance in rats as represented by (one over) the LD_{50} (the statistical estimate of the exposure concentration that is associated with a 50% increase in mortality rate in a laboratory setting), typically multiplied by the quantity of active substance applied. Notable examples include studies on long term toxicity change in herbicide usage in the United States (Kniss 2017; 2016). As with the more complex multi-component measures, these single dimensional values are usually scaled with respect to the QA and represent a simple attempt to scale the amount of application by an index of impact. The authors are not aware of any policy instruments drawing directly on such single dimensional indices and most usage has been restricted to academic discussion of changing patterns of risk.

Environmental Impact Quotient

One of the oldest and most well-known examples of a weighted multi-component PRI is the Environmental Impact Quotient (EIQ) (Kovach et al. 1992). The EIQ is the only indicator developed outside of Europe reviewed here having been originally developed by members of the New York State Integrated Pest Management Program and represents older versions of weighted multi-component indexes developed prior to the advent of widespread computation and data resources (Maud, Edwards-Jones, and Quin 2001). Measures included in the EIQ include dermal toxicity w.r.t. rodents; any evidence of chronic human toxicity or 'systemicity'; acute toxicity w.r.t. fish, birds, bees and beneficial arthropods; the potential for leaching and surface run off; and half-life of the substance in soil and on the surface of plants.

For each of these measures 'bins' (numerical ranges or intervals applied to underlying continuous measures) are constructed to represent substances with 'high, 'medium' or 'low' risk, with each bin being assigned a point value of 1, 3 or 5, which will go on contribute to the overall 'score' for the substance. This final value is calculated using a standard formula, which incorporates 'weightings' of the different aspects of risk and multiplies these by the 'scores' for the bins applicable for each substance. A full description of the calculation of the EIQ, including the weighting can be found in (Kovach et al. 1992). In common with many other similar PRI the EIQ is sometimes expressed as set of sub-indices, in this case representing risks to farmworkers, consumers and natural ecosystems respectively. The commonly presented final EIQ score, or whatever relevant subcomponent is the focus of investigation, is calculated individually for each substance applied, and then multiplied by their respective mass of application, with the total value of the interpreted index being equal to the sum over contributing substances.

The EIQ usefully illustrates the core features typical of weighted multi-component PRIs, i.e. use of underlying measures of potential impacts from laboratory studies; which are sometimes 'binned' to assign a value reflecting the inferred risk; are then combined in a 'weighted' formula to give the overall 'score' for a substance; which in turn is typically multiplied by the amount of substance applied; with the final value being the sum of over substances at scale under investigation. Likewise, the EIQ can be used to illustrate some of the weaknesses particularly in the older generation of such indicators. Many of these problems arise from the use of binning to convert continuous underlying measurements into categorical scores representing low, medium and high risk. This system is a deliberate design element of the EIQ intended to help resolve sparse input data when assessing novel compounds (Kovach et al. 1992). However, early and influential criticism of the EIQ (Dushoff, Caldwell, and Mohler 1994) noted that binning means compounds differing in their toxicity or persistence by a factor of over 1000 can be reduced to a five to one scale in terms of their effect on the EIQ score. Likewise, the way in which quantitative information is fitted to categories can, in some circumstances, mean that a higher qualitative risk can be assigned to pesticides even where the quantitative risk is actually lower in a straight comparison (Dushoff, Caldwell, and Mohler 1994). These issues with EIQs were reinvestigated in the simulation study of Kniss and Coburn (2015) which highlighted that, at least for herbicides, different elements of the EIQ can have widely differing impacts on the estimated value, and strongly criticised its suitability to field conditions.

Maud, Edwards-Jones, and Quin (2001) review several other older PRIs with similar underlying designs to the EIQ, assessing their applicability to the UK. Their conclusions were that none of the examined indicators were judged suitable for the then state of UK agriculture, citing lack of transparency around weighting, inability to discriminate between compounds and the political issues that surround the incorporation of human health alongside indicators of environmental impact (see Section 4 for discussion).

Norwegian Environmental Risk Indicator

While the earliest versions of weighted multi-component PRIs are largely considered outdated, the core concept has been readapted in several recent indicators significant for European policy development (Labite, Butler, and Cummins 2011). The Norwegian Environmental Risk Indicator (Stenrød et al. 2008) (NERI, sometimes also NRI as in (Pierlot et al. 2017) and Figure 1 of this report), was developed in Norway and found to be superior to the EIQ in terms of calculated output, particularly for systems dominated by highly toxic or highly persistent chemicals. The value of NERI is determined by a simple linear sum of a collection of ecological measures (thus giving it both a narrower scope than the EIQ but also avoiding the complexity of an EIQ like weighted formula). Like the EIQ, each of these measures are categorical values associated with scores attached to bins with pre-defined threshold values. This binning means that NERI is subject to many of the same criticisms that apply to the EIQ, although it should be noted that unlike the EIQ, where the minimum score for any measure is always 1, the binning for NERI means that compounds which do not present a significant risk with respect to a particular variable will have a score of zero. NERI also has the advantage that the threshold values used have a clear regulatory basis, in that they are defined by the major acceptance criteria used throughout the EU for PPPs (Labite, Butler, and Cummins 2011). This is not the same as saying that criteria have an objective basis in terms of potential non-target impact, but it does naturally place the index within the existing regularly regime and aids in the harmonisation and transparency of an otherwise complex index (OECD 2004). Details of the calculation of NERI can be found in Norwegian Food Safety Authority (2005).

Danish Pesticide Load Indicator (PLI)

The Danish Pesticide Load Indicator (PLI) (Miljøstyrelsen 2012; Kudsk, Jørgensen, and Ørum 2018) [5]was developed as a policy tool and is one of the most significant examples of a recent PRI in terms of impact, in part because it has a direct link to differentiated taxation of pesticides used in Denmark (Pendersen, Helle, and Andersen 2015). Denmark has one of the most sophisticated (and highest) taxation systems for agriculture in the world, and is often cited as a key example in reviews on the subject of pesticide taxation and pesticide policy instruments more generally (Finger et al. 2017; Lee, den Uyl, and Runhaar 2019; Sud 2020); see Section 5 for discussion. The Danish PLI, was developed as a strategic objective for the 2013-2016 Danish Pesticides Strategy (Kudsk, Jørgensen, and Ørum 2018) and replaced the previous system based on the Treatment Frequency Index (see above) over the growing season 2010-2011.

The PLI covers three key policy areas, represented in the calculation as sub-indices, relating to human health, environmental fate and environmental toxicity. Like the EIQ and NERI, the PLI is primarily based on the combination of measures arising from the existing regulatory process (collated in the Pesticide Properties database(PPDB; Lewis et al. 2016). The three sub-indices are calculated in different ways. The human health component represents numerical scores applied to the 'risk phrases' supplied with individual products[6]. Environmental fate is the weighted sum of Soil DT_{50} (the time taken for half of a substance to decay in soil), an index of Bioaccumulation (the ratio at which organisms pick up compounds from their environment), and the SCI-GROW index, itself a combination of a substances rate of aerobic soil degradation and half-life developed by the US EPA Office of Pesticides (Estes, Pai, and Winchell 2016). Environmental toxicity is likewise a weighted sum of measures of acute and chronic toxicity with respect to birds, mammals, fish, water fleas, algae, aquatic plants, earthworms and honeybees.

[5] An older indicator of the same name is reviewed in (Labite, Butler, and Cummins 2011) but is conceptually different from that described here and more closely related to the TFI (Møhlenberg, Gustavson, and Sørensen 2002; Pendersen, Helle, and Andersen 2015).

[6] The PLI considers human health impacts based on risk phrases applicable to operators, and any risks to bystanders and the wider public are represented only indirectly within the calculation.

Interestingly the ‘weighting’ applied to the environmental toxicity component of the PLI varies by type of PPP application, with different organism having different relative weightings for seed treatments compared to other application methods (as an attempt to crudely capture differences in the chance of exposure for different groups between different types of treatment Miljøstyrelsen (2012)). Note that unlike in previous indicators in this section, measures in the PLI are not ‘binned’ when generating the scores (i.e. avoiding many of the problems described for the EIQ). Instead each measure is expressed with respect to a ‘reference substance’, for which all others have a percentage of impact for the relevant measure (Miljøstyrelsen 2012). It is these ‘standardised’ measures that are the target of ‘weighting’, with the final scores for a substance being the sum of these weighted values. These final scores are expressed by multiplication by the ratio of the application dose and some defined standard dose calculated in manner similar to the TFI, with the final indictor value summed all substances. Full calculation details for the PLI can be found in Miljøstyrelsen (2012).

One of the chief features of the PLI compared with the EIQ and NERI lie in both its comprehensive nature, made possible by the convergence of regulatory information in resources such as the PPDB (Lewis et al. 2016), and also its rejection of subjective binning of measures (which makes it much easier to separate compounds in terms of their relative risk). This avoids many of the statistical pitfalls that come from combining binned measures (Kniss and Coburn 2015) (e.g. the score of different substances for a particular measures e.g. toxicity with respect to bees will vary linearly, rather than being categorised into discrete units during calculation). This does not necessarily mean that the PLI is fully ‘objective’ in representing risk as the weighting of different measures (and the selection of reference substances) still reflect the perceived relative importance of different impacts, which are not fully justified in the published methodology (Miljøstyrelsen 2012) (e.g. in the decision to weight organisms differently when considering substances that are applied as seed treatments). Due to having come to prominence only relatively recently, the PLI was not included in any of the comparative studies examined in Section 4, making it difficult to assess-field performance relative to, e.g. the EIQ. There have however been several recent attempts to use the PLI to help ground-truth ‘quantity only’ indicators, which indicate distinct behaviour when compared to these simplistic measures (Möhring, Bozzola, et al. 2020; Möhring, Gaba, and Finger 2019), see Section 4 for further discussion.

PestScreen

PestScreen (Juraske et al. 2007) is an example of a weighted multi-component PRI that is in part driven by underlying explicit models rather than being based only on laboratory measures as in the previous examples. This in some sense makes it transitionary between this group and the Exposure-toxicity ratio methods described in the following section. As with the EIQ or NERI, the numeric final value of the index is constructed from categorical scores associated with different degrees of perceived risk. However, where it differs is that the values used are often themselves the outputs of models and may make use of a much wider array of information regarding local conditions than is used by systems like the PLI. The value of the PestScreen index is determined by three subcomponents representing Fate, Exposure and Toxicity. The Fate component for example is the combination of binned risk scores for values of long-range transport potential and overall persistence, both of which are estimated using the multi-media fate model ‘Simple Box 3.0’. Exposure is based on an estimate of the proportion of emitted substance that enters the human population (either by inhalation or ingestion) based on another multi-media model implemented in USES-LCA 2.0 (van Zelm, Huijbregts, and van de Meent 2009). Only with toxicity do we see something resembling the laboratory derived measures discussed thus far, incorporating the acceptable (human) daily intake, and the LD_{50} values in rats, bees and fish. This blending of modelling and underlying measures is an increasingly important element of many recent PRIs, which like PestScreen are often intended to be calculated through a dedicated software package (in this case an Excel spreadsheet (Labite, Butler, and Cummins 2011)) rather than by hand. This is a notable improvement in terms of ensuring the consistency by which PRIs are calculated but also leads to a tendency for reduced transparency and issues of stakeholder trust when applied under field conditions (Rose et al. 2018).

Exposure–toxicity ratio methods

In the previous section we focused on indices which represent the potential impacts of PPPs using the laboratory properties of the active substance, scaled to the quantity of application for the purposes of aggregation. Such methods are often contrasted with a diverse group of indicators collectively known as the exposure toxicity ratio (ETR) approach which have become increasingly dominant in academic discussion of PRIs (Feola, Rahn, and Binder 2011; Labite, Butler, and Cummins 2011). The defining feature of ETR approaches is that risk is expressed based on the ratio between the predicted exposure (usually based on the predicted concentration of the PPP in a specified environmental compartment) relative to the toxicity values (usually LD_{50}) for relevant organisms. This approach was originally highlighted by Reus et al. (2002) and has since become the basis of a large number of tools, particularly as dedicated software platforms and data resources have become increasingly widespread.

Due to their need to predict concentrations within a specified compartment (e.g. aerial or fresh water) many ETR indices are dependent on high resolution data regarding the local conditions at the time of application, such as local weather conditions, information on slope, and the distances to nearby water bodies (Feola, Rahn, and Binder 2011; Labite, Butler, and Cummins 2011; Pierlot et al. 2017). This has tended to lead to such indicators being primarily associated with decision making at the farm level, although a smaller number have been adapted to regional or national surveillance (see Section 5).

Synoptic Evaluation Model for Plant Protection Agents

An example of an exposure–toxicity ratio indicator that has received increasing attention in policy is the German derived Synoptic Evaluation Model for Plant Protection Agents indicator or SYNOPS (Gutsche and Rossberg 1997). This model was one of the earliest ETR approaches and has been highly influential in the development of the latter OECD (OECD 2000) and harmonised European indicators (Kruijne et al. 2011); see below . Since its original development SYNOPS has been adapted to be a free to use web accessible tool intended to allow users to evaluate field level strategies across a range of realistic European conditions (Strassemeyer et al. 2017). Since 2008, SYNOPS has been integrated into the German National Action Plan and is used for surveillance across a set of sampled localities to establish national trends (Strassemeyer and Gutsche 2010).

The structure of the SYNOPS model (there are three indicators with the SYNOPS name (Strassemeyer et al. 2017) but we focus here on the original ETR indicator) draws on topographic information (e.g. from GIS data), crop and weather parameters, soil properties, surface water type and distance and the properties of the active substance, to fit a series of one dimensional core models that simulate chemical movement around the plant root zone and volatilisation from bare soils (Ferrari et al. 2005). These models in turn predict drift drainage, runoff and erosion, ultimately generating a predicted concentration in field margins, soil and surface water. As with all ETR approaches predicted concentration is then expressed relative to the toxicity (NOEC, LC_{50}) for relevant organisms for each compartment (earthworms in soil; algae, duckweed, *Daphnia*, midge larvae and fish in surface water; and bees in field margins (Strassemeyer et al. 2017)).

As with many more complex PRIs, SYNOPS is integrated into a series of databases designed to store much of the background information used to calculate the index (at least when presented as a web application) (Strassemeyer et al. 2017). A consequence of this is restricted geographic scope based on the availably of underlying data. Recent work by de Baan (2020) adapts the SYNOPS framework to a novel region (in this case Switzerland), as well as for understanding the uncertainty and dependence of the indicator on particular input parameters, notably slope, temperature, precipitation, water distance, crop interception at the date of application, the soil adsorption coefficient, and the soil DT_{50} (time take for half of a substance to decay in soil) .

SYNOPS is one of the stronger performing indicators in the comparative study of Pierlot et al. (2017) although there is evidence for over-prediction of horizontal transfer by runoff compared to comparable ETR indicators (including POCER; see below). Commentary on SYNOPS in the review by

Labite, Butler, and Cummins (2011) highlights the power of the methodology in the assessment of impact over time, something which has been a key focus in the development of the index (for example the development of SYNOPS-TREND; focused on comparing risk across multiple years) but also note the high complexity associated with the calculation and resulting difficulty in adapting it for application outside of the original development scope (although see de Baan (2020) for a recent successful example).

