TARGET MALARIA
A Vector Control Research Alliance

Independent risk assessment for contained laboratory studies on a sterile male strain of *Anopheles gambiae*

Background

In 2015, malaria killed more than 400,000 people, most of them children under 5, most in sub-Saharan Africa, and more than 200 million people were infected[1]. In addition to deaths, the social and economic costs from the illness are huge, estimated at $12 billion a year in Africa alone.

Malaria is transmitted by anopheline mosquitoes. Current interventions such as drug treatments, bed nets, and insecticide spraying have helped to lower the burden of disease, but equitable access and effective implementation are perennial constraints, since they are expensive to maintain, only protect people under certain circumstances, and the development of resistance to drugs and insecticides is a continual and increasing problem. The consensus is that in many places the current interventions are not enough by themselves to eliminate malaria, and new tools are needed.[2]

Target Malaria is a not-for-profit research consortium that aims to develop and share technology for malaria control. Target Malaria started as a university-based research program and has grown to include scientists, stakeholder engagement teams, risk assessment specialists and regulatory experts from Africa, North America and Europe. We are working to develop a durable method to reduce the population of malaria-transmitting mosquitoes in sub-Saharan Africa, and thereby reduce transmission of the disease. Many current measures to control malaria, such as insecticides and bed nets, also rely on reducing populations of malaria mosquitoes, but we believe our method will overcome the cost and distribution challenges faced by these measures.

Our approach involves using enzymes to disrupt specific regions of the mosquito DNA that influence its reproduction or survival. Ultimately, our goal is to reduce the number of malaria mosquitoes to a level where there are too few left to effectively sustain disease transmission. To do that, we propose to make genetically modified mosquitoes in which the genes for making these enzymes are inserted into the mosquito DNA in such a way that they will spread through a population of malaria mosquitoes, disrupting their reproduction in a self-sustaining way. The technology we are developing will be complementary to other

[1] 10 Facts on Malaria. http://www.who.int/features/factfiles/malaria/en/
[2] Roll Back Malaria http://www.rollbackmalaria.org/microsites/wmd2014/report9.html; Global Technical Strategy for Malaria 2016-2030 http://www.who.int/malaria/publications/atoz/9789241564991/en/

mosquito control methods, and should be relatively inexpensive to implement, because the mosquitoes themselves do the work of stopping malaria (further details available on www.targetmalaria.org).

Safety, for both people and the environment, is paramount as we develop this technology, and our thinking about safety informs much of what we do. Here we briefly mention four such aspects:

1) Our laboratory work is focused on the mosquito species *Anopheles gambiae*, the most harmful mosquito vector of malaria in sub-Saharan Africa, and the initial work has been done in laboratories in the UK, Italy and the US. These facilities meet internationally recognized standards for physical containment (Arthropod Containment Level 2[3]; ACL2). They are also outside the natural species range of *Anopheles gambiae* (which is restricted to sub-Saharan Africa), and in locations that are too cold for the species to persist, as an extra level of safety assurance (ecological containment).
2) We are developing the technology in a step-wise approach consistent with guidance from the World Health Organization[4] to minimize risk at each stage. While our eventual goal is to develop a sustainable technology for reduction of malaria-transmitting mosquitoes by using a 'gene drive system' for spreading the intervention in a self-sustaining way, that is still a long way off. The initial studies with modified mosquitoes in Africa will be done in containment using a construct that causes male mosquitoes to be sexually sterile and which therefore would not persist in mosquito populations. The strain also includes fluorescent markers so it can be identified. While this first strain is unlikely to be useful in itself for malaria control, it will be an important tool in determining how modified mosquitoes behave in an African genetic context, and for enhancing research and regulatory experience in our partner countries. Subject to regulatory permission, the first experiments we want to do are to cross the lab strain with locally-derived African strains of mosquito in containment, to ensure that the sterility and fluorescent traits are maintained in the local genetic background. The proteins expressed by

[3] Arthropod Containment Levels. Vector-Borne and Zoonotic Diseases. June 2003, 3(2): 75-90. doi:10.1089/153036603322163475. http://online.liebertpub.com/doi/abs/10.1089/153036603322163475
[4] The Guidance Framework for testing genetically modified mosquitoes. http://www.who.int/tdr/publications/year/2014/guide-fmrk-gm-mosquit/en/

transgenes in the male sterile strains have been analyzed for similarity with known allergens, using internationally accepted guidelines for assessing allergenicity, and have been shown to present no allergenicity concerns. The expressed proteins have also been shown to have non-toxic modes of action for humans or animals. The project will conduct additional studies to increase the assurance that these expressed proteins are not toxic or allergenic.

3) In preparation for studying this sterile male strain in laboratory conditions, Arthropod Containment Level 2 (ACL2) insectaries have been established at the African partner institutions. These laboratories will be subject to oversight by national regulatory bodies, will maintain standards consistent with national requirements and international guidelines, and will operate equivalently to the European and American labs where the initial development has been undertaken. The extensive experience of working with these mosquito strains in contained conditions in the European and American labs will guide operation of the labs in Africa.
4) To obtain an independent opinion on the safety of the proposed rearing and maintenance of the transgenic male-sterile strain of mosquitoes in containment, our funders (Foundation for the National Institutes of Health - FNIH) commissioned an external risk assessment from the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia's national science organization, which is unaffiliated with the Target Malaria program. Experts in risk evaluation and environmental monitoring from CSIRO, Monash University and Applied Biomathematics conducted the work under the auspices of CSIRO's Biosecurity and Health Flagship, which has broad experience with international health and biosafety issues. The intent of this independent risk assessment was to inform the Target Malaria program's comprehensive biosafety planning. This does not replace risk assessments that will be conducted by the national authorities prior to any decision regarding our application to conduct experiments in containment. Those risk assessments will take into account the physical containment and standard operating procedures under which the work will be done.

The CSIRO risk assessment

The CSIRO team was asked to evaluate risks that might be associated with a hypothetical scenario where an unexpected breach of an African containment facility took place such that all modified mosquitoes within the facility were able to escape to the local

environment. Thus, this risk assessment looked at the possible results of an unlikely "worst case" scenario as a planning precaution, to support consideration of whether any additional risk management measures should be put in place before initiation of contained studies. It is important to remember that the risks evaluated here would only be relevant if the containment measures we have in place were to completely fail while the insectary is at maximum capacity.

CSIRO identified five potential undesirable endpoints resulting from such a loss of containment:

1. Increased transmission of malaria (i.e. increased ability of laboratory strain or genetically modified mosquitoes to transmit malaria as compared to local wild type mosquitoes)
2. Transmission of a novel blood-borne pathogen (i.e., one not normally transmitted by *Anopheles gambiae* mosquitoes)
3. Spread of the transgene construct into non-target eukaryotes
4. Spread of the transgene construct into non-eukaryotes (e.g., bacteria and viruses)
5. Spread of the transgene construct into the *Anopheles gambiae* complex

The first two of these examine the capacity of the escaped genetically modified mosquitoes for transmitting malaria or other pathogens, and are directly linked to human safety. The next three endpoints examine how likely it is that the transgene might get into other organisms or wild populations of *Anopheles gambiae*. These endpoints would not necessarily be harmful *per se*, but would trigger additional risk assessment as to whether it might lead to environmental harm. Using these three endpoints was therefore conservative.

CSIRO conducted a formal, standardized interview of twenty-four experts (8 in the Target Malaria team and 16 outside of it) in order to develop quantitative probabilities of each of the five endpoints occurring in the event of an escape. The full CSIRO risk assessment is attached to this document.

Results and future steps

The results of the CSIRO risk assessment indicate that the risks associated with all five endpoints are sufficiently remote that no additional specific risk mitigation is warranted beyond the physical ACL2 facilities and operational protocols Target Malaria has put in place. For endpoint 5 some experts expressed uncertainty about whether the male sterile construct would continue to function (i.e., cause sterility) once it is introgressed into the

local genetic background, though they did not identify a specific reason why. This uncertainty can be resolved through experimentation – performing the introgressions and testing for sterility – and we will carry out these experiments in containment to assess the functioning of the construct in the new genetic background as a matter of high priority once the strain has been imported into the containment facility. Moreover, we will keep the numbers of genetically modified adults small (<1,000) until such time as we have confirmed that the construct does cause sterility in the new background. In this way new experimental data can resolve the experts' uncertainty.

Target Malaria will continue to conduct risk assessments at each step of the development pathway toward our product, as part of our on-going commitment to safety and responsible research. Target Malaria will provide evidence for national authorities to conduct official risk assessments and will comply with all regulatory requirements in the countries where it works. Research will only be undertaken following national regulatory approvals.

December 2015

CSIRO BIOSECURITY FLAGSHIP
www.csiro.au

Risk Assessment for Controlling Mosquito Vectors with Engineered Nucleases: Sterile Male Construct
Final report

Keith R. Hayes, Simon Barry, Nigel Beebe, Jeffrey M. Dambacher, Paul De Barro, Scott Ferson, Jessica Ford, Scott Foster, Anders Gonçalves da Silva, Geoffrey R. Hosack, David Peel and Ronald Thresher

September 4, 2015

CSIRO Biosecurity Flagship
Castray Esplanade, Hobart, Australia, 7000
Telephone : +61 6232 5222
Fax : +61 6232 5485

Contents

ACKNOWLEDGEMENTS

The CSIRO risk assessment team gratefully acknowledges the assistance of members of the Target Malaria consortium in the development of the Problem Formulation statement used in this report, and for assistance with Figure 1.1.

This risk assessment would not have been possible without the generosity and patience of the members of the Target Malaria consortium, who contributed their time and expertise during the hazard analysis workshops, fault tree construction and elicitation exercises. The CSIRO project team also thanks the following independent experts who graciously agreed to participate in the elicitation exercises and contribute to the fault tree analysis

- George Christophides, Department of Life Sciences, Imperial College London
- Frank Collins, Department of Biological Sciences and Eck Institute for Global Health, Notre Dame University
- Martin Donnelly, Department of Vector Biology, Liverpool School of Tropical Medicine
- Bill Engels, School of Medicine and Public Health, University of Wisconsin-Madison
- Heather Ferguson, Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow
- Kent Golic, Department of Biology, University of Utah
- Fred Gould, Department of Entomology, North Carolina State University
- Stephen Higgs, Biosecurity Research Institute, Kansas State University
- Anthony James, Schools of Medicine and Biological Science, University of California, Irvine
- Greg Lanzarro, School of Veterinary Medicine, University of California, Davis
- Tovi Lehmann, National Institute of Allergy and Infectious Diseases, National Institute of Health
- Steve Lindsey, School of Biological and Biomedical Sciences, Durham University
- Ken Olson, Department of Microbiology, Immunology, and Pathology, Colorado State University
- Jeff Sekelsky, Biology Department, University of North Carolina at Chapel Hill
- Jim Tiedje, Center for Microbial Ecology, Michigan State University
- Michael Turelli, College of Biological Sciences, University of California, Davis

EXECUTIVE SUMMARY

Malaria is a mosquito-borne disease responsible for approximately 600,000 deaths per year, of which 90% occur in Africa. The International Roll Back Malaria partnership has pledged a goal to eradicate malaria worldwide by reducing the global incidence to zero. It is widely acknowledged that this goal will require new control methods.

In February 2014, CSIRO was engaged by the Foundation for the National Institutes of Health (FNIH) to conduct an independent assessment of the risks associated with an escape of mosquitoes genetically modified to be male-sterile by a construct that incorporates the I-PpoI Homing Endonuclease Gene (HEG). The male-sterile mosquitoes are the first stage in the development pathway of a new technology intended to provide a durable and cost-effective tool for reducing the burden of malaria in the Africa by reducing transmission of the malaria parasite.

The CSIRO risk assessment focuses on transgenic male-sterile *Anopheles gambiae* mosquitoes reared under physical containment. Laboratory studies with HEG-bearing, G3 strain mosquitoes, have shown 100% infertility: to date all eggs laid by females mated with these transgenic male mosquitoes have been infertile.

The Target Malaria consortium that developed and tested the I-PpoI construct proposes to examine its effect in different genetic backgrounds, within regulatory compliant African insectaries, by mating transgenic male and female mosquitoes with locally collected wild-type mosquitoes. As part of the due diligence required to support this proposal, the CSIRO was asked to assess the level of risk that might result should the G3 strain of *An. gambiae* mosquitoes, and mosquitoes modified with the I-PpoI male-sterile construct, escape from the insectaries. In this context the G3 strain is used as a comparator for risks that may occur with or without genetic modification of the insectary mosquitoes.

The risk assessment adopted a conservative approach by assuming a complete loss of the maximum planned rearing capacity of an insectary (estimated at 10,000 mosquitoes) composed of G3 mosquitoes or transgenic stock. An unintentional escape of this magnitude would ordinarily be prevented by containment design and carefully designed operating procedures. Nonetheless it might occur, for example, from a large physical breach of containment or human error.

Following a detailed literature review and a systematic, inductive hazard analysis, the CSIRO project team in conjunction with members of the Target Malaria consortium identified 5 risk assessment endpoints:

1. An increase in the malaria vectorial capacity of genetically modified mosquitoes.
2. Transmission of a novel (i.e. not previously known to be vectored by *An. gambiae*) blood-borne pathogen to human or vertebrate host.
3. Spread of the I-PpoI construct in non-target eukaryotes.
4. Spread of the I-PpoI construct in non-eukaryotes.
5. Spread of the I-PpoI construct in the *An. gambiae* complex.

These risk assessment endpoints were chosen to ensure that the analysis was complete and meaningful, but also tractable. The first and second endpoints are intermediate steps (precursors) in event chains that may lead to human mortality due to additional cases of malaria or novel blood-based pathogens.

The third and fourth endpoints focus on potential for horizontal transfer of the I-PpoI construct.

Horizontal transfer would not necessarily be harmful, but could be a pre-cursor to harm, and therefore using these as endpoints is conservative.

The fifth endpoint is not a pre-cursor to a harmful event, indeed reducing populations of species in the *An. gambiae* complex is one way to reduce the transmission of the malaria parasite. This endpoint is, however, considered undesirable at this stage of the development pathway of this new genetic technology because the current construct is supposed to be inherently self-limiting.

During the scoping stage of the assessment four potential endpoints were deliberately excluded. Socio-economic and political endpoints were considered to be outside the project's terms of reference because the risk generating mechanisms focus on a one-time escape of a self-limiting construct. Similarly the risk of project failure was excluded because the Target Malaria consortium addresses this as part of its long term, and day to day, project management processes. Allergic and/or toxic responses are not included because these endpoints will be assessed by established protocols, and the incidental impacts associated with insecticide spraying in the event of a complete loss will be addressed by the consortium as part of its internal risk mitigation planning.

The CSIRO study quantifies the risk associated with the assessment endpoints by using direct elicitation with domain experts to develop subjective prior probability distributions for key malaria transmission parameters, and basic events within fault trees. The elicitation was performed with 8 members of the Target Malaria consortium, and 16 external experts approached for independent opinions. In total 1068 subjective probability density functions and 67 constants were retained for subsequent analysis, together with 1588 comments, covering 352 and 544 basic events or gates (including vectorial capacity parameters) respectively.

The assessment of risk of an increase in malaria vectorial capacity was based on a direct elicitation of parameters that are widely accepted to be the most relevant in this context. If we make the assumption of strong positive dependence between strains, then the mean intrinsic transmission risk index for G3 and I-PpoI mosquitoes is -0.41 and -0.23 respectively. The results indicate a 41% and 23% reduction in the risk of G3 or I-PpoI transgenic mosquitoes producing an infectious bite compared to local wild type mosquitoes.

The results of the fault tree analysis (Table 7.1) indicate that the median value of the risk of G3 strain mosquitoes transmitting a novel blood-based pathogen in a year following a complete escape of 10,000 mosquitoes is 5.2×10^{-7}, while the 90th percentile of this risk is 1×10^{-4}. Comments provided by the experts during the elicitation suggest that a linear pool estimate of the risk of the I-PpoI mosquitoes vectoring a novel pathogen would be the same as, or lower than, the values for the G3 mosquitoes because most experts believe that all of the events in the fault tree would remain unchanged or be lower. Some experts questioned whether the construct might compromise the mosquito immune system, but they expected that this effect would be counteracted by the anticipated higher mortality of the genetically modified mosquitoes.

The fault tree analysis indicates that the median risk of the HEG spreading in non-target eukaryotes or non-eukaryotes in a year following the complete loss of 10,000 I-PpoI modified mosquitoes is 1.2×10^{-10} and 6.7×10^{-7} respectively. The 90th percentile of these risks ranges from 3.1×10^{-6} for spread in non-target eukaryotes to 7.8×10^{-4} for spread in non-eukaryotes. We infer that the probability of consequential impacts on populations of non-target species, over the course of a year, will be no higher than these values, and would be lower if: (i) the probability of any of the events in the causal chain between spread of the construct and any specific

type of detrimental impact was less than one; and/or (ii) only a sub-set of the spread pathways quantified here could lead to a specific impact.

The median value of the risk of the construct spreading in local populations of the related mosquito species *An. coluzzii* or *An. arabiensis* in a year, following the complete loss of 10,000 I-PpoI modified mosquitoes, was estimated to be 1.1×10^{-6}. The median value for the *An. coluzzii* or *An. arabiensis* risk, however, is sensitive to the analysis method used in the fault tree, and this rises to 1.5×10^{-5} under an alternative strategy because the beliefs of one expert have a weaker influence on the risk estimate under this alternative strategy. The 90th percentile of this risk is 0.0024 and 0.01 under the two strategies.

The median value of the risk of the construct spreading in local populations of *An. gambiae* in a year following the complete release of 10,000 I-PpoI mosquitoes was estimated to be 0.0014. The 90th percentile of this risk was estimated to be 0.25.

These results suggest that the risks of the first, second, third and fourth endpoints are sufficiently remote so as not to warrant any additional specific risk mitigation at this stage in the development pathway. We recognise that the potential benefits of the technology are significant, the planned containment and operating procedures will protect against, and further reduce these risks, and that the experiments conducted on the mosquitoes under containment will provide an opportunity to improve confidence, and reduce the uncertainty, associated with the first endpoint.

The risks associated with the fifth endpoint are sufficient to warrant the risk mitigation strategies that are currently being planned by the Target Malaria consortium. The bulk of the risk in this instance occurs because of experts' uncertainty about the potential for the construct to fail when introgressed into wild types mosquitoes. Again, contained experiments provide an opportunity to quantify this risk and reduce the uncertainty associated with the estimates provided here.

KEY POINTS: INTRODUCTION

- Homing Endonuclease Genes (HEGs) are a class of highly specific DNA endonucleases found naturally in some viruses, bacteria and eukaryotes.
- HEGs have recently been identified as a means to disrupt essential genes within populations of pest species, including mosquito malaria vectors, to provide a stable and cost effective mechanism for controlling these species.
- In 2008 the Target Malaria consortium demonstrated that it was possible to genetically modify male *Anopheles gambiae* mosquitoes with a construct containing the I-PpoI HEG that (during spermatogenesis) encodes an endonuclease protein that cleaves ribosomal repeats on the X chromosome.
- Cage experiments involving crosses between transgenic I-PpoI males and wild type females from the *An. gambiae* complex have subsequently confirmed complete transgenic male sterility, and a concomitant decline in stable cage populations into which adult HEG males are introduced.
- The consortium proposes to import the I-PpoI sterile male strain into Africa and conduct experimental crosses with wild type mosquitoes from the *An. gambiae* complex under conditions of physical containment in regulatory compliant laboratories and insectaries.
- In February 2014 CSIRO was engaged by the Foundation for the National Institute of Health (FNIH) to conduct an assessment of the human and ecological risks associated with this proposal.

1 INTRODUCTION

1.1 Project background

In February 2014 CSIRO was engaged by the Foundation for the National Institutes of Health (FNIH) to conduct a risk assessment related to research on a new approach to control mosquito vectors by genetic engineering with Homing Endonuclease Genes (HEGs). HEGs are a class of highly specific DNA endonucleases ("site-specific selfish genes") found naturally in some viruses, bacteria and eukaryotes. HEGs are able to spread rapidly through populations because they are inherited at ratios much higher than the 50% Mendelian ratio (Burt, 2003; Burt and Koufopanou, 2004).

HEGs have recently been identified as a means to disrupt essential genes within populations of pest species, including mosquito malaria vectors, to provide a stable and cost effective mechanism for controlling these species (Gould et al., 2006; Alphey, 2014). HEGs could accomplish this by reducing the number of female mosquitoes in the population or by reducing the fertility of female mosquitoes. In either case the result would substantially lower overall mosquito numbers, which should have the effect of decreasing malaria transmission.

In 2008 a consortium of scientists led by Professor Austin Burt from Imperial College London (hereafter the "Target Malaria consortium"), demonstrated that it was possible to genetically modify *Anopheles gambiae* mosquitoes with a construct containing the I-PpoI HEG – a member of the His-Cys box family of endonucleases from the slime mould *Physarum polycephalum* (Windbichler et al., 2008).

The I-PpoI HEG is known to encode an endonuclease protein that selectively cleaves the ribosomal DNA repeats that, in at least two members of the *An. gambiae* complex, are found near the centromere of Chromosome X (Windbichler et al., 2007). By combining the HEG with a β2-tubulin promoter (Figure 1.1) the Target Malaria consortium were able to induce the HEG expression during spermatogenesis and destroy X-chromosome bearing sperm. Experimental crosses between wild type females and I-PpoI males, however, led to complete sterility rather than the expected sex ratio bias. The complete male sterility was caused by the persistence of "surplus" I-PpoI protein in early embryos attacking the maternally inherited (as well as the paternally inherited) X chromosome (Windbichler et al., 2008). Cage experiments involving crosses between transgenic I-PpoI males and wild type females from the *An. gambiae* complex have subsequently confirmed complete transgenic male sterility, and a concomitant decline in stable cage populations into which adult HEG males are introduced (Klein et al., 2012).

