# EpiBasket: how e-commerce tools can improve epidemiological preparedness.

Weijia Xing, Gilles Hejblum, Alain-Jacques Valleron

▶ **To cite this version:**

Weijia Xing, Gilles Hejblum, Alain-Jacques Valleron. EpiBasket: how e-commerce tools can improve epidemiological preparedness.. Emerg Health Threats J, 2013, 6, pp.19748. inserm-00923828

HAL Id: inserm-00923828
https://inserm.hal.science/inserm-00923828

Submitted on 5 Jan 2014

# EpiBasket: how e-commerce tools can improve epidemiological preparedness

## Weijia Xing[1,2], Gilles Hejblum[1,3] and Alain-Jacques Valleron[1,3]*

[1]Institut National de la Santé et de la Recherche Médicale, Paris, France; [2]Division of Infectious Disease, Key Laboratory of Surveillance and Early-warning on Infectious Disease, Chinese Center for Disease Control and Prevention, Beijing, China; [3]Université Pierre et Marie Curie-Paris 6, UMR-S 707, Paris, France

*Background*: Should an emerging infectious disease outbreak or an environmental disaster occur, the collection of epidemiological data must start as soon as possible after the event's onset. Questionnaires are usually built *de novo* for each event, resulting in substantially delayed epidemiological responses that are detrimental to the understanding and control of the event considered. Moreover, the public health and/or academic institution databases constructed with responses to different questionnaires are usually difficult to merge, impairing necessary collaborations. We aimed to show that e-commerce concepts and software tools can be readily adapted to enable rapid collection of data after an infectious disease outbreak or environmental disaster. Here, the 'customers' are the epidemiologists, who fill their shopping 'baskets' with standardised questions.

*Methods*: For each epidemiological field, a catalogue of questions is constituted by identifying the relevant variables based on a review of the published literature on similar circumstances. Each question is tagged with information on its source papers. Epidemiologists can then tailor their own questionnaires by choosing appropriate questions from this catalogue. The software immediately provides them with ready-to-use forms and online questionnaires. All databases constituted by the different EpiBasket users are interoperable, because the corresponding questionnaires are derived from the same corpus of questions.

*Results*: A proof-of-concept prototype was developed for Knowledge, Attitudes and Practice (KAP) surveys, which is one of the fields of the epidemiological investigation frequently explored during, or after, an outbreak or environmental disaster. The catalogue of questions was initiated from a review of the KAP studies conducted during or after the 2003 severe acute respiratory syndrome epidemic.

*Conclusion*: Rapid collection of standardised data after an outbreak or environmental disaster can be facilitated by transposing the e-commerce paradigm to epidemiology, taking advantage of the powerful software tools already available.

Keywords: *data collection*; *disasters*; *disease outbreaks*; *emergency preparedness*; *epidemiology*; *information systems*; *questionnaires*

To access the supplementary material for this article, please see the Supplementary files under Article Tools online

When an outbreak of an emergent disease or an environmental disaster occurs, public health officials and researchers are expected to set up in real time the best possible epidemiological investigations to understand what happens, what the health consequences of the event are and how to mitigate them as quickly as possible (1).

The three conditions for this real-time epidemiological intelligence are real-time data collection, real-time methods of data analysis and real-time dissemination of results. The main obstacle of real-time epidemiology, addressed herein, is the first step in the process: data collection. Indeed, survey design and the elaboration of the corresponding questionnaires take time (2, 3): unsatisfactory questionnaires can be obtained quickly, but have to be modified a few days later; conversely, excellent questionnaires can be generated after lengthy meetings, but become available too late. Moreover, any important public health problem inherently involves a variety of actors in public health and research. All these players have their own agendas, hypotheses to test, stakeholders to comply with. Each of them designs a different survey,

**1**

with different questionnaires (4). Finally, as a consequence, the information collected in the different surveys usually cannot be merged to provide a unified database, because the definitions and questionnaires used are not standardised (5, 6).

The best solution to overcome the above-outlined difficulties would be that a respected international or national agency, or a scientific society, provides THE questionnaire that everybody would agree to use in the case of an epidemiological disaster. That 'top-down' best solution approach is obviously unrealistic. One of the reasons is that epidemiological investigation is a research activity. Researchers like to design their own tools, and good quality research needs a diversity of approaches.

Herein, we describe software that achieves a 'bottom-up' approach aimed at accelerating and standardising data collection, while allowing epidemiologists to tailor their own questionnaires. The basic concept underlying the software was to revisit with an epidemiologist's eye the consumer's e-commerce model used daily. Indeed, not so long ago, the process of getting same-day delivery to your door of the best product, after an informed choice about all those available, was unimaginable. Today, when a customer wants to buy, for example, a new home-entertainment audiovisual system, the internet quasi-instantaneously provides all the necessary information on all possible TVs, DVD players, hard disks, speakers, and so on from a variety of manufacturers. He can choose his desired elements from each of these categories, put them in a shopping basket, and review and modify his choices before making his final decision. Then, he checks out and his system can be delivered in hours. The success of e-commerce led to the development of powerful software, some of it free, that we used in this project.

We show herein that this e-commerce model and its software can be readily adapted to epidemiology. In the proposed EpiBasket concept, epidemiologists are the 'customers'. The 'categories of products' are the categories of questionnaire questions (e.g., the questions on the outcome of the household, life satisfaction, pre-validated scales for scoring stress). Information on which questions were used in the best high-profile papers published on similar events, information on all 'products' (i.e., questionnaire questions), will be available immediately. The epidemiologist devising his questionnaire will therefore be able to select from the 'catalogue' the best possible questions to meet his needs, put them in his 'shopping basket' and, once he has made up his mind, he 'buys' them and immediately gets his order, here the paper and/or the online version of the questionnaire. If the catalogue provides the best products on the 'epidemiological market', and if a community of researchers starts to collectively build and use this system, a positive byproduct should be the increased possibility of database

interoperability because, in the future, epidemiologists working on the same subject will choose the questions for their questionnaires from the same catalogue.