The Environmental Yardstick for Pesticides

Just as SYNOPS has been highly influential in the development of policy in Germany, the equivalent role in the Netherlands has been played by the Environmental Yardstick for Pesticides (EYP) (Reus and Leendertse 2000). Like SYNOPS, calculation of EYP draws on an underlying prediction of compound concentration within a compartment and is also associated with a web interface[7], albeit one which explicitly assumes characteristic Dutch soil types in the calculation. Compared to SYNOPS, the EYP is often presented in highly simplified manner, such that estimated spray drift percentage is given as an input parameter (defaulting to 1%) multiplied by a fixed value based on the method of application, as opposed to being estimated from local conditions. This results in greatly reduced data requirements when compared with SYNOPS, although the two are shown to closely correlate in compound rankings in the comparative study by Reus et al. (2002). An oddity of the EYP is that results are presented as environmental impact points, a set of threshold values that has a highly specific meaning under Dutch law, but which are not obviously extendable to other national contexts (Labite, Butler, and Cummins 2011). This, along with the strong underlying assumptions regarding soil type, are reasons why the EYP has not been widely adopted outside of the Netherlands. Within the Netherlands a wider environmental labelling and green accounting programme, of which the EYP forms a component, has been associated with reductions in pesticide use and toxic load during the period up to 2000 (Levitan 1997; Halberg, Verschuur, and Goodlass 2005), although data is sparse regarding more recent trends. See Section 5 for further discussion. Due to the difficulties in adapting the index outside of the Netherlands, the EYP has rarely been included in comparative assessments of different PRI, such as those reviewed in Section 4. There is however some data relating to the use of EYP to compare alternate cropping procedures in potato which could be adapted for benchmarking purposes if required (De Jong and De Snoo 2002).

p-EMA

Another simplified index that shares some properties with the exposure–toxicity ratio approach is the UK derived p-EMA method developed by Brown et al. (2003) and incorporated into the Environmental Management for Agriculture software package (Lewis et al. 2003; Dubus and Surdyk 2006). p-EMA is designed to predict environmental concentrations associated with spray using the soil water partition coefficient, normalised for organic carbon content of the soil, half-life (normally first-order) for degradation in soil (DT_{50}) determined either in the laboratory or the field (preferred); and half-lives for aqueous photolysis, neutral hydrolysis and dissipation from the water phase of a water-sediment system. Input data consists of initial concentration applied, and the growth stage of relevant crops and a UK postcode, to assess the prespecified database of soil and hydraulic properties. Calculation is based on the MACRO model (Dubus and Brown 2002) and is based on underlying simulations across four risk classes. Due to the lack of a toxicity component for relevant organisms this method is not on its own considered a true ETR approach, although the authors note this is a natural extension that could be easily adapted (Brown et al. 2003). Like the environmental yardstick, discussion of p-EMA has been largly restricted by its specified geographic scope, although Labite, Butler, and Cummins (2011) and Reus et al. (2002) both comment its practical applicability, particularly as a self-monitoring tool for farmers and advisors.

[7] https://www.milieumeetlat.nl/en/hoe-werkt-het-open-teelt.html

The Pesticide Occupational and Environmental Risk Indicator

The Pesticide Occupational and Environmental Risk Indicator- POCER (Vercruysse and Steurbaut 2002; Claeys et al. 2005)), was original developed in Belgium and is highlighted in the review by Labite, Butler, and Cummins (2011) for its depth of coverage and dynamic nature. POCER consists of modules covering various aspects of both human health and environmental risk selected based on criteria outlined in EU Directive 91/414/EC (Vercruysse and Steurbaut 2002). The ten modules covered by POCER, each representing either a model or collection of laboratory measures, are as follows: risk to pesticide operators, risk to workers and secondarily exposed persons, risk to bystanders, persistence in soil, risk of groundwater contamination, risk to aquatic organisms (fish, Daphnia and algae), acute risk to birds, acute risk to bees, acute risk to earthworms and risk to beneficial arthropods. A full description of the calculation of POCER can be found in Claeys et al. (2005).

The aggregate overall value for POCER is obtained by transformation of each of the underlying measures relative to specified limit values which results in a 0 to 1 indicator of risk. These can them be summed using a specified weighting scheme (Claeys et al. 2005) to generate the total indicator for the formulation.

Unusually for such a complex indicator, POCER has been adapted for multiple situations most notably to include non-agricultural pesticide usage (Claeys et al. 2005). POCER was considered to be one of the comprehensive indicators in the review by Labite, Butler, and Cummins (2011), although underperforming overall relative to other complex indicators in the comparative study of Pierlot et al. (2017); Figure 1 (p. 21).

Harmonized environmental indicators for pesticide risk

The EU FP6 funded Harmonized environmental indicators for pesticide risk (HAIR) was one of the most ambitious projects around PRIs of recent years, and attempted to bring together many of the ideas developed across ETR indicators into a cohesive structure across a range of European conditions (Strassemeyer et al. 2007). Despite the development of a number of approaches, including an online platform developed to support calculation (https://www.pesticidemodels.eu/hair/home (Kruijne et al. 2011), the HAIR indices have seen remarkably little discussion in recent reviews and appear to have been largly neglected in the development of policy (although see Pivato et al. 2015)). One element of HAIR (DRAINAGE-HAIR) was included among the indicators examined by Pierlot et al. (2017) and is considered both the most complex and performing indicator among those examined (see Section 5).

DRAINAGE-HAIR is described as being derived from a classical meta-modelling approach within an underlying MACRO flow model In practical use, HAIR defines a scenario comprising the region, compounds, crop(s), and time period of interest, which is linked to underlying databases to supply the parameters for the model. Modelling focusses on four components: terrestrial, aquatic, groundwater, and workers/bystanders, and the wider software system also includes a number of more general indicators linked to overall protection goals (Kruijne et al. 2014). Unlike some other ETR approaches, chronic and acute risk are both considered in the calculation of toxicity and there is more support for filling of missing values than in other comparable systems (Kruijne et al. 2011). For details of calculation see Kruijne et al. (2011). Despite being the culmination of a major European research effort, evidence for the integration of HAIR into policy instruments is sparse. Likely this is in part because those countries placing the greatest emphasis on PRI development often have their own competing National PRI (Section 5), and the highly data intensive nature of the indicator make it less suitable for countries without a robust existing reporting network. Overall HAIR is a highly developed and complex system of different indicators that mostly lacks a clear functional role within current EU policymaking. It may be that overtime, as broader issues in EU wide collection of usage data are resolved (see Section 5) this indicator will return to prominence.

OECD indicators

Another influential alternative approach to try to build a harmonised and comprehensive indicator was led by the OECD (OECD 2000) culminating in the indicators known as REXTOX, ADSCOR and SYSCOR (Gutsche and Carley, n.d.). These three indicators are all aimed at estimating how much of an applied pesticide migrates from the site of application to surface waters, and its significance for aquatic organisms

- REXTOX (Ratio of EXposure to TOXicity) is defined as the total usage divided by the toxicity for relevant organisms multiplied by two indexes of exposure, representing the % of material entering the surface water by spray drift and the % entering the surface water by run off. Calculating these measures incorporates local information on slope, precipitation, soil type, width of buffer zones required for risk mitigation and pesticide characteristics, as well as mode and intensity of application. Criticism of REXTOX tends to centre on large data requirements, as well as on the relative dominance of spray drift relative to surface run off (Møhlenberg, Gustavson, and Sørensen 2002). The same authors also note that under field conditions the value can be approximated as a weighted sum load index (ratio of QA to aquatic toxicity) between two categories of pesticides: those with a spray buffer zone and those without. In one of the few comparative studies that include this measure (Møhlenberg, Gustavson, and Sørensen 2002) REXOR was found to show similar patterns to the much simpler Treatment Frequency Index (called Frequency of Application), which may account for why it is not more widely discussed in recent PRI reviews.
- ADSCOR (ADditiveSCORing) is a categorised version of the same core concept as REXTOX. Here the exposure parameter is represented by predefined scores linked to the method of application, the dose rate, the frequency of application, and observance of buffer zones (values on page 23 of Møhlenberg, Gustavson, and Sørensen (2002)). As with previous comments on the EIQ, this use of fixed scores to represent differing degree of risk has been subject to criticism, particularly when compared to more sophisticated modelling approaches, and leads to relatively poor performance in the comparative analysis by Pierlot et al. (2017) although still well beyond those achieved by quantity only measures.
- SYSCOR (SYnergistic SCORing) takes this concept of categorisation even further by using predefined tables to combine scorers a for all exposure-related and hazard variables (including for area treated). This final indicator was designed in part to try to resolve the issue of synergistic responses between variables, although there is limited information regarding relative performance with respect to other indicators.

As with HAIR despite having been the product of major transnational research effort there is limited evidence for the adoption of OECD indicators in the development of policy. Those countries that were involved have tended to favour their own national indicators, presumably as these better reflect local decision making and data availability (Møhlenberg, Gustavson, and Sørensen 2002). Hence while significant for understanding the development of PRIs internationally there appears to be little practical significance associated with these indicators in terms of policy development.

Other well characterised approaches

Other well characterised examples of indicators drawing on the exposure toxicity ratio approach include the Environmental Potential Risk Indicator for Pesticides; EPRIP (Padovani, Trevisan, and Capri 2004; Tsaboula et al. 2016), the Pesticide Impact Rating Index; PIRI (Kookana, Correll, and Miller 2005) and the Ecological Relative Risk; EcoRR (Sánchez-Bayo, Baskaran, and Kennedy 2002). Each of these has seen some practical use in various contexts and EPRIP is recognised as one of the best performing indices in the comparative analysis by Pierlot et al. (2017).

Mechanistic and complex models

The line between what constitutes a complex indicator as opposed to a 'model' is inherently subjective and dependent of the scale of the question and sophistication of the user. For ease of discussion we will use 'model' to refer to either a statistical estimate of toxicity (or more commonly estimating exposure within a compartment), or to a mathematical descriptions of a mechanistic process related to how a compound travels or is broken down within a compartment (most commonly soil and or ground water). Our review of this subject will, by necessity, be brief as this is a large area of technical research with many different approaches. Interested readers are referred to work by Siimes and Kämäri (2003); Dubus and Surdyk (2006) for a general introduction to fate modelling and discussion of some of the most influential approaches.

Environmental fate modelling began in the mid 1980's with an emphasis on the simulation of pesticide transfer through a soil column. Since then a huge variety of different models have been developed, with varying degrees of integration into PRIs and the characterisation of risk to non-target systems. Data requirements vary wildly but most incorporate elements of local climate, properties of the soil, and crop cover, as well as the application rates, sorption and degradation characteristics of the active compounds applied in PPPs (Dubus and Surdyk 2006). In many cases, implementation of the various models involves an extensive calibration phase to parametrise the model for local conditions and may involve extensive assumptions based on standardised scenarios (Dubus and Surdyk 2006). As with other elements of PRI development there has been a historic tendency for fate models to be developed in isolation to reflect particularly conditions and experimental set ups (Siimes and Kämäri 2003).

Compared to PRIs, the regulatory importance of fate models has led to a greater emphasis on validation and prediction of field or lysimeter datasets. It should however be noted that validation status of models is complicated by the requirement for calibration and there are often extensive uncertainty and subjectivity in the value of underlying parameters when applied to novel conditions outside of the original devlopement context (Dubus, Brown, and Beulke 2003). Comparative studies of different approaches are rarer, with one the largest being that of Vanclooster et al. (2000) which explored 36 combinations of models and scenarios across four major EU Member States, and included general recommendations for improving the quality and repeatability of datasets. The general conclusions of validation studies summarised in Dubus and Surdyk (2006), indicate that the level of fit obtained with pesticide fate models against lysimeter or field data is usually within one order of magnitude of the observed concentrations (assuming adequate calibration) and that the potential (or lack of) for leaching is usually adequately represented (Dubus and Surdyk 2006). In Europe the FOCUS group of experts from regulatory authorities, registrants and government institutes/universities has played a key role in helping to harmonise differing approaches (Labite et al. 2013) and several fate-models have been adapted for regulatory and authorisation practice at a transnational level.

MACRO (Jarvis 1995; Dubus and Brown 2002; Dubus, Brown, and Beulke 2003; Larsbo et al. 2005) is a one-dimensional dual-permeability model of water flow and solute transport in macro-porous soil and has a long history of development and evaluation across different European contexts (Dubus and Surdyk 2006). Importantly for the development of PRIs, MACRO is the underlying model for drainage output into surface water in both the SYNOPS and HAIR frameworks and has thus been highly influential in the development of complex ETR indicators. In more recent works, MACRO is often paired with the crop growth model STICS (Lammoglia et al. 2017; 2018) to simulate pesticides fate in complex cropping systems and to consider some agricultural practices such as fertilization, mulch, or crop residues management. MACRO is the only example of a fate-model included in the comparative analysis of Pierlot et al. (2017), where it was parameterised using multiple sets of calibration parameters for two of the study sites in France. The input soil variables needed to run the model were, i) for each soil layer, depth, texture, stoniness, pH, organic matter content, ii) bedrock nature, and daily weather data, rain, evapotranspiration, minimum and maximum temperature, collectively given some sense of the extent of localised information required by typical fate models. Unsurprisingly MACRO was one of the most successful tools examined by this study for predicting local concentrations in downstream freshwater (See Section 4, Figure 1). However, the huge costs associated with required data acquisition make it unlikely that it will ever be possible to scale such a procedure to a general use tool at the farm level (see discussion in Section 4).

Other fate models relevant for understanding the PRI outlined above include SCI-GROW, PRZM and PESTLA. The first two were developed by the US Environmental Agency Office of Pesticides as screening tools to estimate drinking water pesticide exposure concentrations in groundwater (Estes, Pai, and Winchell 2016). SCI-GROW, the older of the two indices, is based on a pesticide's aerobic soil degradation, half-life, and linear adsorption coefficient normalized for soil organic carbon content, in combination with the maximum application rate and number of annual applications. This concept is routinely derived for standardised soils (or averaged over multiple soil types) and appears in the environmental fate component of the Danish PLI (Kudsk, Jørgensen, and Ørum 2018). PRZM is a more sophisticated and localised one-dimensional finite-difference model that accounts for pesticide and nitrogen fate in the crop root zone[8] and has now largely replaced SCI-GROW for regulatory assessment of environmental impact of PPPs in the USA[9]. PESTLA (van den Berg and Boesten 1998; Boesten and Gottesbüren 2000) is another influential fate model developed originally in the Netherlands, and integrated into the Belgian POCER indicator (Claeys et al. 2005). Like MACRO, PESTLA is a deterministic model, although its core structure is quite different and emphasises water flow, soil temperature and pesticide behaviour as separate elements.

Finally it is worth noting that a small number of very recent studies have begun to explore the potential role of 'big data' and machine learning approaches in the area of PPP risk modelling e.g. Trajanov et al. (2018). The challenge for these methods thus far has been obtaining adequate training data, particularly for 'end-points' such as the occurrence of "risky" applications (Trajanov et al. 2018), as well as issues of generality and 'overfitting' to potential unrepresentative populations (see discussion in Vanclooster et al. (2000)). As the extent of data science in the agricultural sphere continues to grow we should expect that these kinds of more general 'big data' tools will see increasing use (Wan 2015; Kamilaris, Kartakoullis, and Prenafeta-Boldú 2017), although they are unlikely to ever fully replace the deterministic frameworks outlined above, and any eventual policy implications are difficult to predict at the present time.

