RESEARCH ARTICLE

Risk assessment of drinking water intake contamination from agricultural activities using a Bayesian network

Raja Kammoun1*, Natasha McQuaid1, Vincent Lessard2, Eyerusalem Adhanom Goitom1, Michèle Prévost1, Françoise Bichai1, Sarah Dorner1

1 Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montreal, Quebec, Canada, 2 Conseil des bassins versants des Mille Îles, Saint-Eustache, Quebec, Canada

* raja.kammoun@polymtl.ca

Abstract

Agricultural activities can result in the contamination of surface runoff with pathogens, pesticides, and nutrients. These pollutants can enter surface water bodies in two ways: by direct discharge into surface waters or by infiltration and recharge into groundwater, followed by release to surface waters. Lack of financial resources makes risk assessment through analysis of drinking water pollutants challenging for drinking water suppliers. Inability to identify agricultural lands with a high-risk level and implement action measures might lead to public health issues. As a result, it is essential to identify hazards and conduct risk assessments even with limited data. This study proposes a risk assessment model for agricultural activities based on available data and integrating various types of knowledge, including expert and literature knowledge, to estimate the levels of hazard and risk that different agricultural activities could pose to the quality of withdrawal waters. To accomplish this, we built a Bayesian network with continuous and discrete inputs capturing raw water quality and land use upstream of drinking water intakes (DWIs). This probabilistic model integrates the DWI vulnerability, threat exposure, and threats from agricultural activities, including animal and crop production inventoried in drainage basins. The probabilistic dependencies between model nodes are established through a novel adaptation of a mixed aggregation method. The mixed aggregation method, a traditional approach used in ecological assessments following a deterministic framework, involves using fixed assumptions and parameters to estimate ecological outcomes in a specific case without considering inherent randomness and uncertainty within the system. After validation, this probabilistic model was used for four water intakes in a heavily urbanized watershed with agricultural activities in the south of Quebec, Canada. The findings imply that this methodology can assist stakeholders direct their efforts and investments on at-risk locations by identifying agricultural areas that can potentially pose a risk to DWIs.
1. Introduction

Agricultural practices are significant sources of chemical and microbial pollution to surface waters [1–5], posing a threat to public health. Runoff from agricultural fields and infiltration into groundwater leads to the transport of pesticides and nutrients into receiving waters [6]. The potential for contamination of surface waters depends on the type, quantity, adsorption, and absorption capacity of applied pesticides, nitrogen and phosphorus [7–9]. In addition, soil characteristics [10], rainfall rates [11–13], and agricultural management practices [14–16] are essential factors that determine the susceptibility of runoff to contamination. An increase in nutrient levels from animal and crop production can contribute to the proliferation of cyanobacteria in drinking water sources [17–20]. Some cyanobacteria blooms can release toxins that pose challenges in terms of drinking water safety [21–26]. Animal excrements and manure are also a source of microbial contamination in surface waters [27–30]. Factors influencing the survival and transport of pathogens in water resources include: water temperature, pH, biotic interactions, manure types and characteristics, and farm management practices [31].

Risk assessments can raise awareness of the threats posed by agricultural contamination to drinking water sources. Several methods to assess the risks associated with agricultural activities are available. In recent years, many risk assessment techniques have used deterministic modeling approaches focused on ecological, ecotoxicological, and human health issues [3, 32–37]. Deterministic modeling is costly, requires extensive calibration, and needs a large amount of site-specific data on water quality and hydrodynamics. The central problem with this approach is that it cannot be implemented by all drinking water suppliers due to a lack of financial resources and data. On the other hand, the deterministic approach could lead to an erroneous risk management decision, as it evaluates the impact of only one risk scenario at a time. Some of these limitations can be addressed using a probabilistic approach that simultaneously considers many risk assessment factors, such as various threat exposure scenarios, uncertainty of the different input variables, as well as uncertainty in the relationship between the parameters defining the system and the risk magnitude [38].

Bayesian networks (BN) are a valuable tool for conducting a probabilistic risk evaluation under uncertainty, and integrating different types of knowledge, such as quantitative and qualitative data, expert opinion, and prior experiences [39]. BNs have been extensively applied in different fields, such as in the fields of ecological risk assessment [40–43], human health risk assessment [40–44], climate change risk assessment [45], disaster risk assessment [46, 47], water quality and treatment [48, 49], best management practices of non-point source pollution [50], and water resource management [51]. To the best of our knowledge, these approaches have not been applied to vulnerability assessments of drinking water sources in agricultural watersheds.

The aim of this study was to develop a probabilistic model that decision-makers could use to swiftly identify animal and crop production areas that generate high chemical and microbial risks at DWIs. Our specific objectives were to: (1) develop a BN that combines the three components of risk assessment: hazard, exposure, and vulnerability; (2) test the applicability of existing methods used to define the conditional independence relationship among variables constituting BNs for drinking water risk assessments; (3) develop a novel method to define an efficient Bayes independence of the input variables for water resource risk assessments; (4) assess the probability distribution of the overall hazard of agricultural activities to pinpoint the drainage basins that require more detailed investigation for source water protection.

2. Materials and methods

2.1 Study area

The study was conducted for four DWIs, serving a population of approximately 440,000, drawing water from a river in Southern Quebec, Canada (Fig 1). Drinking water intakes 1 (DWI_1)
and 4 (DWI_4) are located in the most upstream and downstream sections of the river, respectively. DWI_2 and DWI_3 are located in the middle of the river reach. This river drains a lake fed by the Ottawa River, as described in Jalliffier-Verne et al. [52]. The Ottawa River receives runoff from the largest (146,300 km$^2$) watershed in Eastern Canada, and its average discharge is around 1,950 m$^3$/s [53]. This watershed is dominated by forests (75%) of which 40% is dense mixed wood. Agricultural activities take place over 6% of this watershed surface area. The adjacent watersheds that have a local influence on the 42-km long river that supplies the DWIs of this study cover an area of 1,008 km$^2$. This area is zoned as follows: 30% agricultural, 28.69% urban, 20.63% forest land.