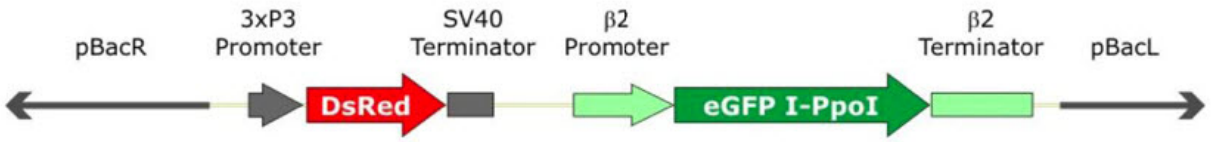


Figure 1.1: Schematic of the I-PpoI sterile male construct and its function showing the I-PpoI effector gene, under the control of β2 tubulin promoter, flanked by non-autonomous piggyBac sequences.

The successful integration of a HEG that targets the rDNA repeats in *An. gambiae* leading to male sterility represents the first proof of concept that HEGs might be utilised to distort sex ratios in populations of malaria vectors, and ultimately eliminate those populations. The "1st generation" sterile male technology described above, however, is unlikely to be used for malaria control in Africa because inundative release of the sterile males – as in a traditional Sterile Insects Technology (SIT) approach – would require prohibitively high resources. Nonetheless, this first generation technology is an important stepping stone towards the ultimate goal of a self-sustaining "Y-drive" technology that would provide a durable and cost-effective tool for malaria eradication. Y-drive requires additional construct design, however, because it requires the HEG to be located and expressed on the Y chromosome, and expression must occur at the right developmental time.

In light of this, and because of its inherently self-limiting properties, the consortium propose to import the I-PpoI sterile male strain into Africa and conduct experimental crosses with wild type mosquitoes from the *An. gambiae* complex in laboratories and insectaries with biosafety level BSL1 and containment level ACL2 conditions (American Committee of Medical Entomology and the American Society of Tropical Medicine and Hygiene, 2003). The purpose of these experiments is to build capacity among local institutions to work with the modified mosquitoes, and to study the phenotypic properties of the construct when it is integrated into wild genotypes. The results from these studies will inform development of the next version of the HEG technology.

As part of its due diligence efforts in preparation for seeking permission to import mosquitoes for contained use, the consortium sought to understand any risks that might arise from an accidental breach of containment during the study period.

1.2 Objectives and structure

The objectives of this risk assessment are to identify and quantify the ecological and human-health risks associated with an accidental release of the I-PpoI mosquitoes from African insectaries in Burkina Faso, Mali and Kenya. The independent risk assessment is part of a much larger initiative under the umbrella of the Target Malaria consortium that includes: construct design and testing, laboratory and insectary trials, field observations of mosquito life-cycle and behaviour, and extensive community consultation.

The overall structure, and components, of the risk assessment are summarised in Figure 1.2. After an initial problem formulation, developed with the assistance of members of the Target Malaria consortium, the risk assessment focuses on three main activities:

- **Hazard analysis and endpoint selection**. The CSIRO project team completed a Boolean literature search and implemented an inductive hazard analysis technique known as Hierarchical Holographic Modelling (HHM). The result of these hazard analysis activities are discussed and documented in Section 2.2, together with their role in the selection of the risk assessment endpoints, and the rationale for this selection.
- **Fault tree analysis**. The risks associated with four assessment endpoints are quantified using fault tree analysis. The probabilities of the basic events within each tree were developed by direct elicitation of experts within the Target Malaria consortium, together with a group of experts that are independent of the project. The probability of a fifth endpoint (the potential for a change in vectorial capacity) is also addressed based on expert beliefs about potential changes in malaria transmission parameters associated between wild

type, G3 and I-PpoI modified mosquitoes. The results of the vectorial capacity analysis are presented in Section 3, and the fault tree analysis in Sections 4 and 5.

- **Null event inference**. The risk assessment also reviews methods for establishing the sample size necessary to establish confidence intervals in the probability of rare events, when these events are not witnessed in experimental settings (known as the "null event" inference problem). The results of the review are presented in Section 6. This analysis has been included in the assessment to complement the experiments and analysis recommended by individual experts during the elicitation, and to assist the Target Malaria consortium in designing experiments for rare events.

Section 7 of the report concludes with a discussion of the risk assessment results and a summary of important issues that were identified by the analysis. The report also includes a number of Appendices that present additional supporting material.

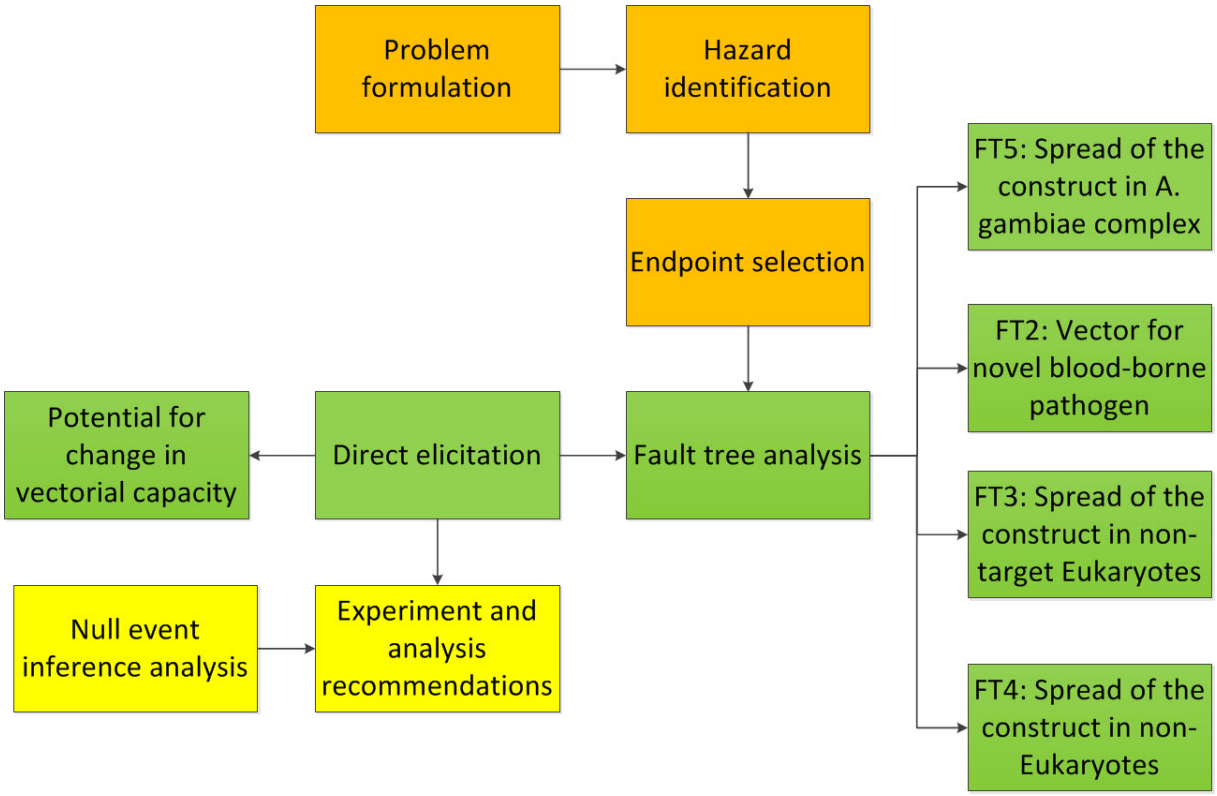


Figure 1.2: Summary of the overall structure and components of the CSIRO risk assessment. After an initial problem formulation, the assessment comprises of three main activities: (i) a hazard analysis that informs the endpoint selection (coloured orange); (ii) a fault tree analysis and an analysis of the potential change in vectorial capacity, supported by direct elicitation of expert opinion (coloured green); and, (iii) a review of the sample size necessary to establish confidence in the probability of rare events that are not observed during experiments (coloured yellow).

KEY POINTS: PROBLEM FORMULATION, HAZARD ANALYSIS AND ENDPOINT SELECTION

- Estimates of malaria-related deaths in 2010 ranged from 655,000 (World Health Organisation, 2011) to over 1.2 million (Murray et al., 2012), with the majority of deaths occurring among African children under five years of age.
- A HEG-based genetic control technology should provide area-wide, durable and low-cost protection against malaria, and be a valuable aid for malaria eradication when used in conjunction with other malaria control tools.
- The CSIRO risk assessment focuses on an accidental escape or unsanctioned release of 10,000 G3 strain mosquitoes and mosquitoes modified with the I-PpoI construct from three African insectaries.
- The CSIRO project team performed a systematic and targeted search of the scientific literature using Google Scholar to identify hazards associated with HEG-based genetic control methods that may have already been proposed in the literature.
- The project team also used an inductive hazard identification tool known as Hierarchical Holographic Modelling (HHM), as a complement to the literature review, to help identify hazardous outcomes that may occur following interactions between genetically modified mosquitoes and all relevant human, biological and environmental systems.
- The hazard analysis indicated that a quantitative risk assessment could be made tractable by focusing on five assessment endpoints:
    1. An increase in the malaria vectorial capacity of genetically modified mosquitoes.
    2. Transmission of a novel (i.e. not previously known to be vectored by *An. gambiae*) blood-borne pathogen to human or vertebrate host.
    3. Spread of the I-PpoI construct in non-target eukaryotes.
    4. Spread of the I-PpoI construct in non-eukaryotes.
    5. Spread of the I-PpoI construct in the *An. gambiae* complex.

2 PROBLEM FORMULATION, HAZARD ANALYSIS AND END-POINT SELECTION

2.1 Problem formulation

Malaria is a mosquito-borne parasitic disease that exacts an enormous public health toll despite ongoing and intensive control efforts. Estimates of malaria-related deaths in 2010 ranged from 655,000 (World Health Organisation, 2011) to over 1.2 million (Murray et al., 2012), with the majority of deaths occurring among African children under five years of age. The International Roll Back Malaria partnership, which includes WHO, UNICEF, UNDP and the World Bank, has pledged a goal to eradicate malaria worldwide by reducing the global incidence to zero through progressive elimination in countries. Yet it is acknowledged widely that this goal will require developing new tools and control methods (Alonso and Tanner, 2013).

One of the most effective ways to reduce the transmission of malaria in endemic areas is to reduce populations of the major insect vectors. In sub-Saharan Africa the main malaria vectors are species of the *An. gambiae* complex. The Target Malaria consortium is focusing on the use of nuclease-based genetic approaches to dramatically diminish the population size of *An. gambiae* mosquitoes in Africa.

The consortium ultimately aims to develop a self-sustaining population suppression technology that will reduce numbers of vector mosquitoes over successive generations, until they are unable to sustain malaria transmission (Deredec et al., 2011; North et al., 2013). The homing endonucleases will be spread by wild type mosquitoes mating with a relatively small population of genetically modified mosquitoes. This HEG-based technology offers a number of features not found in existing control methods – it will be able to target difficult-to-reach segments of the vector population and will protect people without requiring them to change their behaviour or have access to health care. Such a tool should provide area-wide, durable and low-cost protection, and be a valuable aid for malaria eradication when used in conjunction with other malaria control tools.

Although other mechanisms are under investigation within the Target Malaria consortium, the approach that is currently progressing most rapidly employs a homing endonuclease to diminish transmission of the X chromosome from male mosquitoes to the next generation, thus biasing the sex ratio toward production of male mosquitoes (which do not bite or transmit malaria) and reducing overall reproductive potential within the population. This transgenic approach takes its example from naturally occurring meiotic drive systems for sex ratio distortion found in other insects, including some mosquitoes (Mori et al., 2004).

The CSIRO risk assessment focuses on an accidental escape or unsanctioned release of G3 strain mosquitoes and mosquitoes modified with the I-PpoI construct from an African insectary. The different conditions under which the study might take place are exemplified by insectaries located in Bobo Dioulasso (Burkina-Faso), Bamako (Mali) and Mbita Point (Kenya). Bamako is the capital of Mali, it is highly urbanised with a population of approximately 1.8 million. Bobo Dioulasso is Burkina Faso's second largest city but it is much smaller than Bamako with a population of approximately 537,000. Mbita Point is a largely rural location, situated on the shores of Lake Victoria, with a population less than 10,000.

The insectaries are anticipated to house less than 10,000 mosquitoes, consisting primarily of two types of male and female mosquitoes: wild types and G3 strains modified with the I-PpoI

construct. Male and female wild type mosquitoes will be collected from the vicinity of the insectaries. Female G3 laboratory strain mosquitoes, modified with the I-PpoI construct, will be crossed with wild type males to study the phenotypic effects of the construct in local genotypes. Male G3 laboratory strain mosquitoes, modified with the I-PpoI construct will also be crossed with wild type females to study the male sexual sterility effect of the construct in local genotypes.

Empirical observations of anopheline populations in the vicinity of the insectaries are quite scarce. Populations of species from the *Anopheles* complex are thought to be relatively low in the vicinity of the Bamako insectary because it is a highly urbanised area. In Bobo Dioulasso *An. arabiensis* has become the major malaria vector, but populations are still relatively low compared to more rural areas. Around the Mbita Point insectary healthy populations of *An. gambiae* and *An. arabiensis* are known to exist. Additionally, the presence and diversity of a *Plasmodium* reservoir in the local human population is also unclear, particularly within Bamako and Bobo Dioulasso.

Overall the literature suggests that species from the *An. gambiae* complex are unlikely to proliferate in urban areas because they lack appropriate larval habitats. Some species in the complex, however, appear to be adapting to the relatively polluted aquatic environments found in urban areas, and isolated populations can be maintained around the periphery of cities (Fournet et al., 2010), hence interactions between escaped genetically modified mosquitoes and wild type *An. gambiae, An. coluzzii* and *An. arabiensis* cannot be ruled out.

2.2 Hazard analysis

A hazard can be defined as a situation that in particular circumstances could lead to harm (The Royal Society, 1983), or alternatively, considered as the propensity for risk posed by a substance or activity. Hazards are sometimes perceived to be solely a function of a substance's intrinsic properties but, as emphasised in the definition above, they are more usefully conceptualised as a function of both the intrinsic properties of a substance or activity and circumstances surrounding its use.

Hazard analysis is a structured process to identify the circumstances surrounding an activity that might lead to harm. A good hazard analysis acknowledges the intrinsic properties of a substance or activity but also makes transparent the specific set of circumstances – the causal chains – required for harm to be realised. It may also provide a mechanism to rank potential hazards against a variety of criteria (including but not limited to, the likelihood and severity of harm) where this is useful and relevant to the objectives of the analysis.

The identification of potential hazards is a critical step in any risk assessment. The hazard identification and prioritisation process helps to define the scope for the risk assessment and supports the development of post-assessment monitoring programs. Hazards that are not identified during the risk assessment process, either because the problem formulation is too narrow or the hazard identification is cursory and insufficient, may ultimately lead to an underestimation of risk. Conversely, a poorly structured formulation of the risk problem can lead to immeasurable and untestable risk hypotheses (Wolt et al., 2010) and a proliferation of ambiguous hazards.

Hazard identification techniques can be broadly categorised as either inductive or deductive. Inductive hazard analysis postulates the condition of a system component (an initiating event) and identifies the risks that this may entail. Deductive techniques postulate a failed system state (a risk endpoint) and identify the chain of more basic faults that contribute to this failure

(Vesely et al., 1981). This analysis uses a targeted literature review, Hierarchical Holographic Modelling (HHM – an inductive technique) and Fault Tree Analysis (FTA – a deductive technique) to identify potential hazards, and quantify risks, associated with the accidental escape of the I-PpoI mosquitoes from African insectaries.

Inductive hazard analysis methods can evoke large hazard lists because these methods encourage analysts to think "outside the box". It is important to emphasise, however, that while these potential hazards must be at least plausible, some will be so extremely improbable within the context of the overall assessment, that they are not carried forward into a full risk assessment.

2.2.1 Literature review

The CSIRO project team performed a systematic and targeted search of the scientific literature using Google Scholar to identify hazards that may have already been reported in the literature. We used the following Boolean search terms to identify potentially relevant references (with the number of hits noted in parentheses) after excluding most patents, citations and hits within the Encyclopedia:

- "homing endonuclease" AND "horizontal gene transfer" AND "vector-control" (7)
- "homing endonuclease" AND "human health" AND "vector-control" -Encyclopedia (30)
- "homing endonuclease" AND hazard* AND human* AND Environment AND GM* -Encyclopedia (26)
- "homing endonuclease" AND eukaryote AND health AND (hazard* OR risk OR danger OR adverse)- Encyclopedia (18)
- "homing endonuclease" AND GM AND "human health" -Encyclopedia (49)
- "homing endonuclease" AND "horizontal gene transfer" AND "risk" -Encyclopedia (33)

A total of 120 unique references were identified with this search. Each of the references was reviewed by a CSIRO risk team member to check for hazards or information relevant to the risk assessment more generally. The references are listed, together with comments (if any) recorded by the team member, in Appendix A.

The literature review noted potential gaps in the international regulatory frameworks that govern the release of gene-drive systems (Marshall, 2010, 2011). There is nonetheless extensive and widespread support for detailed and transparent risk analysis, together with pre- and post-release monitoring, to be included within any recombinant program, particularly those that are designed to, or could result in, the release of Living Modified Organisms (LMOs) into the environment, and the flow of LMOs across sovereign borders (UNEP, 2010). These recommendations, together with calls to close regulatory gaps, have subsequently been re-iterated on several occasions (Oye et al., 2014; Esvelt et al., 2014).

A hazard that is commonly identified in the literature is the possible "escape" of a genetic construct via Horizontal Gene Transfer (HGT) into non-target taxa. For example, Joardar et al. (2012) notes that HEGs are common in the genome of nine important fungal species, and that their mitochondrial genomes exhibit high levels of interspecies variation due to "significant" intraspecies horizontal transfer and recombination in mitochondrial DNA. Similarly Aguileta et al. (2014) suggests that elements of the mitochondrial genome of fungus could facilitate uptake of exogenous DNA, possibly including HEGs.

In this context it is important to emphasise that HGT does not necessarily lead to harmful outcomes, and terms such as "significant rates of HGT" are over evolutionary times scales, and that the equivalent annual rates are very low. Goddard and Burt (1999), for example, estimate that the average rate of HGT of the ω HEG within 20 species of fungus is 1.6×10^{-6} per annum. Nonetheless the literature acknowledges that the fungal mitochondrial genome could constitute an avenue of escape for the construct into the environment, and this could be particularly relevant for mosquito control strategies using entomopathogenic fungi (Scholte et al., 2004; Khan et al., 2012).

Viruses and environmental bacteria are also identified within the literature as presenting potential routes of HGT (e.g. Keese, 2008), and in the developing world human exposure to bacteria is much higher than in the developed world. Vohra and Blakely (2013) for example notes that diarrhoea caused by the uptake of bacteria from contaminated water is the second major cause of death in children under five in Africa and South East Asia.

The literature review also recovered a series of patents that use germline cells as a mechanism for integrating nucleases and other genetic constructs into new genomes. Lavitrano et al. (2006), for example, note the intrinsic ability of sperm cells to bind and internalise exogenous DNA, and discuss this mechanism ("sperm mediated gene transfer") as a means to create transgenic animals. This suggests the possibility of HGT to non-target organisms that reproduce by shedding eggs and sperm within aquatic environments.

In summary the literature appears to emphasise well established potential hazards and in some instances points to the possibility of specific mechanisms that are pertinent to endonucleases, but overall the review did not highlight any hazards that were not already identified either through the FTA or HHM analysis.

2.2.2 Hierarchical Holographic Model

Hierarchical Holographic Modelling (HHM) is an inductive hazard analysis technique that examines the possible outcomes when risk generating activities interact with complex systems. It facilitates hazard analysis in large systems by adopting different perspectives on the problem and decomposing the system in sub-systems based on each of these perspectives. A complex landscape for example may be examined from a biological perspective and a physical perspective, and then decomposed into sub-systems such as the abiotic processes and the biotic components. These perspectives are then further sub-divided in a hierarchical fashion. The number of divisions and the resulting levels of the hierarchy is a subjective decision guided by the scope of the assessment and the resources available to the analysis. Once the hierarchy of different perspectives has been constructed the hazard analysis proceeds by examining, in a pair-wise manner, all possible interactions between the lowest elements of the hierarchy and the risk generating activity, and from this postulate possible cause and effect pathways that lead to undesirable outcomes (Haimes, 1981, 1998; Hayes et al., 2004).

In this risk assessment we used HHM, as a complement to the literature review and FTA, to help us identify hazardous outcomes that may occur following interactions between genetically modified mosquitoes and all relevant human, biological and environmental systems. In this context the HHM provides a transparent, structured and logical framework that helps to identify the cause-effect relationships that lead to potential hazards in each of these systems across multiple scales. In effect it provides a framework that encourages us to ask the question "what could go wrong?" from as many conceivable perspectives or points of view as possible.

The first step of the HHM analysis is to decide on the perspectives to be adopted and the subsequent decomposition of the system into a hierarchy of sub-systems. This step is critical because it defines the level of resolution for the ensuing identification of hazards. This choice of perspectives is guided by the terms of reference and geographic/temporal scope of the risk assessment. In this case the terms of reference were limited to the potential environmental and human health impacts associated with the accidental escape of transgenic mosquitoes, and that hazards must be interpreted in the context of the limited number of mosquitoes than can be housed within the African insectaries and further contextualised by their location (Section 2.1). The HHM perspectives were therefore limited to environmental and human health. Other potential risk assessment endpoints, such as political risk, socio-economic risks or risks to the project itself - i.e., the project technically failing to achieve it's goals or to being terminated before meeting these goals - were explicitly excluded from this analysis because the risk generating mechanism examined here is a one-time release of a self-limiting construct.

In consultation with members of the Target Malaria consortium, the environmental and human health perspectives were sub-divided into eight components and processes, which were then further divided into four to seven elements (Table 2.1). This hierarchical decomposition of the ecosystem and human health perspective results in a matrix of 729 pair-wise interactions between the elements of each category (Figure 2.1).