4. Comparing Pesticide Risk Indicators

When considering how to compare different PRI approaches and methods there are several factors that need to be considered. The first is the obvious question of how well various methods predict the environmental and human health impacts associated with PPPs. As noted in our introduction to Section 3, calculating the 'true' impact of a given set of PPP inputs is usually prohibitively expensive in a real-world context, and hence the majority of studies that have compared the predictive power of PRIs have focused either on understanding the consistency of different PRIs to predict the same relative ordering of compounds (e.g. (Möhring, Bozzola, et al. 2020; Möhring, Gaba, and Finger 2019; Reus et al. 2002)), or have explored the power of methods to predict concentration of compounds in a specified compartment as an indirect proxy for risk (e.g. Oliver et al. 2016; Pierlot et al. 2017)). The authors were unable to find any examples of studies where PRIs have been assessed on their ability to predict the value of an ecologically relevant end point, e.g. the change in population of a vulnerable species, and there are significant challenges towards conducting such a study under realistic field conditions particularly if multiple potential impacts are to be assessed simultaneously (Dubus and Surdyk 2006), (although see Woodcock et al. (2016) for a partial, single risk component, example). Given this limitation it is important to acknowledge that much of the validation and comparative work around PRIs has focused on the question of whether simpler and less data intensive indicators (most notably quantity only approaches) can be used in place of the more complex, more data intensive methods. It is also worth noting that comparative studies to date have tended to be geographically restricted, so it is not necessarily clear if the 'best' indicators identified are the most accurate under all

[8] https://www.epa.gov/ceam/przm-version-index

[9] https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/about-water-exposure-models-used-pesticide

circumstances, e.g. due to variation in underlying soil types (Pierlot et al. 2017), and this limits our ability to rank different approaches consistently.

The other aspects of comparing PRIs relates to the usability of the various indices for practical (and especially large scale) decision making (see discussion in Section 5 on the various uses to which of PRI have been applicable). The most common framework for addressing usability relates to complexity of the indicator and the extent of data requirements, particularly focusing on information which must be collected at a field or application level (such as slope, distance to water bodies, soil properties and weather conditions at time of spraying (Pierlot et al. 2017)).This reflects the important role of costs of data acquisition/standardisation in the practical application of PRIs, particularly when used at scale (Feola, Rahn, and Binder 2011). Figure 1, taken from Pierlot et al. (2017), shows the indices we have considered in terms other predictive qualities relative to complexity. It shows, to no great surprise, that more complex indices, up to and in including mechanistic models, will generally outperform simple indices, particularly those based on 'quantity only' measures. However, for practical implementation, calculating these more complex indices tends to draw on localised data sources, which can be expensive to collect and aggregate, particularly at the national or transnational scale.

In addition to the data 'overhead', consideration of usability needs to account for the intrinsic complexity of the calculation associated with different PRIs. In most cases the intended audience for PRIs are non-technical, and thus particularly sensitive to the burden of trust required to calculate some of the more complex indices (Rose et al. 2018). Increasingly there has been a trend towards the development of dedicated software packages to help support calculation, although many of these are still dependent on an in-depth understanding of, for example, databasing software to use effectively (Feola, Rahn, and Binder 2011). In general, there is a strong trade-off between the usability of PRIs and their quality of prediction ((Dubus and Surdyk 2006); Figure 1) , which is reflected in the persistence of conceptually simple approaches in the area of policy development concerning large spatial scales, such as national or international monitoring efforts, or where transparency and stakeholder trust are key concerns (Rose et al. 2018). The remainder of this section reviews the classes of indicator described in Section 3 with some general comments on performance before moving in detail onto the small number of field validation studies that have attempted to examine the performance of PRIs under realistic field conditions.

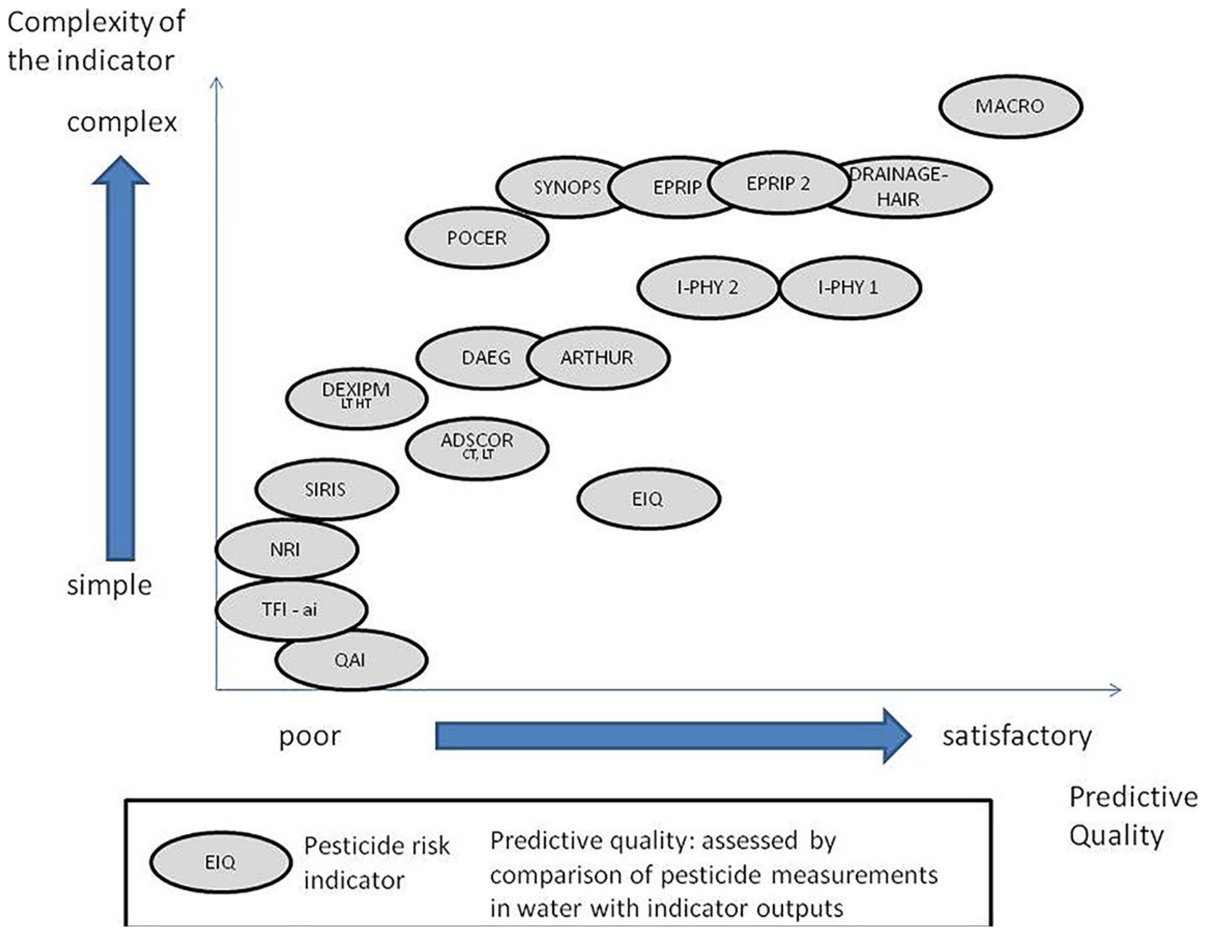

Figure 1 Conceptual schematic of results of the comparative study of PRIs conducted by Pierlot et al. (2017). Figure taken from the cited reference.

General criticisms of PRIs

Before reviewing the differences between the various forms of PRI and the evidence for their quality of prediction, it is worth noting some general criticisms that apply to almost all existing PRIs. In their original criticism of the EIQ, Dushoff, Caldwell, and Mohler (1994) made several key points regarding the general structure of PRI methodologies, some of which are still relevant despite the development of increasingly sophisticated methodologies. The first is that, in almost all cases, when multiple substances are considered within a PRI the aggregate risk of the collective group is represented as the simple sum of the risks for each individual substance. This is somewhat problematic given the increasing concern for and research into synergistic effects linked to the impact of mixtures of PPP on non-target organisms (Laetz et al. 2009; Chen et al. 2015; Bopp et al. 2019; More et al. 2019) but is also a natural consequence of a regulatory regime where substances are approved in isolation as opposed to as part of specified spray regime (Labite, Butler, and Cummins 2011; Milner and Boyd 2017; More et al. 2019). The ongoing Euromix project[10] represents one of the first systemic attempts to develop processes for component-based mixture risk assessment (van der Voet et al. 2019), alongside work by EFSA on the development of cumulative assessment groups on the basis of toxicological profiles (EFSA 2013; Bopp et al. 2019; More et al. 2019). As yet this work has not filtered into the wider conversation regarding development of PRIs but may become increasingly relevant as methods develop. It should also be noted that the study of PPP mixtures remains a large and complex field of research, including various approaches to the devlopement of appropriate null models (reviewed in Schäfer and Piggott (2018)) which may come to provide a conservative approximation of expected impact in the absence of explicit synergistic effects (see e.g. Belden, Gilliom, and Lydy (2007)) and that

[10] https://www.euromixproject.eu/

current guidance from the EFSA identifies the 'concept of dose addition as a pragmatic and precautious default assumption, unless there are indications that the alternative concept of response addition is more appropriate' and that the 'competing concept of response addition often under-predicts observed mixture effects'(More et al. 2019).

What is perhaps a more troubling criticism of PRIs is the distorting effects that having a single indicator value can have on decision making, particularly when combined with legislative thresholds which can create perverse incentives and/or marginalise the cumulative risks of 'safe' applications (Dushoff, Caldwell, and Mohler 1994). As with many areas of agricultural policy, clarity of presentation and knowledge of uncertainties are key components of PRI design and strongly influence how stakeholders would be expected to interact with the indicators (Möhring, Wuepper, et al. 2020). By design, many PRIs mask uncertainty in the underlying information to better present a simple decision support tool or to reflect the needs of non-technical audiences. Various authors have called for the entire concept of PRIs to be abandoned in favour of more tabular presentations of the underlying risk estimates for various different organisms or compartments (Dushoff, Caldwell, and Mohler 1994), although without necessarily a clear consideration for how these trade-offs would then be managed by stakeholders. The current authors take the position that having a single value or classification is a vital component of what makes PRIs functional as decision support tools, but acknowledge that there may be cases where having a broader framework of information is useful, particularly when assessing what and when to apply or for comparing change at the landscape level (Section 5). Beyond the small number of methods that explicitly attempt to characterise uncertainty (most notably I-Phy (van der Werf and Zimmer 1998)) there is very limited discussion of the role that uncertainty might play in implementation of PRIs, with many authors favouring the use of conservative values or worse case scenarios, in place of any explicit consideration of underlying uncertainty in measures or parameter estimates (Dubus and Surdyk 2006).

Conceptual strengths and weaknesses of different approaches

Perhaps the most controversial question in comparative analysis of PRIs is how well do the simplistic, but easily calculated, 'quantity only' indicators approximate the risks identified by more sophisticated models. The idea that with an appropriate representation, approximate risk could be estimated using data collected only from spray records is very appealing to policy makers, as it negates many of the difficulties around data acquisition and representation that limit some of the more sophisticated tools.

A recent study explicitly compared the predictive power of two quantity only indicators (QA and TFI) with the more complex Danish PLI, across a multi-year sample of Swiss farms (Möhring, Gaba, and Finger 2019). The results suggest these indicators, while having overall positive significant correlation with one another, have no statistical dependence with one another at the upper tails of the distribution, in other words; QA and TFI can be interpreted as having low power in the prediction of those places and times where the PLI indicates a high risk of non-target impact (Möhring, Gaba, and Finger 2019). A similar conclusion was also reached in a follow-on study that also considered the apparent impacts on agricultural productivity (Möhring, Bozzola, et al. 2020). Perhaps of most concern from a regulatory point of view, there is evidence that the apparent direction of trends in pesticide impact can be a function of indicator choice, with quantity only indicators diverging dramatically from those which have an inbuilt concept of toxicity (Kniss 2017; Möhring, Gaba, and Finger 2019). This presents a serious challenge to regulators and policy makers as it suggests that changing the indicators used may change the apparent direction of travel in National Action Plans. At present there is limited evidence to suggest that 'quantity only' indicators can approximate the behaviour of more complex and more scientifically justified indicators, and hence their persistence in policy decisions is more a reflection on data availability (particularly at the trans-national scale), procedural inertia, and concerns over the transparency of any competing approaches, as opposed to a consistent science and evidence-based case for use.

Criticism of weighted multi-component PRIs has tended to focus on two issues. The first is the subjective binning or point allocations associated with splitting underlying continuous measures into categories as previously outlined in our discussion of the EIQ (Dushoff, Caldwell, and Mohler 1994), see above). This is less of an issue in some of the more recent indicators, notably the PLI, which tend to be more consistent in how they represent differences between chemicals during calculation. What is however a more general feature of weighted multi-component approaches is the inherent subjectivity that arises from using arbitrary numerical weights to represent the relative 'importance' of different aspects of risk, as opposed to having an explicit estimate of concentration in a given compartment (as would be the case in a ETR based index) (Maud, Edwards-Jones, and Quin 2001). It should be noted that in a comparative analysis such as a time series, subjectivity in weighting is not inherently damaging so long as the calculation is treated consistently (i.e. you can still identify trends of increasing or decreasing risk), but it does introduce an inherent political and opaque aspect to the calculation, with the potential for conflict and challenge among different stakeholders. Guidance on how to establish such measures of relative 'importance' is conspicuously absent from discussions of index definition development (e.g. OECD (2004)), and represents a more general failure to engage with stakeholders in tool development (Rose et al. 2018).

There are other concerns over the scope of various PRIs, e.g. with Maud, Edwards-Jones, and Quin (2001) explicitly rejecting the idea that human health and environmental impacts should be considered within a single measure (on the basis that this prejudices the index when considering trade-offs), whereas many more recent authors would see the human health aspect as one of the central purposes of having an indicator for non-target risk (e.g. Kudsk, Jørgensen, and Ørum (2018)). It is important to recognise that neither of these positions is unjustified, but they do reflect that developing an indicator, particularly when used for policy, cannot be cleanly separated from the implicit value judgements and trade-offs implicit in agricultural policy. The key to further development therefore lies less in trying to develop some idealised multi-component PRI, which perfectly describes risk, and more in the conversations around ensuring transparency and trust in whatever measures best reflect local policy concerns (Feola, Rahn, and Binder 2011).

When looking at ETR derived methods there is broad consensus that these are the most 'realistic' and 'objective' approaches for the assessment of risk. Most benchmark studies, e.g. Feola, Rahn, and Binder (2011); Pierlot et al. (2017), have tended to favour exposure–toxicity ratio approaches, and the former study makes a strong case that the tested non-ETR approaches (including the EIQ and PestScreen) are inadequate to represent the implied risks associated with specific compounds under the tested ETR methods. However, as previously stated there are challenges of data availability, reliance on models with specific geographic scope and general user friendliness associated with many of these approaches and may limit their application outside of their original development context (Feola, Rahn, and Binder 2011). The increasing availability of freely accessible user-friendly software for indicators such as EPRIP (Trevisan, Di Guardo, and Balderacchi 2009), SYNOPS (Strassemeyer et al. 2017) and HAIR (Kruijne et al. 2014)) makes it much easier for indicators be adopted in practice at the farm or application level, but there are still many outstanding challenges with aggregation and the user experience of technical tools. The transition away from Excel spreadsheets and/or Access databases (Labite, Butler, and Cummins 2011) towards more online tools helps to support improved user experience, but also presents challenges around access, data usage and recording that need to be overcome. At present the best that can be said is that, within the EU, and particularly those countries that have well defined and active engagement in pesticide risk, the direction of travel is towards more integrated and centralised tools for calculating PRI, although at present such efforts are highly restricted by a lack of standardised frameworks and national priorities (see Section 5 and discussion in Nicholson et al. (2020)).

Validation: How well do PRIs perform in field conditions?

As noted above when we talk about the practical performance of PRIs, we are usually addressing the ability of tested indicators to predict the observed concentration (or some metric thereof) in a specific compartment (typically surface water) based on known applications. Such an approach naturally favours ETR like indicators, which are explicitly designed to predict such exposure, but can still be informative for the performance of simpler approaches. A typical validation study such as that of Oliver et al. (2016), compares field measures from a small number of sites, in this case two Australian orchards over a two year period, and examines the correlation of a small number of risk indices, in this case PIRI and EPRIP, based on spray records. In this particular instance both indicators corresponded loosely with each other and with the empirical measurements of environmental PPP occurrence, with the authors ascribing discrepancies to a failure to appropriately account for buffer width (a common source of uncertainty in many ETR approaches due to high variation between methods, application and compounds), as well as questions over the generality of the tested indicators (Oliver et al. 2016).