Delineating protection zones is a critical step to assess non-point source pollution risks to drinking water sources, reduce the risk of contamination, and support drinking water safety. In this study, the delineation of immediate and intermediate protection areas was carried out to specify the land area that contributes water to the drinking water supply. The immediate protection area is a zone close to a DWI where pollution sources are deemed of concern and where contaminants may be readily found in the raw water with minimal dilution [54]. The intermediate protection area is a different zone where the travel time to the DWI is too short to allow intervention in the case of an accidental spill or a spike in contaminant concentration [54].

In Quebec, provincial regulation requires the delineation of surface water protection zones using a fixed distance approach [54]. This approach comprises only a strip of land on both sides of a river and does not include land surfaces whose relief leads to the drainage of land, lakes, and streams upstream of the river under study. For more information on regulatory methods for delineating surface water protection areas, see the S1 Text. To control diffuse
pollution, it is necessary to consider the drained non-point source pollution lands upstream of the tributary outlets included in the protection zones of the DWI. Therefore, this study develops an improved methodology of delineation based on levels 1 and 2 sub-watersheds. Level 1 sub-watersheds are the large watersheds whose outlet is the study site river. Level 2 sub-watersheds are those that drain into level 1 sub-watersheds and have the river in level 1 sub-watersheds as their outlet.

The geographic information system software (ArcGIS) was used to delineate the protection zones using the following data layers: location of DWIs, topological hydrographic network lines, and high-water mark. The immediate protection zone was delimited within 50 m downstream and 500 m upstream of the water intake as required by Quebec regulation [54]. This area includes any surface water, portions of tributaries, and a 10-m strip of land measured from the high-water mark. Surface and water portions of tributaries were also included in the intermediate protection zone and incorporate a 120-m strip of land measured from the high-water mark. This protection zone was delimited within 50 m downstream and 10 km upstream of the DWI [54]. This delineation was used to identify all level 1 and 2 sub-watersheds with outlets within the boundaries of these protection zones and to include them in the study site. Every sub-watershed selected is a drainage basin (DB) (S1 Fig). We define a drainage basin as a territory including a set of parcels used for one or more anthropogenic activities presenting a source of diffuse pollution and whose drained water is discharged into a watercourse through an outlet. Each drainage basin presents a potential source of microbial and/or chemical contamination to the studied DWIs.

2.2 Bayesian network framework

A probabilistic and spatial risk assessment of agricultural activities impacting DWIs was performed at a regional scale using a BN and ArcGIS. The risk assessment framework was built to integrate the three components of risks: hazard, vulnerability, and exposure (Fig 2).

A BN is a hierarchical representation and a probabilistic graphical model representing a relationship between a range of variables. Every single variable is represented as a node, and relationships between variables are presented by arcs defined by conditional dependencies [55]. This directed, acyclic graph allows the modeling of various uncertain events, facts, and systems [56]. More definitions of BN can be found in other studies [38, 47, 50, 57–60]. Bayesian networks are also known as conditional probability networks, casual probability networks, probabilistic networks, belief networks, Bayesian belief networks, or Bayesian reliability networks.

In this study, the BN was developed using the software Netica 6.09 (Norsys Software Corp), applying the Bayes’ theorem [61]. This software is an appropriate tool for creating, changing, and saving networks, as reviewed by Uusitalo [39].

2.2.1 Defining the nodes. The first step to structure the BN model was defining a set of random variables (nodes) and their states, which can be discrete or continuous. Discrete variables have a limited set of fixed values or states that they can take. Continuous variables are classified as real numbers since they may have any value within a certain range [62]. Continuous variables were discretized by breaking them up into intervals by assigning thresholds to define each interval. Input nodes are labeled “parent” nodes, and “child” nodes have input connections from one or more parent nodes [40, 58].

The determination of the critical nodes constituting the BN was carried out based on the problem and the objective of the study, which considers the DWI as the main subject of the risk assessment from agricultural activities. These nodes were chosen to represent all components of the risk assessment model based on the interaction of vulnerability to microbiological...
contamination (VMC), vulnerability to chemical contamination (VCC), vulnerability to water scarcity (VWS), exposure (EXP), and probability of hazard (PH). In the Working Group II (WGII) sixth assessment report from Intergovernmental Panel on Climate Change [63], the hazard is described in terms of the potential occurrence of the chemical and microbial contaminants that may cause degradation of water resource. The exposure depends on the presence of DWI in locations that could be adversely affected. The vulnerability is evaluated as the propensity or predisposition of the DWI to be adversely affected.

In this study, the intrinsic vulnerability of the intake was assessed by examining the sensitivity of the raw water to potential pollutants and the physical characteristics of the intake. Three key variables were used to assess the vulnerability to microbiological contamination of water intakes: (i) the geometric mean of *Escherichia coli* concentrations (GMEC) was used to present the central tendency of the microbial contamination indicator while being robust to extreme values; (ii) the 99th percentile of *E. coli* concentrations (PKEC) was used to account for the peaks of microbial contamination that are usually responsible for waterborne diseases [64]; and (iii) the seasonal variability of microbial contamination (SVMC) that was calculated based on a consecutive disparity index that estimates the variability of *E. coli* concentrations while considering the order of their chronological distribution, and without being independent of the data set’s average [65]. All variables were calculated for the four seasons of the year [66] (Winter, Spring, Summer, and Autumn) to study microbial contamination distribution and evaluate the sensitivity to seasonal variations.
Assessment of the vulnerability to chemical contamination was carried out based on the total phosphorus in raw water, which is routinely analyzed at drinking water treatment plants in Quebec. As data for other chemicals, such as nitrate and nitrogen, were not available, the vulnerability to chemical contamination was also estimated by analyzing the land use of activities that may release chemical contaminants in the immediate and intermediate DWIs’ protection zones. Raw water quality data available over a period of five years (January 2013 to December 2017) was used to assess the microbiological and chemical vulnerability of DWIs as required by Water Withdrawal and Protection Regulation [54].

The results of the vulnerability to water scarcity from the Levesque [67] study were used to assess vulnerability based on (i) the water levels in the studied river, which depend on the physical characteristics of the DWI and (ii) the water demand assigned to each catchment. The exposure was estimated based on the location of the outlet of each drainage basin in relation to the intake and its protection areas. This node allows assessing the non-point source pollution exposure pathways to the raw water supply.