In this article, we describe the general outline of the 'EpiBasket' concept in three sections: the construction of the catalogue of questions by the developer, the user's point of view and the software that was used. The proposed concept is illustrated with a prototype based on the Knowledge, Attitudes, Practices (KAP) studies conducted on the 2003 severe acute respiratory syndrome (SARS) epidemic (7). This prototype, whose detailed characteristics are described in the Supplementary file, can be assayed at http://www.epibasket.org

## Methods

The methodology used to stock the catalogue with questions and to tag each question with information of potential usefulness for the users is described in this section. The user's point of view and the software used are addressed in Results section.

### Construction of the catalogue of questions

The final goal is to provide users with a 'catalogue' of the 'best' questions needed to explore a given epidemiological issue (e.g., 'the burnout of health workers during the outbreak' or 'the changes of attitudes towards immunisation in the general population during an outbreak'). A basic assumption underlying the EpiBasket project is that most of the essential questions that are needed today were posed in the past, and that those providing the results published in the best scientific journals should certainly be considered as the best possible candidate questions. Three steps were defined to stock the catalogue with questions.

The first was to identify the 'variables of interest' that must be acquired to investigate a given issue. This identification relied on a systematic review of the pertinent papers published in the domain. To facilitate the extraction of those variables, we supposed that they were mentioned in the tables and figures of the papers analysed in the systematic review. The manual extraction of those variables that we performed to build the prototype could subsequently be automated, with natural language-analysis tools (8) 'reading' the tables, figures and their corresponding legends. A relational database comprising all extracted variables was constituted. It allows tracking of the context of the variable, for example, to immediately find the sets of variables that are usually used together.

In a perfect world, the questionnaires used in each reviewed study would be made available to readers. In the real world, very few authors provide their questionnaires, hence the 'question' had to be inferred from the 'variable'. Therefore, the second stage was to deduce the 'question' from the 'variable' identified in the previous step.

For example, the questionnaire may have asked the date of birth, while the paper may report the corresponding results in a table showing age-class frequencies. In addition, EpiBasket can propose a recommendation for a preferred coding, with the aim of maximising the chances of interoperability among databases constituted with EpiBasket-derived questionnaires and their responses.

Third, the formulation of questions was generalised, when possible, to make them reusable in broader situations than those in the source papers. It was accomplished by replacing words or patterns found in the reviewed papers by 'root terms' that the user would replace with the specific term that fits the event of interest. For example, the question 'In the case of a SARS outbreak, would you avoid going to the cinema?' provides a template for asking questions regarding avoidance behaviours during an outbreak. Now, using 'root terms', it is formatted as 'In the case of #$DISEASE outbreak, would you avoid going to #$PLACE?', where #$DISEASE and #$PLACE are the root terms to be instantiated by the user selecting this question template, for example, 'In the case of a cholera outbreak, would you avoid going to a restaurant?' So far, six root terms have been implemented in the current EpiBasket prototype: '#$DISEASE', '#$DRUG', '#$NUMBER', '#$PERIOD', '#$PLACE' and '#$SOMEONE'. For example, the specific questions 'Did you wear a mask in a public place?', 'Did you wear a mask at work?' and 'Did you wear a mask during a flight?' were reframed as a single generic formulation, 'Did you wear a mask #$PLACE?'.

In addition to the questions extracted from the literature, a 'User suggestions' category was added to store user-proposed questions.

### Tagging of the questions
Each question in the catalogue has tags aimed at helping the user construct his questionnaire.

### General tags
Categories to which the question belongs, time of its use (before, during and/or after the epidemiological event considered), type(s) of population(s) studied (general population, patients, risk groups, healthcare workers), MeSH description of the term(s) used in the question.

### References in literature
List of the published papers analysed that used the question (full article references, links to PubMed and Google Scholar, journal impact factors [IF]).

### 'Value' of a question
Each question is also tagged with a 'value', the equivalent of a price in e-commerce. The underlying assumption is

that, when the user will have to choose among several questions (e.g., choosing among several scales available for scoring burnout), the one with the best 'value' will be retained. Since such a logical view is likely to be followed by most EpiBasket users, EpiBasket should facilitate better standardisation of questionnaires. At present, we arbitrarily decided to give to a question the value defined as the 'Best IF', that is, the highest journal IF encountered among all papers using it. This choice assumes that, if the user has to choose between two burnout scales, the one previously published in a high-profile journal will be preferred. As a consequence of this valuing system, all demographic variables (e.g., age, sex) get a high value (because they are used in all papers, including those published in journals with very high IF). IF can be systematically obtained from the Journal Citation Reports database, for which a fee must be paid. However, most journals provide their IF on their websites and we used those values in the prototype (see the Supplementary file). The value of 0.000 was arbitrarily assigned to questions published in journals not indexed in the Journal Citation Reports database. Researchers not wanting to take these IF into account can just ignore the 'values' indicated when they build their questionnaire.

### Keywords
Each question was associated with at least one keyword. These keywords are given on the EpiBasket homepage as a tag cloud (with term size proportional to term frequency). Whenever the user clicks on a given cloud term, all questions associated with it are listed.

### Time of study/population studied
Each paper from which the questions were extracted concerns a study conducted at a given time, either during the outbreak or after it, and a given population type. This information is available to the user to help him to identify the relevant questions for his own study.

### Related questions
Each EpiBasket question provides the list of the other questions jointly asked in the papers that used it. EpiBasket also provides the list of questions in the catalogue that share at least one of the keywords associated with question.

### Information on the questionnaire developers
To maximise the likelihood of compatibility among questionnaires, we paralleled the e-commerce approach, according to which shoppers know when they buy, for example, a book, what other books other customers also bought. Here, we implemented the function 'users who selected this question also selected ... <*names of the other questions*>'.

*Fig. 1.* Example of an EpiBasket window (see http://www.epibasket.org for a demonstration).
In this example, the user chose the subcategory 'Travel history' in the 'Personal history' category (left column). Five questions are proposed in this subcategory (middle of the window). If the user is also interested in the related question dealing with 'past flight history', clicking on question 4 will open a window giving the reference to the paper with the highest impact factor (IF) in the database that posed that question, and the exact question formulation used in that paper.