	Human processes	Biological components	Biological processes	Physical components	Physical processes	Chemical processes	Human hierarchy
Biological components	49						
Biological processes	42	42 (7)					
Physical components	28	28 (7)	24 (6)				
Physical processes	35 (7)	35 (7)	30 (6)	20 (4)			
Chemical processes	28 (7)	28 (4)	24 (4)	16 (4)	20 (4)		
Human hierarchy	28 (7)	28 (4)	24 (4)	16 (4)	20 (4)	16 (4)	
Human components	28 (7)	28 (4)	24 (4)	16 (4)	20 (4)	16 (4)	16 (4)

Figure 2.1: Hazard matrix listing number of potential pair-wise interactions between the elements of the ecosystem and human-health systems (Table 2.1) considered in the HHM analysis. Numbers in parentheses indicate grouped interactions where elements from one category of components or processes in Table 2.1 were grouped and compared simultaneously with individual elements from another category

The identification of hazards is structured around these pair-wise interactions. Each comparison considered during the elicitation was discussed by the workshop participants and recorded as a brief statement that described a specific hazard scenario.

The elicitation was conducted by members of the CSIRO Risk Team in a workshop setting with members of the Target Malaria consortium on the 28th and 29th April 2014. During the first day of the elicitation, however, it became clear that it would be more efficient to group elements from one category of components or processes and simultaneously compare them with individual elements from another category, rather than consider all 729 interactions individually. This grouping

Biological components	Physical components	Human hierarchy	Human components
1. genes	1. atmospheric- surface waters	1. individual	1. buildings
2. *Plasmodium*	2. soil	2. family	2. transport & infrastructure
3. bacteria, viruses, fungi	3. vegetative cover	3. social & community	3. permanent & temporary standing water
4. mosquitoes	4. topography	4. governance	4. crops & pasture
5. plants			
6. predators, prey & competitors			
7. other invertebrates & vertebrates			
Biological processes	**Physical processes**	**Chemical processes**	**Human processes**
1. survival & growth	1. wind movement	1. gene expression	1. lab quality control, monitoring
2. reproduction	2. water movement	2. bio- accumulation	2. mosquito vector control
3. movement	3. storms, weather, temperature	3. synthesis	3. unauthorized procedures, criminal activity & war
4. predation, competition	4. fire	4. degradation	4. manufacturing & trade
5. disease transmission	5. earthquake		5. agriculture
6. mutation & selection			6. media
			7. annual & seasonal cycles

Table 2.1: Components and processes of the human health and ecological systems considered in the HHM analysis

of the interaction elements resulted in a total of 244 comparisons from the revised matrix of interactions (Figure 2.1).

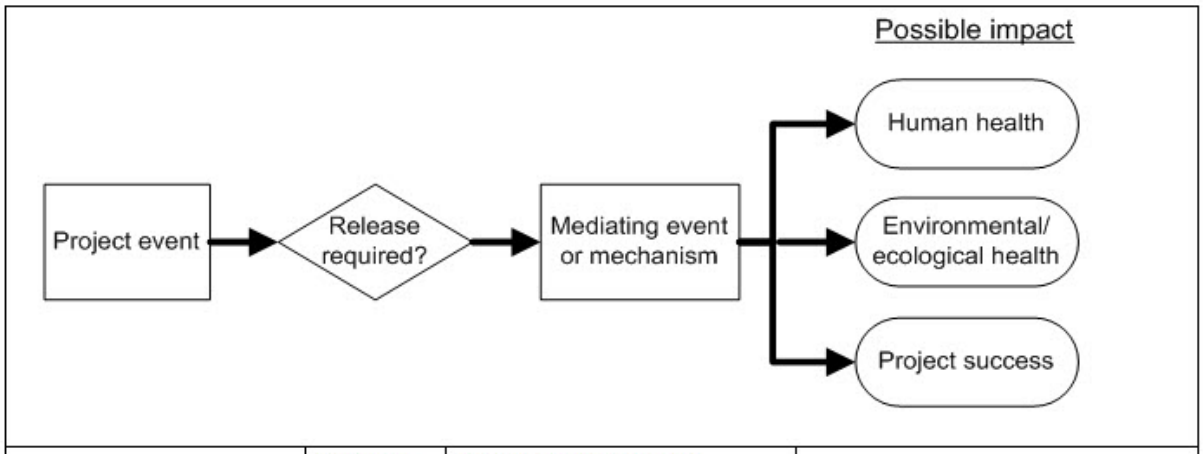
An example of a question posed from a grouped set of interactions was: *"Is there a conceivable harm to human or environmental and ecological health that could arise following an escape of 10,000 laboratory-reared mosquitoes from interactions between any of the four chemical processes of 1) gene expression, 2) bio-accumulation, 3) synthesis, or 4) degradation, somehow interacting with buildings (a human component)."* The actual response for this question is hazard scenario 96 in Appendix B.

Following the workshop, members of the CSIRO risk team examined the list of elicited hazard scenarios to identify generalities in the risk endpoints implied by the hazard, such as risks to human health, ecological impacts and project risks, and commonalities between descriptions of the causal pathways described in each scenario. The generalities identified in the implied risk endpoints were then used as a basis for the subsequent scope of the fault tree analysis (Section 4).

While potential threats to project success were explicitly excluded from the risk assessment's terms of reference, and therefore outside of the project scope, many of the hazard scenarios were interdependent with project operations, and were carried forward in the analysis. In some scenarios aspects of project success, such as the ability to monitor or control local populations of mosquitoes, were included as an intermediate step leading to an impact on human health, while in other scenarios an impact to project success was itself identified as a final outcome.

The next step of the analysis was to collate the elicited list of potential hazard scenarios based on similarities in their cause-effect relationships. Each potential hazard scenario was parsed into a four-step causal-effect template (Figure 2.2) and given attributions for project event, requirement of an escape of laboratory-reared mosquitoes, mediating event or mechanism, and possible impact to human health, environmental and ecological health, or project success. The list of scenarios was then hierarchically sorted and grouped based on: (i) risk assessment endpoint (impact), (ii) project-related initiating event; and, (iii) whether or not an escape was required. The sorting exercise helped to reveal common cause-effect relationships that arise from the various (and sometimes disparate) component-process interactions that catalysed the identification of the scenarios in the first place.

Particular cause-effect relationships were sometimes invoked in multiple comparisons, but with elaborations based on novel circumstances suggested from the HHM analysis. These hazard scenarios were kept separate where there were distinct differences in associated events or contexts. Finally, the mediating events or mechanisms associated with each group of potential hazard scenarios were used to describe a hazard summary statement that emphasises the commonalities in the cause-effect relationships between the initiating event and the undesired outcome.



Project event	Release require?	Mediating event or mechanism	Possible Impact
1. Release of mosquitoes (many) 2. Laboratory practices and environment, & laboratory reared mosquitoes (which are somehow different) 3. Mosquitoes genetically modified 4. Laboratory presence	yes/no	1.0. Increase vectorial capacity (FT1) 1.1. Increase in local abundance of mosquitoes 1.2. Increase in female mosquito life expectancy 1.3. Increase in female mosquito biting rate 2. Increased plasmodium infection 3. Mosquitoes infected with other disease (FT2) 4. Mosquitoes become cryptic 5. HGT (FT3 & FT4) & VGT (FT5) 6. Pesticide resistance 7. Humans change behaviour/susceptibility 8. Increased use of pesticides 9. Decreased vector control 10. Change in reproductive behaviour of mosquito	1.0. Human health 1.1. Malaria (FT1) 1.2. Novel disease (FT2) 1.3. Allergic reaction or toxicity 1.4. Economic & emotional stress 2.0. Environmental/ecological health 2.1. Reduced invertebrate population (FT4) 3.0. Project success 3.1. HGT (FT3 & FT4) 3.2. Mosquitoes difficult to monitor 3.3. Mosquitoes difficult to control 3.4. Male sterility/transgene spread (FT5)

Figure 2.2: Template and list of cause-effect categories that were used to attribute, group and interpret hazard scenarios identified by the HHM analysis. Mediating effects and possible impacts (endpoints) that were subsequently addressed in the fault tree analysis are shown by tree references in parenthesis (Note: FT1 refers to change in vectorial capacity)

.

Summary of cause-effect relationships for each group of potential hazard scenarios	No. scenarios	Project event	Release required	Mediating events or mechanisms	Impact
1. Release of large number of mosquitoes leads to increase chance of malaria, especially where environment enhances vectorial capacity, or humans are more susceptible or change their behaviour and interaction with mosquitoes.	22	1	yes	1.1, 1.3, 2, 7	1.1
2. Insectary practices contribute to malaria infection either from change in phenotype (e.g., better adapted to receiving environment) or malaria introduced into the insectary population.	6	2	yes	1.1, 2	1.1
3. Release of GM mosquitoes leads to increase chance of malaria in humans due to GM mosquitoes having 1) a greater vectorial capacity, by virtue of enhanced survival, life expectancy, or biting habit, or 2) an increased rate of infection with malaria *Plasmodium*, or 3) a change in behaviour in humans that either increases their interaction with infected GM mosquitoes or 4) decreases the effectiveness of vector control measures.	15	3	yes	1.0, 1.2, 1.3, 2, 7, 9, 10	1.1
4. Presence of insectary itself causes change in human behaviour that leads to increased interaction with mosquitoes, or aspects of local environment or seasonal weather patterns lead to increased vectorial capacity or human-mosquito interactions.	12	4	no	1.0, 1.1, 1.3, 7, 9	1.1
5. Factors exogenous to project or operation of insectary cause an increase of malaria in humans due to increased vectorial capacity or interaction between humans and mosquitoes.	8	none	no	1.1,7, 9	1.1
6. Insectaries may be located in environments where local human population has an increased risk of exposure to novel pathogens, which may be transmitted by escapee mosquitoes.	3	1	yes	3, 7	1.2
7. GM mosquitoes have a different biting habit or susceptibility to other pathogens leading to spread of non-malaria disease, and this effect can be enhanced where there has been local depletion of invertebrates from spraying program.	6	3	yes & no	1.3, 3, 5, 10	1.2
8. GM mosquito bite has enhanced allergic reaction or is toxic if ingested by humans.	1	3	yes		1.3
9. Presence of insectary or spraying program negatively affects local economy.	2	1, 4	yes & no	8	1.4
10. Spraying program associated with release reduces local invertebrate populations.	2	2	yes	7, 8, 9	2.1
11. Invertebrate population suppressed due to spraying program that is intensive specifically because it is associated with GM mosquitoes that have changed behaviour, or from HGT eradicating other members of a species complex.	3	3	yes	1.3, 5, 8	2.1
12. Miscellaneous factors possibly associated with HGT.	17	1	yes & no	5	3.1

Summary of cause-effect relationships for each group of potential hazard scenarios	No. scenarios	Project event	Release required	Mediating events or mechanisms	Impact
13. Social observances or customs, or seasonal change in weather or change in climate affect ability to effectively monitor mosquito populations.	2	1	yes	1, 7	3.2
14. Insectary reared mosquitoes have different behaviour or life history attribute that makes them difficult to monitor if released.	2	2	yes	1, 4	3.2
15. Monitoring of GM mosquitoes is confounded by use of fluorescent dyes or other aspects of local industrial or urban environments.	1	3	no	1, 5	3.2
16. Monitoring of mosquitoes is made difficult due to local environment, weather conditions or civic unrest.	4	4	no	7	3.2
17. Monitoring and control program compromised by local governance, unlawful inference with the insectary program or aspects of local environment associated with access or native mosquito population.	9	4	yes & no	1.1, 7, 9	3.2, 3.3
18. Escaped GM mosquitoes could backcross with wild-type mosquitoes that have become insecticide resistant due to spraying program or other intensive uses of insecticides (e.g., agriculture). Seasonal variation in use of insecticide could lead change in level of this resistance. Offspring of this backcross could have increased fitness.	4	3	yes & no	6, 9	3.3
19. Change in local environmental, biological or social conditions lead to breakdown in effectiveness of spraying program to control escaped mosquitoes.	5	4	no	6, 9	3.3
20. Spread of transgene due to favourable receiving environment.	1	1	yes	1	3.4
21. Insectary populations of mosquitoes have non-sterile males due to knockout mutation, a breakdown or change in laboratory procedures that is undetected, or enhanced fitness that leads to VGT.	5	2	yes & no	1.1, 5, 10	3.4
22. Miscellaneous scenarios with no discernible impacts to human health, environment or project.	9	2, 3, 4	yes & no	9	none

Table 2.2: Summary of the HHM Hazard scenarios aggregated into 22 groups based on a hierarchical sorting of: (i) impact, (ii) project event; and, (iii) release required, which are elements of the cause-effect template of Figure 2.2.

The HHM analysis identified a total of 139 potential hazard scenarios that might be possible risk sources (Appendix B). The most frequently identified potential hazards involved impacts on human health (75 scenarios) and then on project success (50 scenarios). Surprisingly, only five scenarios were identified that led to environmental impacts, while nine others were described that had no discernible impact to human health, the environment or project success. Eighty nine of the scenarios required the release of mosquitoes from the insectary, whereas 49 did not; one scenario could have occurred with or without a release. The potential hazard scenarios were subsequently aggregated into 22 groups based on similar cause-effect relationships. For these we derived summaries of the cause-effect relationships common to each group (Table 2.2). The following is a brief description of these groups.

The HHM analysis identified 63 scenarios that led to the possibility of an increased incidence of malaria in human populations surrounding the insectaries. These scenarios were typically predicated on the release of a large number of mosquitoes leading to an increase in the rate of transmission of malaria, especially where changes in the local environment or changes in the behaviour of the human population acted to increase vectorial capacity.

Laboratory practices within the insectaries were also noted as having the potential to increase transmission of malaria, either through changes to the mosquito's phenotype or genotype (i.e., via genetic modification) or via the inadvertent introduction of malaria into the insectary population of mosquitoes by a technician. In this context, however, it is important to recognise that potential hazards were identified freely and without reference to the containment, biosafety and *Plasmodium* screening policies that will be implemented within the insectaries and would mitigate against this possibility.

The mere presence of the insectary itself, even without an accidental release of mosquitoes, was noted as having the potential to increase the incidence of malaria if, for example it created a perception that a cure or solution was imminent among the local population which in turn encouraged a change in their behaviour, which subsequently led to an increase in human-mosquito interactions.

The analysis also noted that the local incidence of malaria might change for many reasons, totally independent of the project. These scenarios typically involved changes to the local environment (e.g. changes in agricultural practices, street lighting, and weather) or human behaviour (e.g. use of bed nets and effect of other disease vectors).

The analysis identified nine scenarios that led to the potential increase in transmission of novel pathogens (that is novel with respect to transmission by species in the *An. gambiae* complex). These scenarios arise as a result of the location of the insectary leading to an increased risk of exposure to a novel pathogen, or from the genetic modifications made to the mosquitoes that creates the ability to transmit a novel pathogen.

One scenario describes the potential for bites from a genetically modified mosquito to invoke an allergic reaction in humans or is otherwise toxic if ingested. Other types of impacts on human health were also described in two scenarios where the presence of the insectary or a post-escape spraying program negatively affects the local economy.

Potential ecological impacts were involved in only five potential hazard scenarios, and all of these describe the possibility of a reduction in local populations of invertebrates, either as an incidental consequence of a post-escape spraying program designed to kills escapees, or by species being eradicated as a consequence of horizontal gene transfer. Gene-by-environment

(G×E) interactions were described as possibly being involved in three hazard scenarios, but none could be ascribed to any definitive mechanism or impact.

Of the factors that were identified as potentially compromising project success, 17 scenarios included horizontal gene transfer, 18 included the inability to effectively monitor mosquitoes (which included mediating effects from the local environment, local governance, social customs and aspects of laboratory practices), and 18 more included the loss of the ability to control local mosquito populations. The last included interference with, or diminishment of, the project's monitoring and control programs, which included mosquitoes gaining resistance to insecticides. Another impact to project success involved vertical gene transfer, which in six scenarios was described as being associated with environmental conditions that facilitated hybridisation or the occurrence of non-sterile males in insectary populations due to the failure of the construct when integrated into local genotypes.

The potential hazard scenarios summarised above provide a systematic appraisal of the potential sources of risk for the project. The likelihood of endpoints directly relevant to five of the impacts identified in this process – an increase in malaria transmission, vectoring novel blood-based diseases, spread of the construct in non-target eukaryotes and non-eukaryotes via horizontal gene transfer and spread of the construct in the *An. gambiae* complex via vertical gene transfer – are quantified in the subsequent risk assessment. These five endpoints account for the impacts identified in 95 (73%) of the 130 potential hazard scenarios for which a specific impact was described.

Of the 62 scenarios that are not carried forward in the risk assessment, all are considered to be out of the scope of assessment. In particular impacts associated with allergic reaction or toxicity to humans (one scenario), economic or emotional stress (two scenarios), reduced invertebrate populations following a post-escape spraying programme (five scenarios) and a decreased ability to control (nine scenarios) or monitor (nine scenarios) escaped mosquitoes, have been deliberately excluded. Again here the containment policies implemented within the insectaries are designed to prevent escape and mitigate the need for post-escape control or monitoring.

One potential hazard scenario, from which no impact was discerned (Number 132, Appendix B), describes a mechanism whereby wild type mosquitoes brought into the insectaries introduce unknown genes, say through viral gene sequences, into the insectary population. If these sequences become integrated in the genome of genetically modified mosquitoes, then that could be a means for subsequent HGT. This issue is addressed through fault trees 3 and 4 (Section 4)

While the primary mechanism for horizontal gene transfer is the process of remobilisation of the transgene, for example through exogenous transposase, the HHM (and the literature review) noted a variety of environmental and biological factors that might facilitate this process. These factors included the effects of temperature and humidity on DNA degradation rates, prokaryote (soil or other microbial communities) and eukaryote (e.g. fungi) facilitated mechanisms, digestion of GM mosquitoes by invertebrates and vertebrates and local concentrations of GM mosquitoes in water bodies. These factors have been incorporated into the structure of fault tree 3.

A final cause-effect relationship described during the HHM workshop (number 119, Appendix B) involves the possibility of inbreeding and selection differentially affecting laboratory strains of mosquitoes, which over time could give rise to heterotic males that compromise the sterility of the insectary population by "rescuing male reproductive characteristics" (see for example Baeshan et al., 2014).

2.3 Risk assessment endpoints

In risk assessments for complex systems it is not unusual for a natural tension to arise between the assessment endpoints that are closest to community values and concerns, and the assessment endpoints that can be quantified or otherwise incorporated into the assessment with the available resources. Suter (1990) for example notes that risk assessment endpoints must be valued by society but they need not be the "ultimate" values rather they are the "highest" values that can be formally assessed. Similarly Hayes (2003a) emphasizes that endpoints that are further away from the start of risk-generating event chains are likely to be more difficult to assess and will likely have higher levels of uncertainty. Hence the choice of assessment endpoints must judiciously balance community concerns and the need to control the size and complexity of the overall assessment.

The hazard analysis completed here suggests that the risk assessment can be made considerably more tractable by focusing on five assessment endpoints:

1. An increase in the malaria vectorial capacity of genetically modified mosquitoes.
2. Transmission of a novel (i.e. not previously known to be vectored by *An. gambiae*) blood-borne pathogen to human or vertebrate host.
3. Spread of the I-PpoI construct in non-target eukaryotes.
4. Spread of the I-PpoI construct in non-eukaryotes.
5. Spread of the I-PpoI construct in the *An. gambiae* complex.

The first endpoint addresses the hazard of a potential increase in malaria cases attributable to mosquitoes that have escaped from the insectary. The second endpoint addresses the hazard that the genetic modification and/or laboratory rearing practices somehow enable the genetically modified mosquitoes to transmit a novel blood-based pathogen. The third and fourth endpoints address the potential for horizontal gene transfer leading to the acquisition and spread of the construct in non-target eukaryotes and non-eukaryotes.

An important assumption here is that spread of the construct in non-target eukaryotes and non-eukaryotes is a pre-cursor to harmful consequences irrespective of the species concerned, and these endpoints in particular make the risk assessment tractable. Conversely any attempt to model the consequences of either of these two events in terms of, for example, the impact on the population abundance of any particular species, would lead to an impractical and overly complex assessment. We must recognise, however, that these endpoints are conservative because HGT will not always lead to harmful outcomes.

The last endpoint addresses the potential hazard that the construct fails to induce sterility in male mosquitoes, leading potentially to its persistence and spread in populations of species in the *An. gambiae* complex. This is not a pre-cursor to a harmful event, indeed reducing populations of species in the *An. gambiae* complex is one way to reduce the transmission of the malaria parasite. This endpoint is, however, considered undesirable at this stage of the development pathway of this new genetic technology because the current construct is supposed to be inherently self-limiting.

In the analysis that follows this endpoint is clarified by separating it into: (i) spread of the I-PpoI construct in the *An. gambiae*; and (ii) spread of the I-PpoI construct in *An. coluzzii* or *An. arabiensis* because the latter are thought to be the only species in the vicinity of the insectaries

that *An. gambiae* can hybridize with (*pers. comm.* A. Burt).

During the scoping stage of the assessment four potential endpoints were deliberately excluded. Socio-economic and political endpoints were considered to be outside the project's terms of reference because the risk generating mechanisms focuses on a one-time escape of a self-limiting construct. Similarly the risk of project failure was excluded because the Target Malaria consortium addresses this as part of its long term, and day to day, project management processes. Allergic and/or toxic responses are not included because these endpoints will be assessed by established protocols, and the incidental impacts associated with spraying in the event of a complete loss will be addressed by the consortium as part of its internal risk mitigation planning.

KEY POINTS: CHANGE IN VECTORIAL CAPACITY

- The primary objective of the vectorial capacity analysis is to assess the probability that newly escaped unmodified and modified (with the I-PpoI construct) G3 laboratory strain female mosquitoes have a higher malaria vectorial capacity than wild type female mosquitoes.
- The assessment is based on a direct elicitation of the parameters in the derivation of vectorial capacity given by Garrett-Jones (1964a) which we generalise by incorporating transmission efficiencies to improve the ability to capture how vectorial capacity may change with different vectors.
- To complete the assessment we define an index of intrinsic malaria risk, defined as the probability of increased risk of malaria transmission following a bite by a G3 or I-PpoI mosquito compared to a bite by a wild type mosquito.
- We provide bounds on the risk by comparing estimates that assume independence between G3, I-PpoI and wild type strains versus estimates that are derived after imposing strong positive dependence on the magnitudes of the vectorial capacity parameters across strains, but we suggest that strong positive dependence is the more reasonable assumption based on comments made by experts during the direct elicitation.
- If we assume strong positive dependence, the analysis indicates a 41% or 23% reduction in the risk of G3 or I-PpoI mosquitoes (respectively) producing an infectious bite compared to local wild type mosquitoes.