The probability of hazard was estimated based on the proportion of the drainage basins occupied by crop and animal production activities. Crop classes and animal classes were used as input variables to estimate surface water contamination levels by pesticides and microbial contaminants.

The final node presents the overall risk probability distribution across three possible states (low, medium, and high) of agricultural activities occupying a watershed located upstream of a DWI by combining vulnerability, exposure, and hazard as outlined below.

The input variables, their states, and the discretization methods are described in Table 1. Once the nodes were set up, links (dependencies) between them were defined using conditional probability tables (CPTs) (Section 2.3.2). Finally, the values for each input parameter were computed using the data associated with the study site to produce a probability distribution for each state associated with the nodes.

The determination of states for the remaining nodes and directed links was accomplished based on scientific knowledge (the literature), the Quebec regulations, and the availability of data in provincial and federal databases. The choice of nodes was discussed with experts and stakeholders to provide a suitable BN, rectify, refine, and fill the potential gaps associated with the variables and their status.

2.2.2 Filling the CPTs. Filling the CPTs of the Bayesian network is the most challenging step in setting up the framework. To accomplish this step, different known methods were tested: (i) expert judgment [38, 58, 59], (ii) the interpolation method [88], and (iii) the Cain method [89, 90]. Methods that did not generate reliable outcomes in terms of probabilities were rejected under the expert decision.

It was challenging to estimate the quantity of chemical and microbiological contaminants released during agricultural production activities, as well as their diffusion, dispersion, and advection into the DWIs. Thus, employing expert judgment and scientific knowledge from the literature were the best methods for determining the conditional distribution of the quantities of pollutants that may be released in the immediate and intermediate protection zones. The exposure to pollutants released by upstream agricultural activities was also assessed using this method. In these situations, completing the CPTs was based on experts’ judgments and their interpretation by the extensive research team (rather than through formal interviews). This is due to the nature of the project: experts’ judgments were interpreted throughout four years of work on a collaborative project between six municipal partners, a watershed organization, and a research team at Polytechnique Montreal. Thus, the probabilities for each status of each child node were elucidated through discussion meetings with researchers, professionals, and municipal water stakeholders and were enhanced by scientific knowledge.
Table 1. Summary of input variables' properties.

<table>
<thead>
<tr>
<th>Parent node</th>
<th>Name</th>
<th>Description</th>
<th>Type of discretization</th>
<th>Risk component</th>
<th>State discretization methodology</th>
<th>States (from lowest to highest level)</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DBA</td>
<td>Drainage basin (DB) area</td>
<td>Continuous</td>
<td>Hazard</td>
<td>Statutes were defined based on percentiles of the area of 4723 sub-watersheds (level 2) in Quebec.</td>
<td>0 to 5 km$^2$ / 5 to 12 km$^2$ / 12 to 45 km$^2$ / $&gt; = 45$ km$^2$</td>
<td>Quebec portal of knowledge on water [68]</td>
</tr>
<tr>
<td></td>
<td>PPCO</td>
<td>Plant production cover</td>
<td>Continuous</td>
<td>Hazard</td>
<td>Statutes were defined based on Water Withdrawal and Protection Regulation (WWPR) (schedule IV) [54]; land use-based method for assessing water vulnerability.</td>
<td>0 to 20% / 20 to 50% / $&gt; = 50$%</td>
<td>Mapping of land use in the St. Lawrence Lowlands [69]</td>
</tr>
<tr>
<td></td>
<td>APC</td>
<td>Animal production cover</td>
<td>Continuous</td>
<td>Hazard</td>
<td>[54]: land use-based method for assessing water vulnerability.</td>
<td>0 to 20% / 20 to 50% / $&gt; = 50$%</td>
<td>Data are provided by the Ministère des Affaires municipales et de l'Habitation de Quebec (MAMH).</td>
</tr>
<tr>
<td></td>
<td>CPC</td>
<td>Crop production classes</td>
<td>Discrete</td>
<td>Hazard</td>
<td>Environmental impacts depend on the crop class. The statutes were specified based on the data layer available in the ministry database (Ministère de l'environnement et de la lutte contre les changements climatiques (MELCC)), and their classification was made based on the following references: Beaudin et al. [70], Michaud et al. [71], Gasser et al. [72], Giroux [73], and Giroux [74].</td>
<td>No crop production / Non-cultivated agricultural area / Undefined culture / Specialized culture / Perennial culture / Annual crop</td>
<td>Quebec land use [75]</td>
</tr>
<tr>
<td></td>
<td>APC</td>
<td>Animal production classes</td>
<td>Discrete</td>
<td>Hazard</td>
<td>Animal manure is a source of microbial contamination for surface waters and contains pathogens. These pathogens are specific to the types of animals from which manure is derived. The statutes were specified based on the data layer available in the MAMH database, and their classification was made based on the following references: Soupîr et al. [76], Jaffrezic et al. [77], Lalancette et al. [78], and Haack et al. [79].</td>
<td>No animal production / Class VII (2) / Classes V (3) - VI (4) / Classes II (5) - III (6) / IV (7) / Class I (8)</td>
<td>Data are provided by MAMH.</td>
</tr>
<tr>
<td></td>
<td>PZ</td>
<td>Protection zone</td>
<td>Discrete</td>
<td>Exposure</td>
<td>Statutes were defined based on WWPR (Art. 70, 72, and 74) [54]; protection zone delineation.</td>
<td>Extended protection zone / Intermediate protection zone / Immediate protection zone</td>
<td>Data are provided by the watershed organization.</td>
</tr>
<tr>
<td></td>
<td>GMEC_X (1)</td>
<td>Geometric mean of E. coli concentrations</td>
<td>Continuous</td>
<td>Vulnerability</td>
<td>Statutes were defined based on the regulation of the drinking water quality [80], and the development of an index of bacteriological and physicochemical quality (IQBP) [81].</td>
<td>0 to 15 CFU/100 ml / 15 to 150 CFU/100 ml / $&gt; = 150$ CFU/100 ml</td>
<td>Raw water quality data are provided by drinking water treatment plants managers.</td>
</tr>
<tr>
<td></td>
<td>PKEC_X (1)</td>
<td>99th percentile of E. coli concentrations</td>
<td>Continuous</td>
<td>Vulnerability</td>
<td>Statutes were defined based on the regulation of the drinking water quality [80], and the development of an index of bacteriological and physicochemical quality (IQBP) [81].</td>
<td>0 to 200 CFU/100 ml / 200 to 2000 CFU/100 ml / $&gt; = 2000$ CFU/100 ml</td>
<td>Raw water quality data are provided by drinking water treatment plants managers.</td>
</tr>
<tr>
<td></td>
<td>SVMC_X (1)</td>
<td>Seasonal variability of microbial contamination</td>
<td>Continuous</td>
<td>Vulnerability</td>
<td>Statutes were defined based on the results of the index of seasonal variability to microbial contamination developed by Kammoun et al. [65].</td>
<td>0 to 1 / 1 to 1.5 / $&gt; = 1.5$</td>
<td>Appendix (Continued)</td>
</tr>
<tr>
<td></td>
<td>VWS</td>
<td>Vulnerability to water scarcity</td>
<td>Discrete</td>
<td>Vulnerability</td>
<td>Statutes were defined based on the water scarcity vulnerability index developed by Leveque [67].</td>
<td>Low / Medium / High</td>
<td>Leveque [67] study</td>
</tr>
</tbody>
</table>