By far the most comprehensive validation study for PRIs is one we have cited multiple times in this review that of Pierlot et al. (2017). These authors took run-off measurements from three localities in France and compared the ability of the 26 examined indicators to predict a) frequency of exceeding a 0.1µg/L water quality threshold, b) maximum concentration over the monitoring period c) maximum flux of active ingredient, d) cumulative flux over the measurement period and e) weighted average concentration. Measurements were taken at the field scale and represented a range of soil types and application periods with 20 different compounds being applied as appropriate to cropping, which included maize, spring peas, winter peas, broad beans and winter wheat. A summary of their overall findings is shown in Figure 1. Pierlot et al. (2017) also note that the performance of many PRIs including some fairly sophisticated approaches is relatively poor is absolute terms, which they assign to a combination of necessary simplifications imposed by the experimental setup, such as not including daily climatic data and water status of soil, as well as the fact that most indicators (by-design) do not include any effect of random variation in the underlying measures. It should also be noted that the authors themselves outline the restrictive geographic scope of their analysis and emphasise that wider and more inclusive studies would be a valuable addition to the development of PRIs (Pierlot et al. 2017).

5. Pesticide Risk Indicators in practice

The principle context around the practical use of PRIs (outside of an academic setting) can be broken down into five major areas. In this section we briefly review the roles played by PRIs in each of these cases with some practical examples from across Europe.

Tools for decision making by farmers and advisors

The first area in which PRIs have seen practical application is as tools for decision making by farmers or their advisors regarding which compounds to apply and when. The increasing development of software tools, and especially online platforms, are allowing elements of risk to be incorporated into management decisions, such as, when or what products to apply, although in many cases these are restricted to specific geographic scope due to the need for representative underlying datasets (Nicholson et al. 2020).

This style of tool and its impacts on pesticide management is the subject on the ongoing EU FAIRWAY project, the report of which (Nicholson et al. 2020) identifies over 150 decision support tools (DST) associated with reductions in nitrate and pesticide inputs. Of the nine pesticide management decision support tools examined in detail by these authors several have direct ties to the PRIs discussed above, most notably the Dutch EYP (Reus and Leendertse 2000). Other farm level tools for pesticide decisions include FARMSCOPER (developed in the UK (Gooday et al. 2014; Price et al. 2011)), and DST Plant Protection Online (developed in Denmark[11]), with a number of others designed to be applicable at the catchment/regional level. Uptake of decision support tools has been patchy (Rose et al. 2016), with users of DST Plant Protection Online in Denmark highlighting several issues, including tools being too time consuming and complex for practical use (see also Rose et al. (2018)), as well as citing competition

[11] https://plantevaernonline.dlbr.dk/cp/menu/Menu.asp?id=djf&subjectid=1&language=en

from consultants, and lack of confidence in underlying data or algorithms (Nicholson et al. 2020; Möhring, Wuepper, et al. 2020).

Language barriers have been identified as significant restrictions on the adoption of tools outside of their original county of development, as are the integration of tools into a specific countries' legal framework (which can make it difficult for tools to transcend national boundaries) (Nicholson et al. 2020). In a similar way to PRIs in general, this has led to a proliferation of tools, and considerable redundancy, as researchers and policy makers have developed unique solutions to their specific circumstance rather than adapting general-purpose frameworks. Projects like FAIRWAY are a useful first step towards the development of more harmonised and general use approaches but also highlight the outstanding issues around developing an intuitive user experience, developing trust (both in the tools themselves and the correctness of the underlying data and models), and maintaining compliance to the ever shifting legislative and policy environment surrounding PPP use (Rose and Bruce 2018).

Tools for surveillance and monitoring

The second context for PRIs is tied to monitoring and the assessment of change within the landscape. This is probably the area of greatest concern to policy makers and is the principle function that the described PRIs play within National Action Plans or other related monitoring activity (Barzman and Dachbrodt-Saaydeh 2011). As noted above, several of the indicators described are integrated to a greater or lesser degree with National Action Plans and serve as key indicators of change within the landscape (Barzman and Dachbrodt-Saaydeh 2011). Historically, 'quantity only' indicators dominated surveillance activity such as volume consumption in the Netherlands (up until 2010, when it was replaced with the Environmental Yardstick for Pesticides), the TFI in Denmark (up until 2013 when it was replaced by the Pesticide Load Indicator) and the NUD in France (often used in conjunction with TFI) (Barzman and Dachbrodt-Saaydeh 2011; Sud 2020). Among the more complex indicators, SYNOPS has been the basis for reduction targets in Germany since 2009, based on a representative sample of farms taken from across the various regions, and Belgium makes use of an indicator called the *Pesticide Risk Indicator for Belgium;* which is derived from POCER, to help set national impact reduction targets (Vergucht and Steurbaut 2007; Van Bol and Pussemier 2005; Barzman and Dachbrodt-Saaydeh 2011). Measured using SYNOPS reductions in in impact in Germany in 2006 to 2008 (relative to the mean of the reference period of 1996 to 2005) show a mixed pattern across different classes of PPP, with fungicides generally lagging behind targeted reductions compared to those of insecticides and herbicides (despite relatively stable trends in total PPP sales across all three groups, see[12]) (Strassemeyer and Gutsche 2010). Perhaps the most significant example of PRI in surveillance occurs in Denmark where, due to a long standing and comprehensive system of registration of pesticide usage, it is possible to visualise, at fine geographic scale the estimated impact on the landscape based on the Danish Pesticide Load indicator (Kudsk, Jørgensen, and Ørum 2018; Pedersen and Nielsen 2017), which is now integrated both into taxation and reduction targets (Pendersen, Helle, and Andersen 2015). By contrast, the UK National Action Plan (Defra 2013), in common with the majority or EU countries, has not included clear quantitative end points, but used a more distributed approach discussed in Section 6 (Barzman and Dachbrodt-Saaydeh 2011).

At the broader EU scale crude 'quantity only' sales-based data tends to underpin policy developments due to the limited data requirements and the need to compare across many regimes with very different approaches to PPP surveillance. At the heart of EU monitoring of PPP use are a series of data challenges historically associated with the inconsistent way in which regulations have been handled/interpreted across Member States. The result has been a great deal of confusion as to precisely which sets of crops, statistical units and active substances are compiled across different countries, which has compromised efforts for consistent surveillance of change (European Court of Auditors 2020). This is made more problematic by the fact that under current regulations data is only required to be reported for a

[12] http://www.ceureg.com/18/docs/presentations/19_Joern%20Strassemeyer_Germany.pdf

reference period of maximum 12 months at any time within a five year period[13] (European Court of Auditors 2020), making it very difficult to assess trends in a rapidly changing market or to understand the impacts of changing authorisation regulations and product withdrawal. While Member States have been obliged to monitor a specified set of 36 active substances with respect to surface water under the environmental quality standards since 2008[14], data on agricultural PPP use have only been recorded since 2015 and is subject to problems of statistical aggregation, which have prevented the development of harmonised values for the total amount of PPP application (European Commission 2019). The 2019 Commission report (European Commission 2019) highlights a number of areas requiring improvement in the generation of regular and consistent transnational statistics on PPP usage. The commission report also notes that due to the extensive time lags associated with the data collection it will take some time for such harmonisation to have a measurable impact and it is likely to be several years before consistent and coherent transnational records of pesticide usage become routine available for policy development (European Commission 2019).

Neither of the EU's current harmonised risk indicators are considered to be strongly indicative of risk to non-target systems, with HR1 being particularly criticised for its crude aggregation, failure to account for the context of PPP usage, and a lack of scientific rationale for the chosen weightings (European Court of Auditors 2020). A recent review of data practice (European Commission 2020) outlines the upcoming strategy around collation of Statistics on Agricultural Input/Output (SAIO) which indicates that for pesticides, national statistics of total usage by crop are expected to be reported every fifth year to supplement annual sales records. Note that this contrasts with the UK where usage data is reported biennially for major crop groups on based on the Pesticide Usage Survey, e.g. Garthwaite et al. 2019; see Section 6 for discussion. It remains to be seen if an adequate harmonised approach can be established at an EU level and what future role such statistics can play in the development of more sophisticated PPP surveillance at the transnational scales and in associated policy devlopement.

It should be noted that while some PRI are treated as monitoring and surveillance tools in their own right (Kudsk, Jørgensen, and Ørum 2018), integration with other surveillance of relevant endpoints is often lacking (Mancini, Woodcock, and Isaac 2019). For example, while limited surveillance of water bodies is routine in many Member States, there are relatively few examples outside of dedicated validation studies (see Section 4) that have attempted to link these to local spray records to help understand how well predictive tools represent exposure risk in the field. Likewise, only a small number of studies have explicitly examined the linkage between changes in monitored populations of organisms of concern (such as the standard reference organism used in laboratory studies) and pesticide risk indicators, largely due to a lack of data availability. In one rare example from the UK, Woodcock et al. (2016) used a PRI called the foliar insecticide impact index (derived from part of the EIQ and subject to the same issues of binned high, medium and low toxicity scores) to predict population change in bees over the English landscape based on a well-established voluntary national survey between 1994 and 2011. Their findings, which correlated neonicotinoid exposure with regional declines in species known to habitually forage on treated crops, provides a model for how PRI measures might be integrated into wider ecological studies on changes in wildlife populations (see discussion in Mancini, Woodcock, and Isaac (2019)). On the human health side the work of Larsen, Gaines, and Deschênes (2017) and others provides a well-defined framework for analysis of concerns, all be it one which has largly been conducted using 'quantity only' measures of impact. As surveillance becomes increasingly sophisticated and improved data resources become available we expect the integration of PRIs into studies of non-target impacts will continue to grow, although at present the

[13] Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics.

[14] Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 as amended by Directive 2013/39/EU of the European parliament and of the Council of 12 August 2013

lack of coherent and accessible data resources make this challenging outside of a small number of intensive surveillance regimes.

Tools for administration of policy instruments

As well as being tools for monitoring change some PRIs are also linked to the implementation of practical policy instruments (Lee, den Uyl, and Runhaar 2019). Perhaps the most notable examples of such instruments are the various approaches to pesticide taxation used in different European countries many of which have direct ties to specific PRIs (Böcker and Finger 2016; Finger et al. 2017). Norway, Denmark and France have all at various times used differentiated pesticide taxation schemes linked to the estimation of 'risk' for specific PPPs (Böcker and Finger 2016; Finger et al. 2017). These include relatively simple systems such as Norwegian Banded Pesticide Tax Scheme (Norwegian Food Safety Authority 2005), as well as much more complex instruments such as the Danish Pesticide load indicator (Böcker and Finger 2016; Pedersen and Nielsen 2017). The success of these instruments in achieving impact reduction objectives has been mixed, with the very low taxation rates implemented in France being associated with a failure to achieve objectives and a near complete revision of the scheme in 2008, some substitution for lower risk compounds reported in Norway (Bragadóttir et al. 2014), and substantial (40% between 2013-2015) reductions in non-target impact based on the PLI reported in Denmark, although without corresponding (and targeted) reduction in the amount of PPP use as expressed using the TFI (Sud 2020). The consistent failure of Danish pesticide action plans to achieve their use reduction objectives has been attributed to failures of the underlying economic model to accurately reflect stakeholder incentives, particularly around perceptions of risk and motivations for behaviour, with wider implications for the deployment of policy instruments (Pedersen and Nielsen 2017)

France has had a particularly complex history with pesticide policy instruments reviewed in detail in Sud (2020). Pesticide taxation was originally introduced to France in 2000 as part of the *taxe générale sur les activités polluantes*, which represented potential non-target impacts under seven risk categories based on the associated 'Risk phrases[15]' (similar to the Norwegian system outlined above). This was replaced in 2008 with a simplified three category system as part of the Ecophyto national action plan, notable for its overall lack of success in achieving targeted PPP reductions (based on the number of unit doses index, the number of treatments per hectare increased by 29% between 2008 and 2014 (Sud 2020)). Ecophyto Plan II, introduced on 2015, revised the national targets and introduced so called 'pesticide saving certificates' (CEPP), which are aimed at pesticide distributors, who must encourage farmers to adopt practices associated with lower pesticide use, with associated penalties if reductions are not demonstrated over a five year period (OECD 2017). As with other elements of French surveillance the key indicator for these reductions is the Number of Unit doses, a quantity only indicator criticised during stakeholder consultation for its failure to account for toxicity (OECD 2017; Potier 2014; Huyghe and Blanck 2017). The Ecophyto Plan II text mentions replacement of the NUD with an unspecified set of more finely resolved indicators, but it is unclear at the time of writing how these would be structured in practice (Sud 2020). Both Ecophyto schemes are also worth noting for their emphasis on training and technical guidance provided to farmers, including the DEPHY network of 1,900 demonstration farms (planned to be expanded to 3,000 in Ecophyto Plan II; Ministère de l'Agriculture et de l'Alimentation (2015)) , which has been highly successful in development and communication of approaches towards the reduction of PPP use while maintaining productivity and are considered one of the most successful elements of the framework (Sud 2020).

Tools for supporting the approval of PPP

In the UK and EU, the first tier of the approval of a novel pesticide product requires that the predicted environmental exposure concentration for a given compound is lower than a maximum concentration considered safe for non-target organisms (Schäfer et al. 2019). Current standards of pesticide risk assessment are tied closely to the ETR framework, with many focusing on the comparisons between

[15] As defined in Annex III of European Union Directive 67/548/EEC

so called toxicity exposure ratios (closely related to the estimated compartment exposure in the ETR framework) relative to trigger values based on LC_{50} and/or NOEC values of relevant organisms (Silva et al. 2019; Boivin and Poulsen 2017). Given this close conceptual similarity it is not surprising that many approaches for pesticide risk assessment have been adapted for PRIs under field conditions (Labite, Butler, and Cummins 2011). One of the key questions around authorisation and risk assessment is the triggering of higher tier investigation of a compound referring to testing of a prospective PPPs within increasingly intricate and 'realistic' experimental settings (Schäfer et al. 2019). Traditionally PRIs, known in context as 'Plant Protection Product Ranking Tools' played a key role in determined when higher tier assessment was required, a topic reviewed in detail in Labite, Butler, and Cummins (2011). As authorisation increasingly shifts towards a being more focused on longer term monitoring and field conditions (Schäfer et al. 2019), it is expected that the role played by PRIs will increase, particularly in the assessment of product substitution (Steingrímsdóttir, Petersen, and Fantke 2018) and the impact of mixtures (Bopp et al. 2019). At present it is not clear which specific indicators will be adapted for this expanded role and how these will filter into monitoring and decision support. Nevertheless, it is important to acknowledge the authorisation process as a key context for development of PRIs and one of the principal research areas where the next generation of indicators are likely to originate.