3 CHANGE IN VECTORIAL CAPACITY

The primary objective of the vectorial capacity analysis is to assess the probability that newly escaped unmodified and modified (with the I-PpoI construct) G3 laboratory strain female mosquitoes have a higher vectorial capacity than wild type female mosquitoes. Only one generation of escapees is considered here. Vectorial capacity is composed of several parameters that likely vary among the three strains (wild type, G3 and modified G3) of mosquitoes. The likely magnitudes of these parameters were therefore elicited from domain experts. The independent experts were asked a series of questions developed in conjunction with the Target Malaria consortium. The questions were designed to help the consortium parameterise a model of additional malaria cases due to an accidental escape from an insectary which is a more detailed model than vectorial capacity. The subset of these questions that are relevant to the vectorial capacity analysis conducted here are discussed below.

3.1 Relevance of vectorial capacity

The seminal text of Ross (1911, Addendum 66) provides the foundation for a mathematical description of malaria dynamics. His analysis distils the relationship between a competent malaria vector and its human hosts down to a small set of essential parameters. Macdonald (1952) used these parameters to propose a threshold index $\mathcal{R}_0$, which we refer to as the "basic reproduction number" (although this term is not universally applied in the literature). Malaria persists in the long term if $\mathcal{R}_0 > 1$ but not otherwise. Garrett-Jones (1964a) simplified this index by defining "vectorial capacity" $\mathcal{V}$, which ignored the duration of infectiousness in the host and assumed perfect transmission. It was argued that this index was more relevant than $\mathcal{R}_0$ when assessing how vector control impacts malaria transmission (Garrett-Jones, 1964a,b). These mathematical analyses formed the basis for early efforts by the World Health Organization to control malaria (Bailey, 1982) and continue to form the basis of modern models of vector-borne disease dynamics (Reiner et al., 2013).

Despite the continuing importance of $\mathcal{R}_0$ and $\mathcal{V}$, Reiner et al. (2013) note several problems with the indices. There are several convenient but perhaps overly simplistic assumptions, such as homogeneity of populations and constant mortality rate. Departures from these assumptions introduce uncertainty into how $\mathcal{V}$ corresponds to malaria incidence (e.g., Dye, 1986; Smith et al., 2010) and perhaps explain why $\mathcal{R}_0$ historically does not perform well in endemic areas (Dietz et al., 1974; Smith and McKenzie, 2004).

The desire to encompass greater realism has lead to much more complicated models of malaria dynamics, and yet the difficulty of assessing the uncertainty in the underlying parameters means that it is difficult (and uncommon) to empirically estimate even the comparatively few parameters that compose $\mathcal{R}_0$ and $\mathcal{V}$. The problem is more difficult for more complex models that arguably encompass greater realism but at the cost of having to specify many more parameters. It is perhaps the simplicity and generality of $\mathcal{R}_0$ and $\mathcal{V}$ that explains its continued foundational role in malaria modelling.

Here we take $\mathcal{V}$ to be an index of transmission risk for a newly escaped population of female mosquitoes that are free of malaria. We do not attempt to incorporate the effects of heterogeneity of populations, different survival models and so on, by proposing a more complex model structure. Instead, we retain the simple well-known structure of $\mathcal{R}_0$ and $\mathcal{V}$ and elicit the uncertainty in the underlying parameters from experts who are familiar with these entomological

Parameter	Definition
a	Number of bites on humans each day per mosquito
b_{yx}	Transmission efficiency from humans to female mosquitoes
b_{xy}	Transmission efficiency from female mosquitoes to humans
μ	mortality rate of mosquitoes (day^{-1})
n	duration of extrinsic incubation period in days
m	Ratio of numbers of female mosquitoes to humans
r	rate of recovery from infectiousness by infected humans (day^{-1})

Table 3.1: Parameters and variables for the classic models of malaria dynamics and statics. The experts were not asked to elicit the vector-host ratio m or the human recovery rate r.

features. $\mathcal{V}$ is, however, generalised to account for uncertainty in the transmission efficiencies.

The derivation of vectorial capacity by Garrett-Jones (1964a), which only considers female mosquitoes, follows the heuristic derivation by Macdonald (1952, 1957, Appendix I) for a human population free of malaria

$$\begin{aligned} \mathcal{V} = \quad & am && \text{Number of human bites per day} \\ & \times e^{-\mu n} && \text{Proportion surviving } n \text{ days} \\ & \times 1/\mu && \text{Expected remaining life} \\ & \times a && \text{Mosquito bites per day over expected remaining life,} \end{aligned} \tag{3.1}$$

where the parameters are given in Table 3.1. The product am is also referred to as the human biting rate. The proportion of female mosquitoes that survive n days, which is the length of the extrinsic incubation period, is sometimes presented as p^n, where p is the probability of a mosquito surviving one day. In that case, the expected remaining life (also referred to as mean residual life in survival analysis) is given as $-1/\log(p)$. Both presentations assume an exponential survival function. From Equation (3.1), $\mathcal{V}$ gives the number of infectious bites by female mosquitoes after an accidental insectary escape assuming perfect transmission efficiencies between vector and host.

Incorporating transmission efficiencies can improve the ability to capture how vectorial capacity may change with different vectors (Dye, 1986). Thus we generalise $\mathcal{V}$ such that,

$$\mathcal{V}_0 = b_{xy} b_{yx} \mathcal{V}, \tag{3.2}$$

where the additional transmission efficiency parameters are defined in Table 3.1. From Equations. (3.1) and (3.2), $\mathcal{V}_0$ gives the number of infectious bites by female mosquitoes after an accidental insectary escape assuming imperfect transmission efficiencies between vector and host.

3.2 Probability of increased relative malaria risk

$\mathcal{V}_0$ is arguably more relevant to the case of an accidental insectary escape than $\mathcal{R}_0$ because the human recovery rate, which only appears in $\mathcal{R}_0$ is probably not very important in this scenario. For example, this could be the case for a constant background prevalence of malaria in human hosts over the course of the first generation of escaped insectary mosquitoes. However, the two indices are closely related. If $\mathcal{V}_0$ is multiplied by the average recovery time of a human malaria case ($1/r$, see Table 3.1), then the number of new infections is given by,

$$\mathcal{R} = (1 - y_p)\frac{\mathcal{V}_0}{r}, \tag{3.3}$$

which goes to the basic reproduction number $\mathcal{R}_0$, or the number of new infections delivered by newly infected mosquitoes, as the proportion of previously infected mosquitoes y_p goes to zero (Macdonald, 1957, Appendix I). Equation. (3.3) generalises the definition of (Macdonald, 1957, Appendix I) to allow for imperfect transmission efficiency from an infected human host to a vector. Clearly $\mathcal{R}_0$ is a function of vectorial capacity.

There is some disagreement in the literature, however, on whether $\mathcal{R}_0$ as defined above is really the square of the basic reproduction number (defined as the number of secondary infections arising from a primary individual) (e.g., see Diekmann and Heesterbeek, 2000; van den Driessche and Watmough, 2002). Smith and McKenzie (2004), for example, give a set of coupled ordinary differential equations that result in the threshold condition such that $\mathcal{R}_0$ must be greater than 1 for malaria to become endemic following the introduction of an infected host to a susceptible population. Using the definition of the basic reproduction number given by van den Driessche and Watmough (2002), these equations then result in an alternative basic reproduction number that is defined as $\mathcal{R}_0' = \sqrt{\mathcal{R}_0}$ (Hosack et al., 2008). This alternative definition could have a large effect on the magnitude of the resulting estimate of the basic reproduction number, although the threshold condition is unchanged.

Here we define an endpoint, in Equation (3.4), such that the complication of competing definitions for $\mathcal{R}_0$, and $\mathcal{V}_0$ by extension, is unimportant. Both definitions of the basic reproduction number lead to the same probability of increased risk of malaria transmission compared to wild type. The probability of increased risk of malaria transmission by a G3 mosquito compared to wild type is given by,

$$\begin{aligned} P\left(\mathcal{R}_0^G > \mathcal{R}_0^W\right) &= P\left(\frac{{a^G}^2 b_{xy}^G b_{yx}^G m e^{-\mu^G n^G}}{r\mu^G} > \frac{{a^W}^2 b_{xy}^W b_{yx}^W m e^{-\mu^W n^W}}{r\mu^W}\right) \\ &= P\left(\frac{{a^G}^2 b_{xy}^G b_{yx}^G e^{-\mu^G n^G}}{\mu^G} > \frac{{a^W}^2 b_{xy}^W b_{yx}^W e^{-\mu^W n^W}}{\mu^W}\right), \end{aligned} \tag{3.4}$$

and a similar relation for I-PpoI is given by substituting I for G in the above equation.

The relevance of Equation (3.4) to an accidental insectary escape is as follows. The probability that G3 or I-PpoI produce more infectious bites compared to wild type mosquitoes avoids considering the human recovery rate r because this does not depend on the strain of mosquito. This probability also does not depend on the vector-host ratio m because the scenario is conditioned on the same number of escaped mosquitoes for each strain. The probability similarly does not depend on the baseline number of human hosts infected by malaria because this also does not

depend on the strain of mosquito; recall that the basic reproduction number is defined for the situation where the number of infected human hosts approaches zero (e.g., Macdonald, 1957). The indices proposed are thus closely aligned with the risk endpoints, perhaps more closely than the use of $\mathcal{R}_0$ in areas of ongoing malaria transmission in any case.

A related endpoint is the difference in risk of an infectious bite by G3 or I-PpoI compared to wild type. Define the intrinsic index of malaria risk for strain t by $\mathcal{I}^t = {a^t}^2 b_{xy}^t b_{yx}^t e^{-\mu^t n^t}/\mu^t$, where the superscript $t = W, G, I$ indicates whether a parameter is specified for a female wild type, G3 strain or I-PpoI mosquito respectively. Then the difference in risk between G3 and wild type strains, for example, is given by,

$$\mathcal{R}_0^G - \mathcal{R}_0^W = (m/r) \times \left[\mathcal{I}^G - \mathcal{I}^W\right] \tag{3.5}$$

$$\propto \mathcal{I}^G - \mathcal{I}^W, \tag{3.6}$$

such that the difference in the intrinsic indices $\mathcal{I}^G - \mathcal{I}^W$ is proportional to the difference in risk of malaria transmission between strains. Moreover, because $\mathcal{V}_0^t = m\mathcal{I}_0^t$, the same difference is also proportional to the difference in the risk of infectious bites by escaped insectary female mosquitoes.

3.3 Expert elicitation

The CSIRO risk team elicited subjective probability distribution for the parameters appearing in Equation (3.4) for the three strains of mosquitoes (wild type, G3 and I-PpoI) with a number of independent experts. In total, 48 elicitations were contributed by 5 experts for the target parameters in Equation (3.4). Experts are only included in these results if they responded to all three strains for a given parameter.

The number of bites made by a mosquito each day is given by the parameter a. The elicited probability density functions are shown in Figure 3.1. In Figure 3.1 and subsequent figures, the plotted curves are derived from a kernel density estimate based on 10^5 Monte Carlo samples drawn from each of the elicited probability density functions. In addition, a linear pool or equally weighted mixture density of experts is also plotted.

The transmission efficiency from humans to mosquitoes is given by the parameter b_{yx}. The elicited probability density functions are shown in Figure 3.2. The transmission efficiency from mosquitoes to humans is given by the parameter b_{xy}. The elicited probability density functions for this parameter are shown in Figure 3.3.

The mortality rate is given by the parameter μ and has units of day^{-1}. The experts, however, answered the question that targeted this parameter in different ways. Experts 1 and 4 chose to elicit the probability of daily mortality p_m, which defines the mortality rate as $\mu = -\log(1 - p_m)$. Expert 2 chose to elicit the probability of daily survival p_s, which defines the mortality rate as $\mu = -\log(p_s)$. By specifying constant daily survival rate (albeit with uncertainty), these elicitations are consistent with the assumption of an exponential survival model in $\mathcal{R}_0$ and $\mathcal{V}_0$. Expert 5, however, chose to elicit lifespan under a lognormal distribution. The lognormal distribution implies a lognormal survival model with non-constant mortality rate. Thus this expert's implied survival model had greater complexity than assumed by the logic that underlies the indices $\mathcal{V}$ and $\mathcal{R}_0$. To conform with the assumed exponential survival model and constant mortality rate, this elicitation is interpreted as average lifespan or life expectancy, which corresponds to assuming that the inverse of the mortality rate follows a lognormal distribution with location m and scale σ, $1/\mu \sim LN(m, \sigma^2)$. The resulting elicitations for mortality rate are shown in Figure 3.4.

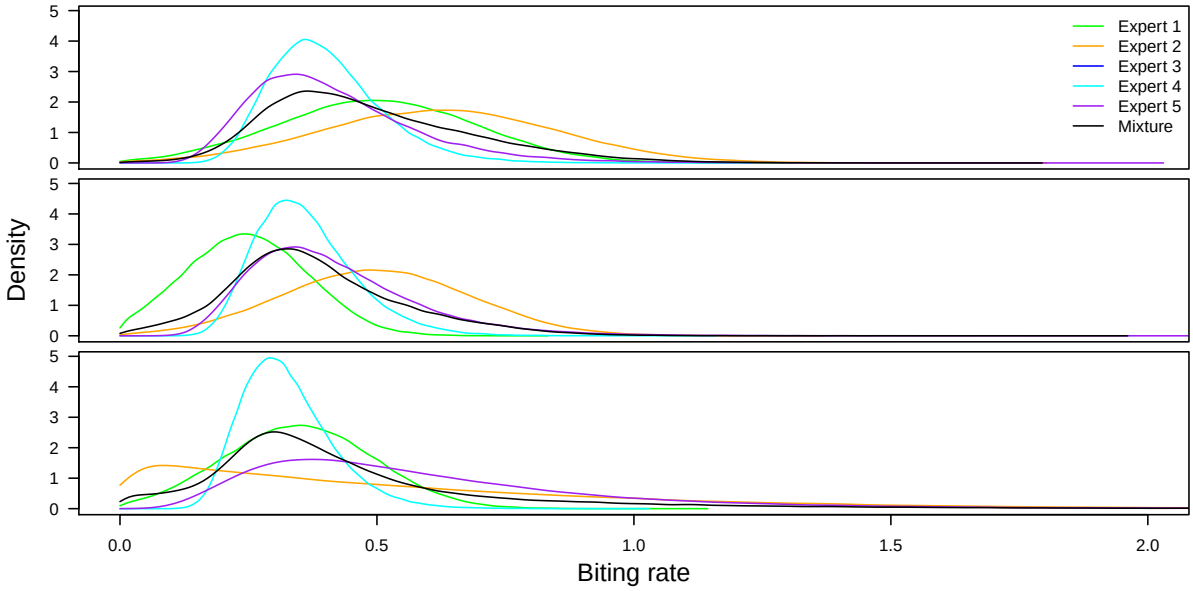


Figure 3.1: Biting rate (parameter a) for wild type (top), G3 (middle) and I-PpoI (bottom). Expert 3 did not contribute assessments to these questions. The linear pool, which is an equally weighted mixture density of the contributing experts, is shown in black.

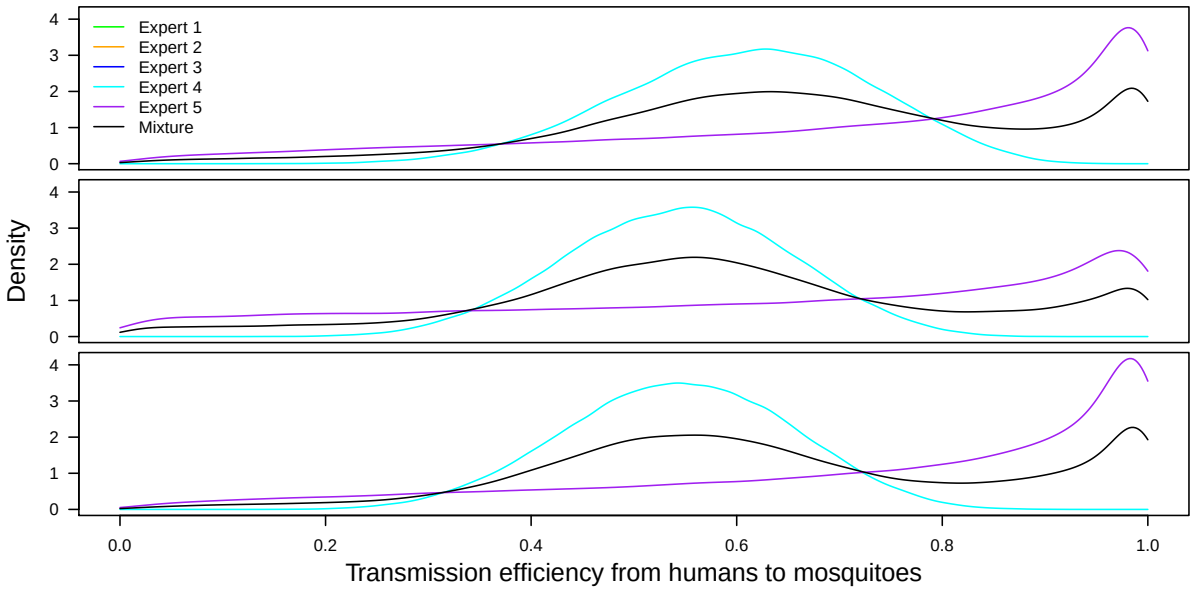


Figure 3.2: Transmission efficiency from humans to mosquitoes (parameter b_{yx}) for wild type (top), G3 (middle) and I-PpoI (bottom). Only experts 4 and 5 contributed assessments to all strains. The mixture density is a linear pool of these contributing experts.

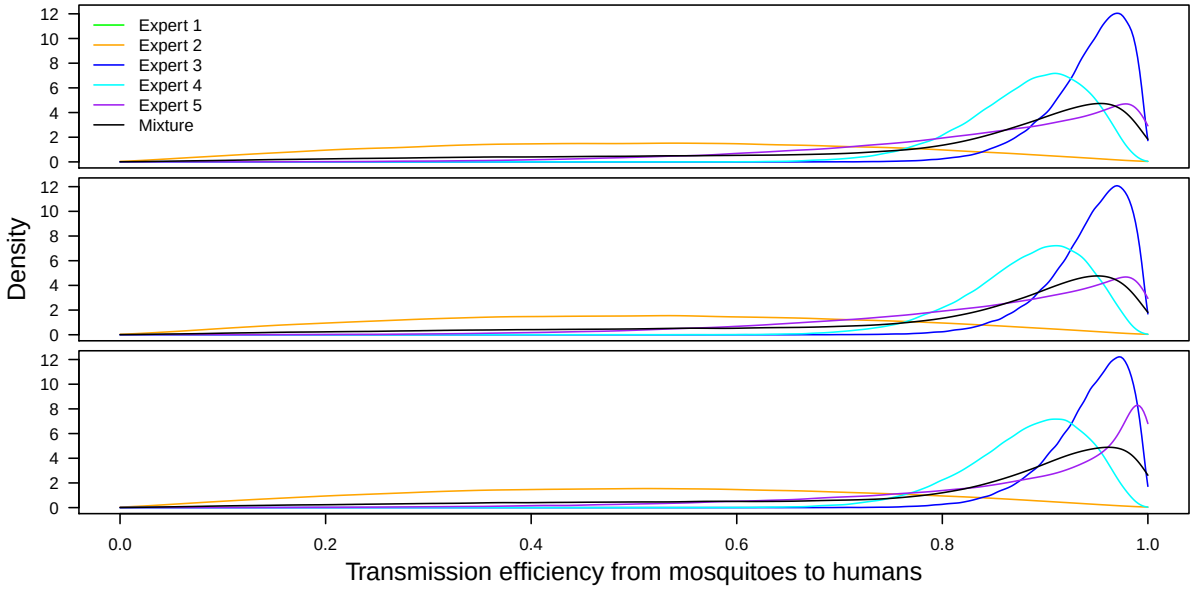


Figure 3.3: Transmission efficiency from mosquitoes to humans (parameter b_{xy}) for wild type (top), G3 (middle) and I-PpoI (bottom). The mixture density is a linear pool of the experts that contributed assessments to all strains.

The length of the extrinsic incubation period in days is given by the parameter n and has units of days. The elicitations for this parameter are shown in Figure 3.5.