(Continued)
However, it is challenging to ensure that the definition of probability distributions of several tables is coherent with each other rather than using simply the judgment of experts. This is mostly because experts lack the rigid regularity of machines [88, 91]. During a long process, boredom and fatigue are enough to introduce errors, uncertainties, and a lack of standardization in the distributions.

One solution would be to standardize the method of obtaining probabilistic information from the experts by providing a much smaller selection of probability distributions. Therefore, the interpolation method was used to obtain the remaining probability distributions in a reasonable time [88]. This method requires assigning a weight to each parent node to quantify the relative strength of its influence over the child node. This step was done based on expert judgment. The sum of the weights must equal one. Then, conditional probability distribution (CPD) elicitation was done for three cases: when all parents nodes are in extreme states (high and low states) and medium states. The determination of the probability distribution for the remaining combinations of the various parents’ states was done by interpolation. This method was allied to all CPTs of the Bayesian network. However, it was deemed to be not adequate for our study because an examination of the probability distributions revealed that the resulting probabilities were inconsistent. It was difficult to maintain consistency between the probability distributions, especially in cases where a parent node’s state was highly different from the one used for interpolation. In this instance, this method does not account for the influence of variation bias in the interpolation. Therefore, the interpolation method was not pursued further.

The Cain method was also tested. It involves determining CPDs for cases where the relationships between the parent nodes are known; these obtained conditional probability distributions are termed anchor CPDs. Anchors were defined according to the number of parents and the number of statuses associated with each parent. The distribution of missing conditional probabilities was derived by interpolating those corresponding to the anchor CPDs and using interpolation factors. The latter were calculated using the CPDs at anchors. Interpolation

Table 1. (Continued)

<table>
<thead>
<tr>
<th>Parent node</th>
<th>Risk component</th>
<th>Type of discretization</th>
<th>State discretization methodology</th>
<th>States (from lowest to highest level)</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKTP 99th percentile of total phosphorus concentrations</td>
<td>Vulnerability</td>
<td>Continuous</td>
<td>Statutes were defined based on the WWPR [54], and the development of an index of bacteriological and physicochemical quality (IQBP) [91].</td>
<td>0 to 50 μg/l / 50 to 100 μg/l / &gt; = 100 μg/l</td>
<td>Raw water quality data are provided by drinking water treatment plants managers.</td>
</tr>
<tr>
<td>MTP Median of total phosphorus concentrations</td>
<td>Vulnerability</td>
<td>Continuous</td>
<td>Statutes were defined on the basis of the WWPR [54].</td>
<td>0 to 30 μg/l / 30 to 50 μg/l / &gt; = 50 μg/l</td>
<td>Drinking water intakes vulnerability assessment reports [82-87]</td>
</tr>
<tr>
<td>AIMPZ Area of immediate protection zone</td>
<td>Vulnerability</td>
<td>Continuous</td>
<td>Statutes were defined on the basis of the WWPR [54].</td>
<td>0 to 1km² / 1 to 5 km²</td>
<td></td>
</tr>
<tr>
<td>AINPZ Area of intermediate protection zone</td>
<td>Vulnerability</td>
<td>Continuous</td>
<td>Statutes were defined on the basis of the WWPR [54].</td>
<td>0 to 10 km² / 10 to 30 km² / &gt; = 30 km²</td>
<td></td>
</tr>
<tr>
<td>PIMPA / PINPA Percentage of the immediate/ intermediate protection zone used for activities that may release chemical contaminants</td>
<td>Vulnerability</td>
<td>Discrete</td>
<td>No / 0 to 20% / 20 to 50% / &gt; = 50%</td>
<td>Land use data are provided by MAMH.</td>
<td></td>
</tr>
</tbody>
</table>

(1) X refers to season: Winter/ Spring/ Summer/ Autumn
(2) Class VII: other types of animal production/ experimental farm/ other agricultural activities; (3) Class VI: apiculture; (4) Class V: poultry and egg production; (5) Class IV: sheep and goat farming; (6) Class III: pig farming; (7) Class II: equine farming; (8) Class I: beef farming/ dairy cattle/ grazing land and pasture

https://doi.org/10.1371/journal.pwat.0000073.t001
factors of a parent node were calculated for each change from one state to another to quantify the positive to negative state reduction. The processes of defining anchors and calculating interpolation factors are explained in more details in Cain [89] and Mkrtchyan et al. [90].

Only two CPTs, relating to crop and animal activity levels, were completed using the Cain approach because this method is applicable for CPTs in which the effect of the parent state on the child state has a constant direction: the increase in the area occupied by animal and crop production activities results in an increase in the level of agricultural activities in the drainage basin. The results of filling in the probability tables for the other child nodes were judged to be unacceptable since the interpolation is dependent on the number of favourable and unfavourable states for the child state of interest. As a result, there will be the same probability distributions for different child nodes even though the parent nodes have varying degrees of influence on them [92].