## Results

Composing a questionnaire is viewed, from the user's point of view, as 'shopping for' questions, that is, filling a 'basket' with the questions he wants to eventually have in the questionnaire. However, while the order in which the items composing the shopping basket were added does not matter in a usual consumer's purchase, here the questions composing a questionnaire must be appropriately ordered. The EpiBasket software helps the user draft a questionnaire and to exchange his draft (that we called a 'wish list') with his collaborators until he decides to obtain a ready-to-use paper questionnaire, for a paper-and-pencil survey and/or an electronic file for an online survey. In addition, EpiBasket includes several tools enabling user–user and user–Editorial Committee exchanges to promote collaborative epidemiology. Fig. 1

gives a snapshot of one of the windows. A demonstration of the different EpiBasket features of can be tried at http://www.epibasket.org.[1] These features are described below in more detail.

### Building the draft of the questionnaire

First, the user must select the candidate questions pertaining to the event of interest: questions according to categories of the questions, time of study, study population and so on. For example, the user may want to consult all the questions belonging to the 'Knowledge' category, concerning healthcare workers and to be collected during the outbreak, which would require three

[1]A password can be obtained upon request from the corresponding author (A-J V).

clicks. Because questions are labelled with keywords and linked to papers in which they were initially mentioned, suggested associations of potential appropriate questions can be easily accessed.

### Building a wish list and filling a shopping basket

The user prepares a 'wish list' that will eventually be transformed into a final 'buy'. A given user may constitute different wish lists, which may be shared with colleagues. These wish lists allow epidemiologists to discuss questionnaire contents upstream before impression of the final product. Finally, the user selects questions in the wish list to fill his 'shopping basket', which becomes the final questionnaire. The order in which selected questions are added to the basket will be the order of the questions in the final document.

### Obtaining the questionnaire

When the user validates his basket, EpiBasket automatically generates the corresponding questionnaire in three formats: XML, pdf and csv. XML is the most generic output of EpiBasket, as it is the most popular encoding standard for machine-readable document files. Appropriate parsing of an EpiBasket XML output file greatly

facilitates further importation of the questions (including the proposed answers) constituting the generated questionnaire in any popular relational database-management system. The pdf format is available for epidemiologists wanting to conduct a paper-and-pencil survey. The csv format is proposed to those planning an online survey: users can directly import their csv file and initiate the corresponding survey online in EpiBasket (http://www.epibasket.org/limesurvey/admin/index.html). The user can also install on his own server LimeSurvey (9), freeware designed for building and deploying online surveys, and run the survey on this server. The electronic questionnaire imported in LimeSurvey may be modified by the user, if desired.

The generated questionnaires are stored in the user's 'My questionnaires' section (Fig. 2). The user has permanent access to any of the questionnaires ordered since his initial registration in the EpiBasket site.

### Promoting interactions among epidemiologists

Several tools were developed in EpiBasket to encourage collaborations among epidemiologists and facilitate users' comments and suggestions on the methods. More specifically, users can comment on each catalogue question,
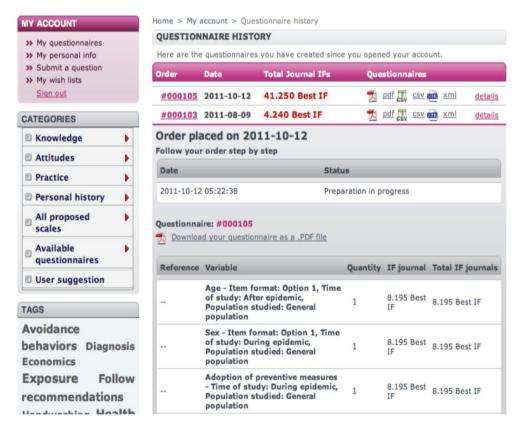


*Fig. 2.* The questionnaire manager in EpiBasket.
A user can store and edit his successive questionnaires or drafts of questionnaires listed at the top of the middle panel. Each of the questionnaires is available in pdf (for a paper questionnaire), in csv (for an online survey using the link to LimeSurvey), or in XML (the most generic EpiBasket output). The preview of the questions is available immediately.

just as e-commerce customers express their opinions on the products being sold. Users can also propose new questions by filling out an electronic form, recommend word changes in a question and suggest coding of the responses. They can also save questionnaires that they find relevant for outbreak or environmental disaster epidemiology in the 'Epibasket store'. This possibility was illustrated in the available EpiBasket prototype with a selection of questionnaires on coronavirus outbreaks.

### Software

For the present proof-of-concept demonstration of EpiBasket, we used free open-source code. The main components of the EpiBasket-prototype development are described below.

The EpiBasket application was built in the Eclipse 3.6 software-development environment (10), using PHP5 language. Development was based on the importation and adaptation/customisation of the freeware part of PrestaShop (11), an open-source PHP5-based developed application devoted to the building of online e-commerce sites. PrestaShop, a full-featured cross-platform shopping basket package, was written in PHP5 with an underlying MySQL relational database-management system. PrestaShop can be deployed in the server-side supporting PHP5 and can be customised through the inclusion of modules proposed at the PrestaShop site (free or not, many supplemental module blocks were developed and are available for enhancing PrestaShop functions), or through a developer's initial code modification (i.e., code suppression/correction/addition).

The main adaptation/customisation features for developing EpiBasket were the following. An initial group of modifications was made to cope with the epidemiological specificity of EpiBasket: the catalogue's offered 'e-commerce products' are 'epidemiological questions'. All 'orders' and 'invoices' are replaced by 'questionnaires'. Moreover, user account-information features used in e-commerce, such as ('my credit slips', 'my vouchers', 'my personal info', and so on) required substantial modifications/suppressions/additions. For example, in EpiBasket, the 'orders' attached to a given user are 'my questionnaires', that is, those ordered since his initial registration in the EpiBasket site. We also developed a specific source code to allow a user account to propose the implementation of a new question in the EpiBasket catalogue. All these adaptations required numerous and substantial code-source changes of the original e-commerce software.