Environmental life cycle assessment

Life cycle assessment (LCA) is an increasingly important tool in economic analysis focused around reporting potential environmental loads and resources consumed in each step of a product or service supply chain (Notarnicola et al. 2017). Its application to PPPs specifically is tied to the need to characterises the effects on compounds applied deliberately to parts of the biosphere usually in the context of wider understanding of agricultural systems (Margni et al. 2002). The basic model for Pesticide Life Cycle Analysis resembles closely the ETR approach to PRIs, although there often isn't the same degree of mathematical formalism in the structure of the resulting indicator. In general, such assessments include a combination of fate modelling, related to the exposure within a given compartment, and impact modelling which relates to the potential effects of a given exposure of relevant organism, and in practice are closely tied to the toxicity concept in ETR methods (Apostol et al. 2009). The earliest LCA methods used broad stroke assumptions to characterise fate (e.g. Margni et al. 2002). However more recently approaches have become increasingly sophisticated and will now often include multiple pathways and underlying parameters (e.g. Birkved and Hauschild 2006) which may in turn be derived from mechanistic models (e.g. Pest Tox; Felix, Holst, and Sharp (2019), PestLCI 2.0; Dijkman, Birkved, and Hauschild (2012)). The most influential approach linked to LCA is the so called USEtox model, developed and promoted by the United Nations Environment Program and the Society for Environmental Toxicology and Chemistry (Rosenbaum et al. 2008; Berthoud et al. 2011; Henderson et al. 2011). This models the ecotoxicity impacts of PPPs through the simulation of the release of chemicals based on emissions from or into six main urban and continental environmental compartments. In many ways the role played by LCA is overlapping with that of other PRIs although its focus tends to be on the larger economic picture, rather than the details of decision support or in the specification of monitoring or policy targets (Saouter et al. 2019). The only one of the listed indicators developed with a LCA focus is PestScreen, which draws heavily on the LCA approaches in its conception of fate (Margni et al. 2002).

6. Pesticide Risk Indicators in a UK context

Data availability

To discuss the role of PRIs in current and future UK PPP policy and decision making we must return to the theme of the availability of data to calculate different indices. In terms of recording the usage of PPP, the UK sits in a intermediate phase between the robust and near universal recording systems used in Denmark and California (Eurostat 2008), and the extremely aggregated and sparse records of sales records available for many other EU Member States. UK usage of PPPs is reported via extrapolation from a stratified sample of holdings visited during the biennial (for arable crops) Pesticide Usage Survey administered by FERA science limited on behalf of UK government, and in collaboration with the Science & Advice for Scottish Agriculture (SASA) and the Department of Agriculture and Rural Development, Northern Ireland (DARD).

The calculation of usage via the PUS begins with the DEFRA 'June' agricultural survey which attempts to characterise the number and cropping behaviour of UK holdings, and grouped based on overall holding size[16]. These groups are then combined with regional information to establish a stratified sample of holdings, which are then the target for visits to collect maintained spray records. The average rates of application calculated from these records are then rescaled to the overall population to create the reported statistics on use (Thomas 1999). Almost all discussion of observed trends in pesticide impacts within the UK, e.g. Cross and Edwards-Jones (2006, 2011) derives ultimately from the PUS and associated national statistics. One of the challenges implicit in using a sample-based approach to surveillance is the introduction of structural uncertainty associated with how well the sample represents the overall population. The most recent PUS reports, e.g. Garthwaite et al. (2019) have incorporated confidence intervals on their estimates of overall usage, although these have not always been appropriately incorporated into downstream estimates of potential impact e.g. The Pesticides Forum (2020). This intrinsic uncertainty, combined with the observation that rules regarding the confidentially for provided spray records, dramatically restricts the spatial resolution of available outputs (Mancini, Woodcock, and Isaac 2019) and has led to some calls for a more universal system of recording (Thomas 1999; Eurostat 2008). An important supplement to the PUS is provided by long term monitoring of sites such as the Game Conservancy Trust's Sussex Study on a site in Sussex Downs, which since 1970 has provided data on PPP use and change in the population of various indicator species and is used in the development of statistics under the National Action Plan (Ewald et al. 2016; Defra 2013).

In terms of supporting data, information on toxicity and other laboratory properties of active substances relevant to the UK are maintained in the Pesticide Properties Data Base (Lewis et al. 2016), the same resource which supports the Danish PLI and other international systems. What is lacking from current UK surveillance is for example the sophisticated GIS and scenario based tools underpinning for example the German SYNOPS indicator (Strassemeyer et al. 2017), which provides links to soil and elevation maps, as well as providing estimates for the minimal distance from the field to the edge of the surface water via high resolution land cover data. Given the aggregation level generated by the PUS it is unclear the extent to which such information would be useful under the current UK surveillance regime, but it is worthy of consideration in the context of decision support tools targeting farmers and advisors. Land cover maps for the UK are maintained by (among others), the Centre for Ecology and Hydrology[17] and the national soil map is published by Cranfield University[18]. These and other similar resources could, in principle, be integrated into a SYNOPS like tool suitable for the UK, although the authors are currently not aware of any such efforts at the time of writing.

Also relevant for the wider context of PPP usage in the UK is the government's 25-year environment plan, published in 2018 (HM Government 2018) includes an indicator (H4) which is designed to measure 'Exposure and adverse effects of chemicals on wildlife in the environment'(Defra 2019). At the time of writing construction of datasets to support H4 are largely driven by integration of existing surveillance and 'data are currently available for some chemicals and some invertebrates, fish, shellfish, crustaceans, mammals, and birds of prey' (Defra 2019; p 110). There is however known to be a wider research effort around H4 which could in future come to encompass work around PRI, and/or consideration of how PRI derived from the PUS could be integrated into wider national monitoring of PPP impacts. At the time of writing there are no explicit policy targets associated with H4, although

[16]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/182206/defra-stats-foodfarm-landuselivestock-june-junemethodology-20120126.pdf
[17] https://www.ceh.ac.uk/services/land-cover-map-2015
[18] http://www.landis.org.uk/data/natmap.cfm

the related H3 (focused on impacts of persistent organic pollutants and specifically mercury) aims to achieve a 50% reduction in land-based emissions of mercury by 2030 (Defra 2019).

Index selection

Of the indicators listed above three have been used in the context of UK agriculture. The EIQ has been used for long term trend analysis based on the PUS in arable (Cross and Edwards-Jones 2006b; 2011) vegetable (Cross and Edwards-Jones 2006a) and orchard crops (Cross 2013), with evidence suggesting modest declines in impact under this measure, varying by crop type. The p-EMA index was originally developed in the UK but has limited practical application and has no associated time series analysis. Most recently, a team including the lead author have adapted the Danish Pesticide Load indicator to a UK context, incorporating a novel procedure for dealing with structural uncertainty in the indicator (Rainford et al. In preparation). This underlying approach, which is based on using bootstrap resampling of the holdings visited by the PUS in order to estimate confidence intervals on the amount of PPP applied, could in theory be adapted for use with other PRIs. What is unclear at present is the scope for adapting indices which incorporate more localised information, such as SYNOPS or POCER, to the sampled holdings from the PUS. Such indicators, when used as national targets, tend to utilise a 'worst case scenario' approach to even out the underlying variation between holdings, timings of application etc. Thus, the lack of high-resolution information of e.g. the distance to nearby water bodies, is less of an impediment than might be first thought for adaption to the UK. There would however be a need for carefully managed stakeholder engagement in how such scenarios are constructed given the traditional scepticism of UK stakeholders around indicators in general, and particularly uncertainty arising from known sampling issues in the PUS.

Political context and targets

The focus of political change in the area of PPPs in the UK is a) the development of an updated National Action Plan, scheduled for consultation in 2020, and b) ongoing change in the regulatory process associated with exit from the EU. At the time of writing there appears to be no discussion of a novel national PRI, although various NGO's have suggested the possibility of adopting the TFI or NUD indices used in France and (formerly, prior to the adoption of the PLI) in Denmark(Pesticide Action Network UK 2018), although there is little evidence that these are consistently superior to the measures of overall mass currently in place (Möhring, Gaba, and Finger 2019; Lamichhane et al. 2015). At present it appears unlikely that the revised NAP will include specific reduction targets either in terms of 'quantity only' use or an estimate of potential impact[19]. Defra is however exploring some of these indicators, e.g. (Rainford et al. In preparation), which may indicate a greater willingness to consider such instruments in future. The position of the UK government has traditionally been to reject an interpretation of the PPP Regulation which mandates a specific national reduction target, and there is no clear indication that this will be revised following the UK's exit from the EU regulatory regime.

7. Discussion and Conclusions

PRIs are diverse metrics that have been adapted for a wide range of different policy and decision context. The goal of this review was to provide an overview of some of the most significant approaches to the approximation of risk to non-target systems and to review the context and development of PRIs in Europe and the UK. Broadly the major advances in PRI development over the last decade have tended to focus on refinement and application of methodologies most suitable for use as decision support tools for individual farmers and their advisors, while much of the policy focus has been at larger scales and the issues of how to harmonise across widely divergent national monitoring regimes. This disconnect between the requirements of PRIs at different scales has led to a disjointed discussion of the role played by PRIs, which makes general conclusions difficult to draw. On the one hand, the adoption by the EU of a common harmonised risk indicator for PPP use is clearly a major advance in terms of comparative analysis within the Union, however the statistical flaws in the calculation, as well

[19] https://publications.parliament.uk/pa/cm201719/cmselect/cmeuleg/301-ix/30112.htm

as the fundamentally 'quantity only' nature of HR1 places it well behind academic discussion of PRIs, which is increasingly focused on the development of semi-automated ETR based approaches (e.g. SYNOPS Web (Strassemeyer et al. 2017)).

Of the national PRIs in use across Europe, perhaps the most interesting from a UK perspective is the Danish Pesticide Load indicator (Kudsk, Jørgensen, and Ørum 2018), in part because its calculation is relatively independent of scale and easily adapted to a specific national context. In contrast with indicators like SYNOPS or POCER, the location specific information requirements around the PLI are minimal and to a large extent already collated for the UK in resources like the Pesticide Properties database. It is however worth noting that the PLI was not included in the set of indicators reviewed in the validation study of Pierlot et al. (2017), and that other weighted multi-component approaches such as EIQ were identified as relatively weak performers under field conditions (Figure 1). Of the indicators discussed here the PLI is probably the one most easily translated to a UK context, without major revisions to existing infrastructure, although recent studies have shown that with expanded effort indicators like SYNOPS can be adapted outside of their original development context (de Baan 2020). It should however be noted that the widespread acceptance of such an indicator is by no means assured, particularly given the structural uncertainty intrinsic in the PUS, and there is considerable work that would be required to integrate the ETR PRIs and alternatives such as the TFI and NUD into the existing framework of indicators used in the UK NAP.

The key question in the adoption of a novel PRI is understanding who the stakeholders are and what their information requirements are for making relevant decisions (Schäfer et al. 2019). Recently there has been a lot of discussion about the challenges of monitoring the post-authorisation impacts of PPPs (Milner and Boyd 2017), particularly in the context of the recent withdrawal of neonicotinoid substances over concerns regarding bee health (Mancini, Woodcock, and Isaac 2019). Such changes would necessitate a widescale revision to existing monitoring networks and potentially to the role of PRIs as decision support tools. To date the authors are unaware of any national scale post authorisation monitoring effort in use or planning within EU Member States, which explicitly attempts to link environmental observations of PPPs to PRI proxies estimated from spray records. The scope for such an undertaking remains largly limited by data availability even within the sophisticated monitoring networks of California (see e.g. Epstein and Bassein 2003; Moran et al. 2020) and Denmark[20], but may represent the ultimate direction of travel as authorisation and monitoring systems become increasingly sophisticated at the landscape scale (Schäfer et al. 2019).

Pesticide risk indicators are a valuable tool, applicable at multiple operational levels, and with potential to inform a wide range of decisions around the use of PPPs. The key challenge however lies in moving beyond specific indicators for a specific decisions and towards a more generalised and comparable approach which will help different stakeholders achieve a consistency in how they relate risk to non-target systems, as well as building trust in the developed methods. As modern farming becomes increasingly sophisticated and as the scope for automated and standardised big data continues to grow, many PRIs may find a new role in the next generation of decision support tools, national monitoring networks and/or any revisions to the authorisation process (Kamilaris, Kartakoullis, and Prenafeta-Boldú 2017; Streissl, Egsmose, and Tarazona 2018; McGrath et al. 2019). There is thus a clear need for transparency and clear communication in what indices are adopted and a need for further exploration into future requirements from proxies of risk around PPPs and how these perform in practice.

[20] https://pesticidbelastning.dk/#/

8. References

Apostol, Laura, Raluca Hlihor, Camelia Betianu, Lucian Vasile Pavel, Brindusa Mihaela Sluser, Florentina Anca Caliman, and Maria Gavrilescu. 2009. ‘Life Cycle Impact Assessment of Pesticides: Current Issues and Perspectives’. *Bulletin of the Polytechnic Institute of Iasi; Section Chemistry and Chemical Engineering*, 67–83.

Baan, Laura de. 2020. ‘Sensitivity Analysis of the Aquatic Pesticide Fate Models in SYNOPS and Their Parametrization for Switzerland’. *Science of The Total Environment* 715 (May): 136881. https://doi.org/10.1016/j.scitotenv.2020.136881.

Barzman, Marco, and Silke Dachbrodt-Saaydeh. 2011. ‘Comparative Analysis of Pesticide Action Plans in Five European Countries’. *Pest Management Science* 67 (12): 1481–85. https://doi.org/10.1002/ps.2283.

Belden, Jason B., Robert J. Gilliom, and Michael J. Lydy. 2007. ‘How Well Can We Predict the Toxicity of Pesticide Mixtures to Aquatic Life?’ *Integrated Environmental Assessment and Management* 3 (3): 364–72. https://doi.org/10.1002/ieam.5630030307.

Berg, Femke van den, and J. J. T. I. Boesten. 1998. ‘Pesticide Leaching and Accumulation Model (PESTLA) Version 3.4; Description and User’s Guide’. https://research.wur.nl/en/publications/pesticide-leaching-and-accumulation-model-pestla-version-34-descr.

Berthoud, Amandine, Pauline Maupu, Camille Huet, and Antoine Poupart. 2011. ‘Assessing Freshwater Ecotoxicity of Agricultural Products in Life Cycle Assessment (LCA): A Case Study of Wheat Using French Agricultural Practices Databases and USEtox Model’. *The International Journal of Life Cycle Assessment* 16 (8): 841. https://doi.org/10.1007/s11367-011-0321-7.

Böcker, Thomas, and Robert Finger. 2016. ‘European Pesticide Tax Schemes in Comparison: An Analysis of Experiences and Developments’. *Sustainability* 8 (4): 378. https://doi.org/10.3390/su8040378.

Bockstaller, Christian, Laurence Guichard, David Makowski, Anne Aveline, Philippe Girardin, and Sylvain Plantureux. 2008. ‘Agri-Environmental Indicators to Assess Cropping and Farming Systems. A Review’. *Agronomy for Sustainable Development* 28 (1): 139–49. https://doi.org/10.1051/agro:2007052.

Boesten, J. J. T. I., and B. Gottesbüren. 2000. ‘Testing PESTLA Using Two Modellers for Bentazone and Ethoprophos in a Sandy Soil’. *Agricultural Water Management* 44 (1): 283–305. https://doi.org/10.1016/S0378-3774(99)00096-7.

Boivin, Arnaud, and Véronique Poulsen. 2017. ‘Environmental Risk Assessment of Pesticides: State of the Art and Prospective Improvement from Science’. *Environmental Science and Pollution Research* 24 (8): 6889–94. https://doi.org/10.1007/s11356-016-8289-2.

Bopp, Stephanie K., Aude Kienzler, Andrea-Nicole Richarz, Sander C. van der Linden, Alicia Paini, Nikolaos Parissis, and Andrew P. Worth. 2019. ‘Regulatory Assessment and Risk Management of Chemical Mixtures: Challenges and Ways Forward’. *Critical Reviews in Toxicology* 49 (2): 174–89. https://doi.org/10.1080/10408444.2019.1579169.

Bourguet, Denis, and Thomas Guillemaud. 2016. ‘The Hidden and External Costs of Pesticide Use’. In *Sustainable Agriculture Reviews: Volume 19*, edited by Eric Lichtfouse, 35–120. Sustainable Agriculture Reviews. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-26777-7_2.