3.4 Endpoint definition

Let $\boldsymbol{\theta}$ denote the vector of unknown parameters with an arbitrary entry denoted θ_j that corresponds to a parameter from Table 3.1 for one of the three strains. We indicate the conditioning on expert by writing the expert's prior belief for θ_j as $p(\theta_j|i)$, where $i = 1, \ldots, I_j$ for the I_j experts that contributed assessments to all strains for that parameter. For each parameter we create a linear pool of expert opinion (see review by Genest and Zidek, 1986) such that

$$p(\theta_j) = \sum_{i}^{I_j} w_i^{(j)} p(\theta_j|i)$$

where $\sum_i^{I_j} w_i^{(j)} = 1$ for non-negative weights $w_i^{(j)}$. We give equal weight to each expert so that $w_i^{(j)} = \frac{1}{I_j}$. We assume the expert elicitations on different parameters are independent, such that the joint prior distribution on all parameters is

$$p(\boldsymbol{\theta}) = \prod_{j}^{J} p(\theta_j) = \prod_{j}^{J} \left[\sum_{i}^{I_j} w_i^{(j)} p(\theta_j|i) \right]. \tag{3.7}$$

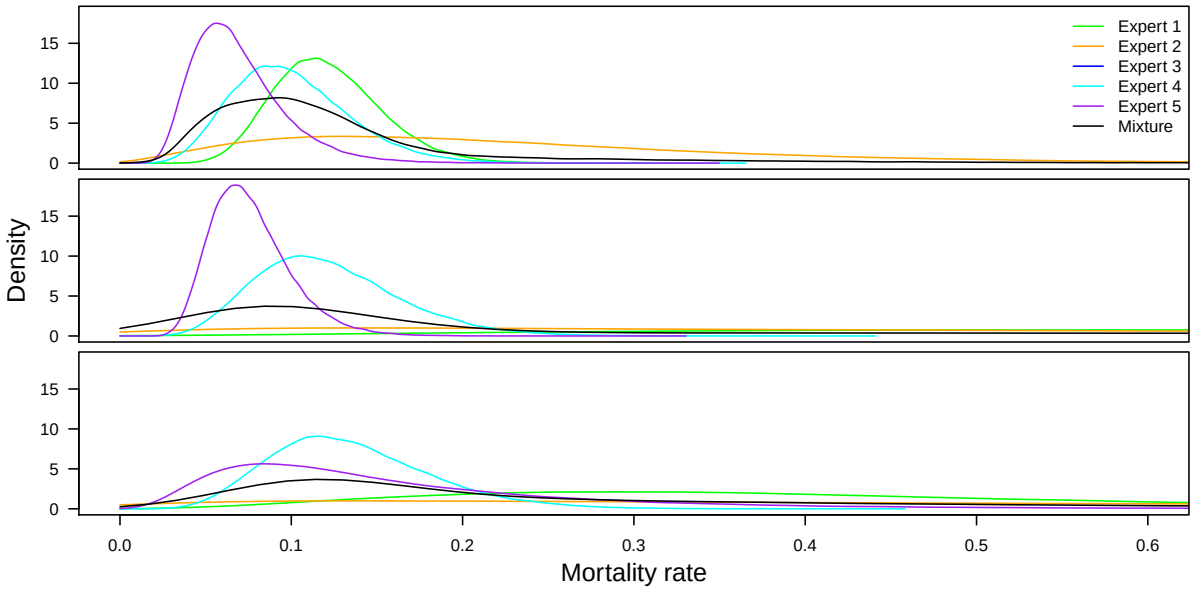


Figure 3.4: Mortality rate (parameter μ) for wild type (top), G3 (middle) and I-PpoI (bottom). The mixture density is a linear pool of the experts that contributed assessments to all strains.

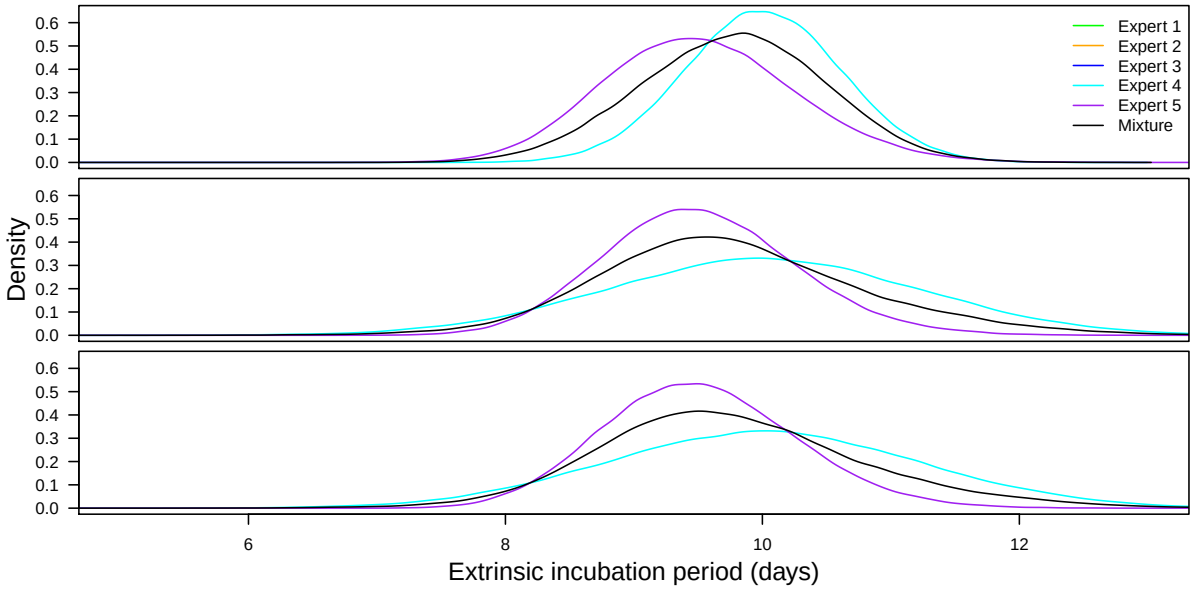


Figure 3.5: Extrinsic incubation period (parameter n) for wild type (top), G3 (middle) and I-PpoI (bottom). The mixture density is a linear pool of the experts that contributed assessments to all strains.

Integrating over the uncertainty of the experts for a given function $f(\cdot)$ gives the expectation,

$$\mathbb{E}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta}) \prod_{j}^{J} \left[\sum_{i}^{I_j} w_i^{(j)} p(\theta_j|i) \right] d\boldsymbol{\theta} \tag{3.8}$$
$$\approx \frac{1}{K} \sum_{k}^{K} f\left(\boldsymbol{\theta}_k\right),$$

where the second line is a Monte Carlo approximation based on K simulated parameter vectors drawn from the prior given by, $p(\boldsymbol{\theta}) = \prod_j^J \sum_i^{I_j} w_i^{(j)} p(\theta_j|i)$. This is obtained by first sampling an expert by a random draw from a multinomial distribution, then drawing a value from that expert's prior for θ_j. This is done for all j for each simulation k. The Monte Carlo approximation in the last line, denoted $\bar{f}$, for large K yields a standard normal distribution $(\bar{f} - \mathbb{E}\left[f(\boldsymbol{\theta})\right])/\sqrt{v_{mc}}$, where v_{mc} is the variance obtained from the Monte Carlo samples (Robert and Casella, 2005). The approximation converges at the rate of $O(\sqrt{N})$. For these simulations, $N = 10^5$.

Here we compare G3 and I-PpoI to wild type using two different functions for $f(\cdot)$ in Equation (3.8). The first function is based on the difference in intrinsic malaria transmission risk, given by Equation (3.6) for the comparison between G3 and wild type strains. We also consider the same function applied to the difference between I-PpoI and G3 strains (with the parameter vector $\boldsymbol{\theta}$ in Equation (3.8) modified accordingly). The second function that we consider is the probability of increased risk, given by Equation (3.4) for the case of G3 versus wild type. Again, we also consider the comparison between I-PpoI and wild type. These two functions are the risk endpoints considered for evaluating the risk of increased malaria transmission that may result from the first generation of G3 or I-PpoI female mosquitoes after an accidental escape from an insectary.

3.5 Risk calculation

The estimated difference in the transmission risk of escaped G3 mosquitoes and I-PpoI mosquitoes, relative to wild type, is sensitive to assumptions about dependence between the uncertain parameters in the $\mathcal{V}$ index. In reality at least some of these parameters are expected to be dependent across the wild type, G3 strain and I-PpoI modified mosquitoes. Comments made by the experts during the elicitation indicated that parameters such as survival rate, biting frequency or transmission effectiveness might all be negatively impacted by the laboratory rearing practices. This in turn implies a positive dependence between these parameters across these strains. This type of dependence is known to have an important effect on risk calculations (Ferson, 1995, 1996).

One approach to this issue is to ask questions such that target parameters can be taken as independent, while functions of these target parameters induce dependence between the wild type, G3 and I-PpoI strains. Other techniques are also possible but require more intensive elicitation approaches than used in this study; that is, more information and effort is required from the experts to obtain estimates of dependencies among unknown parameters. The questions specified for the elicitation, however, did not provide for either approach but instead addressed each strain independently.

To address this issue we provide bounds on the risk estimates by comparing risk estimates that assume independence between strains versus estimates that are derived after imposing strong positive dependence among the expert's subjective beliefs about the magnitudes of a

parameter across each strains. We retain, however, the assumption of independence between experts. The analysis uses a standard approach to impose positive dependence: for each expert a sequence of random variates are drawn from a standard uniform distribution. This common set of variates is then transformed into the probability density functions elicited from an expert for a given parameter for each of the three strains using the probability integral transform. This is a variance reduction technique for the Monte Carlo estimate of the expected difference between two quantities. Here it is used to impose strong positive dependence among the various strains for each parameter. For instance, a relatively high realisation for biting rate in the wild type will be paired with a relatively high biting rate in G3 and I-PpoI, although the absolute magnitudes can still be quite different (if, for instance, one strain has a much longer upper tail than the others).

If we impose strong positive dependence, then the estimated probability of an increase in the transmission risk index of escaped G3 mosquitoes relative to wild type is 0.007 with Monte Carlo standard error 2.64768×10^{-4}. The equivalent probability for escaped I-PpoI mosquitoes relative to wild type is 0.052 with Monte Carlo standard error 7.03898×10^{-4}. The results indicate a substantial reduction in the risk of G3 and I-PpoI mosquitoes producing an infectious bite compared to wild type if we make the reasonable assumption of positive dependence between the strains.

The difference in the intrinsic indices $\mathcal{I}_0^t - \mathcal{I}_0^W$, which is proportional to the difference in malaria risk for the strains (Section 3.2), has mean -0.414 (Monte Carlo standard error: 0.00847) for G3 versus wild type and mean -0.23 (Monte Carlo standard error: 0.0278) for I-PpoI versus wild type with the positive dependence among strains. These results indicate a 41% and 23% reduction in the risk of G3 or I-PpoI transgenic mosquitoes producing an infectious bite compared to local wild type mosquitoes.

If we assume independence between strains, then the estimated probability of an increase in the transmission risk index of escaped G3 mosquitoes relative to wild type is 0.196 with Monte Carlo standard error 0.00125. The equivalent for escaped I-PpoI mosquitoes relative to wild type is 0.197 with Monte Carlo standard error 0.00126.

The difference in intrinsic risk with independence among strains is presented for both comparisons in Figure 3.6, which has mean -0.391 (Monte Carlo standard error: 0.01851) for G3 versus wild type and mean -0.289 (Monte Carlo standard error: 0.01414) for I-PpoI versus wild type. These results indicate a 39% and 29% reduction in the risk of G3 or I-PpoI transgenic mosquitoes producing an infectious bite compared to local wild type mosquitoes. The estimated density is presented for both comparisons in Figure 3.7.

It is clear that the shapes of the distributions are altered by imposing strong positive dependence (Figures 3.7 and 3.6). Assessing dependence can have a particularly important role in the vicinity of no difference. Several experts noted that the three strains share genetic material and we believe that experts would have imposed positive dependence if they were given an opportunity to do so during the elicitation procedure. As noted above, however, this will require an elicitation procedure designed to efficiently capture this information.

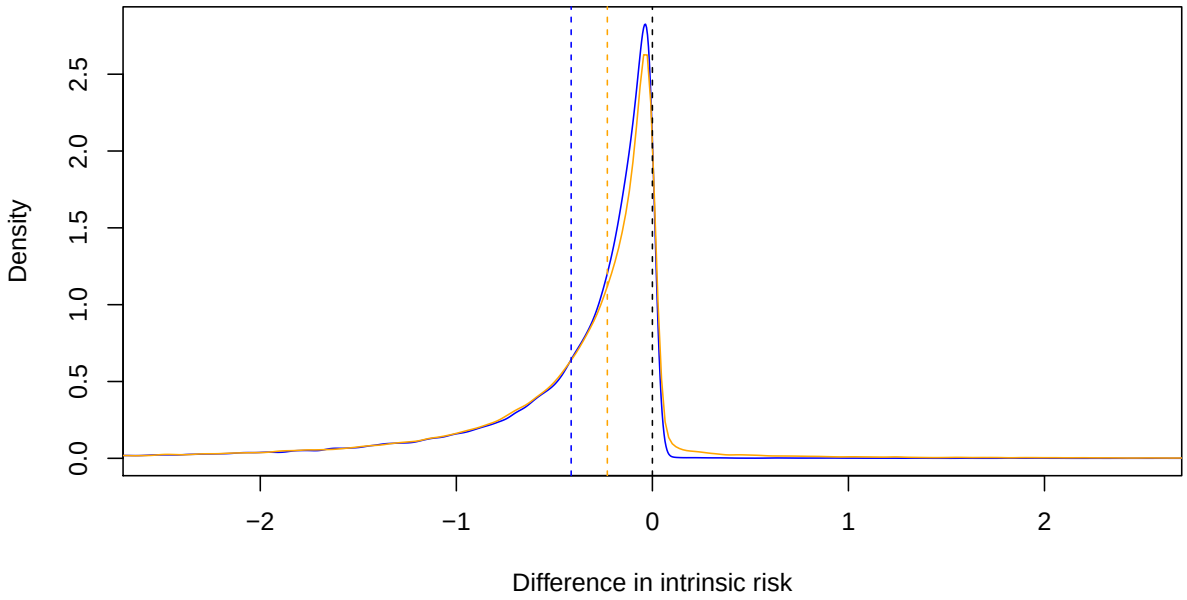


Figure 3.6: The difference in intrinsic risk for G3 versus wild type (blue line) and I-PpoI versus wild type (orange line) with positive dependence between strains induced by correlated sampling. The blue and orange dotted lines are the mean difference in intrinsic risk for each group. The area under each each curve to the right hand side of the black dotted line corresponds to the probability of increased risk compared to wild type (see Equation (3.4)).

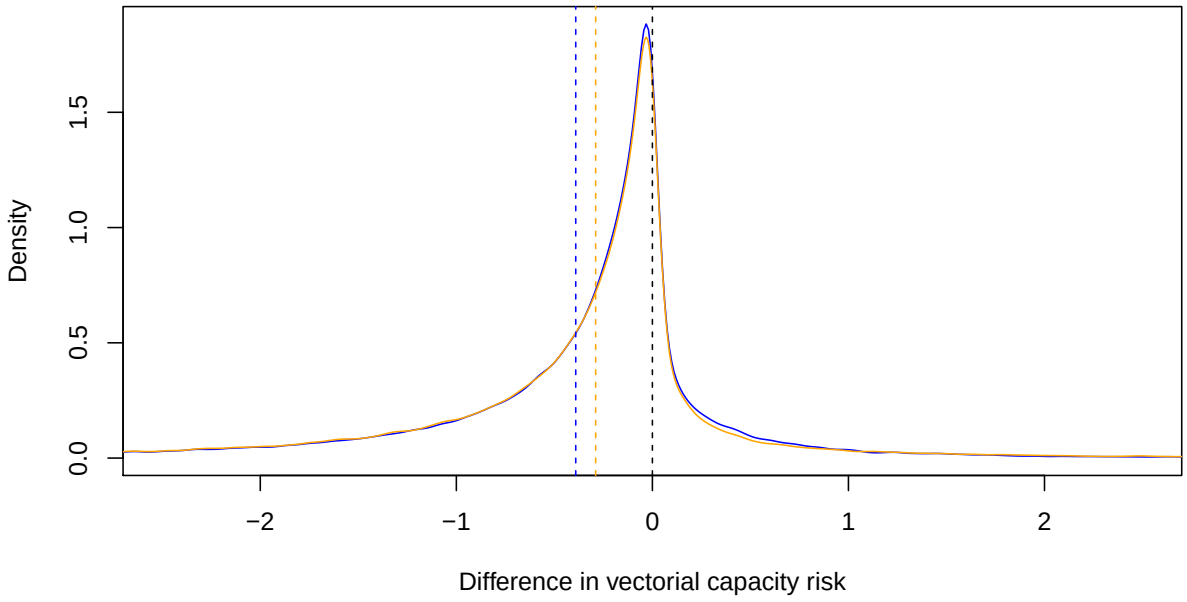


Figure 3.7: The difference in intrinsic malaria risk for G3 versus wild type (blue line) and I-PpoI versus wild type (orange line) assuming independence between strains. The blue and orange dotted lines are the mean difference in intrinsic risk for each group. The area under each curve to the right hand side of the dotted line corresponds to the probability of increased risk compared to wild type (see Equation (3.4)).

KEY POINTS: FAULT TREE ANALYSIS METHODOLOGY

- Fault tree analysis serves two important roles in this risk assessment: (i) it forces the analyst to think systematically and logically about all the steps that are necessary for the risk assessment endpoint to occur, and thereby helps the analyst uncover all risk generating pathways; and, (ii) it provides an opportunity to quantify the probability of the risk assessment endpoint.
- In this analysis we characterised the probability of the basic events in four fault trees (one for each of the remaining endpoints) using subjective conditional probability distributions elicited from 8 members of the Target Malaria consortium and 16 independent experts.
- In total 1068 subjective probability density functions and 67 constants were retained for subsequent analysis, together with 1588 comments, covering 352 and 544 basic events or gates (including vectorial capacity parameters) respectively.
- The qualitative heuristic benefits of fault tree analysis accrued immediately in this analysis but we encountered a number of challenges when completing the quantitative risk assessment.
- These challenges forced us to write our own analysis software that implements the fault tree equation symbolically, step wise through the tree, using the fault tree logic provided by the SAPHIRE software that we used to draw the trees.
- The advantages of our software is that it allows us to compare two different analysis approaches (Aggregate First Then Convolute versus Convolute First Then Aggregate), retain elicitations made by experts at the fault tree gates, complete a coherent probabilistic uncertainty analysis, and provide an initial sensitivity analysis to identify important events in the fault tree.
- The disadvantage of our software is that the step-wise symbolic calculations require large amounts of memory and on occasion we were forced to truncate terms in the probability equation at a gate that contained the product of more than ten or twelve basic events.

4 FAULT TREE ANALYSIS METHODOLOGY

4.1 Fault tree construction and roles

Fault tree analysis is a deductive hazard analysis and risk assessment tool. The analyst specifies a system failure (the "top-event"), and then within the context of the system's environment identifies the events that lead to the specified failure. The events that contribute to this failure are drawn as a tree with branches connected by a series of logical functions, known as gates. Taken together the events and gates describe a graphical model of the various parallel and sequential combinations of faults that result in the occurrence of the top event (Vesely et al., 1981).

Following the completion of the hazard analysis, the risk assessment progressed by conducting a fault tree analysis for each of the assessment endpoints described above. Five fault trees were constructed with members of the Target Malaria consortium over a period of 3 days from the 30th of April to the 2nd May, and then subsequently refined following review by CSIRO and the consortium prior to the elicitation of the probabilities of the basic events with US-based independent experts (Section 4.3). During the initial construction stage the first fault tree, increased incidence of malaria following a complete release of all mosquitoes from the insectary, was discarded in favour of a direct elicitation of an essential set of the parameters that describe the dynamics of malaria transmission (Section 3).

Minor amendments were subsequently made to the fault trees during the basic event elicitation with the US-based independent expert group. Further amendments were made during the elicitation exercises completed with the UK-based independent expert group, and members of the Target Malaria consortium, in September and October 2014 respectively (Section 4.3).

Fault trees play two very important roles in a risk assessment. Firstly, the construction of the trees requires the analyst to think systematically and logically about all the steps that are necessary for the top event to occur. In this way they serve an important heuristic role, helping the analyst to uncover all the pathways that lead to the top event. Fault trees have proven to be a useful tool in this sense, particularly in complex industrial systems, but also in environmental systems (Barnthouse et al., 1986; Hayes, 2002, 2003b)

Secondly, fault trees provide an opportunity to quantify the probability of the top event. The gates that link the branches of a fault tree represent logical functions. The two most important functions are logical union, represented by the "AND" gate, and logical intersection represented by the "OR" gate. The basic probability laws for union and intersection allow the analyst to calculate the probability of an event above the gate given the probability of the events below the gate (see for example Sherwin and Bossche, 1993, for further details).

4.2 Dependency

The quantification of the probability of an event above a gate, and by extension through out a fault tree, is rudimentary in the special case that the events below the gates are unique (appear only once in the logical function) and independent. These calculations become more complicated, however, when events within the fault tree are dependent. If the probabilistic laws of intersection and union are propagated step-wise through the tree, converting the probability of the events below a gate to the probability of the event above it, all other information about the events that enter the gate, including their dependency with other events elsewhere in the tree, is lost. This can lead to incorrect results.

Dependency between the basic events of a fault tree can occur for a variety of reasons:

- **Repeated events**. Repeated events, or multiple occurring events, are identical events that occur in multiple locations in a tree. There are numerous examples of repeat events in this analysis. A complete list is provided in Table D.1 in Appendix D.
- **Mutually exclusive events**. Mutually exclusive events are events that cannot occur at the same time. There are no examples of this type of dependency within this analysis. This type of dependency, however, arises in engineering contexts where the top event occurs because, for example, a sensor gives a false positive or a false negative alarm, but clearly it cannot give both at the same time.
- **Common cause failures**. Common causes are conditions internal, or external, to the system under consideration that simultaneously influence the probability of multiple events within the tree that were otherwise thought to be independent (Ericson, 2011). The classic example is extreme external conditions, such as the heat from a fire, that cause independent safety systems to simultaneously fail. Common cause failures that are identified during the course of the analysis appear as repeated events within the tree (see above). Common cause failures that are not identified during the analysis cause partial dependency (see below).
- **Partial dependency**. Partial dependency between events occurs when the fault tree fails to include mechanistic (or in our case biological) processes that influence the probability of multiple events within the tree. Partial dependency implies further sub-events within the systems under consideration that are either independent, repeated or mutually exclusive (Sherwin and Bossche, 1993).
- **"State of knowledge" dependency**. Pedroni and Zio (2013) define this to be the dependency between estimates of the probability of the same basic event caused by dependent information sources, e.g. the subjective probability of two expert's who have seen the same, perhaps limited, information.

All of the dependencies listed above will influence the probability of the top event in a fault tree. Some, however, are easier to account for than others. The basic events in a fault tree are Boolean variables: switches and valves are either open or closed, alarms operate or fail, and events either occur or they do not. The top event of a fault tree can therefore be represented as a Boolean equation, and the laws of Boolean algebra can be used to simplify it – in a process known as the "minimum cut sets method" – to eliminate repeat events and mutually exclusive events from the calculations (see for example Ericson, 2011, and Appendix C for details).