A second group of modifications concerns the relational associations between tags and questions, and between related questions to facilitate the user's navigation interface to explore questions. To do so, the following PrestaShop modules were implemented: REPOSITION MODULE to allow a drop-down presentation of question categories and subcategories; FILTER SEARCH COMMUNITY EDITION to filter questions according to our desired criteria; PRODUCT BOUGHT BY OTHER PEOPLE to enable users to examine which questions were associated in questionnaires in previous questionnaires built by EpiBasket users; and CY-RELATED PRODUCTS which provides a list of those questions sharing keyword similarities with the question currently displayed.

A third group of specific developments concerned the EpiBasket questionnaire outputs. The open-source QUEXML tool (12) was implemented in the Eclipse EpiBasket project and was used to generate a questionnaire according to the encoding XML 1.0 standard scheme. The XML questionnaire template was in turn used to generate pdf or csv versions. A questionnaire written in XML can easily be converted to paper (pdf format), enabling rapid deployment of paper-and-pencil surveys, whereas the converted csv file can be directly imported into the LimeSurvey freeware, enabling rapid deployment of an online survey. LimeSurvey (13) is an open-source online-survey application written in PHP, based on a MySQL, PostgreSQL or MSSQL database, and distributed under the GNU General Public License. It is particularly user friendly, as it enables users to develop and publish surveys, and collect responses, without doing any coding (14). LimeSurvey also offers the possibility of designing telephone surveys with the computing-assisting telephone interviewing (CATI) standard through the use of the QUEXS tool that was not considered in the current EpiBasket prototype.

### Discussion

Herein, we described a prototype that, by adapting WEB tools and e-commerce concepts to epidemiology, should accelerate the creation of an epidemiological questionnaire at the time of a public health emergency. The strength of our approach is that it relies on and takes full advantage of past knowledge and experiences in similar events. The goal is not only avoiding that epidemiologists 'reinvent the wheel' when they are faced with situations similar, or partly similar, to those the past. It also strives to indirectly encourage more standardised and better quality questionnaires. Another indirect attribute of the proposed approach is to promote interactions among epidemiologists. Indeed, despite informatics advances, which make it easy, at least in theory, to devise databases easily sharable and accessible from any place, the fact is that, in practice, this is not yet reality (15). The main reason is likely to be found in human factors. However, the tool we present has integrated the potential to stimulate collaborative approaches. For example, the 'EpiBasket store' can host a public depository for existing questionnaires, which are very rarely made available by authors in their papers.

In this prototype, we chose to exemplify the EpiBasket concept with KAP studies after an outbreak because they

are quite similar to those studies done after an environmental disaster. They represent a small part of the epidemiological investigations done during/after an outbreak and a full-size EpiBasket should provide questions for all fields concerned. After the 2003 SARS epidemic, we identified (7) at total 10 fields other than KAP belonging to three categories: Investigation and Surveillance (four fields: description of the outbreak, search for causative agents, transmission studies, risk factors), case management (five fields: clinical presentation, diagnosis, treatment and medical interventions, prognosis, medical decision-making), and prevention and control. Thus, the EpiBasket prototype described herein is far from what a full-size 'EpiBasket' could be. The design and computer implementation of that ideal project would require a workforce (i.e., epidemiologists and software engineers) that exceeds the resources of a single research group. We think the software aspect is likely the easiest, as it was shown here that the efforts already made in the profitable field of e-commence can be relatively easily adapted to epidemiology. The future editorial and organisational aspects are much more difficult to foresee. The editorial work covers the definitions of the means applied to 'value' the questions. For the prototype, we chose the best journal IF among published papers on similar events that used the question. It is also possible, for example, that an Editorial Committee would prefer a mechanism according high values to questions submitted by experts in the field. In that scenario, coding recommendations should certainly be made by subcommittees. One of the organisational issues concerns copyrighted questionnaires, particularly those that are available only for a fee. The current EpiBasket version does not address such issues.

Finally, because a major objective of EpiBasket is to support collaborative epidemiology, additional tools could be added to the users' community at very little informatics cost: for example, a registry of epidemiological investigations on outbreaks or public health disasters, where epidemiologists could register their new studies before publication with a brief description, authors' names, affiliations and addresses. The difference with the current database on clinical trials (16) would be that epidemiologists would be free to register or not.

The key message of the EpiBasket prototype is that the informatics tools that are used daily and successfully by lay people, like e-commerce and WEB 2.0 applications, have great potential to enhance the quality and timeliness of epidemiological studies, especially in the context of investigation of outbreaks or other public health events requiring a real-time, quality approach. We also think that our proposed approach can help the desired trend towards more transparency, more data sharing and more collaborations among epidemiologists.

## References

1. Fineberg HV, Wilson ME. Epidemic science in real time. Science. 2009;324:987.
2. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ. 2004;328:1312–5.
3. Burns KE, Duffett M, Kho ME, Meade MO, Adhikari NK, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ. 2008;179:245–52.
4. Schilling LM, Kozak K, Lundahl K, Dellavalle RP. Inaccessible novel questionnaires in published medical research: hidden methods, hidden costs. Am J Epidemiol. 2006;164:1141–4.
5. Kleppner D, Sharp PA. Research data in the digital age. Science. 2009;325:368.
6. Olsen J. Epidemiology deserves better questionnaires. IEA European Questionnaire Group. International Epidemiological Association. Int J Epidemiol. 1998;27:935.
7. Xing W, Hejblum G, Leung GM, Valleron A-J. Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: a time-stratified revie. PLoS Med. 2010;7: e1000272.
8. Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE: an architecture for development of robust HLT applications. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 2002. Stroudsburg, PA: Association for Computational Linguistics; 2002:168–175.
9. LimeSurvey download page. [downloaded 2010 August 16]. Available from: http://www.limesurvey.org/en/download
10. Eclipse. [downloaded 2010 May 16]. Available from: http://www.eclipse.org/downloads/packages/eclipse-classic-362/heliossr2
11. PrestaShop. [downloaded 2010 June 2]. Available from: http://www.prestashop.com/
12. queXML. [downloaded 2010 July 8]. Available from: http://quexml.sourceforge.net/
13. LimeSurvey. [cited 2010 Aug 16]. Available from: http://limesurvey.org/
14. Turner BL, Haygarth PM. Biogeochemistry. Phosphorus solubilization in rewetted soils. Nature. 2001;411:258.
15. Editorial. Dreams of flu data. Nature. 2006;440:255–6.
16. Clinical Trials. [cited 2011 May 26]. Available from: http://clinicaltrials.gov/