Bragadóttir, Hrafnhildur, Carl von Utfall Danielsson, Roland Magnusson, Sampo Seppänen, Amanda Stefansdotter, and David Sundén. 2014. *The Use of Economic Instruments: In Nordic Environmental Policy 2010-2013*. Nordic Council of Ministers.

Brown, Colin D., Andy Hart, Kathleen A. Lewis, and Igor G. Dubus. 2003. ‘P-EMA (I): Simulating the Environmental Fate of Pesticides for a Farm-Level Risk Assessment System’. *Agronomie* 23 (1): 67–74. https://doi.org/10.1051/agro:2002074.

Bürger, Jana, Friederike de Mol, and Bärbel Gerowitt. 2008. 'The "Necessary Extent" of Pesticide Use—Thoughts about a Key Term in German Pesticide Policy'. *Crop Protection* 27 (3): 343–51. https://doi.org/10.1016/j.cropro.2007.06.006.

Carpentier, A., and X. Reboud. 2018. 'Why Farmers Consider Pesticides the Ultimate in Crop Protection: Economic and Behavioural Insights'. *AgEcon Search*. https://doi.org/10.22004/ag.econ.277528.

Carson, R. 1964. *Silent Spring*. Boston: Houghton Mifflin.

Chen, Chen, Yanhua Wang, Yongzhong Qian, Xueping Zhao, and Qiang Wang. 2015. 'The Synergistic Toxicity of the Multiple Chemical Mixtures: Implications for Risk Assessment in the Terrestrial Environment'. *Environment International* 77 (April): 95–105. https://doi.org/10.1016/j.envint.2015.01.014.

Claeys, Sara, Bénédicte Vagenende, Bart De Smet, Liesbeth Lelieur, and Walter Steurbaut. 2005. 'The POCER Indicator: A Decision Tool for Non-Agricultural Pesticide Use'. *Pest Management Science* 61 (8): 779–86. https://doi.org/10.1002/ps.1062.

Claydon, Sam. 2020. '2020: A Massive Year for UK Pesticides'. *Pesticide Action Network UK* (blog). 31 January 2020. https://www.pan-uk.org/2020-a-massive-year-for-uk-pesticides/.

Coll, Moshe, and Eric Wajnberg. 2017. *Environmental Pest Management: Challenges for Agronomists, Ecologists, Economists and Policymakers*. John Wiley & Sons.

Cooper, Jerry, and Hans Dobson. 2007. 'The Benefits of Pesticides to Mankind and the Environment'. *Crop Protection* 26 (9): 1337–48. https://doi.org/10.1016/j.cropro.2007.03.022.

Cox, Louis Anthony (Tony), Djangir Babayev, and William Huber. 2005. 'Some Limitations of Qualitative Risk Rating Systems'. *Risk Analysis* 25 (3): 651–62. https://doi.org/10.1111/j.1539-6924.2005.00615.x.

Cross, Paul. 2013. 'Pesticide Hazard Trends in Orchard Fruit Production in Great Britain from 1992 to 2008: A Time-Series Analysis'. *Pest Management Science* 69 (6): 768–74. https://doi.org/10.1002/ps.3436.

Cross, Paul, and Gareth Edwards-Jones. 2006a. 'Variation in Pesticide Hazard from Vegetable Production in Great Britain from 1991 to 2003'. *Pest Management Science* 62 (11): 1058–64. https://doi.org/10.1002/ps.1272.

———. 2006b. 'Variation in Pesticide Hazard from Arable Crop Production in Great Britain from 1992 to 2002: Pesticide Risk Indices and Policy Analysis'. *Crop Protection* 25 (10): 1101–8. https://doi.org/10.1016/j.cropro.2006.02.013.

———. 2011. 'Variation in Pesticide Hazard from Arable Crop Production in Great Britain from 1992 to 2008: An Extended Time-Series Analysis'. *Crop Protection* 30 (12): 1579–85. https://doi.org/10.1016/j.cropro.2011.08.003.

De Jong, Frank M. W, and Geert R De Snoo. 2002. 'A Comparison of the Environmental Impact of Pesticide Use in Integrated and Conventional Potato Cultivation in The Netherlands'. *Agriculture, Ecosystems & Environment* 91 (1): 5–13. https://doi.org/10.1016/S0167-8809(01)00262-6.

Defra. 2013. 'UK National Action Plan for the Sustainable Use of Pesticides (Plant Protection Products)'. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/221034/pb13894-nap-pesticides-20130226.pdf.

———. 2019. 'Measuring Environmental Change: Outcome Indicator Framework for the 25 Year Environment Plan'. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/802094/25-yep-indicators-2019.pdf.

Dereumeaux, Clémentine, Clémence Fillol, Philippe Quenel, and Sébastien Denys. 2020. 'Pesticide Exposures for Residents Living Close to Agricultural Lands: A Review'. *Environment International* 134 (January): 105210. https://doi.org/10.1016/j.envint.2019.105210.

Dijkman, T. J., M. Birkved, and M. Z. Hauschild. 2012. 'PestLCI 2.0: A Second Generation Model for Estimating Emissions of Pesticides from Arable Land in LCA'. *The International Journal of Life Cycle Assessment* 17 (8): 973–86. https://doi.org/10.1007/s11367-012-0439-2.

Dubus, Igor G., and Colin D. Brown. 2002. 'Sensitivity and First-Step Uncertainty Analyses for the Preferential Flow Model MACRO'. *Journal of Environmental Quality* 31 (1): 227–40. https://doi.org/10.2134/jeq2002.2270.

Dubus, Igor G., Colin D. Brown, and Sabine Beulke. 2003. 'Sensitivity Analyses for Four Pesticide Leaching Models'. *Pest Management Science* 59 (9): 962–82. https://doi.org/10.1002/ps.723.

Dubus, Igor G., and Nicolas Surdyk. 2006. 'Deliverable DL4; State of the Art on Pesticide Fate Models and Environmental Indicators'. DL#4 of the FP6. FOOTPRINT project. http://julienmoeys.info/assets/pdf/FOOTPRINT/FOOTPRINT_DL04.pdf.

Dushoff, Jonathan, Brian Caldwell, and Charles L. Mohler. 1994. 'Evaluating the Environmental Effect of Pesticides: A Critique of the Environmental Impact Quotient'. *American Entomologist* 40 (3): 180–84. https://doi.org/10.1093/ae/40.3.180.

ECOTEC, CESAM, CLM, University of Gothenburg, UCD, and IEEP. 2001. 'Study on Environmental Taxes and Charges in the EU. Chapter 8: Pesticide Taxes and Charges'. https://ec.europa.eu/environment/enveco/taxation/pdf/ch8_pesticides.pdf.

EFSA. 2013. 'Scientific Opinion on the Identification of Pesticides to Be Included in Cumulative Assessment Groups on the Basis of Their Toxicological Profile'. *EFSA Journal* 11 (7): 3293. https://doi.org/10.2903/j.efsa.2013.3293.

Epstein, Lynn, and Susan Bassein. 2003. 'Patterns of Pesticide Use in California and the Implications for Strategies for Reduction of Pesticides'. *Annual Review of Phytopathology* 41 (1): 351–75. https://doi.org/10.1146/annurev.phyto.41.052002.095612.

Estes, Tammara L., Naresh Pai, and Michael F. Winchell. 2016. 'Comparison of Predicted Pesticide Concentrations in Groundwater from SCI-GROW and PRZM-GW Models with Historical Monitoring Data'. *Pest Management Science* 72 (6): 1187–1201. https://doi.org/10.1002/ps.4097.

European Commission. 2017. 'Overview Report on a Series of Audits Carried out in EU Member States in 2016 and 201: In Order to Evaluate the Systems in Place for the Authorisation of Plant Protection Products'. https://ec.europa.eu/food/audits-analysis/overview_reports/act_getPDF.cfm?PDF_ID=1021.

———. 2019. 'Statistics on Agricultural Use of Pesticides in the European Union'. https://ec.europa.eu/eurostat/documents/749240/0/Statistics+on+the+agricultural+use+of+pesticides+in+the+EU.

———. 2020. 'Strategy for Agricultural Statistics for 2020 and Beyond'. https://ec.europa.eu/eurostat/documents/749240/749310/Strategy+on+agricultural+statistics+Final+version+for+publication.pdf/9c7787ca-0e00-f676-7a64-7f56e74ec813.

European Commission, and Statistical Office of the European Union. 2019. *Methodology for Calculating Harmonised Risk Indicators for Pesticides under Directive 2009/128/EC: 2019 Edition.* https://op.europa.eu/publication/manifestation_identifier/PUB_KSGQ19009ENN.

European Court of Auditors. 2020. 'Special Report 05/2019: Sustainable Use of Plant Protection Products: Limited Progress in Measuring and Reducing Risks'. https://www.eca.europa.eu/Lists/ECADocuments/SR20_05/SR_Pesticides_EN.pdf.

Eurostat. 2008. 'A Common Methodology for the Collection of Pesticide Usage Statistics within Agriculture and Horticulture'. https://ec.europa.eu/eurostat/documents/3859598/5902633/KS-RA-08-010-EN.PDF/f88b0b37-571a-43e0-8b6a-da8d1b43d521?version=1.0.

Ewald, J.A., C.J. Wheatley, N.J. Aebischer, S. Duffield, and D. Heaver. 2016. 'Investigation of the Impact of Changes in Pesticide Use on Invertebrate Populations'. NECR182. Natural England. http://publications.naturalengland.org.uk/publication/4858366522818560.

Feola, G., E. Rahn, and C. R. Binder. 2011. 'Suitability of Pesticide Risk Indicators for Less Developed Countries: A Comparison'. *Agriculture, Ecosystems & Environment* 142 (3): 238–45. https://doi.org/10.1016/j.agee.2011.05.014.

Ferrari, Federico, Michael Klein, Ettore Capri, and Marco Trevisan. 2005. 'Prediction of Pesticide Volatilization with PELMO 3.31'. *Chemosphere* 60 (5): 705–13. https://doi.org/10.1016/j.chemosphere.2005.01.043.

Finger, Robert, Niklas Möhring, Tobias Dalhaus, and Thomas Böcker. 2017. 'Revisiting Pesticide Taxation Schemes'. *Ecological Economics* 134 (April): 263–66. https://doi.org/10.1016/j.ecolecon.2016.12.001.

Foote, Natasha. 2020. 'Controversial Risk Indicator to Be Basis for Pesticide Reduction Targets'. *Www.Euractiv.Com* (blog). 4 March 2020. https://www.euractiv.com/section/agriculture-food/news/controversial-risk-indicator-to-be-basis-for-pesticide-reduction-targets/.

Foster, Vivien, and Susana Mourato. 2000. 'Valuing the Multiple Impacts of Pesticide Use in the UK: A Contingent Ranking Approach'. *Journal of Agricultural Economics* 51 (1): 1–21. https://doi.org/10.1111/j.1477-9552.2000.tb01206.x.

Garthwaite, David G., Lee Ridley, Abigale Mace, G. Parrish, Ian Barker, James L. Rainford, and Roy MacArthur. 2019. 'Pesticide Usage Survey Report 284 Arable Crops in the United Kingdom 2018'. Pesticide Usage Survey. FERA Science Limited on behalf of DEFRA. https://secure.fera.defra.gov.uk/pusstats/surveys/documents/arable2018.pdf.

Gooday, R. D., S. G. Anthony, D. R. Chadwick, P. Newell-Price, D. Harris, D. Duethmann, R. Fish, A. L. Collins, and M. Winter. 2014. 'Modelling the Cost-Effectiveness of Mitigation Methods for Multiple Pollutants at Farm Scale'. *Science of The Total Environment* 468–469 (January): 1198–1209. https://doi.org/10.1016/j.scitotenv.2013.04.078.

Gravesen, L. 2003. 'The Treatment Frequency Index: An Indicator for Pesticide Use and Dependency as Well as Overall Load on the Environment'. In *Reducing Pesticide Dependency in Europe to Protect Health, Environment and Biodiversity*. København.

Gutsche, Volkmar, and John Carley. n.d. 'OECD Aquatic Risk Indicators Computer System: User Guide'. http://www.oecd.org/env/ehs/pesticides-biocides/2753225.pdf.

Gutsche, Volkmar, and Dietmar Rossberg. 1997. 'SYNOPS 1.1: A Model to Assess and to Compare the Environmental Risk Potential of Active Ingredients in Plant Protection Products'. *Agriculture, Ecosystems & Environment*, Integrated Crop Protection: Towards Sustainability?, 64 (2): 181–88. https://doi.org/10.1016/S0167-8809(97)00037-6.

Halberg, Niels, Gerwin Verschuur, and Gillian Goodlass. 2005. 'Farm Level Environmental Indicators; Are They Useful?: An Overview of Green Accounting Systems for European Farms'. *Agriculture, Ecosystems & Environment* 105 (1): 195–212. https://doi.org/10.1016/j.agee.2004.04.003.

Henderson, Andrew D., Michael Z. Hauschild, Dik van de Meent, Mark A. J. Huijbregts, Henrik Fred Larsen, Manuele Margni, Thomas E. McKone, Jerome Payet, Ralph K. Rosenbaum, and Olivier Jolliet. 2011. 'USEtox Fate and Ecotoxicity Factors for Comparative Assessment of Toxic Emissions in Life Cycle Analysis: Sensitivity to Key Chemical Properties'. *The International Journal of Life Cycle Assessment* 16 (8): 701. https://doi.org/10.1007/s11367-011-0294-6.

HM Government. 2018. 'A Green Future: Our 25 Year Plan to Improve the Environment'. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf.

Hossard, Laure, Laurence Guichard, Céline Pelosi, and David Makowski. 2017. 'Lack of Evidence for a Decrease in Synthetic Pesticide Use on the Main Arable Crops in France'. *Science of The Total Environment* 575 (January): 152–61. https://doi.org/10.1016/j.scitotenv.2016.10.008.

Huyghe, C., and M. Blanck. 2017. 'Pesticide saving certificates: context and implementation.' *6e COMAPPI, Conférence sur les Moyens Alternatifs de Protection pour une Production Intégrée, Lille, France, 21-23 mars 2017*, 13–24.

Jarvis, N. J. 1995. 'Simulation of Soil Water Dynamics and Herbicide Persistence in a Silt Loam Soil Using the MACRO Model'. *Ecological Modelling*, Modelling of Geo-Biosphere Processes, 81 (1): 97–109. https://doi.org/10.1016/0304-3800(94)00163-C.

Jepson, Paul C., Katie Murray, Oliver Bach, Maria A. Bonilla, and Lars Neumeister. 2020. 'Selection of Pesticides to Reduce Human and Environmental Health Risks: A Global Guideline and Minimum Pesticides List'. *The Lancet Planetary Health* 4 (2): e56–63. https://doi.org/10.1016/S2542-5196(19)30266-9.

Johnson, Branden B., and Caron Chess. 2006. 'Evaluating Public Responses to Environmental Trend Indicators'. *Science Communication* 28 (1): 64–92. https://doi.org/10.1177/1075547006291346.

Juraske, Ronnie, Assumpció Antón, Francesc Castells, and Mark A. J. Huijbregts. 2007. 'PestScreen: A Screening Approach for Scoring and Ranking Pesticides by Their Environmental and Toxicological Concern'. *Environment International* 33 (7): 886–93. https://doi.org/10.1016/j.envint.2007.04.005.

Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. 2017. 'A Review on the Practice of Big Data Analysis in Agriculture'. *Computers and Electronics in Agriculture* 143 (December): 23–37. https://doi.org/10.1016/j.compag.2017.09.037.

Kim, Ki-Hyun, Ehsanul Kabir, and Shamin Ara Jahan. 2017. 'Exposure to Pesticides and the Associated Human Health Effects'. *Science of The Total Environment* 575 (January): 525–35. https://doi.org/10.1016/j.scitotenv.2016.09.009.