In this analysis, we characterise the probability of the basic events in the tree using subjective conditional probability but we wish to retain information collected at the gates of the tree. We therefore apply the probability laws of intersection and union step-wise through the tree symbolically. By doing this symbolically we retain information associated with the events, such as whether they belong to a set of repeat events or mutually exclusive events. The dependency induced by these types of events can be corrected by applying the joint probability laws $\Pr(AA) = \Pr(A)$ for repeat events, and $\Pr(AB) = 0$ for mutually exclusive events A and B (Appendix C).

Common cause failures that are identified in the tree are handled as repeated events. Common cause failures that are not identified in the tree contribute to the sources of partial dependency within the system under consideration. It is practically impossible, however, to provide a com-

plete and accurate account of the effect of partial dependencies in fault tree analysis. At best the analyst aims to minimise this source of potential error by constructing the trees carefully, using the best available knowledge, in a manner that respects well-tested theories and current understanding.

Similarly, it is difficult to account for the potential effects of state of knowledge dependency. In this analysis we have attempted to minimise the potential for this type of dependency by consulting widely and using experts that are independent of the Target Malaria consortium. The nature of biotechnology risks, however, and our desire to consult experts that are part of, and independent of, the Target Malaria consortium, introduced a number of computational challenges (Section 4.4).

4.3 Eliciting basic event probabilities

For the majority of the basic events in the fault tree there is little, if any, empirical data to estimate their frequency or annual probability. We therefore turn to expert opinion to perform a quantitative analysis but must do so in a careful and a coherent fashion to ensure that the analysis conforms to the axioms of probability theory. We do this by using direct elicitation methods to elicit the beliefs of experts as subjective conditional probability density functions.

The scientific literature recommends various (sometimes contradictory) methods for direct elicitation. The issues under contention stem from the way in which expert's beliefs are elicited, and the extent to which these methods protect against well known heuristics and biases such as overconfidence, defined as the tendency for experts to overstate their ability to predict uncertain events, resulting in intervals and/or distribution functions that do not capture true values. Morgan and Henrion (1990) state that there is "mounting evidence" that methods that assess the probability that an uncertain variable lies within a specified range produce more diffuse distributions (and hence are usually better calibrated) than methods that assess a value of the uncertain quantity that lies within or below specified percentiles (such as tertiles or quartiles). O'Hagan et al. (2006), however, notes that the latter approach is one of the most widely used elicitation methods, and Garthwaite and O'Hagan (2000) recommend tertiles rather than quartiles because this leads to larger standard deviations and more diffuse distributions.

In light of this debate, the CSIRO risk team has tested two alternative methods:

1. eliciting tertiles, 10th and 90th percentiles and the median - the expert is asked to give values for the uncertain quantity x that they believe there is a 10%, 33%, 67% and 90% chance that the true value of x is less than, and then a value that they think there is a 50% chance that the true value is less than; and
2. eliciting "free-choice" central credible intervals - the expert is asked to give values that the uncertain quantity is probably lower than (upper bound), and probably higher than (lower bound), together with their confidence (probability) that the true value lies within this interval.

The first approach provides more information, but even after training in subjective probability and quantiles, and practise elicitation exercises, some experts found it difficult to provide coherent answers – e.g., their elicited quantiles were sometimes out of order such that the value for a lower quantile was greater than the value elicited for a higher quantile. Some experts also found tertiles difficult for rare events, but when offered the second approach they were more willing to provide upper and lower bounds for these events together with a probability that the truth lies

within these bounds. Moreover, the second approach substantially eased the expert's task of providing coherent quantiles. Based on these observations we chose to use the free-choice central credible interval method throughout the remainder of the analysis.

In risk assessments for novel genetic technologies it is also important that the upper and lower tails of any distribution that is fitted to data elicited from experts is verified by the expert concerned. In our experience the upper tails of the final risk distribution are often contentious and the probability of extreme events can feature significantly in the decision making process. For this reason our elicitation procedure includes immediate feedback wherein the distribution fitted to the expert's belief is shown to the expert, and the analyst confirms that he or she is comfortable with the overall shape and tail probabilities (shown in terms of the fitted 90th and 10th percentiles) of the distribution fitted to their data.

To facilitate this feedback step, we developed an interactive Graphical User Interface (GUI). The GUI enabled the risk team to input elicited values or probabilities into a model, and for experts to instantly visualize the probability distribution generated using this model (e.g. a beta distribution to represent uncertainty about the probability of an event). The immediate feedback enables experts to gauge how their opinion has been interpreted, and make any amendments in order for the model to more accurately reflect their opinion.

The ease with which data could be uploaded, fitted, visualised and changed, coupled with time for additional discussion and comment, we believe, has resulted in a more in-depth and thorough analysis. While the GUI succeeded in facilitating the elicitation process, this was not its only purpose. The GUI was also designed to assist in documenting the process. All probabilities given by the experts are registered along a time line, in addition to any comments the experts thought relevant. The GUI also helped to facilitate downstream analyses, as the data were recorded in a format that is useful for our analyses.

The direct elicitation was performed with 16 independent (of the Target Malaria consortium) experts, and with 8 members of the consortium, during individual (one-on-one) interviews. On two occasions individual elicitations were not possible for logistical reasons, and two experts were interviewed together at the same time. Elicitations were completed in three phases – from the 3rd to the 15th July 2014, the 1st to the 17th of September 2014 and from the 16th to the 23rd of October – usually at the expert's own institution. Experts were divided into areas of expertise domain, then directed to events within the fault trees that fell within their domain, and were actively discouraged from responding to questions that they did not feel qualified to answer. In total 1068 subjective probability density functions and 67 constants were retained for subsequent analysis, together with 1588 comments, covering 352 and 544 basic events or gates (including vectorial capacity parameters) respectively.

After each elicitation the experts were provided with individual reports that represent a formal record of their elicitation. The reports document the central credible intervals elicited from the experts, the 5th, 10th, 50th, 90th and 95th percentiles of the fitted distribution, together with the original basic event question and any comments made by the expert when addressing that basic event. Each expert was given an opportunity to amend either the central credible interval or their comments.

4.4 Analysis challenges

Fault tree analysis (FTA) was developed in 1962 to study failure rates in the Minuteman missile launch control systems. Since then it has been widely and successfully applied to complex industrial systems such as nuclear power plants, petro-chemical plants, aerospace systems and communication satellites. The application of FTA to biological systems, however, is relatively novel.

The qualitative, heuristic benefits of FTA accrue immediately in biological systems, and this has been recognised for almost thirty years (Barnthouse et al., 1986). The benefits and challenges of a quantitative analysis within these systems, however, has hardly been explored. In our opinion the transparency provided by a quantitative analysis is essential, but in completing such an analysis for the I-PpoI construct we encountered a number of challenges:

- **Multi-disciplinary analysis**. The biological processes involved in biotechnology risks are complex. They cover multiple scales – from intracellular processes to population dynamics – in diverse domains – from the eukaryote to non-eukaryote domains. Consequently the basic events within the fault trees cut across a range of scientific disciplines, such that no single person is likely to be expert in all of the basic events within a single tree. In our analysis this led to different sample sizes at the basic events because no single expert felt confident in providing subjective estimates of the conditional probabilities at every event in any single tree.
- **Aggregation and convolution**: By canvassing the opinion and subjective probabilities of a large group of experts, to protect against motivational bias and partial dependency, the sample size at many of the basic events is greater than 1. This raises two analysis avenues: The "Aggregate First Then Convolute" (AFTC) method involves taking the linear pool of expert opinion at each basic event prior to performing the probability calculations necessary to calculate the top event. The second approach, "Convolute First Then Aggregate" (CFTA), performs the probability calculations for each expert, before taking the linear pool of top event probabilities. Under certain circumstances these two different approaches can lead to quite different answers.
- **Missing data AFTC**: For a small proportion of the basic events in fault trees 3, 4, 50 and 51 (5%, 17.5%, 4.4% and 5% respectively) we were unable to attain at least one subjective probability estimate. In these instances we assumed a uniform prior distribution on the range $[0, 1]$.
- **Missing data CFTA**: When attempting to implement the CFTA analysis method we inevitably encountered cases of missing data because of the first challenge. No single expert felt able to answer all the conditional probability questions within any one tree hence it is not possible to implement the CFTA approach without encountering basic events that the expert did not respond to. This raises further alternative options for how to handle missing data at a basic event. We explored three of these: (i) using the linear pool of expert belief at the event; (ii) using a non-informative (uniform) prior; and, (iii) using a prior that assumes $\Pr(\text{Event with missing data}) = 1$. We found the last two options provided no beneficial insights, and in some cases led to demonstrably incorrect results. We do not therefore report these results here.
- **Elicitations at gates and model variations**. A few experts chose to provide elicitations at a gate because they believed this to be a more sensible approach. This is equivalent

to modifying the fault tree model by turning a gate into a basic event. Furthermore, two experts maintained quite different conceptual models of the failure pathways associated with the top event, and chose to slightly amend parts of the tree, or provide a different statistical model for the event represented by one of the gates.

- **Effective sample size**. To perform the calculations for a linear pool, whilst accommodating different sample sizes at each base event, and elicitations at gate, we needed a strategy to compute the effective sample size for the AFTC and CFTA methods. To do this we added a switching parameter S to the equation at a gate G with basic event inputs, say A and B, such that

$$\Pr(G) = S \times \Pr(G_e) + (1 - S) \times f(\Pr(A), \Pr(B)), \tag{4.1}$$

  where $f(\cdot)$ denotes the probability equation for intersection or union, and $\Pr(G_e)$ is the probability of the event at the gate (as provided by expert e). Setting $S = 1$, for example for the CFTA method for expert e, means expert e's data is used to calculate $\Pr(G)$, and setting $S = 0$, means the data elicited at the basic events A and B is used. To take a linear pool of expert opinion to calculate $\Pr(G)$ during the AFTC method we set $\Pr(S = 1) = \frac{n_g}{(n_A + n_B)}$ where n_g denotes the number of experts who provided elicitations at the gate, and n_A the number at event A, and so on.

As far as we aware, there is no currently available FTA software that can accommodate all of the challenges listed above. We therefore wrote dedicated software in the R statistical computing language, using as input the fault tree logic provided by the SAPHIRE software (https://SAPHIRE.inl.gov/) that we used to draw the trees. The SAPHIRE fault tree logic encodes the fault tree equation as a text file showing the gate reference number, type, and the reference numbers of the gate inputs. We used this to compile the fault tree equation symbolically using the Ryacas library (http://cran.r-project.org/web/packages/Ryacas/index.html), and perform the probabilistic operations of intersection and union step-wise through the tree equation. Performing the tree calculations symbolically enables us to retain information about the events and the gates, including dependency induced by repeat events and elicitations at a gate, throughout the probability calculations (Appendix C).

One disadvantage of our step-wise, symbolic approach to the probability calculations, is that they require large amounts of memory. The number of terms in the expanded form of the step-wise equation grows exponentially as each new "OR" gate adds additional terms according to the inclusion-exclusion rule, and each new "AND" gate multiplies these terms by another set of terms. This can be controlled to some extent by keeping the equation for probabilistic intersection in its factored form – i.e. $\Pr(\cup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} (1 - \Pr(A_i))$ but this equation must be expanded to remove repeat events if any of its terms are dependent.

For the larger trees in this analysis, and those with high numbers of repeat events, the memory allocation requirements of the symbolic expanded fault tree equation proved excessive. To handle this we were forced on occasion to separate sub-sections of the tree that contained no repeated events, for those that did, and sometimes truncate terms in the probability equation at a gate that contained the product of more than ten or twelve basic events. Truncation of large product terms is standard practise in fault tree analysis on the grounds that these terms can be expected to contribute virtually nothing to the overall probability of the top event.

4.5 Uncertainty analysis

One of the benefits of a quantitative approach is that it enables a coherent, probabilistic analysis of uncertainty and allows the analyst to conduct sensitivity studies to identify which basic events are the most important to the overall probability of the top event. In this analysis we use Monte Carlo methods to propagate the uncertainty in the base events through the fault trees. Section 3.2 demonstrates the application of these methods to the malaria transmission risk endpoint.

To complete the uncertainty analysis we simulate a large number of base event probabilities from the experts' elicitations, and calculate the probability of the top event. If we denote the top event T then $\Pr(T) = f(\boldsymbol{\theta})$ where $f(\cdot)$ is the fault tree equation and $\boldsymbol{\theta}$ is a vector of the uncertain base events and gates within the tree, with an arbitrary event (or gate) denoted θ_j.

Using the same notion as Section 3.2 we indicate the conditioning on expert by writing the expert's prior belief for θ_j as $p(\theta_j|i)$, where $i = 1, \ldots, I_j$ for the I_j experts that contributed to the base event or gate. For the AFTC analysis method, we first create a linear pool of expert opinion $p(\theta_j) = \sum_i^{I_j} w_i^{(j)} p(\theta_j|i)$ where $\sum_i^{I_j} w_i^{(j)} = 1$ for non-negative weights $w_i^{(j)}$. We place the superscript j on the weight to indicate that it varies between parameter because the number of experts who addressed each gate or event varies. Again we give equal weight to each expert, and assume the expert elicitations on different parameters are independent, such that the joint prior distribution on all parameters is the same as Equation 3.7. The expected value of the top event is now defined as

$$\mathbb{E}_{AFTC}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta}) \prod_j^J \left[\sum_i^{I_j} w_i^{(j)} p(\theta_j|i) \right] d\boldsymbol{\theta} \tag{4.2}$$

$$\approx \frac{1}{K} \sum_k^K f\left(\boldsymbol{\theta}_k\right), \tag{4.3}$$

where $K = 10^5$ is the total number of Monte Carlo simulations, and $\boldsymbol{\theta}_k$ denotes the k^{th} of N realisations from the parameter vector $\boldsymbol{\theta}$ drawn from the joint prior distribution (linear pool) of the base event or gate. The error in the Monte Carlo approximation (the last line of this equation) can be obtained using standard techniques (Robert and Casella, 2005).

For the CFTA analysis method the order of operations is changed. The probability of the top event is calculated separately for each expert and then aggregated, such that

$$\mathbb{E}_{CFTA}[f(\boldsymbol{\theta})] = \sum_i^I w_i \int f(\boldsymbol{\theta}) \prod_j^J p(\theta_j|i) d\boldsymbol{\theta} \tag{4.4}$$

$$\approx \sum_i^I w_i \frac{1}{K} \sum_k^K f\left(\boldsymbol{\theta}_{k|i}\right), \tag{4.5}$$

where now $\sum_i^I w_i = 1$ for the I experts that participated and the realisations $\boldsymbol{\theta}_{k|i}$ are drawn from the prior $p(\boldsymbol{\theta}|i) = \prod_j^J (\theta_j|i)$ for each expert.

It is important to note, however, that Equation 4.5 represents the ideal case where each expert has contributed to all of the basic events gates. In reality the CFTA approach is complicated by missing data at each gate - i.e. $p(\theta|i) = \varnothing$. Unfortunately none of the potential solutions to this problems, such as using a non-informative uniform prior distribution, a constant, or the linear pool of other expert opinion at the gate, are particularly satisfactory, and for these reasons the AFTC method is our preferred approach. Nonetheless we report the CFTA linear pool results for each of these strategies as a comparison.

4.6 Sensitivity analysis

The objective of a fault tree sensitivity analysis is to identify which of the basic events in the tree exert the most influence on the probability of the top event. The relative importance of the different basic events is typically measured by an importance measure, such as the Birnbaum index (Birnbaum, 1969; Zio, 2009). The Birnbaum index expresses the importance of basic events within the tree in terms of the likely increase in risk that the events could impart. Formally, it is defined as the derivative of the probability of the entire fault tree with respect to each basic event assessed sequentially.

One of the shortcomings of the original Birnbaum index is that it is only defined for systems where the probabilities of the base events are known with certainty. In this analysis, however, the probabilities of the basic events are presented as subjective probability density functions. There have been a number of attempts to amend the Birnbaum index to incorporate uncertainty. Wang et al. (2004) use simulation of the uncertain probabilities as a way to "integrate out" uncertainty. We developed a similar strategy for this analysis, but here we propagate the probability of an event, rather than the event itself. We do this because many of the probabilities of the basic events in this analysis are very small, and any finite set of event simulations (as Boolean terms) is unlikely to contain enough failure occurrences to calculate the Birnbaum index.

To complete the sensitivity analysis we simulate a large number of base event probabilities from the expert elicitations. Each of the $i = 1 \ldots N$ simulated base event probabilities is denoted by $\boldsymbol{\theta}_i$, which is composed of elements θ_{ij} for each of the $j = 1 \ldots M$ base events. For each set of base events probabilities ($\boldsymbol{\theta}_i$), we calculate the probability of the fault tree top event, denoted by $f(\boldsymbol{\theta}_i)$. We also calculate the rate of change of the probability of the top event with respect to the probability of each base event, that is $\frac{\partial f(\boldsymbol{\theta}_i)}{\partial \theta_{ij}}$. The importance measure is then the median rate of change of the N sets of simulated base event probabilities. Formally, the importance measure I_j for base event j is given by

$$\bar{I}_j = \underset{i=1\ldots N}{median}\left(I_{ij}\right) = \underset{i=1\ldots N}{median}\left(\frac{\partial f(\boldsymbol{\theta}_i)}{\partial \theta_{ij}}\right). \tag{4.6}$$

The variation in this measure represents the spread in plausible importance. During the course of this analysis we found the variance of the simulated importance measures to be large for all of the fault trees. There are two possible reasons for this: (i) the sensitivity index is inherently noisy for the fault tree equations $f(\cdot)$; and (ii) the sensitivity index encompasses variation between events (by virtue of the position of the event within the tree) as well as the variation within the event (by virtue of the inter- and intra-expert uncertainty). It may be possible to separate the "within tree" and "within event" variation, but doing so would depend on the purpose of the analysis as well as developing some purpose-specific methodologies. For now, we present the median estimate only.

KEY POINTS: RESULTS OF THE FAULT TREE ANALYSIS

- The objective of fault tree 2 is to estimate the risk of G3 strains of mosquitoes transmitting novel blood-based pathogen to a human or vertebrate host, and then to ascertain how and why G3 strains modified with the I-PpoI construct might be different in this context.
- The results of the FT2 analysis indicate that the median value of the risk of G3 strain mosquitoes transmitting a novel blood-based pathogen in a year following a complete escape of 10,000 mosquitoes is 5.2×10^{-7}, while the 90th percentile of this risk is 10^{-4}.
- A linear pool estimate of the risk of the I-PpoI mosquitoes vectoring a novel pathogen would be the same as or lower than the values reported here for the G3 mosquitoes because most experts believe that all of the events in the fault tree would remain unchanged or be lower. Some experts questioned whether the construct might compromise the mosquito immune system, but they expected that this effect would be counteracted by the anticipated higher mortality of the genetically modified mosquitoes.
- The objectives of fault trees 3 and 4 are to estimate the risk of spread of the HEG to non-target eukaryotes and non-eukaryotes respectively. Horizontal transfer would not necessarily be harmful, but could be a pre-cursor to harm, and therefore using these as endpoints is conservative.
- The analysis indicates that the median risk of the HEG spreading in non-target eukaryotes or non-eukaryotes in a year following the complete loss of 10,000 I-PpoI modified mosquitoes is 1.2×10^{-10} and 6.7×10^{-7} respectively.
- The objective of fault tree 5 is to estimate the risk of the HEG spreading in local populations of species in the *An. gambiae* complex because this is deemed *a priori* to be undesirable for a supposedly self-limiting construct.
- The analysis indicates that the median value of the risk of the construct spreading in local populations of the related mosquito species *An. coluzzii* or *An. arabiensis* in a year, following the complete loss of 10,000 I-PpoI modified mosquitoes is 1.1×10^{-6}, however, this rises to 1.5×10^{-5} under the alternative (Convolute First Then Aggregate) analysis strategy.
- The analysis indicates that the median value of the risk of the construct spreading in local populations of *An. gambiae* in a year following the complete release of 10,000 I-PpoI mosquitoes is 0.0014.

5 FAULT TREE ANALYSIS RESULTS

5.1 FT2: Vector novel blood-borne pathogen

5.1.1 Fault tree structure

The top event of fault tree 2 is defined as *"The probability that a (non-GM) insectary mosquito will transmit a novel (not previously known to be vectored by An. gambiae) blood-based pathogen in a year following an escape of 10,000 mosquitoes"*. The objective of this tree is to estimate the risk of G3 strains of mosquitoes transmitting a novel blood-based pathogen to a human or vertebrate host, and then ascertain how and why G3 strains modified with the I-PpoI construct might be different in this context.

The structure of the fault tree is illustrated in Figures 5.1 to 5.3. The structure of the tree recognises two basic mechanisms for transmission of novel blood-based pathogens. The first, called here "Biological transmission", refers to the malaria-like mechanism whereby the pathogen is delivered to the human or vertebrate host via the saliva of the mosquito. The second mechanism, called here "Mechanical transmission", occurs either by adhesion of the pathogen to the mosquito's proboscis, or via simple transport by the mosquito of contaminated blood from an infected to uninfected individual.

Mechanical transmission of blood borne pathogens via adhesion to the proboscis, and via delivery of contaminated blood, has been previously documented in mosquitoes and hematophagous flies (see for example Vickerman, 1973; Hoch et al., 1985; Desquesnes and Dia, 2003), but as far as we are aware there are no documented cases involving *An. gambiae*. The fault tree recognises three possible exposure mechanisms for the delivery of contaminated blood: (i) ingestion of the mosquito, (ii) mixing of blood if the mosquito is squashed on an open wound; and, (iii) inhalation of contaminated mosquito faeces. Again, however, we are unaware of any documented cases of this transmission route within *An. gambiae*.