Therefore, a novel approach to filling CPTs was developed that combines a mixed aggregation method [93] with expert judgment. The mixed aggregation was used to aggregate indicators in the ecological assessments field using a deterministic approach [93]. It was adapted to be applied in the case of a probabilistic approach and for water resource assessment. First, the different state of each parent node \( P_n \) was standardized to a common scale from 0 (best possible state) to 1 (worst possible state). Second, the assignment of weights \( W_n \) for each parent node \( P_n \) was performed to reflect its importance relative to the child node. Third, causal dependencies were performed with aggregation. In this step, additive and maximal aggregation were combined (Eqs (1), (2), (3), (4), (5)). Additive aggregation \( f_{add} \) compensates for the poor status of a parent node by other nodes with a good status. On the other hand, maximal aggregation \( f_{max} \) leads to pessimistic results since only the parent node with the worst status is considered. The combination of the additive and maximal aggregation allows for some compensation while penalizing the very poor state to not neglect an impact that requires more attention to protect water resources.

\[
f_{add-max}(P_1, P_2, \ldots, P_n) = \alpha f_{add}(P_1, P_2, \ldots, P_n) + (1 - \alpha) f_{max}(P_1, P_2, \ldots, P_n)
\]  

Where:

\[
f_{add}(P_1, P_2, \ldots, P_n) = \sum_{i=1}^{n} W_i P_i = W_1 P_1 + W_2 P_2 + \cdots + W_n P_n \tag{2}
\]

\[
\sum_{i=1}^{n} W_i = 1 \tag{3}
\]

\[
f_{max}(P_1, P_2, \ldots, P_n) = \max(P_1, P_2, \ldots, P_n) \tag{4}
\]

\[
\alpha + (1 - \alpha) = 1 \tag{5}
\]

\( \alpha \) and \((1-\alpha)\) were chosen according to the relative importance to be given to each aggregation method. The more alpha tends to 0, the more the maximum aggregation is emphasized. Otherwise, additive aggregation tends to be prioritized.

The most favorable weight \( W_n \) set of parent nodes and aggregation method was determined by performing a calibration that requires knowledge and experience and by building on previous work [82–87] conducted to assess the potential risk of agricultural activities on water resources using a deterministic approach. This step was performed by analyzing the outputs resulting from each possible combination of \( \alpha \) and \( W_n \). After validating the most appropriate variables and depending on the number of parent nodes, a 2D or 3D graph was generated to extract the CPDs. These data were used to fill the CPTs. The weights selection and graph
The generation process was codified in R (V.1.1.456, RStudio, Inc.). S1 Table presents an example of CPT associated with the “Global Risk Level (GRL)” node that was determined by combining three parent nodes: probability of hazard (PH), exposure (EXP), and DWI vulnerability level (DWIVL). All probability distributions were calculated based on the 3D graph (S2 Fig) generated by combining all possible cases related to the parent nodes.

Table 2 shows the methods applied to fill 18 CPTs. This table also lists the methods that were tested out but were rejected because of unreliable outcomes as explained above.

2.2.3 Sensitivity analysis. A sensitivity analysis was performed to quantify the magnitude of the sensitivity of the changes that occur in the endpoint probabilities when the parent nodes are changed and to identify which of the variables are most relevant [40, 60]. This analysis is necessary to understand where more information is needed to reduce uncertainty.

Table 2. Overview of the CPTs filling methods.

<table>
<thead>
<tr>
<th>Child node</th>
<th>Description</th>
<th>Parent node</th>
<th>Weight of parent node</th>
<th>α</th>
<th>Nominated CPTs filling method</th>
<th>Tested and rejected CPTs filling methods</th>
</tr>
</thead>
</table>
| CAL        | Crop activities level | Hazard | PPCO DBA | NA(*) | NA | Cain method | • Expert judgment  
• Interpolation method  
• Mixed aggregation method combined to expert judgment |
| AAL        | Animal activities level | Hazard | APCO DBA | NA(*) | NA | Expert judgment | • Cain method  
• Interpolation method  
• Mixed aggregation method combined to expert judgment |
| CRIM       | Contaminants that could be released in the intermediate protection zone | Vulnerability | PINPA AIMPZ | NA(*) | NA | Expert judgment | • Cain method  
• Interpolation method  
• Mixed aggregation method combined to expert judgment |
| CRIN       | Contaminants that could be released in the intermediate protection zone | Vulnerability | PINPA AIMPZ | NA(*) | NA | Expert judgment | • Cain method  
• Interpolation method  
• Mixed aggregation method combined to expert judgment |
| EXP        | Exposure | Exposure | PZ | 0.9 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| CAH        | Crops activities hazard | Hazard | CAL CPC | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| APH        | Animal production hazard | Hazard | AAL APC | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| PH         | Probability of hazard | Hazard | APH CAH | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| MQRW_(x)(**) | Microbiological quality of raw water | Vulnerability | PKEC_(x) GMEC_(x) SVMC_(x) | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| VMC        | Vulnerability to microbiological contamination | Vulnerability | MQRW_W MQRW_SP MQRW_S MQRW_A | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| RWQLTP     | Raw water quality level for total phosphorus | Vulnerability | PKTP MTP | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| CVLU       | Chemical vulnerability level based on land use | Vulnerability | CRIM CRIN | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| VCC        | Vulnerability to chemical contamination | Vulnerability | RWQLTP CVLU | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| DWIVL      | DWI vulnerability level | Vulnerability | VMC VWS VCC | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |
| GRL        | Global risk level | Risk | PH DWIVL EXP | 0.1 | 1 | Mixed aggregation method combined to expert judgment | • Interpolation method  
• Mixed aggregation method |

(*) Not applicable  
(**) x refers to season: winter/ Spring/ Summer/ Autumn  
https://doi.org/10.1371/journal.pwat.0000073.t002
This sensitivity analysis was executed by calculating the mutual information ($I$) between two discrete variables $X$ and $Y$ [94]:

$$I(X, Y) = H(X) - H(X|Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

(6)

where the sum is over all states $x$ and $y$ of variables $X$ and $Y$, respectively, $H(X)$ is the entropy reduction of $X$ before any new finding, $H(X|Y)$ is the conditional entropy of $X$ at given $Y$. In the case of computing the $I$ between continuous random variables, the double sum has to be replaced by a double integral.