**\*Alain-Jacques Valleron**
Batiment Pincus
Hôpital du Kremlin Bicêtre
80 rue du Général Leclerc, 94276 Le Kremlin Bicetre, France
Email: alain-jacques.valleron@inserm.fr

**Supplemental material:** Description of the EpiBasket prototype

**Abbreviations**

**CATI:** computing-assisting telephone interviewing

**IF:** impact factor

**KAP:** knowledge, attitudes, practice

**MeSH:** medical subject headings

**SARS:** severe acute respiratory syndrome

## Supplemental material

### Description of the EpiBasket prototype

The EpiBasket prototype can be visited at http://www.epibasket.org/. This prototype takes the

example of Knowledge, Attitudes, Practice (KAP) surveys[1] conducted at the time of the 2003

severe acute respiratory syndrome (SARS) epidemic. Such studies are also frequently

conducted before or after environmental disasters.

With this prototype, we illustrate the different EpiBasket features described in the paper.

The database of questions was based on the 59 KAP papers (listed below as references (1-59))

identified in a previous analysis of the SARS epidemiology literature (60). Authors of two of

the 59 studies supplied complete versions of their questionnaires; another study provided an

incomplete version of its questionnaire in the paper's appendix. We sent emails to the

corresponding author of all articles requesting the corresponding questionnaires and finally

obtained 15 original questionnaires among the 59 (n.b., these 15 papers actually used 7

different questionnaires, used respectively in 5, 3, 2, 2, 1, 1 and 1 papers). For these 15

publications, the questionnaire questions and the variables that were only mentioned in the

report were included in the database. For the 44 remaining papers, all variables mentioned in

the articles' tables and figures were collected with a data-collection grid (Table SM1-1). A

total of 799 variables were recorded in the database. The generic formulation approach,

described in the paper, was used in the prototype with the following six root terms:

---

[1]According to the Medical Subject Headings (MeSH) thesaurus, such studies refer to "Knowledge, attitudes, and associated behaviors which pertain to health-related topics such as pathologic processes or diseases, their prevention, and treatment. This term refers to non-health workers and health workers."

"#$DISEASE", "#$DRUG", "#$NUMBER", "#$PERIOD", "#$PLACE" and

"#$SOMEONE". For example, the questions, "Did you wear a mask in a public place?", "Did

you wear a mask at work?", "Did you wear a mask during a flight?" were reframed as a single

generic formulation as "Did you wear a mask #$PLACE?" Finally, this question processing

yielded 327 questions. Among them, 200 (61%) questions were used only in one paper; 52

(16%) were used in two papers, 64 (20%) were asked in 3 to 10 papers, 11 (3%) were posed

in more than 10 papers. Finally, 130 questions were included for the present demonstration of

the EpiBasket prototype.

For each question, the different information described in the paper was acquired and is

readily available to the EpiBasket user (Figure SM1-1). In particular, each question was

labeled with at least one of the four following categories: Knowledge, Attitudes, Practice

(KAP), Personal history (e.g., demographic or occupational environment questions, contact

history, travel history…). In each category, subcategories were defined. Each question could

be assigned to several categories or subcategories. For example, the question "cleaning

patient's room" belonged to the "Practice" and "Personal history" categories. As scales based

on a series of questions from which a score is derived were widely used in KAP studies

(stress scales, burnout scales, life-satisfaction scales...), a special "All proposed scales"

category was added. The full description of these scales (lists of questions and scoring

methods) is provided. Another special "User suggestions" category was also added to store all

the questions proposed by users.

The prototype user can construct a new questionnaire as indicated in the paper, and then obtain his paper questionnaire in a pdf (see example in Figure SM1-2) or csv format to prepare an online survey (see the section "My questionnaire" where examples can be found).

**References**

1. Blendon RJ, Benson JM, DesRoches CM, Raleigh E, Taylor-Clark K. The public's response to severe acute respiratory syndrome in Toronto and the United States. Clin Infect Dis. 2004;38:925-31.

2. Caves ND, Irwin MG. Attitudes to basic life support among medical students following the 2003 SARS outbreak in Hong Kong. Resuscitation. 2006;68:93-100.

3. Chan BC, Lee CP, Tang GW. Universal SARS preventive measures in an obstetrics unit: experience of health care staff. Am J Infect Control. 2004;32:417-20.

4. Chan SS, Leung GM, Tiwari AF, Salili F, Leung SS, Wong DC, et al. The impact of work-related risk on nurses during the SARS outbreak in Hong Kong. Fam Community Health. 2005;28:274-87.

5. Chan SS, So WK, Wong DC, Lee AC, Tiwari A. Improving older adults' knowledge and practice of preventive measures through a telephone health education during the SARS epidemic in Hong Kong: a pilot study. Int J Nurs Stud. 2007;44:1120-7.

6. Cheng C, Ng AK. Psychosocial factors predicting SARS-preventive behaviors in four major SARS-affected regions. J Appl Soc Psychol. 2006;36:222-47.

7.  Cheng SK, Sheng B, Lau KK, Wong CW, Ng YK, Li HL, et al. Adjustment outcomes in Chinese patients following one-month recovery from severe acute respiratory syndrome in Hong Kong. J Nerv Ment Dis. 2004;192:868-71.

8.  Cheng SK, Tsang JS, Ku KH, Wong CW, Ng YK. Psychiatric complications in patients with severe acute respiratory syndrome (SARS) during the acute treatment phase: a series of 10 cases. Br J Psychiatry. 2004;184:359-60.

9.  Cheng SK, Wong CW, Tsang J, Wong KC. Psychological distress and negative appraisals in survivors of severe acute respiratory syndrome (SARS). Psychol Med. 2004;34:1187-95.