Kniss, Andrew. 2016. 'Trends in Diversity and Relative Toxicity of Herbicide Use in the United States'. *Proceedings of the Integrated Crop Management Conference*, December. https://lib.dr.iastate.edu/icm/2016/proceedings/12.

———. 2017. 'Long-Term Trends in the Intensity and Relative Toxicity of Herbicide Use'. *Nature Communications* 8 (1): 1–7. https://doi.org/10.1038/ncomms14865.

Kniss, Andrew, and Carl W. Coburn. 2015. 'Quantitative Evaluation of the Environmental Impact Quotient (EIQ) for Comparing Herbicides'. *PLOS ONE* 10 (6): e0131200. https://doi.org/10.1371/journal.pone.0131200.

Kookana, Rai S., Raymond L. Correll, and Rosalind B. Miller. 2005. 'Pesticide Impact Rating Index – A Pesticide Risk Indicator for Water Quality'. *Water, Air, & Soil Pollution: Focus* 5 (1): 45–65. https://doi.org/10.1007/s11267-005-7397-7.

Kovach, Joseph, Curtis Petzoldt, Janice Degni, and James Tette. 1992. 'A Method to Measure the Environmental Impact of Pesticides'. https://ecommons.cornell.edu/handle/1813/55750.

Kruijne, R., J. W. Deneer, J. Lahr, and J. Vlaming. 2011. 'Hair 2010 Documentation: Calculating Risk Indicators Related to Agricultural Use of Pesticides within the European Union'. 2113.1. Wageningen: Alterra. https://library.wur.nl/WebQuery/wurpubs/409362.

Kruijne, R., J. Vlaming, J. W. Deneer, Riikka Nousiainen, and K. Rasaen. 2014. 'HAIR2014 Software Manual'. Wageningen: Alterra Wageningen UR. https://www.pesticidemodels.eu/sites/default/files/downloads/HAIR/HAIR2014%20Software%20Manual.pdf.

Kudsk, Per, Lise Nistrup Jørgensen, and Jens Erik Ørum. 2018. 'Pesticide Load—A New Danish Pesticide Risk Indicator with Multiple Applications'. *Land Use Policy* 70 (January): 384–93. https://doi.org/10.1016/j.landusepol.2017.11.010.

Labite, Herve, F. Butler, and E. Cummins. 2011. 'A Review and Evaluation of Plant Protection Product Ranking Tools Used in Agriculture'. *Human and Ecological Risk Assessment: An International Journal* 17 (2): 300–327. https://doi.org/10.1080/10807039.2011.552392.

Labite, Herve, Nicholas M. Holden, Karl G. Richards, Gaelene Kramers, Alina Premrov, Catherine E. Coxon, and Enda Cummins. 2013. 'Comparison of Pesticide Leaching Potential to Groundwater under EU FOCUS and Site Specific Conditions'. *Science of The Total Environment* 463–464 (October): 432–41. https://doi.org/10.1016/j.scitotenv.2013.06.050.

Laetz, Cathy A., David H. Baldwin, Tracy K. Collier, Vincent Hebert, John D. Stark, and Nathaniel L. Scholz. 2009. 'The Synergistic Toxicity of Pesticide Mixtures: Implications for Risk Assessment and the Conservation of Endangered Pacific Salmon'. *Environmental Health Perspectives* 117 (3): 348–53. https://doi.org/10.1289/ehp.0800096.

Lamichhane, Jay Ram, Silke Dachbrodt-Saaydeh, Per Kudsk, and Antoine Messéan. 2015. 'Toward a Reduced Reliance on Conventional Pesticides in European Agriculture'. *Plant Disease* 100 (1): 10–24. https://doi.org/10.1094/PDIS-05-15-0574-FE.

Lammoglia, Sabine-Karen, François Brun, Thibaud Quemar, Julien Moeys, Enrique Barriuso, Benoît Gabrielle, and Laure Mamy. 2018. 'Modelling Pesticides Leaching in Cropping Systems: Effect of Uncertainties in Climate, Agricultural Practices, Soil and Pesticide Properties'. *Environmental Modelling & Software* 109 (November): 342–52. https://doi.org/10.1016/j.envsoft.2018.08.007.

Lammoglia, Sabine-Karen, Julien Moeys, Enrique Barriuso, Mats Larsbo, Jesús-María Marín-Benito, Eric Justes, Lionel Alletto, et al. 2017. 'Sequential Use of the STICS Crop Model and of the MACRO Pesticide Fate Model to Simulate Pesticides Leaching in Cropping Systems'. *Environmental Science and Pollution Research* 24 (8): 6895–6909. https://doi.org/10.1007/s11356-016-6842-7.

Larsbo, Mats, Stephanie Roulier, Fredrik Stenemo, Roy Kasteel, and Nicholas Jarvis. 2005. 'An Improved Dual-Permeability Model of Water Flow and Solute Transport in the Vadose Zone'. *Vadose Zone Journal* 4 (2): 398–406. https://doi.org/10.2136/vzj2004.0137.

Larsen, Ashley E., Steven D. Gaines, and Olivier Deschênes. 2017. 'Agricultural Pesticide Use and Adverse Birth Outcomes in the San Joaquin Valley of California'. *Nature Communications* 8 (1): 302. https://doi.org/10.1038/s41467-017-00349-2.

Lee, Rhiannon, Roos den Uyl, and Hens Runhaar. 2019. 'Assessment of Policy Instruments for Pesticide Use Reduction in Europe; Learning from a Systematic Literature Review'. *Crop Protection* 126 (December): 104929. https://doi.org/10.1016/j.cropro.2019.104929.

Levitan, Lois. 1997. 'An Overview of Pesticide Impact Assessment Systems (a.k.a. "Pesticide Risk Indicators") Based on Indexing or Ranking Pesticides by Environmental Impact'. Background Paper Prepared for the OECD Workshop on Pesticide Risk Indicators 21-23 April, 1997. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.3449&rep=rep1&type=pdf.

———. 2000. '"How to" and "Why": Assessing the Enviro–Social Impacts of Pesticides'. *Crop Protection*, XIVth International Plant Protection Congress, 19 (8): 629–36. https://doi.org/10.1016/S0261-2194(00)00083-1.

Levitan, Lois, Ian Merwin, and Joe Kovach. 1995. 'Assessing the Relative Environmental Impacts of Agricultural Pesticides: The Quest for a Holistic Method'. *Agriculture, Ecosystems & Environment* 55 (3): 153–68. https://doi.org/10.1016/0167-8809(95)00622-Y.

Lewis, Kathleen A., Colin D. Brown, Andy Hart, and John Tzilivakis. 2003. 'P-EMA (III): Overview and Application of a Software System Designed to Assess the Environmental Risk of Agricultural Pesticides'. *Agronomie* 23 (1): 85–96. https://doi.org/10.1051/agro:2002076.

Lewis, Kathleen A., John Tzilivakis, Douglas J. Warner, and Andrew Green. 2016. 'An International Database for Pesticide Risk Assessments and Management'. *Human and Ecological Risk Assessment: An International Journal* 22 (4): 1050–64. https://doi.org/10.1080/10807039.2015.1133242.

Lindahl, Anna M. L., and Christian Bockstaller. 2012. 'An Indicator of Pesticide Leaching Risk to Groundwater'. *Ecological Indicators* 23 (December): 95–108. https://doi.org/10.1016/j.ecolind.2012.03.014.

Maggi, Federico, Daniele la Cecilia, Fiona H. M. Tang, and Alexander McBratney. 2020. 'The Global Environmental Hazard of Glyphosate Use'. *Science of The Total Environment* 717 (May): 137167. https://doi.org/10.1016/j.scitotenv.2020.137167.

Mancini, Francesca, Ben A. Woodcock, and Nick J. B. Isaac. 2019. 'Agrochemicals in the Wild: Identifying Links between Pesticide Use and Declines of Nontarget Organisms'. *Current*

Opinion in Environmental Science & Health, Environmental Pollution: Wildlife, 11 (October): 53–58. https://doi.org/10.1016/j.coesh.2019.07.003.

Margni, M., D. Rossier, P. Crettaz, and O. Jolliet. 2002. 'Life Cycle Impact Assessment of Pesticides on Human Health and Ecosystems'. *Agriculture, Ecosystems & Environment* 93 (1): 379–92. https://doi.org/10.1016/S0167-8809(01)00336-X.

Maud, Jackie, Gareth Edwards-Jones, and Fraser Quin. 2001. 'Comparative Evaluation of Pesticide Risk Indices for Policy Development and Assessment in the United Kingdom'. *Agriculture, Ecosystems & Environment* 86 (1): 59–73. https://doi.org/10.1016/S0167-8809(00)00258-9.

McGrath, Gavan, P. Suresh C. Rao, Per-Erik Mellander, Ivan Kennedy, Michael Rose, and Lukas van Zwieten. 2019. 'Real-Time Forecasting of Pesticide Concentrations in Soil'. *Science of The Total Environment* 663 (May): 709–17. https://doi.org/10.1016/j.scitotenv.2019.01.401.

Meek, M. E. Bette, Alan R. Boobis, Kevin M. Crofton, Gerhard Heinemeyer, Marcel Van Raaij, and Carolyn Vickers. 2011. 'Risk Assessment of Combined Exposure to Multiple Chemicals: A WHO/IPCS Framework'. *Regulatory Toxicology and Pharmacology: RTP*, April. https://doi.org/10.1016/j.yrtph.2011.03.010.

Miljøstyrelsen. 2012. 'The Agricultural Pesticide Load in Denmark 2007–2010. Environmental Review 2, 2012'. København. https://www2.mst.dk/Udgiv/publikationer/2012/03/978-87-92779-96-0.pdf.

Milner, Alice M., and Ian L. Boyd. 2017. 'Toward Pesticidovigilance'. *Science* 357 (6357): 1232–34. https://doi.org/10.1126/science.aan2683.

Ministère de l'Agriculture et de l'Alimentation. 2015. 'Ecophyto Plan II –20 October 2015'. https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_sup_nap_fra-ecophyto-2_en.pdf.

Møhlenberg, Flemming, Kim Gustavson, and Peter B Sørensen. 2002. 'Pesticide Aquatic Risk Indicators; an Examination of the OECD Indicators REXTOX, ADSCOR and the Danish Indicators FA and LI Based on Danish Sales Data from 1992-2000'. Denmark: OECD.

Möhring, Niklas, Martina Bozzola, Stefan Hirsch, and Robert Finger. 2020. 'Are Pesticides Risk Decreasing? The Relevance of Pesticide Indicator Choice in Empirical Analysis'. *Agricultural Economics* 51 (3): 429–44. https://doi.org/10.1111/agec.12563.

Möhring, Niklas, Sabrina Gaba, and Robert Finger. 2019. 'Quantity Based Indicators Fail to Identify Extreme Pesticide Risks'. *Science of The Total Environment* 646 (January): 503–23. https://doi.org/10.1016/j.scitotenv.2018.07.287.

Möhring, Niklas, David Wuepper, Tomke Musa, and Robert Finger. 2020. 'Why Farmers Deviate from Recommended Pesticide Timing: The Role of Uncertainty and Information'. *Pest Management Science* n/a (n/a). https://doi.org/10.1002/ps.5826.

Moran, Kelly, Brian Anderson, Bryn Phillips, Yuzhou Luo, Nan Singhasemanon, Richard Breuer, and Dawit Tadesse. 2020. 'Water Quality Impairments Due to Aquatic Life Pesticide Toxicity: Prevention and Mitigation in California, USA'. *Environmental Toxicology and Chemistry* 39 (5): 953–66. https://doi.org/10.1002/etc.4699.

More, Simon John, Vasileios Bampidis, Diane Benford, Susanne Hougaard Bennekou, Claude Bragard, Thorhallur Ingi Halldorsson, Antonio F. Hernández-Jerez, et al. 2019. 'Guidance on Harmonised Methodologies for Human Health, Animal Health and Ecological Risk Assessment of Combined Exposure to Multiple Chemicals'. *EFSA Journal* 17 (3): e05634. https://doi.org/10.2903/j.efsa.2019.5634.

NAPAN. 2014. 'Belgium Action Plan to Reduce the Risks and Impacts Linked to Pesticides 2013 2017'. https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_sup_nap_bel_en.pdf.

Nicholson, Fiona, Rikke Krogshave Laursen, Rachel Cassidy, Luke Farrow, Linda Tendler, John Williams, Nicolas Surdyk, and Gerard Velthof. 2020. 'How Can Decision Support Tools Help Reduce Nitrate and Pesticide Pollution from Agriculture? A Literature Review and Practical Insights from the EU FAIRWAY Project'. *Water* 12 (3): 768. https://doi.org/10.3390/w12030768.

Norwegian Food Safety Authority. 2005. 'Guidelines for a Banded Pesticide Tax Scheme, Differentiated According to Human Health and Environmental Risks'. https://www.mattilsynet.no/language/english/plants/plant_protection_products/guidelines_for_a_banded_pesticide_tax_scheme_differentiated_according_to_human_health_and_environmental_risks.19283/binary/Guidelines%20for%20a%20Banded%20Pesticide%20Tax%20Scheme,%20Differentiated%20According%20to%20Human%20Health%20and%20Environmental%20Risks.

Notarnicola, Bruno, Serenella Sala, Assumpció Anton, Sarah J. McLaren, Erwan Saouter, and Ulf Sonesson. 2017. 'The Role of Life Cycle Assessment in Supporting Sustainable Agri-Food Systems: A Review of the Challenges'. *Journal of Cleaner Production*, Towards eco-efficient agriculture and food systems: selected papers addressing the global challenges for food systems, including those presented at the Conference "LCA for Feeding the planet and energy for life" (6-8 October 2015, Stresa & Milan Expo, Italy), 140 (January): 399–409. https://doi.org/10.1016/j.jclepro.2016.06.071.

OECD. 2000. 'Report of the OECD Pesticide Aquatic Risk Indicators Expert Group'. https://www.oecd.org/env/ehs/pesticides-biocides/2078654.pdf.

———. 2004. *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistance and Long-Range Transport*. OECD Series on Testing and Assessment. OECD. https://doi.org/10.1787/9789264079137-en.

———. 2017. 'The Evolution of the Tax on Pesticides and the Pesticide Savings Certificates in France'. The Political Economy of Biodiversity Policy Reform. OCED. https://dx.doi.org/10.1787/9789264269545-7-en.

———. 2018. 'Considerations for Assessing the Risks of Combined Exposure to Multiple Chemicals, Series on Testing and Assessment No. 296, Environment, Health and Safety Division, Environment Directorate'. http://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf.

Oerke, E.-C. 2006. 'Crop Losses to Pests'. *The Journal of Agricultural Science* 144 (1): 31–43. https://doi.org/10.1017/S0021859605005708.

Oliver, Danielle P., Rai S. Kookana, Jenny S. Anderson, and Beng Umali. 2016. 'Field Evaluation of Two Risk Indicators for Predicting Likelihood of Pesticide Transport to Surface Water from Two Orchards'. *Science of The Total Environment* 571 (November): 819–25. https://doi.org/10.1016/j.scitotenv.2016.07.056.

Padovani, Laura, Marco Trevisan, and Ettore Capri. 2004. 'A Calculation Procedure to Assess Potential Environmental Risk of Pesticides at the Farm Level'. *Ecological Indicators* 4 (2): 111–23. https://doi.org/10.1016/j.ecolind.2004.01.002.

Pedersen, Anders Branth, and Helle Ørsted Nielsen. 2017. 'Effectiveness of Pesticide Policies'. In *Environmental Pest Management*, 297–324. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119255574.ch13.