Entropy reduction refers to how much a target node’s entropy is decreased when the values of its parent nodes are held constant [95–97]. The amount of this entropy reduction is used to gauge how much the parent nodes, also known as findings nodes, influence the target node. Entropy is determined using the probabilities given to the potential states of the target node. It is a measure of uncertainty or randomness associated with the distribution of probabilities in a system. According to this theory, the target node is thought to be more affected by nodes with higher entropy reduction than nodes with lower entropy reduction [95–97].

These measures of entropy reduction and mutual information between various variables and a given endpoint node were conducted with the “sensitivity to finding” command in the Netica software.

### 3. Results

#### 3.1 Bayesian network model

The developed BN model integrates 43 nodes (Fig 3) and 1326 conditional probabilities. The nodes constituting this BN are described in Tables 1 and 2. The probabilities associated with
each node whose sum is equal to 100% are shown on the right side of the status window of the detailed BN (S3 Fig). Grey nodes represent the input variables.

### 3.2 DWIs vulnerability

The findings related to the vulnerability nodes, which are a component of risk, are presented in this section. As illustrated in Fig 4, more than 60% of DWIs were extremely vulnerable. This is most likely because these intakes are in a highly urbanised watershed with agricultural activity (section 2.1). The vulnerability of source water to microbiological contamination was noted at the four intakes. The seasonal variation in microbial contamination of raw water was investigated (S4 Fig), and our results revealed that all intakes were associated with medium to high probability distributions for all seasons. The probabilities describing poor water quality could be due to the transport of microbial contaminants during the late winter and early spring snowmelt period, but also during episodes of heavy rainfall in the summer, and autumn as demonstrated by Sylvestre et al. [29].

DWI_3 and DWI_4 were highly vulnerable to chemical contamination (82.3% and 72.2% probability in the high state, respectively) (Fig 4) as a result of total phosphorus leading to lower quality of the raw water. In addition, a significant area of the intermediate zone of DWI_3 was occupied by anthropogenic activities that may release chemical contaminants (S5 Fig). DWI_1 and DWI_2 were moderately vulnerable to chemical contamination (55.2% and 65% probability in the medium state, respectively). They were also influenced by the land use upstream of the intakes, as illustrated in S5 Fig.

### 3.3 Probability of the hazard

The findings related to the hazard nodes, which represent another risk component, are presented in this section (Eq 1). The protection area and drainage basin delineation results showed that the study site includes 25 DBs that discharge within the boundaries of the
intermediate protection zones of the four DWIs. The immediate protection areas of all intakes and seven drainage basins (DB_1, DB_3, DB_5, DB_16, DB_17, DB_21, and DB_22) were not occupied by agricultural activities, so they do not constitute a risk from agriculture to water quality at DWIs. The remaining drainage basins and their occupancy by animal and crop production activities are shown in the S6–S14 Figs. The minimum areas of crop and animal production activities were 0.08 km$^2$ (in DB_20) and 0.03 km$^2$ (in DB_12 and DB_18), respectively, while the DB_8 has the greatest area covered by these activities. The crop activities were spread over 3488.26 km$^2$, and the animal activities were occupied by 3177.22 km$^2$ of this DB. The total area occupying all the DBs upstream of all the DWIs were 8278 km$^2$ in crop production activities and 7830 km$^2$ in animal activities.

Fig 5 shows the probability distributions of the overall hazard (PH), the crop activity hazard (CAH), and the animal production hazard (APH) for each DB. The probability of hazard was at high level in DB_8, DB_10 and DB_23 (81.7%, 96.6% and 81.6% probability in the high state, respectively). This outcome is due to the area occupied by agricultural activities and the type of production that presents the greatest hazard to water quality: class I for animal production and annual crop for vegetable production. The probability of hazard was moderate in DB_6, where 51.1% was in the medium state. For DB_2, DB_7, DB_15, and DB_24, the probability of hazard was ranged between low and medium levels with close percentages (for example, 48.7% and 49.6% probability in low and medium states, respectively for DB_2). The remaining DBs were all characterised by low hazard probabilities level, which averaged 63.6%.

### 3.4 Global risk

This section reports the global risk results, which drive from the combination of hazard, vulnerability, and exposure. The overall risk from crop and animal activities was relatively high in three drainage basins (DB_8, DB_10, DB_23) whose outlets were located upstream of these DWIs: DWI_2, DWI_3, and DWI_4 (Fig 6). Their probabilities of being in high status vary between 81.9% and 95.5%. Another five DBs (DB_2, DB_6, DB_7, DB_15, and DB_24) represented a moderate overall risk with probabilities ranging from 55.6% to 60.9% (Fig 6). Eight DBs (DB_4, DB_11, DB_12, DB_13, DB_14, DB_18, DB_19, and DB_25) were associated with
a low global risk with probabilities that are almost equal between the low and medium status, and with a variance in probabilities that does not exceed 10%. DB_9 and DB_20 were associated with low global risks with probabilities ranging from 54.8 to 67.4% in the low status.

The crops activities hazard (CAH), animal production hazard (APH), exposure (EXP), DWI vulnerability level (DWIVL), vulnerability to microbiological contamination (VMC), and vulnerability to chemical contamination (VCC) are the nodes of interest in this study, so they were selected as findings nodes. Subsequently, the sensitivity for the endpoint "Global Risk Level (GRL)" in every DB occupied by agricultural activities was calculated and analyzed as explained in section 2.3.3. Results of this analysis (Fig 7) showed that the developed model was most sensitive to the crops activities hazard (CAH) (from 17.8 to 29.5% variance

Fig 6. Global risk distribution for all drainage basins upstream four DWIs.

https://doi.org/10.1371/journal.pwat.0000073.g006

Fig 7. Results of sensitivity of the « Global risk level » variable.

https://doi.org/10.1371/journal.pwat.0000073.g007
reduction) and the animal production hazard (APH) (from 6.54 to 25.6%) for all DWIs. This model was also moderately influenced by the exposure node (from 2.5 to 4.13%), especially for the cases of the drainage basins that were slightly occupied by a single type of production activity, either crop or animal, such as BD_9 and BD_20 that are located upstream DWI_2, DWI_3 and DWI_4.