During the basic event elicitations for Biological transmission an expert suggested that the structure of the tree below FT2010 (*"Given that a mosquito has acquired a novel pathogen through a blood meal, what is the probability that the pathogen survives all of the mosquito's immune systems (cellular, humoral and RNA interference)?"*) was incorrect. In their opinion, once a mosquito has acquired a novel pathogen through a contaminated blood meal, the probability of surviving the digestive enzymes and then entering a gut cell are more important than the immune system in determining the probability that the mosquito becomes infected with the pathogen. This disagreement between experts in this (small) part of the tree is highlighted in Figure 5.2 by colouring the basic events yellow rather than blue.

Finally, whilst reviewing the fault tree structure an independent expert suggested a third potential transmission route whereby the mosquito acquires an entomopathogenic fungus and subsequently transmits this to another vertebrate or human host. This risk pathway was motivated by reports of a field study in Tanzania where an entomopathogenic fungus was successfully used as a biological control agent against *An. gambiae*. All of the experts who subsequently responded to this, however, suggested that this would only be a plausible pathway in the event that such a control programme was being implemented around the insectaries. This is not currently true hence this pathway was not included in the final fault tree structure.

Fault Tree Graphics Expanded Report **Probability that insectary escaped mosquito transmits novel blood based pathogen**
Feb 26, 2015 11:17 AM

FT2

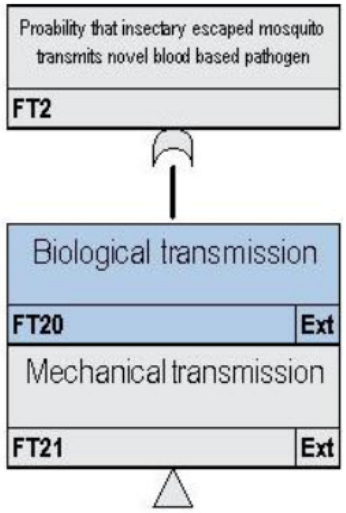


Figure 5.1: Page 1 of fault Tree 2

Fault Tree Graphics Expanded Report **Probability that insectary escaped mosquito transmits novel blood based pathogen**
Feb 26, 2015 11:17 AM
FT20

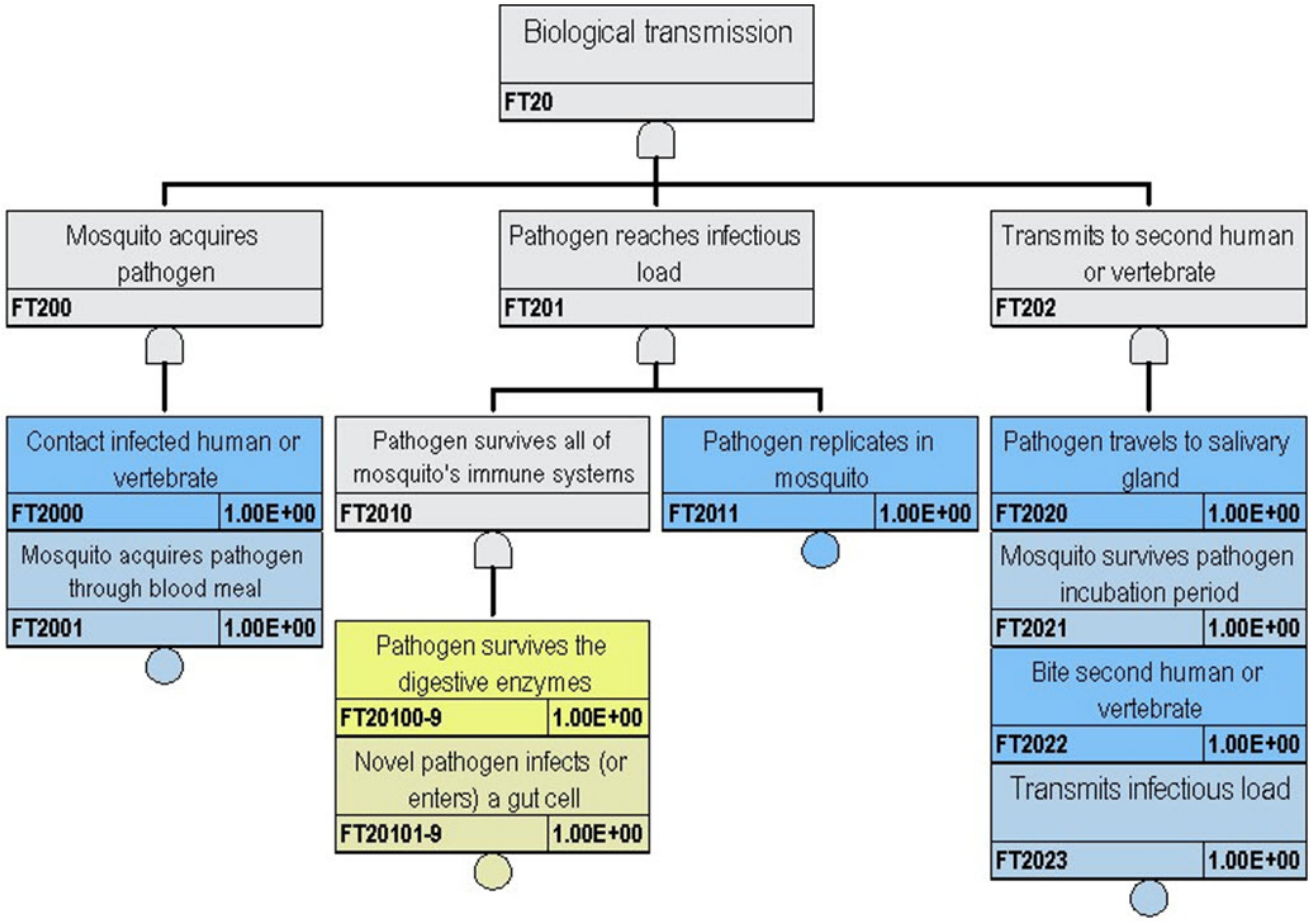


Figure 5.2: Page 2 of fault tree 2

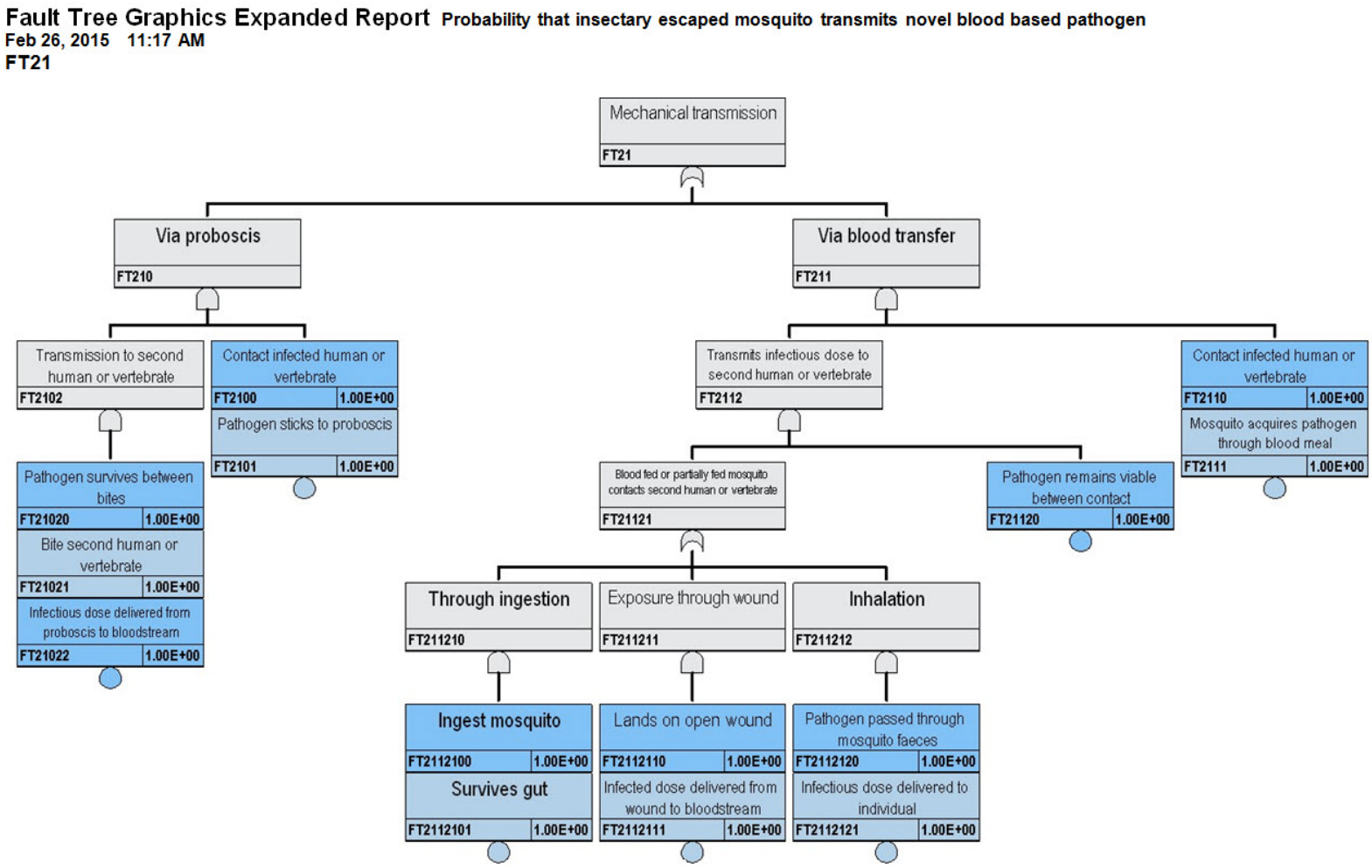


Figure 5.3: Page 3 of fault tree 2

5.1.2 Risk calculations

The results of the quantitative analysis for fault tree 2, using the AFTC and CFTA methods, are summarised in Figure 5.4. Table 5.1 shows the 50th, 90th and 99th percentiles of these results for five computational strategies. The first strategy (Method A) accounts for repeat event dependencies in the tree, and removes squared terms of repeat events from the tree equation by using fully expanded symbolic terms at each gate. This is our preferred approach but it is memory intensive.

The second and third strategies (Method B10 and B12) remove squared terms of repeat events from the tree, but truncate all terms in the fault tree equation with more than 10 and 12 terms respectively during the calculations. In this tree truncating 12 terms has no effect on the 50th percentile or the 99th percentile for the CFTA analysis with linear pool analysis.

The fourth strategy (Method C) is an approach designed to assist with memory allocation issue by identifying independent sub-trees within the calculations. These are sub-sections of the tree that have no repeated events elsewhere in the tree. The probability equations for these sub-trees are kept in an unexpanded factor form which reduces the overall memory requirement of the analysis. Squared terms of repeat events are removed from the dependent parts of the tree using one of the other methods. The results here are identical to the first scenario confirming that there are no coding errors in this strategy.

The last strategy (Method E) shows the effect of ignoring the error that occurs in the calculations if terms of repeat events squared are not removed from the tree equation. In this tree this has a small effect on the overall risk estimates across all the percentiles for most of the analysis methods.

Figure 5.4 indicates that there is no evidence of “motivational bias” in elicitations associated with this fault tree. The results for the members of the Target Malaria consortium do not cluster below those of the independent experts. These results also highlight differences of opinion across individual experts, and suggest three potential “outliers”: two independent experts stand out as having much lower subjective probability beliefs from the overall group, whilst one member of the Target Malaria consortium displays a very high confidence relative to the others.

Figure 5.4 also shows that there is little difference between the AFTC methods and CFTA with linear pool method in this tree. This, together with the difficulties associated with missing data under the CFTA method (see below) suggests that the AFTC method is preferable in this instance.

Method A : Squared terms of repeats handled by full expansion			
	50%	**90%**	**99%**
AFTC	5.2×10^{-7}	10^{-4}	0.0046
CFTA_LP	4.5×10^{-7}	3.9×10^{-4}	0.019
Method B10: Squared terms of repeats handled + truncation of >10 terms			
	50%	**90%**	**99%**
AFTC	4.9×10^{-7}	9.6×10^{-5}	0.0045
CFTA_LP	4.3×10^{-7}	3.5×10^{-4}	0.018
Method B12 : Squared terms of repeats handled + truncation of >12 terms			
	50%	**90%**	**99%**
AFTC	4.9×10^{-7}	9.6×10^{-5}	0.0045
CFTA_LP	4.3×10^{-7}	3.5×10^{-4}	0.018
Method C. Squared terms of repeats handled + sub-trees			
	50%	**90%**	**99%**
AFTC	4.9×10^{-7}	9.6×10^{-5}	0.0045
CFTA_LP	4.3×10^{-7}	3.5×10^{-4}	0.018
Method E. Squared terms of repeats allowed			
	50%	**90%**	**99%**
AFTC	5.9×10^{-7}	10^{-4}	0.0043
CFTA_LP	4.3×10^{-7}	3.5×10^{-4}	0.018

Table 5.1: Median and upper tail percentiles for the probability of the top event in fault tree 2, for two analysis methods and five computational strategies

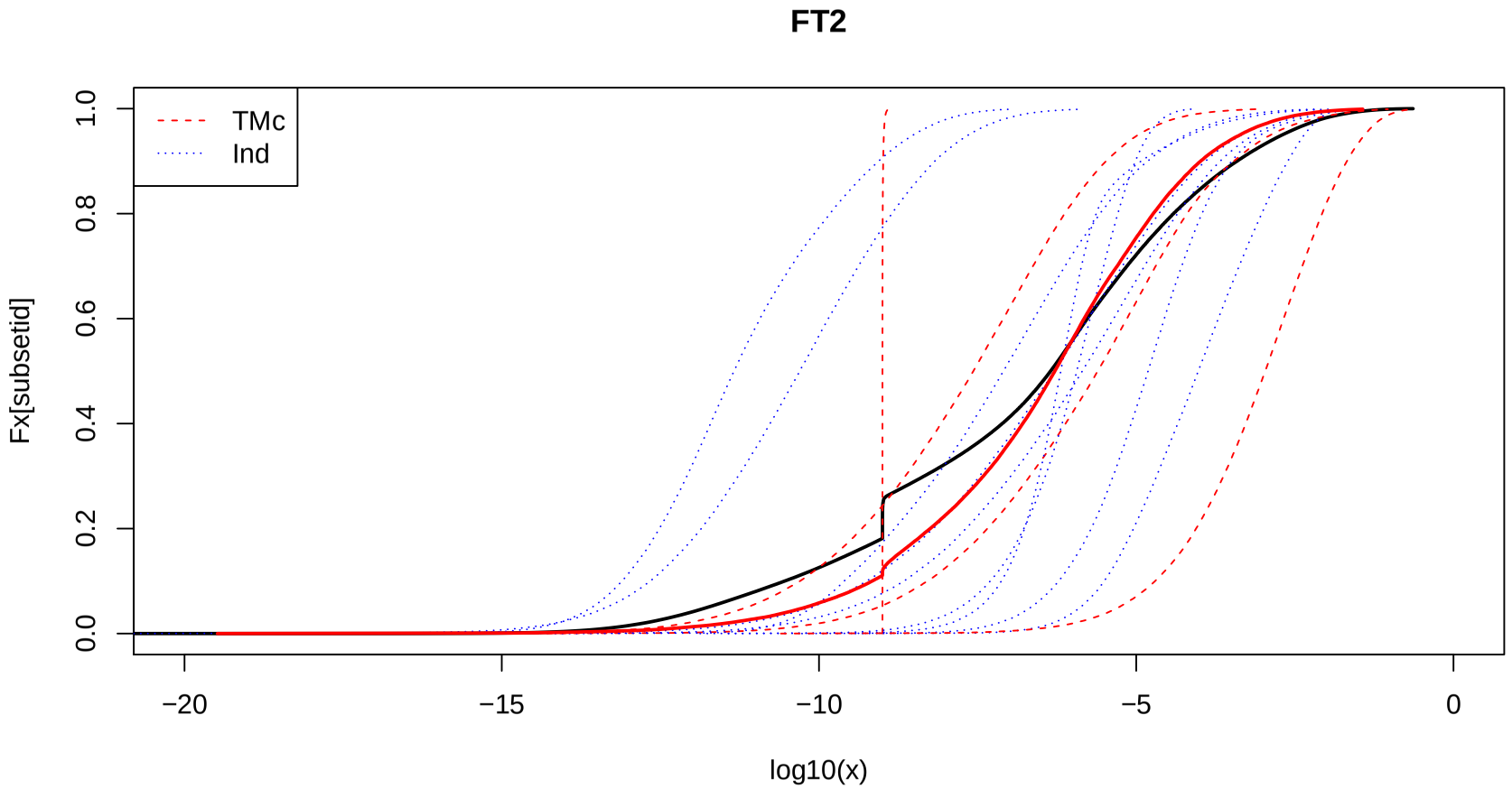


Figure 5.4: Cumulative distribution functions (CDF) of the probability of the top event in fault tree 2. Red solid curve shows the result for the AFTC method. Solid black curve shows the CFTA method using the linear pool of expert opinion for missing values at the basic events. Dashed curves show result for each expert (TMc = Target Malaria consortium, Ind = Independent) under the CFTA strategy using a linear pool for missing values.

5.1.3 Sensitivity analysis

Figure 5.5 shows the results of the sensitivity analysis for the top 15 events. These results suggest that the basic events: FT2000, FT2001, FT21022 and FT2110 have the strongest influence on the probability of the top event of FT2. Only one of these, however, FT21022 features in the cut set of FT2 with the smallest cardinality (which makes it important by virtues of its location in the tree structure). The other events are also not identified as having significant uncertainty hence this analysis is inconclusive. When repeating the analysis we found that FT21020 intermittently emerged in the top 5 events even with a $N = 10^5$ simulations. This event features in the second cut set as well and is an event associated with significant uncertainty. The high "within event" uncertainty may be why it appears intermittently in the top 15 important events when the analysis is repeated.

5.1.4 Effect of the I-PpoI construct

Following the elicitation of the basic events in fault tree 2 for the G3 strain, each expert was asked how they would amend their elicited values if all of the escaped mosquitoes were genetically modified with the I-PpoI construct. One expert provided amended central credible intervals for three of the basic events in the tree. The new intervals indicated that the expert: (i) was confident that the probability of a GM-mosquito surviving the incubation period of a novel pathogen would be lower than the G3 strain (because in their opinion the GM mosquitoes will have a higher mortality rate); (ii) was less certain about the probability that the pathogen would remain viable between contacts (confidence in the same central interval decreased from 0.6 to 0.5); and, (iii) believed the construct could weaken the mosquito's immune system which could increase the probability that a novel pathogen infects the mosquito. The first of these factors would serve to decrease the I-PpoI risk, the second would leave the expected value unchanged, but the last would serve to increase the I-PpoI risk.

In addition to the amended elicitations discussed above, six experts provided additional comments in relation to the I-PpoI modified mosquitoes. Three experts stated that they would not change their elicitations because they see no reason why the I-PpoI construct would increase the risk of novel pathogen transmission. Two of these experts noted that the transmission risk would probably be lower because the mortality rate of the I-PpoI mosquitoes is probably higher (see also Section 3.3). Similarly, a fourth expert suggested that as a rough rule of thumb laboratory raised mosquitoes are 20% less fit than wild types and genetically modified mosquitoes are 10% less fit than unmodified laboratory strains. If this fitness cost results in a higher mortality rate then disease transmission risk would be lower. The two remaining experts expressed high levels of uncertainty on this issue, suggesting that the risk of transmitting novel pathogens could go down (for the same fitness reasons highlighted by the other experts) or they could go up if the genetic construct compromises the mosquito's immune system.

The combined effect of these comments suggests that a linear pool estimate of the risk of the I-PpoI mosquitoes vectoring a novel pathogen would be the same as or lower than the values reported here for the G3 mosquitoes because four out of seven experts believe that all of the events in the fault tree would remain unchanged or be lower. Three out of seven experts highlighted that the construct may comprise the mosquito immune system but this effect would be counteracted by the anticipated higher mortality of the genetically modified mosquitoes. In this context it is worth re-iterating that a separate expert questioned the relevance of the immune system on the transmission of novel pathogens.