10. Cheng SKW, Chong GHC, Chang SSY, Wong CW, Wong CSY, Wong MTP, et al. Adjustment to severe acute respiratory syndrome (SARS): roles of appraisal and post-traumatic growth. Psychol Health. 2006;21:301-17.

11. Chua SE, Cheung V, Cheung C, McAlonan GM, Wong JW, Cheung EP, et al. Psychological effects of the SARS outbreak in Hong Kong on high-risk health care workers. Can J Psychiatry. 2004;49:391-3.

12. Chua SE, Cheung V, McAlonan GM, Cheung C, Wong JW, Cheung EP, et al. Stress and psychological impact on SARS patients during the outbreak. Can J Psychiatry. 2004;49: 385-90.

13. Dimoulas P, Green KA, Shigayeva A, Aquino M, McGeer A, Scales DC. Patient contact recall after SARS exposure. Emerg Infect Dis. 2005;11:625-8.

14. Grace SL, Hershenfield K, Robertson E, Stewart DE. Factors affecting perceived risk of contracting severe acute respiratory syndrome among academic physicians. Infect Control Hosp Epidemiol. 2004;25:1111-3.

15. Grace SL, Hershenfield K, Robertson E, Stewart DE. The occupational and psychosocial impact of SARS on academic physicians in three affected hospitals. Psychosomatics. 2005;46:385-91.

16. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis. 2004;10:1206-12.

17. Ho SM, Kwong-Lo RS, Mak CW, Wong JS. Fear of severe acute respiratory syndrome (SARS) among health care workers. J Consult Clin Psychol. 2005;73:344-9.

18. Lau JT, Yang X, Pang E, Tsui HY, Wong E, Wing YK. SARS-related perceptions in Hong Kong. Emerg Infect Dis. 2005;11:417-24.

19. Lau JT, Yang X, Tsui H, Kim JH. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. J Epidemiol Community Health. 2003;57:864-70.

20. Lau JT, Yang X, Tsui H, Pang E, Kim JH. SARS preventive and risk behaviours of Hong Kong air travellers. Epidemiol Infect. 2004;132:727-36.

21. Lau JT, Yang X, Tsui HY, Kim JH. Impacts of SARS on health-seeking behaviors in general population in Hong Kong. Prev Med. 2005;41:454-62.

22. Lau JT, Yang X, Tsui HY, Pang E. SARS related preventive and risk behaviours practised by Hong Kong-mainland China cross border travellers during the outbreak of the SARS epidemic in Hong Kong. J Epidemiol Community Health. 2004;58:988-96.

23. Lau JT, Yang X, Tsui HY, Pang E, Wing YK. Positive mental health-related impacts of the SARS epidemic on the general public in Hong Kong and their associations with other negative impacts. J Infect. 2006;53:114-24.

24. Lau JT, Yang X, Wong E, Tsui H. Prevalence and factors associated with social avoidance of recovered SARS patients in the Hong Kong general population. Health Educ Res. 2006;21:662-73.

25. Lee AM, Wong JG, McAlonan GM, Cheung V, Cheung C, Sham PC, et al. Stress and psychological distress among SARS survivors 1 year after the outbreak. Can J Psychiatry. 2007;52:233-40.

26. Lee DT, Sahota D, Leung TN, Yip AS, Lee FF, Chung TK. Psychological responses of pregnant women to an infectious outbreak: a case-control study of the 2003 SARS outbreak in Hong Kong. J Psychosom Res. 2006;61:707-13.

27. Lee DT, Wing YK, Leung HC, Sung JJ, Ng YK, Yiu GC, et al. Factors associated with psychosis among patients with severe acute respiratory syndrome: a case-control study. Clin Infect Dis. 2004;39:1247-9.

28. Lee S, Chan LY, Chau AM, Kwok KP, Kleinman A. The experience of SARS-related stigma at Amoy Gardens. Soc Sci Med. 2005;61:2038-46.

29. Lee TM, Chi I, Chung LW, Chou KL. Ageing and psychological response during the post-SARS period. Aging Ment Health. 2006;10:303-11.

30. Lee-Baggley D, DeLongis A, Voorhoeave P, Greenglass E. Coping with the threat of severe acute respiratory syndrome: Role of threat appraisals and coping responses in health behaviors. Asian Journal of Social Psychology. 2004;7:9-23.

31. Leung GM, Ho LM, Chan SK, Ho SY, Bacon-Shone J, Choy RY, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis. 2005;40:1713-20.

32. Leung GM, Lam TH, Ho LM, Ho SY, Chan BH, Wong IO, et al. The impact of community psychological responses on outbreak control for severe acute respiratory syndrome in Hong Kong. J Epidemiol Community Health. 2003;57:857-63.

33. Leung GM, Quah S, Ho LM, Ho SY, Hedley AJ, Lee HP, et al. A tale of two cities: community psychobehavioral surveillance and related impact on outbreak control in Hong Kong and Singapore during the severe acute respiratory syndrome epidemic. Infect Control Hosp Epidemiol. 2004;25:1033-41.

34. Maunder RG, Lancee WJ, Balderson KE, Bennett JP, Borgundvaag B, Evans S, et al. Long-term psychological and occupational effects of providing hospital healthcare during SARS outbreak. Emerg Infect Dis. 2006;12:1924-32.

35. Maunder RG, Lancee WJ, Rourke S, Hunter JJ, Goldbloom D, Balderson K, et al. Factors associated with the psychological impact of severe acute respiratory syndrome on nurses and other hospital workers in Toronto. Psychosom Med. 2004;66:938-42.

36. McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Can J Psychiatry. 2007;52:241-7.

37. Ng SM, Chan TH, Chan CL, Lee AM, Yau JK, Chan CH, et al. Group debriefing for people with chronic diseases during the SARS pandemic: Strength-Focused and Meaning-Oriented Approach for Resilience and Transformation (SMART). Community Ment Health J. 2006;42:53-63.

38. Nickell LA, Crighton EJ, Tracy CS, Al-Enazy H, Bolaji Y, Hanjrah S, et al. Psychosocial effects of SARS on hospital staff: survey of a large tertiary care institution. CMAJ. 2004;170:793-8.