4. Discussion

The probability of DWIs contamination by agricultural activities was assessed using a probabilistic approach in this study. The proposed Bayesian model generates a probabilistic estimation of the DWIs vulnerability level, crop and animal production activities hazard, the exposure, and the overall risk associated with each drainage basin occupied by agricultural activities.

One of the strengths of this model is the ability to embed different types of data in the same system. For example, to fill the input variables, water quality and land use data were used, and the outputs of other studies (vulnerability to water scarcity) can be included. The discretization of all nodes constituting this model allows for more efficient and effective communication with stakeholders on the components of the risk assessment, the interpretation of risks, and the preparation of action plans, since the terminology and statuses of nodes are based on regulatory criteria and lexicons already used in the provincial databases. The graphical model can be used as a communication tool between scientists, municipal officials, government professionals, and members of a watershed organization. It could also be a simple approach for farmer organizations that support surface water pollution reduction to convey potential risks to their members. The acyclic graph facilitates understanding and collaboration among multidisciplinary members, thus increasing the impact in the decision-making process, as discussed in Stritih et al. [98] and Laurila-Pant et al. [99].

In testing several methods to fill CPTs, it was concluded that the chosen methodology can significantly influence the final output of the risk assessment. The additive aggregation was applied through the mixed aggregation by choosing $\alpha = 1$ to fill CPTs related to child nodes of crops activities hazard (CAH), animal production hazard (APH), probability of hazard (PH), and global risk level (GRL). This method is suitable for cases that do not require emphasizing the pessimism bias. For instance, a high level of crop activity (CAL node) may not represent a high hazard (CAH node) in the case where most of the area is covered (CPC node) by undefined crops or non-cultivated land. In another case, the worst-case bias was emphasized by setting $\alpha$ to 0, which applied to the case of assessing vulnerability to chemical contaminants based on land use (CVLU node) in immediate (CRIM node) and intermediate protection zone (CRIN node). This means that the maximum aggregation is applied to penalize the worst of inputs to protect the water resource against the highest hazard to water quality. However, for the rest of the cases, combining both types of aggregation (additive and maximal) was done due to the evaluation of complementary variables, such as the case of chemical and microbiological water quality assessment based on raw water quality analysis. The mixed aggregation method enables the inclusion of inter-correlation between parent nodes, resulting in a risk assessment that is more unbiased and objective compared to other methods. Our case study also revealed that the weight assigned to each parent node might reflect its relative significance and reliance on its respective child node.

As shown in Fig 6 and maps of agricultural activities (from S6-S14 Figs), medium level risk probability dominates for DBs that were slightly occupied by agricultural activities. This was due to the high DWIs vulnerability, and the location of the outlets of the DBs at a distance that does not exceed 10 km upstream from these intakes. This implies that even though the hazard probability is not so significant, the vulnerability of the intake and the exposure of the threat.
could be the factors contributing to a higher risk level (medium to high). Therefore, all nodes in the model are essential for assessing the risk of agricultural activities. As a result, across a basic network graph, this Bayesian model gives a clearer understanding of the relative relationship and interaction between all variables \[100, 101\]. As anticipated, results proved that the BN could be applied to water resource risk assessment with the same efficiency as environmental risk assessment \[38\]. The developed model provides a transparent and systematic way to characterize risk through uncertainty. It reveals all risk states with their estimated probabilities.

The sensitivity analysis results showed that, in order to improve the global risk estimation, it is most impactful to focus more on the CAH, APH, and EXP since they are linked to the highest entropy reduction. For this purpose, the use of data regarding pesticides (types, quantities, method of application and frequency), animal production, and characteristics of the receiving environment (soil type, hydrodynamics of the river, etc.) are beneficial to have a more accurate assessment. These data are not available in provincial databases. To bridge this data gap, targeted analysis are thus needed. Such models are suitable for determining the catchment areas that could potentially result in the highest risk to water intakes. This simple model is handy for authorities and stakeholders to estimate the level of risk on a regional and provincial scale with commonly available data. Therefore, this model is considered a jump-start for stakeholders to channel their efforts and investments properly to the most appropriate target for contamination mitigation efforts. This risk assessment model may benefit from an upgrade over time by including variables that define the modeling of microbial pathogens and chemical contaminants after rain discharge. As shown by Haack et al. \[79\], Soupir et al. \[76\], and Jaffrezic et al. \[77\], water bodies were most sensitive to livestock-derived fecal and chemical marker inputs after rainfall, and bacteria concentrations frequently exceed standards for primary contact.

Discussion of the risk analysis outcomes with stakeholders requires recognizing uncertainties resulting from the lack of data, inaccuracies in laboratory measurements for raw water quality monitoring, gaps in the knowledge, and inexactness of geographic data. Nevertheless, as Uusitalo \[39\] pointed out, the advantage of the BN is that it incorporates uncertainty explicitly and naturally through the probability distribution.

As expected, this model has a limitation concerning its validation since it is not readily possible to perform a quantitative validation. This limitation is in accordance with the findings discussed in reviews by Kaikkonen et al. \[38\] and Phan et al. \[51\], which also exhibited the same issue. For this reason, in this study, a qualitative model validation was performed by examining the outcomes of many scenarios in comparison to the regulatory vulnerability analyses. Results were consistent; however, the BN approach provides more context for risk levels and an approach to provide an aggregated and global assessment of DWI vulnerability.

This model is a potential tool for source water managers to easily apply since it is based on readily available input data and raw water analyses commonly required by drinking water quality regulations \[80\], which specify standards for all water systems designated for human use throughout the province of Quebec. Consequently, it is simple to apply this Bayesian model throughout Quebec by modifying the input values in the model to generate probabilistic results regarding the risk posed by agricultural activities upstream of DWIs. In this case, all model components, including parent and child nodes and their probabilistic connections, are not subject to modification. This Bayesian model can also be applied to other study sites on a larger scale (national and international scale) if all the data defining the input nodes are available. As mentioned by Dorner et al. \[50\], the Bayesian network has the advantage of being used rapidly to investigate outputs under various scenarios. This evidence holds when applying the same Bayesian model that has already been developed. It is crucial to highlight that the groundwork for developing this model, including the generation of CPTs, took a substantial
amount of time and effort. If the necessary data is unavailable for other study sites, adjustments must be made to the model design based on the existing data. As part of these adjustments, it is necessary to decide on the appropriate nodes to use as input to the model while considering the targeted region’s regulations. This decision should be based on the data available, especially those that describe the agricultural activities inventoried at the research site and the quality of surface water at the DWI. The adjustments would also include defining the conditional probability distributions that characterize the probabilistic links of the new nodes added with their child nodes, despite applying the identical CPTs filling approaches described in section 2.2.2. At this scale, the involvement of experts is a must for the qualitative validation of the model.