Pendersen, Anders, Nielson Helle, and Mikael Andersen. 2015. 'The Danish Pesticide Tax'. In *Use of Economic Instruments in Water Policy*, 14:73–87. Global Issues in Water Policy. Springer, Cham. https://doi.org/10.1007/978-3-319-18287-2_6.

Peshin, Rajinder, Rakesh S. Bandral, WenJun Zhang, Lewis Wilson, and Ashok K. Dhawan. 2009. 'Integrated Pest Management: A Global Overview of History, Programs and Adoption'. In *Integrated Pest Management: Innovation-Development Process: Volume 1*, edited by Rajinder Peshin and Ashok K. Dhawan, 1–49. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-8992-3_1.

Pierlot, Frédéric, Jonathan Marks-Perreau, Benoît Réal, Nadia Carluer, Thibaut Constant, Abdeljalil Lioeddine, Paul van Dijk, et al. 2017. 'Predictive Quality of 26 Pesticide Risk Indicators and One Flow Model: A Multisite Assessment for Water Contamination'. *Science of The Total Environment* 605–606 (December): 655–65. https://doi.org/10.1016/j.scitotenv.2017.06.112.

Pivato, Alberto, Alberto Barausse, Francesco Zecchinato, Luca Palmeri, Roberto Raga, Maria Cristina Lavagnolo, and Raffaello Cossu. 2015. 'An Integrated Model-Based Approach to the Risk Assessment of Pesticide Drift from Vineyards'. *Atmospheric Environment* 111 (June): 136–50. https://doi.org/10.1016/j.atmosenv.2015.04.005.

Popp, József, Károly Pető, and János Nagy. 2013. 'Pesticide Productivity and Food Security. A Review'. *Agronomy for Sustainable Development*. https://agris.fao.org/agris-search/search.do?recordID=US201400133114.

Potier, D. 2014. 'Pesticides et Agro-Écologie, Les Champs Du Possible'. Paris, France. https://agriculture.gouv.fr/telecharger/56000?token=7bf92926cba72dbc99beeeef8758248e.

Price, Newell, D. Harris, M. Taylor, J. R. Williams, S. G. Anthony, D. Duethmann, R. D. Gooday, et al. 2011. 'An Inventory of Mitigation Methods and Guide to Their Effects on Diffuse Water Pollution, Greenhouse Gas Emissions and Ammonia Emissions from Agriculture; User Guide'. https://www.cost869.alterra.nl/UK_Manual_2011.pdf.

Rainford, James L., David G. Garthwaite, Glyn Jones, John Tzilivakis, and Kathleen A. Lewis. In preparation. 'Developing a UK Pesticide Load Indicator'. FERA Science limited and the University of Hertfordshire on behalf of DEFRA.

Reus, J, P Leendertse, C Bockstaller, I Fomsgaard, V Gutsche, Kathleen A. Lewis, C Nilsson, et al. 2002. 'Comparison and Evaluation of Eight Pesticide Environmental Risk Indicators Developed in Europe and Recommendations for Future Use'. *Agriculture, Ecosystems & Environment* 90 (2): 177–87. https://doi.org/10.1016/S0167-8809(01)00197-9.

Reus, J, and Peter C Leendertse. 2000. 'The Environmental Yardstick for Pesticides: A Practical Indicator Used in the Netherlands'. *Crop Protection*, XIVth International Plant Protection Congress, 19 (8): 637–41. https://doi.org/10.1016/S0261-2194(00)00084-3.

Rose, David C., and Toby J. A. Bruce. 2018. 'Finding the Right Connection: What Makes a Successful Decision Support System?' *Food and Energy Security* 7 (1): e00123. https://doi.org/10.1002/fes3.123.

Rose, David C., Caroline Parker, Joe Fodey, Caroline Park, William J. Sutherland, and Lynn V. Dicks. 2018. 'Involving Stakeholders in Agricultural Decision Support Systems: Improving User-Centred Design'. *International Journal of Agricultural Management* 6 (3–4): 80–89. https://doi.org/10.5836/ijam/2017-06-80.

Rose, David C., William J. Sutherland, Caroline Parker, Matt Lobley, Michael Winter, Carol Morris, Susan Twining, Charles Ffoulkes, Tatsuya Amano, and Lynn V. Dicks. 2016. 'Decision Support Tools for Agriculture: Towards Effective Design and Delivery'. *Agricultural Systems* 149 (November): 165–74. https://doi.org/10.1016/j.agsy.2016.09.009.

Rosenbaum, Ralph K., Till M. Bachmann, Lois Swirsky Gold, Mark A. J. Huijbregts, Olivier Jolliet, Ronnie Juraske, Annette Koehler, et al. 2008. 'USEtox—the UNEP-SETAC Toxicity Model: Recommended Characterisation Factors for Human Toxicity and Freshwater Ecotoxicity in Life Cycle Impact Assessment'. *The International Journal of Life Cycle Assessment* 13 (7): 532. https://doi.org/10.1007/s11367-008-0038-4.

Sánchez-Bayo, Francisco, Sundaram Baskaran, and Ivan Robert Kennedy. 2002. 'Ecological Relative Risk (EcoRR): Another Approach for Risk Assessment of Pesticides in Agriculture'. *Agriculture, Ecosystems & Environment* 91 (1): 37–57. https://doi.org/10.1016/S0167-8809(01)00258-4.

Saouter, Erwan, Deidre Wolff, Fabrizio Biganzoli, and Donald Versteeg. 2019. 'Comparing Options for Deriving Chemical Ecotoxicity Hazard Values for the European Union Environmental Footprint, Part II'. *Integrated Environmental Assessment and Management* 15 (5): 796–807. https://doi.org/10.1002/ieam.4169.

Sattler, Claudia, Harald Kächele, and Gernot Verch. 2007. 'Assessing the Intensity of Pesticide Use in Agriculture'. *Agriculture, Ecosystems & Environment* 119 (3): 299–304. https://doi.org/10.1016/j.agee.2006.07.017.

Savary, Serge, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. 2019. 'The Global Burden of Pathogens and Pests on Major Food Crops'. *Nature Ecology & Evolution* 3 (3): 430–39. https://doi.org/10.1038/s41559-018-0793-y.

Schäfer, Ralf B., Matthias Liess, Rolf Altenburger, Juliane Filser, Henner Hollert, Martina Roß-Nickoll, Andreas Schäffer, and Martin Scheringer. 2019. 'Future Pesticide Risk Assessment: Narrowing the Gap between Intention and Reality'. *Environmental Sciences Europe* 31 (1): 21. https://doi.org/10.1186/s12302-019-0203-3.

Schäfer, Ralf B., and Jeremy J. Piggott. 2018. 'Advancing Understanding and Prediction in Multiple Stressor Research through a Mechanistic Basis for Null Models'. *Global Change Biology* 24 (5): 1817–26. https://doi.org/10.1111/gcb.14073.

Siimes, Katri, and Juha Kämäri. 2003. 'A Review of Available Pesticide Leaching Models: Selection of Models for Simulation of Herbicide Fate in Finnish Sugar Beet Cultivation'. *Boreal Enviromental Research* 8: 31–51.

Silva, Vera, Hans G. J. Mol, Paul Zomer, Marc Tienstra, Coen J. Ritsema, and Violette Geissen. 2019. 'Pesticide Residues in European Agricultural Soils – A Hidden Reality Unfolded'. *Science of The Total Environment* 653 (February): 1532–45. https://doi.org/10.1016/j.scitotenv.2018.10.441.

Skevas, T., A. G. J. M. Oude Lansink, and S. E. Stefanou. 2013. 'Designing the Emerging EU Pesticide Policy: A Literature Review'. *NJAS - Wageningen Journal of Life Sciences* 64–65 (September): 95–103. https://doi.org/10.1016/j.njas.2012.09.001.

Steingrímsdóttir, María Magnea, Annette Petersen, and Peter Fantke. 2018. 'A Screening Framework for Pesticide Substitution in Agriculture'. *Journal of Cleaner Production* 192 (August): 306–15. https://doi.org/10.1016/j.jclepro.2018.04.266.

Stenrød, Marianne, Heidi E. Heggen, Randi I. Bolli, and Ole Martin Eklo. 2008. 'Testing and Comparison of Three Pesticide Risk Indicator Models under Norwegian Conditions—A Case Study in the Skuterud and Heiabekken Catchments'. *Agriculture, Ecosystems & Environment* 123 (1): 15–29. https://doi.org/10.1016/j.agee.2007.03.003.

Strassemeyer, J., D. Daehmlow, A. R. Dominic, S. Lorenz, and B. Golla. 2017. 'SYNOPS-WEB, an Online Tool for Environmental Risk Assessment to Evaluate Pesticide Strategies on Field Level'. *Crop Protection*, Pesticide use and risk reduction with IPM, 97 (July): 28–44. https://doi.org/10.1016/j.cropro.2016.11.036.

Strassemeyer, J., and Volkmar Gutsche. 2010. 'The Approach of the German Pesticide Risk Indicator SYNOPS in Frame of the National Action Plan for Sustainable Use of Pesticides'. https://www.semanticscholar.org/paper/The-approach-of-the-German-Pesticide-Risk-Indicator-Strassemeyer-Gutsche/6aa1fa688fff1b80c8df68eb214729d3d7874d68?citingPapersSort=relevance&citingPapersLimit=10&citingPapersOffset=0&year%5B0%5D=0&year%5B1%5D=0&citedPapersSort=relevance&citedPapersLimit=10&citedPapersOffset=0.

Strassemeyer, J., Volkmar Gutsche, Colin Brown, Matthias Liess, and Carola Schriever. 2007. 'HArmonised Environmental Indicators for Pesticide Risk Aquatic Indicators'. Federal Biological Research Centre for Agriculture and Forestry, Central Science Laboratory,Centre for Environmental Research Leipzig-Halle. 10.13140/RG.2.2.13571.96806.

Streissl, Franz, Mark Egsmose, and José V. Tarazona. 2018. 'Linking Pesticide Marketing Authorisations with Environmental Impact Assessments through Realistic Landscape Risk Assessment Paradigms'. *Ecotoxicology* 27 (7): 980–91. https://doi.org/10.1007/s10646-018-1962-0.

Sud, Megha. 2020. 'Managing the Biodiversity Impacts of Fertiliser and Pesticide Use: Overview and Insights from Trends and Policies across Selected OECD Countries'. 1555. OECD Environment Working Papers. Paris, France. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/WKP(2020)2&docLanguage=En.

Surgan, Michael, Madison Condon, and Caroline Cox. 2010. 'Pesticide Risk Indicators: Unidentified Inert Ingredients Compromise Their Integrity and Utility'. *Environmental Management* 45 (4): 834–41. https://doi.org/10.1007/s00267-009-9382-9.

The Pesticides Forum. 2020. 'Pesticides in the UK; The 2018 Report on the Impact and Sustainable Use of Pesticides'. https://webcommunities.hse.gov.uk/gf2.ti/f/21154/698373.1/PDF/-/PesticidesForumReport2018.pdf.

Thomas, Miles R. 1999. 'Guidelines for the Collection of Pesticide Usage Statistics within Agriculture and Horticulture'. OECD, Eurostat. https://secure.fera.defra.gov.uk/pusstats/surveys/documents/guide.pdf.

Trajanov, Aneta, Vladimir Kuzmanovski, Benoit Real, Jonathan Marks Perreau, Sašo Džeroski, and Marko Debeljak. 2018. 'Modeling the Risk of Water Pollution by Pesticides from Imbalanced Data'. *Environmental Science and Pollution Research* 25 (19): 18781–92. https://doi.org/10.1007/s11356-018-2099-7.

Trevisan, Marco, Andrea Di Guardo, and Matteo Balderacchi. 2009. 'An Environmental Indicator to Drive Sustainable Pest Management Practices'. *Environmental Modelling & Software* 24 (8): 994–1002. https://doi.org/10.1016/j.envsoft.2008.12.008.

Tsaboula, Aggeliki, Emmanouil-Nikolaos Papadakis, Zisis Vryzas, Athina Kotopoulou, Katerina Kintzikoglou, and Euphemia Papadopoulou-Mourkidou. 2016. 'Environmental and Human Risk Hierarchy of Pesticides: A Prioritization Method, Based on Monitoring, Hazard Assessment and Environmental Fate'. *Environment International* 91 (May): 78–93. https://doi.org/10.1016/j.envint.2016.02.008.

Uthes, Sandra, Ines Heyer, Annemarie Kaiser, Peter Zander, Christian Bockstaller, Yann Desjeux, Szilárd Keszthelyi, et al. 2019. 'Costs, Quantity and Toxicity: Comparison of Pesticide Indicators Collected from FADN Farms in Four EU-Countries'. *Ecological Indicators* 104 (September): 695–703. https://doi.org/10.1016/j.ecolind.2019.05.028.

Van Bol, V., and Luc Pussemier. 2005. 'Pesticide Indicators: A Study Case in Belgium Using the Index of Load Why Do We Need Pesticide Risk Indicators'. 10.13140/RG.2.2.15159.16800.

Vanclooster, M., J. J. T. I. Boesten, M. Trevisan, C. D. Brown, E. Capri, O. M. Eklo, B. Gottesbüren, V. Gouy, and A. M. A. van der Linden. 2000. 'A European Test of Pesticide-Leaching Models: Methodology and Major Recommendations'. *Agricultural Water Management* 44 (1): 1–19. https://doi.org/10.1016/S0378-3774(99)00081-5.

Vercruysse, Fangio, and Walter Steurbaut. 2002. 'POCER, the Pesticide Occupational and Environmental Risk Indicator'. *Crop Protection* 21 (4): 307–15. https://doi.org/10.1016/S0261-2194(01)00102-8.

Vergucht, S., and Walter Steurbaut. 2007. 'Development of Pesticide Risk Indicator for the Evaluation of the Belgian Reduction Plan'. In *Water Pollution in Natural Porous Media at Different Scales: Assessment of Fate, Impact and Indicators* ., 279–86. Groundwater. Madrid, Spain: IGME.

Voet, Hilko van der, Johannes W. Kruisselbrink, Waldo J. de Boer, Marco S. van Lenthe, J.J.B. (Hans) van den Heuvel, Amélie Crépet, Marc C. Kennedy, et al. 2019. 'Draft Paper on the EuroMix Toolbox of Models and Data to Support Chemical Mixture Risk Assessment', May. https://doi.org/10.5281/zenodo.3474943.

Wan, Neng. 2015. 'Pesticides Exposure Modeling Based on GIS and Remote Sensing Land Use Data'. *Applied Geography* 56 (January): 99–106. https://doi.org/10.1016/j.apgeog.2014.11.012.

Werf, Hayo M. G. van der, and Christophe Zimmer. 1998. 'An Indicator of Pesticide Environmental Impact Based on a Fuzzy Expert System'. *Chemosphere* 36 (10): 2225–49. https://doi.org/10.1016/S0045-6535(97)10194-1.

Wilson, Clevo, and Clem Tisdell. 2001. 'Why Farmers Continue to Use Pesticides despite Environmental, Health and Sustainability Costs'. *Ecological Economics* 39 (3): 449–62. https://doi.org/10.1016/S0921-8009(01)00238-5.

Woodcock, Ben A., Nicholas J. B. Isaac, James M. Bullock, David B. Roy, David G. Garthwaite, Andrew Crowe, and Richard F. Pywell. 2016. 'Impacts of Neonicotinoid Use on Long-Term Population

Changes in Wild Bees in England'. *Nature Communications* 7 (1): 12459. https://doi.org/10.1038/ncomms12459.
Zelm, Rosalie van, Mark A. J. Huijbregts, and Dik van de Meent. 2009. 'USES-LCA 2.0—a Global Nested Multi-Media Fate, Exposure, and Effects Model'. *The International Journal of Life Cycle Assessment* 14 (3): 282–84. https://doi.org/10.1007/s11367-009-0066-8.