39. Parker MJ, Goldman RD. Paediatric emergency department staff perceptions of infection control measures against severe acute respiratory syndrome. Emerg Med J. 2006;23:349-53.

40. Poon E, Liu KS, Cheong DL, Lee CK, Yam LY, Tang WN. Impact of severe respiratory syndrome on anxiety levels of front-line health care workers. Hong Kong Med J. 2004;10:325-30.

41. So WK, Chan SS, Lee AC, Tiwari AF. The knowledge level and precautionary measures taken by older adults during the SARS outbreak in Hong Kong. Int J Nurs Stud. 2004;41:901-9.

42. Stinson J, McCartney CJ, Leung A, Katz J. Impact of attending a health care conference in Toronto during the severe acute respiratory syndrome crisis: survey of delegates. Pain Res Manag. 2004;9:137-43.

43. Tam CW, Pang EP, Lam LC, Chiu HF. Severe acute respiratory syndrome (SARS) in Hong Kong in 2003: stress and psychological impact among frontline healthcare workers. Psychol Med. 2004;34:1197-204.

44. Tam KP, Lau IYM, Chiu CY. Biases in the perceived prevalence and motives of severe acute respiratory syndrome prevention behaviors among Chinese high school students in Hong Kong. Asian Journal of Social Psychology. 2004;7:67-81.

45. Tang CS, Wong CY. An outbreak of the severe acute respiratory syndrome: predictors of health behaviors and effect of community prevention measures in Hong Kong, China. Am J Public Health. 2003;93:1887-8.

46. Tang CS, Wong CY. Factors influencing the wearing of facemasks to prevent the severe acute respiratory syndrome among adult Chinese in Hong Kong. Prev Med. 2004;39:1187-93.

47. Tang CS, Wong CY. Psychosocial factors influencing the practice of preventive behaviors against the severe acute respiratory syndrome among older Chinese in Hong Kong. J Aging Health. 2005;17:490-506.

48. Tolomiczenko GS, Kahan M, Ricci M, Strathern L, Jeney C, Patterson K, et al. SARS: coping with the impact at a community hospital. J Adv Nurs. 2005;50:101-10.

49. Wong CY, Tang CS. Practice of habitual and volitional health behaviors to prevent severe acute respiratory syndrome among Chinese adolescents in Hong Kong. J Adolesc Health. 2005;36:193-200.

50. Wong JG, Cheung EP, Cheung V, Cheung C, Chan MT, Chua SE, et al. Psychological responses to the SARS outbreak in healthcare students in Hong Kong. Med Teach. 2004;26:657-9.

51. Wong SY, Wong W, Jaakkimainen L, Bondy S, Tsang KK, Lee A. Primary care physicians in Hong Kong and Canada--how did their practices differ during the SARS epidemic? Fam Pract. 2005;22:361-6.

52. Wong TW, Gao Y, Tam WWS. Anxiety among university students during the SARS epidemic in Hong Kong. Stress and Health. 2007;23:31-5.

53. Wong TW, Tam WW. Handwashing practice and the use of personal protective equipment among medical students after the SARS epidemic in Hong Kong. Am J Infect Control. 2005;33:580-6.

54. Wong TW, Yau JK, Chan CL, Kwong RS, Ho SM, Lau CC, et al. The psychological impact of severe acute respiratory syndrome outbreak on healthcare workers in emergency departments and how they cope. Eur J Emerg Med. 2005;12:13-8.

55. Wong WC, Lee A, Tsang KK, Wong SY. How did general practitioners protect themselves, their family, and staff during the SARS epidemic in Hong Kong? J Epidemiol Community Health. 2004;58:180-5.

56. Wu KK, Chan SK, Ma TM. Posttraumatic stress after SARS. Emerg Infect Dis. 2005;11:1297-300.

57. Wu KK, Chan SK, Ma TM. Posttraumatic stress, anxiety, and depression in survivors of severe acute respiratory syndrome (SARS). J Trauma Stress. 2005;18:39-42.

58. Yip HK, Tsang PC, Samaranayake LP, Li AH. Knowledge of and attitudes toward severe acute respiratory syndrome among a cohort of dental patients in Hong Kong following a major local outbreak. Community Dent Health. 2007;24:43-8.

59. Yu HYR, Ho SC, So KFE, Lo YL. The psychological burden experienced by Hong Kong midlife women during the SARS epidemic. Stress and Health. 2005;21:177-84.

60. Xing W, Hejblum G, Leung GM, Valleron A-J. Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: a time-stratified review. PLoS Med 2010;**7**:e1000272.

**Table SM1-1.** Data-collection grid used to analyze the source papers

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **title** | Title of the paper | Text | Full title |
| **journal** | Title of the journal | Journal (full, or abbreviated name, according to Medline or the Web of Science format) | Medline format (e.g., Emerg Infect Dis) Web of Science format (e.g., Emerging Infectious Diseases) |
| **year** | Publication year of the paper | Number | |
| **volume** | Volume in which the paper was published | Number | |
| **issue** | Issue in which the paper was published | Text | |
| **pages** | The paper's first and last page numbers in the journal | Text | e.g., 1297-300 |
| **authors** | Full names and initials | Text | e.g., Wu KK |

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **auth_order** | The place of signature of the author in the paper | Number | 1=first author, 2=second author, …, 99=last author |
| **doi** | Digital Object Identifier of the paper | Text | |
| **if** | Journal impact factor | Number | 5.756 |
| **pmid** | PubMed Identifier of the paper | Number | e.g., 16102324 |
| **url_pm** | The link to the paper in PubMed database | URL[a] | e.g., http://www.ncbi.nlm.nih.gov/pubmed/16102324 |

[a]Uniform Resource Locator

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **url_gs** | The link to the paper in Google Scholar database | URL[a] | e.g., http://scholar.google.com/scholar?q=Posttraumatic+stress+after+SARS |
| **topic** | The main topic(s) that the paper studied | Check box:<br>− knowledge<br>− attitudes<br>− practice<br>− personal history | |

[a]Uniform Resource Locator

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **timing** | The timing of the survey | Check box: <br> − during the event (e.g., epidemic) <br> − after the epidemic | |
| **population** | The population | Check box: <br> − patients <br> − general population <br> − healthcare workers <br> − other | |

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **questionnaire** | Use of a questionnaire in the study | Yes/No | Yes: If the word "questionnaire", "survey" or "interview" is explicitly used to describe the instrument of data collection in the paper |
| **quest_available** | Availability of the questionnaire | Dropdown menu List:<br>− whole questionnaire given in paper or appendix<br>− a part of questionnaire given in paper or appendix<br>− questionnaire available on request<br>− questionnaire not available | |

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **charac_quest** | Characteristics of questionnaire[a] | Check box:<br><br>− existing questionnaire used[b]<br><br>− "pretested" or "pilot" study[c]<br><br>− none | |

[a]As in the paper by Rosen T et al (Am J Epidemiol. 2006).