5. Conclusions

This study used the Bayesian network (BN) to assess the risk of agricultural activities for four DWIs that draw their water across a river in the south of Quebec, Canada. The findings have led us to conclude that:

- The Bayesian model allowed for the combination and manipulation of data in various formats (numerical, statistical, and geographical data) into a consistent methodology for water resource risk assessment for drinking water supplies.

- The interactions between all risk components (the vulnerability of drinking water intake, the exposure to the threat, and the probability of hazard) and their combined uncertainties can be quantified in a single model. Stakeholders and agricultural organizations may utilize this probabilistic reasoning tool to grasp the causal relationships between variables, allowing them to identify the key factors that demand special attention due to their large impacts on the outcomes.

- The end-user can fill data gaps in the model by referencing expert knowledge. The explicit identification of data gaps and sensitivity analyses facilitates the prioritization of future data collection needs. Furthermore, the model can serve as a basis for integrating local knowledge from various stakeholders in the watershed.

- The mixed aggregation method \( f_{\text{add-max}} \) combined with expert judgment is a simple and straightforward way for populating conditional probability tables, and it is suitable for considering parent-child dependencies. The correlation between the parent nodes is also considered, making the evaluation accurate based on expert judgment. The weights allocated to the additive \( f_{\text{add}} \) and maximum \( f_{\text{max}} \) aggregation allow for a more focused examination of the influence factor in the field of water resource protection. This is a crucial characteristic when the effect of parent nodes functioning concurrently is considered. This method is useful for filling CPTs semi-automatically to help experts calculate reasonable probabilities when data are not available.

- The model developed in this paper is the first element of a risk assessment cycle. This Bayesian model proved successful in identifying drainage basins occupied by agricultural activities that may pose a medium to high risk of contamination of drinking water intakes. These findings can assist managers and authorities in developing targeted sampling programs for microbiological and chemical testing of surface water while reducing costs and prioritizing drainage basins associated with a high level of risk. The analysis results should constitute the input variables of this Bayesian network while incorporating new nodes and updating it to better pinpoint the source of risk.
The availability of the data describing the input nodes determines how easily the model may be exported to different water resource risk assessment scenarios. If the data are readily available and suitable for the Bayesian model, the established CPTs in this study can be applied without adjustments. Nevertheless, developing new CPTs would require a lot of time and effort if a new Bayesian network had to be constructed to accommodate other types of input nodes. It is at least plausible to use the same approach of generating conditional probabilities outlined in this study, as it is relevant to the field of water resource protection.

From a water resource protection perspective, in an urban and agricultural watershed, this research is expected to serve as a basis for future studies on assessing microbial risk from combined sewer overflows (CSOs) using a probabilistic approach.

Supporting information

S1 Text. Protection zone delineation.
(DOCX)

S1 Table. Conditional probability distribution.
(DOCX)

S1 Fig. Example of drainage basin (DB) which have outlet located within the boundaries of intermediate protection zone of a drinking water intake.
(DOCX)

S2 Fig. 3D graph of global risk response under the combination of the probability of hazard, exposure, and drinking water intake (DWI) vulnerability level.
(DOCX)

S3 Fig. Detailed Bayesian network structure for assessing drinking water intake contamination risk from agricultural activities.
(DOCX)

S4 Fig. Results of BN submodel assessing the microbiological quality of raw water for all drinking water intakes (DWIs) overall seasons of the year.
(DOCX)

S5 Fig. Results of BN submodel assessing raw water quality for total phosphorus and the chemical vulnerability level based on land use for all drinking water intakes (DWIs).
(DOCX)

S6 Fig. Maps of agricultural activities within the drainage basins (DB_2 and DB_4), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.
(DOCX)

S7 Fig. Maps of agricultural activities within the drainage basins (DB_6 and DB_7), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.
(DOCX)

S8 Fig. Maps of agricultural activities within the drainage basins (DB_8 and DB_9), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.
(DOCX)
S9 Fig. Maps of agricultural activities within the drainage basins (DB_10 and DB_11), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

S10 Fig. Maps of agricultural activities within the drainage basins (DB_12 and DB_13), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

S11 Fig. Maps of agricultural activities within the drainage basins (DB_14 and DB_15), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

S12 Fig. Maps of agricultural activities within the drainage basins (DB_18 and DB_19), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

S13 Fig. Maps of agricultural activities within the drainage basins (DB_20 and DB_23), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

S14 Fig. Maps of agricultural activities within the drainage basins (DB_24 and DB_25), as well as the probability distribution of the overall hazard, the crop activity hazard, and the animal production hazard.

(DOCX)

Acknowledgments

We thank the partner municipalities and the watershed organization team responsible for the study area for contributing to the data collection and vulnerability assessments.

Author Contributions

Conceptualization: Raja Kammoun, Françoise Bichai.

Data curation: Raja Kammoun.

Formal analysis: Raja Kammoun.

Funding acquisition: Michèle Prévost, Sarah Dorner.

Investigation: Raja Kammoun, Sarah Dorner.

Methodology: Raja Kammoun, Sarah Dorner.

Project administration: Michèle Prévost, Françoise Bichai, Sarah Dorner.

Resources: Michèle Prévost, Françoise Bichai, Sarah Dorner.

Software: Raja Kammoun.

Supervision: Françoise Bichai, Sarah Dorner.
Validation: Raja Kammoun, Natasha McQuaid, Michèle Prévost, Françoise Bichai, Sarah Dorner.

Visualization: Raja Kammoun, Natasha McQuaid.

Writing – original draft: Raja Kammoun.

Writing – review & editing: Raja Kammoun, Natasha McQuaid, Vincent Lessard, Eyeresaalem Adhanom Goitom, Michèle Prévost, Françoise Bichai, Sarah Dorner.

References