[b]Indicates that the paper uses a fully referenced questionnaire that was previously used by others

[c]Indicates that a pilot study of the questionnaire was explicitly mentioned.

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **mode_adm** | Mode of administration used in the study[a] | Check box: <br><br>− interview administered <br><br>  ◦ face to face <br><br>  ◦ by telephone <br><br>  ◦ both telephone and face to face <br><br>  ◦ unspecified <br><br>− self-questionnaire <br><br>  ◦ by filling out a paper questionnaire <br><br>  ◦ by answering by mail or e-mail <br><br>  ◦ by filling out a computer form <br><br>  ◦ unspecified <br><br>− Unknown | |

[a]As in the paper by Rosen T et al (Am J Epidemiol. 2006).

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| scale | Use of a scale in the study | Yes/No | A set of questions summarized with a score |
| var_orig | Name of the variable, as in the original paper | Text | e.g., sex |
| var_WX | Name of the variable | Text | Only in the case when the name of the variable was not given in the paper |
| resp_question | Different instances of the variable | Text | e.g., female, male |
| question | Question formulation of the variable in the questionnaire | Text | Only when the questionnaire is available |

*Table continues on the following page.*

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **answer** | Answers to the question in the questionnaire | Text | Only when the questionnaire is available |
| **quest_type** | Type of the question | Dropdown menu List:<br>− single choice question<br>− multiple choice question<br>− open question | |

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **var_type** | Type of the variable | Dropdown menu List: <br> − quantitative variable <br> − nominal variable[a] <br> − ordinal variable[b] <br> − scale[c] | |

[a]Nominal variable: the values of the variable are the elements of a non-hierarchical category, for example: the name of journals.

[b]Ordinal variable: the values of the variable are the elements of a hierarchical category, they could be ranked in a meaningful sequence, for example: never, occasionally and frequently. The Likert scale is considered an ordinal variable in our study.
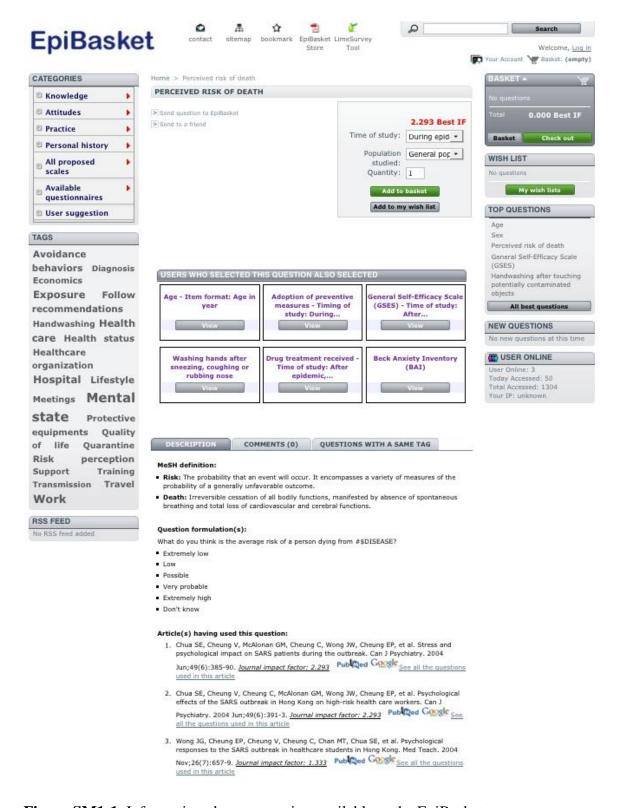
[c]Scale: a set of questions summarized with a score.

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **scale_full** | Full name of the scale | Text | e.g., Perceived Stress Scale |
| **scale_abbr** | Abbreviated name of the scale | Text | e.g., PSS |
| **scale_ref** | Reference(s) of the scale | Text | e.g., Cohen S, Kamarck T, J Health Soc Behav, 1983;24:385-96 |
| **scale_integ** | If the whole scale is used in the study | Yes/No/NA (not applicable) | "No" indicates that only a part of the scale was used in the study |
| **Mesh** | MeSH[a] term(s) corresponding to the variable | Text | |
| **mesh_def** | Definition of the MeSH[a] term(s) | Text | |

[a]Medical Subject Headings.

*Table continues on the following page.*

**Table SM1-1** (Continued)

| Variable Name | Information | Format Specification | Comment |
|---|---|---|---|
| **section** | Section of the paper where the variable was first mentioned | Check box:<br><br>− Materials and Methods<br><br>◦ text<br><br>◦ figure<br><br>◦ table<br><br>− Results<br><br>◦ text<br><br>◦ figure<br><br>◦ table | |

**Figure SM1-1**. Information about a question available to the EpiBasket user.

The example concerns the perceived risk of death. Upper part of the figure: The user gets only

one question in the catalogue (see:http://www.epibasket.org/product.php?id_product=19).

Bottom part of the Figure: The user gets all information on the question, including the

references of the papers that used it, and access to which questions were posed jointly in the

reference papers.

**Figure SM1-2**. Sample of a paper questionnaire generated by EpiBasket.

The complete pdf document can be accessed at http://www.epibasket.org/demo/example.pdf.

The csv format questionnaire may be directly imported into the LimeSurvey tool and

implemented in the EpiBasket site.