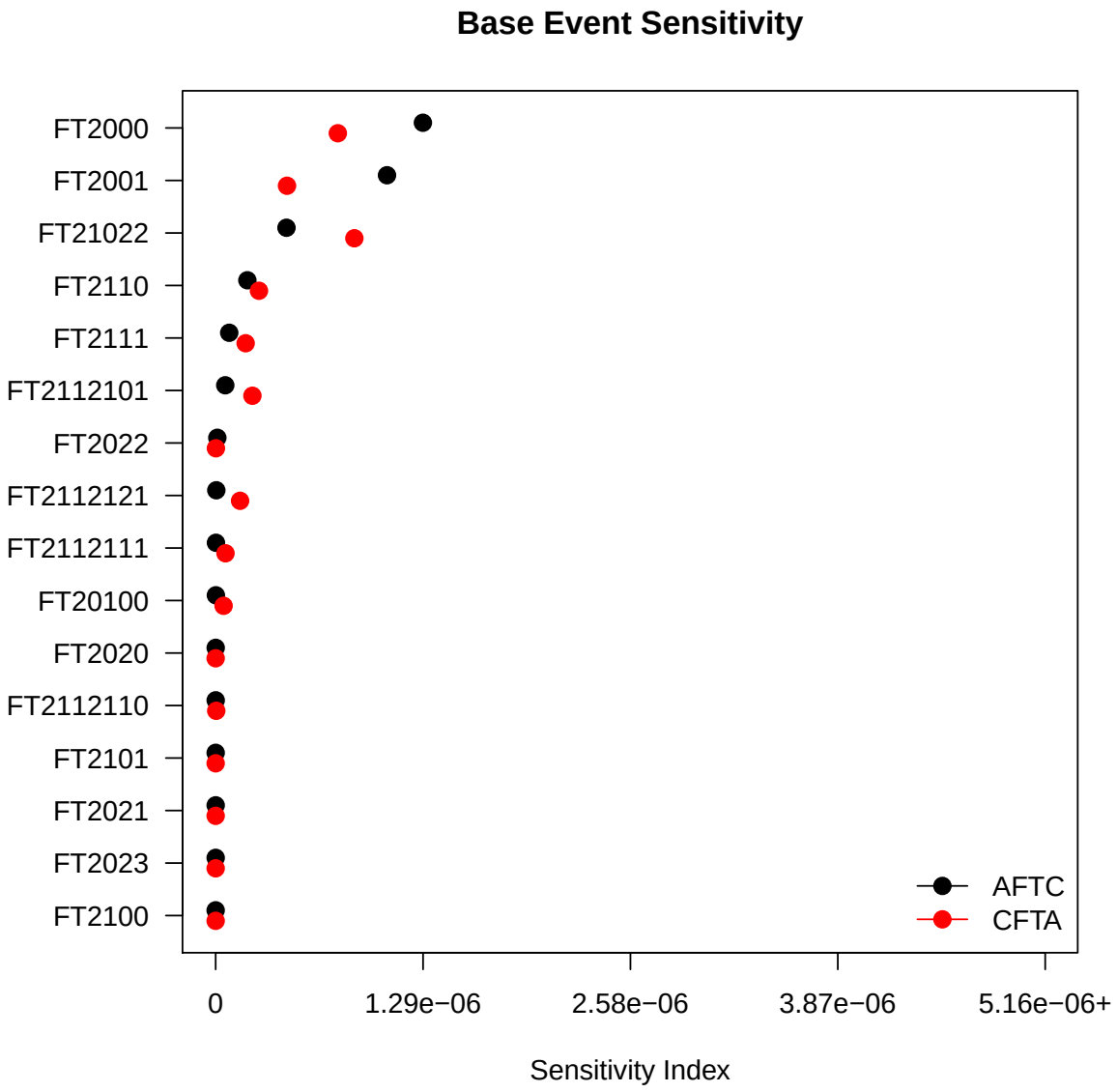


Figure 5.5: Result of the sensitivity analysis for FT2 for the AFTC and CFTA with linear pool analysis methods. Solid circles show the median of the importance measure (Equation 4.6) for the top 15 most important events

5.2 FT3: Spread of construct in non-target eukaryotes

5.2.1 Fault tree structure

The top event in fault tree 3 is defined as *"The probability that the I-PpoI construct will spread in non-target eukaryotes over a year following an escape of 10,000 genetically modified mosquitoes"*. The objective of this analysis is to estimate the likelihood of horizontal gene transfer because this is deemed *a priori* to be undesirable because of the potential adverse environmental consequences (Section 2.3).

The structure of fault tree 3 is shown in Figures 5.7 to 5.15. The tree is large because of the complexity and diversity of HGT mechanisms. The analysis recognises two fundamental conditions for HGT: acquisition of the construct in non-target eukaryotes and spread of the construct. The fault tree identifies four routes of acquisition: unmediated, prokaryote mediated, viral mediated and transposon mediated (Figure 5.7).

Unmediated acquisition entails no intermediate organisms – transfer of the construct proceeds directly from the mosquito to the genome of a germline cell of a non-target eukaryote (Figure 5.8). The analysis recognises that the probability of this event can be quite different between soil environments, aqueous environments and gut environments because contributing factors, such as the probability that the construct remains intact when the mosquito dies (or is ingested), varies depending on environmental conditions. One expert made slight changes to the tree structure by providing separate elicitations for multi-cellular and single-celled eukaryotes below FT300003 and FT300013 (coloured yellow in Figure 5.8). This type of change is accommodated in the analysis by treating all other expert responses as an elicitation at the gate (e.g. at FT300003).

Prokaryote mediated acquisition is similar to unmediated acquisition except that the acquisition of the construct occurs via direct contact between the construct and a competent bacteria (Figure 5.9). As per unmediated acquisition this can occur in three environments, soil, water and the gut of a eukaryote, but also within an additional environment, the genetically modified mosquito itself. The inclusion of an intermediate organism (the competent bacteria) introduces additional steps into the causal event chain – e.g. the construct has to survive the bacteria's restriction enzymes. Once a transformed prokaryote has been created (FT3010) transfer to the non-target eukaryote requires the transformed prokaryote to come into contact with (or enter) the eukaryote, the construct must transfer to the nucleus of a germline cell and recombine into the genome.

Viral mediated acquisition (Figure 5.10) can again occur within the environment or within the I-PpoI modified mosquito. For acquisition to occur within the modified mosquito it must be infected with a virus, the virus has to acquire the construct and still be able to replicate. Viral acquisition occurs either directly or can be mediated by a virion which infects a cell that is already infected with a second virus. On two occasions individual experts amended the tree structure by separating direct acquisition into two steps – transposase excision of the construct and incorporation into the viral genome – and by distinguishing RNA viruses from DNA viruses under FT302002 (transformed virus is able to replicate). These changes are highlighted in yellow in Figure 5.10.

Acquisition within the environment requires the construct to be made available in either soil, water or the gut of a eukaryote. Viral transformation, however, must be mediated by either a transformed competent bacteria (in the case of a bacteriophage) or by some other transformed cellular organism that is infected with the virus (Figure 5.11). Clearly this introduces further steps into the causal event chain, notably the initial transformation of the viral host. Inside the viral host, viral transformation occurs as per the same mechanisms identified within the mosquito.

For the transposon mediated acquisition the fault tree separates piggyBac transposons from other flanking transposons of the same family (Figure 5.12) because the I-PpoI construct is already flanked by piggyBac elements – the transgene is contained within a non-autonomous piggyBac transposable element that has inverted repeats but no transposase-encoding regions (Figure 1.1) – whilst the latter route requires the construct to be flanked by other transposons of the same family in the mosquito (which can occur in three ways, Figure 5.13).

With the exception of the additional steps required for transposons other than piggyBac, the causal events between the two branches of the transposon mediated tree are very similar. Both branches require a source of transposase, reverse transcriptase or integrase to act upon the transposable elements flanking the HEG and excise it from the genome (FT30331 and FT30340). Both branches also recognize that integration into the genome of a non-target eukaryote can occur with and without the active integrase bound to the excised construct. Integration again requires the intact construct (with or without integrase) to leave the mosquito, remain intact in the environment and enter the nucleus of the germline cell of the non-target eukaryote.

Spread of the construct in non-target eukaryotes can occur in three ways: (i) via male or female selection in sexually reproducing eukaryotes; (ii) via one of three potential routes of fitness benefit in non-sexually reproducing eukaryotes; and (iii) via homing or Y-Drive (Figure 5.14).

The I-PpoI gene targets a 15bp sequence. In *P. polycephalum* this sequence is located within the large subunit (26s) rRNA gene that is present in about 300 copies on extrachromosomal nuclear plasmids. In *An. gambiae* this sequence is located on the 28s rDNA gene found in the rDNA repeats of the X chromosome (Figure 5.6). The HE protein, however, tolerates a certain amount of site degeneracy and its target sequence is preserved in the large subunit rRNA gene of all eukaryotes.

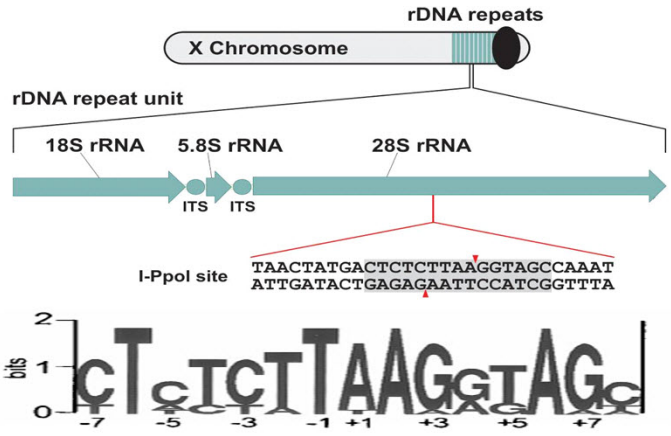


Figure 5.6: Schematic showing the location of the I-PpoI target sequence on the rDNA repeats in the *An. gambiae* genome (top), and the I-PpoI target sequence degeneracy map (bottom).

Hence, in non-target eukaryotes the construct can home in one of three ways: (i) at the same (or slightly degenerated) target site on the ribosomal repeat; (ii) at the same (or slightly degenerated) target site elsewhere in the genome; or (iii) if the HEG mutates, at an alternative target site elsewhere in the genome. For homing to be successful, however, the construct must move into the new recognition site (scenarios (ii) and (iii)), maintain its germline activity and impose a relatively low fitness effect (Figure 5.15).

Fault Tree Graphics Expanded Report
Feb 26, 2015 10:59 AM

FT3- Spread of construct in non-target eukaryotes

FT3

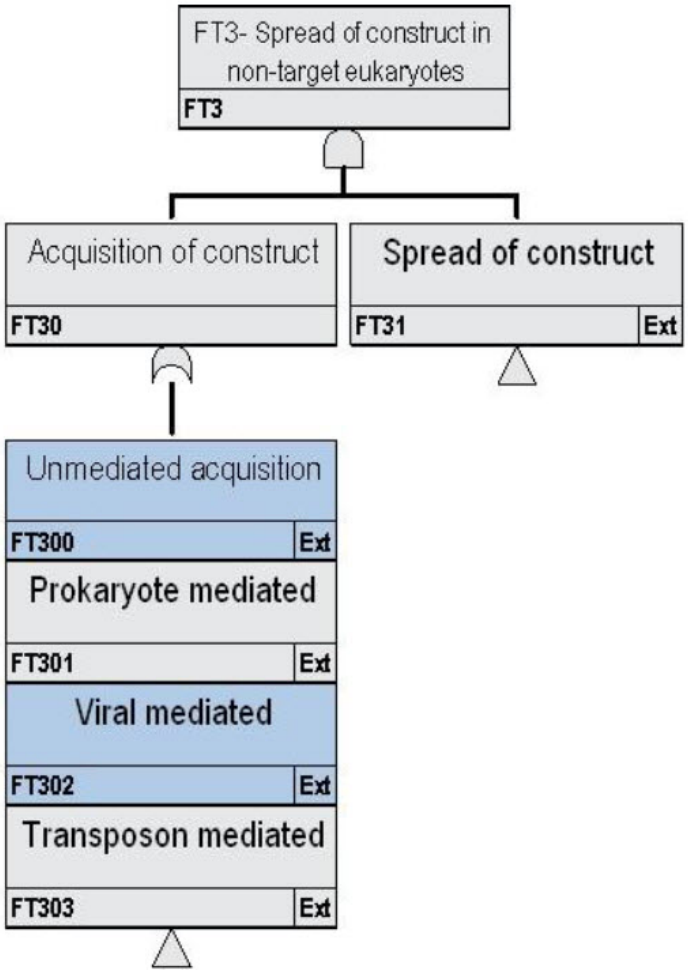


Figure 5.7: Page 1 of fault tree 3

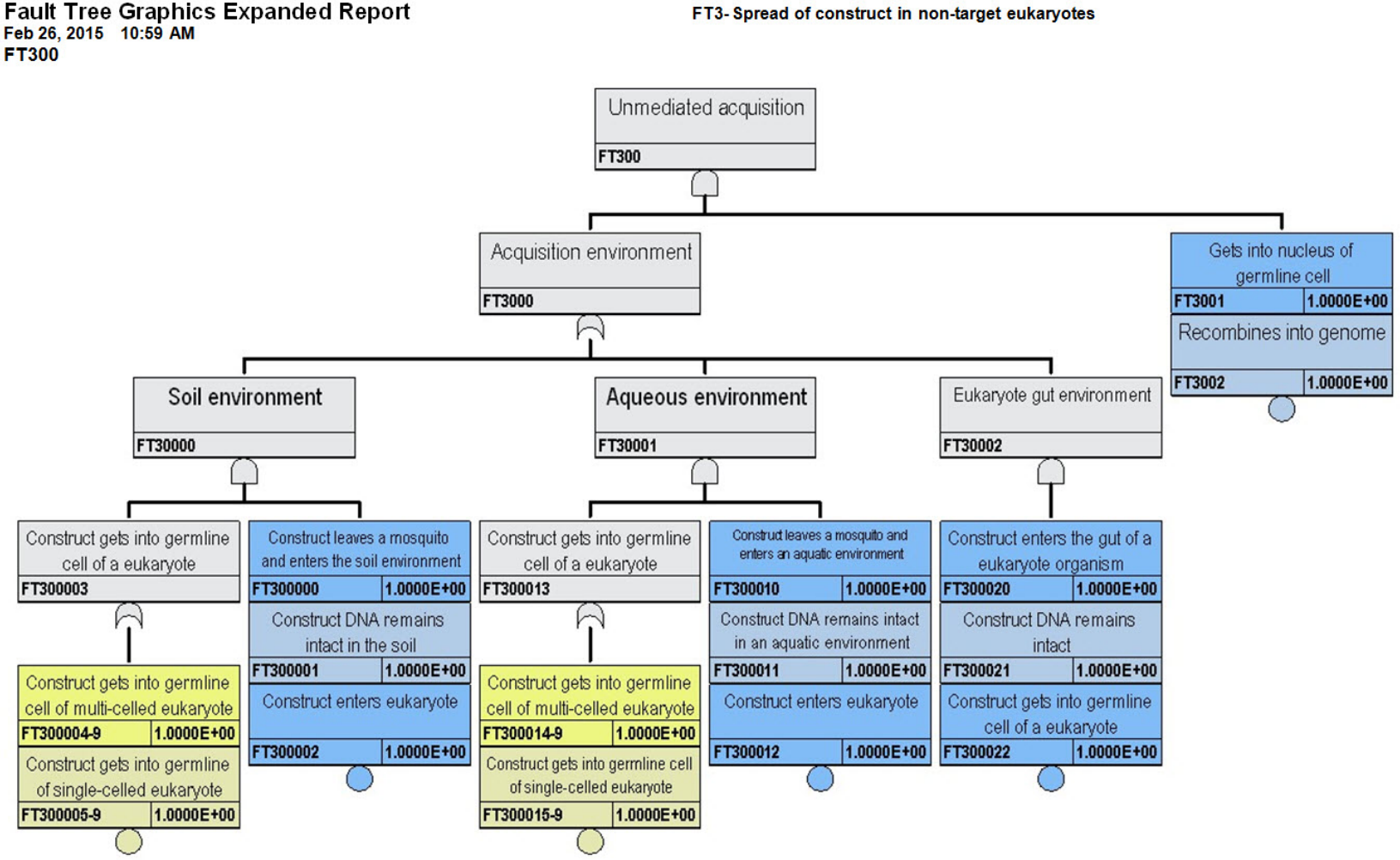


Figure 5.8: Page 2 of fault tree 3

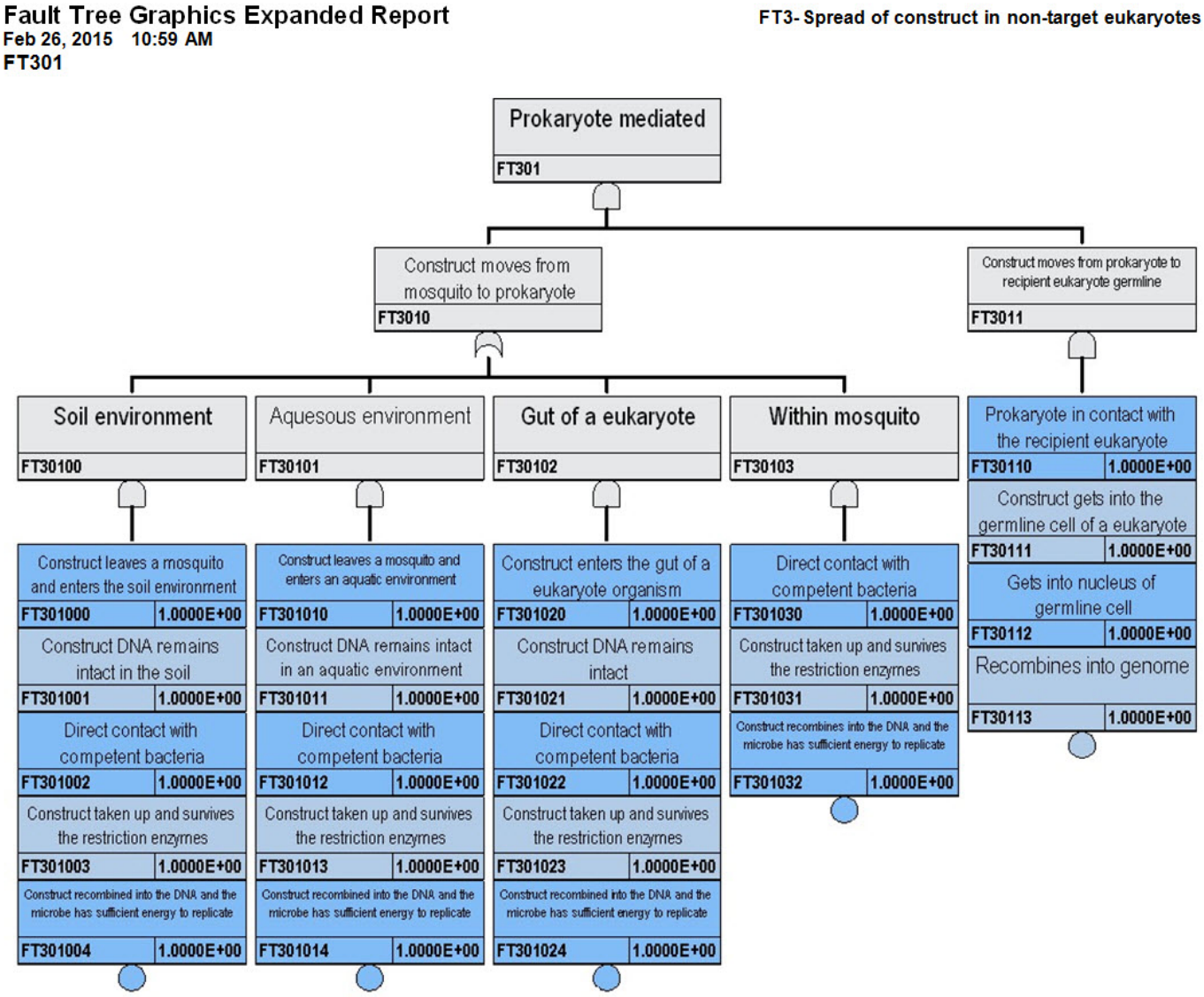


Figure 5.9: Page 3 of fault tree 3

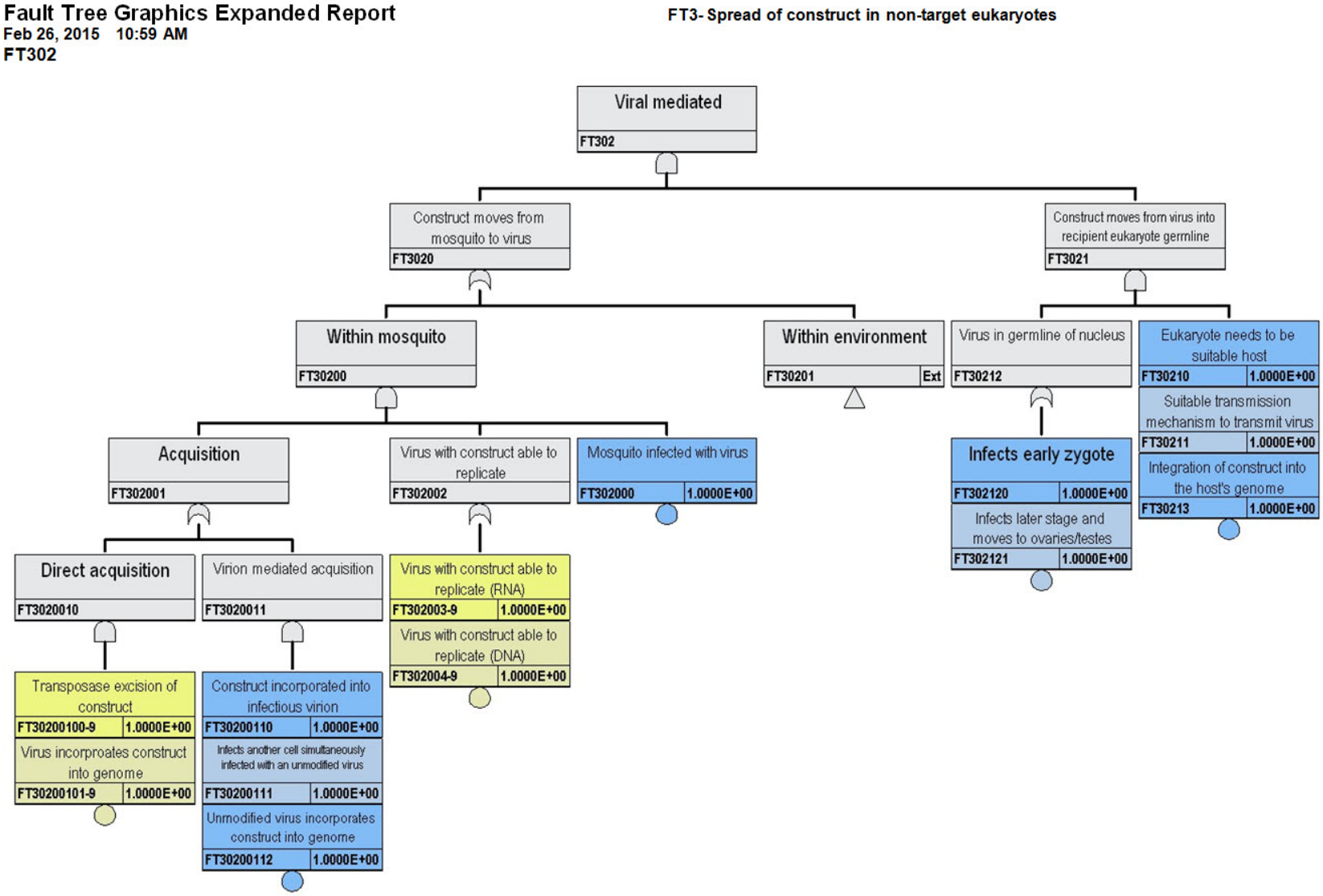


Figure 5.10: Page 4 of fault tree 4

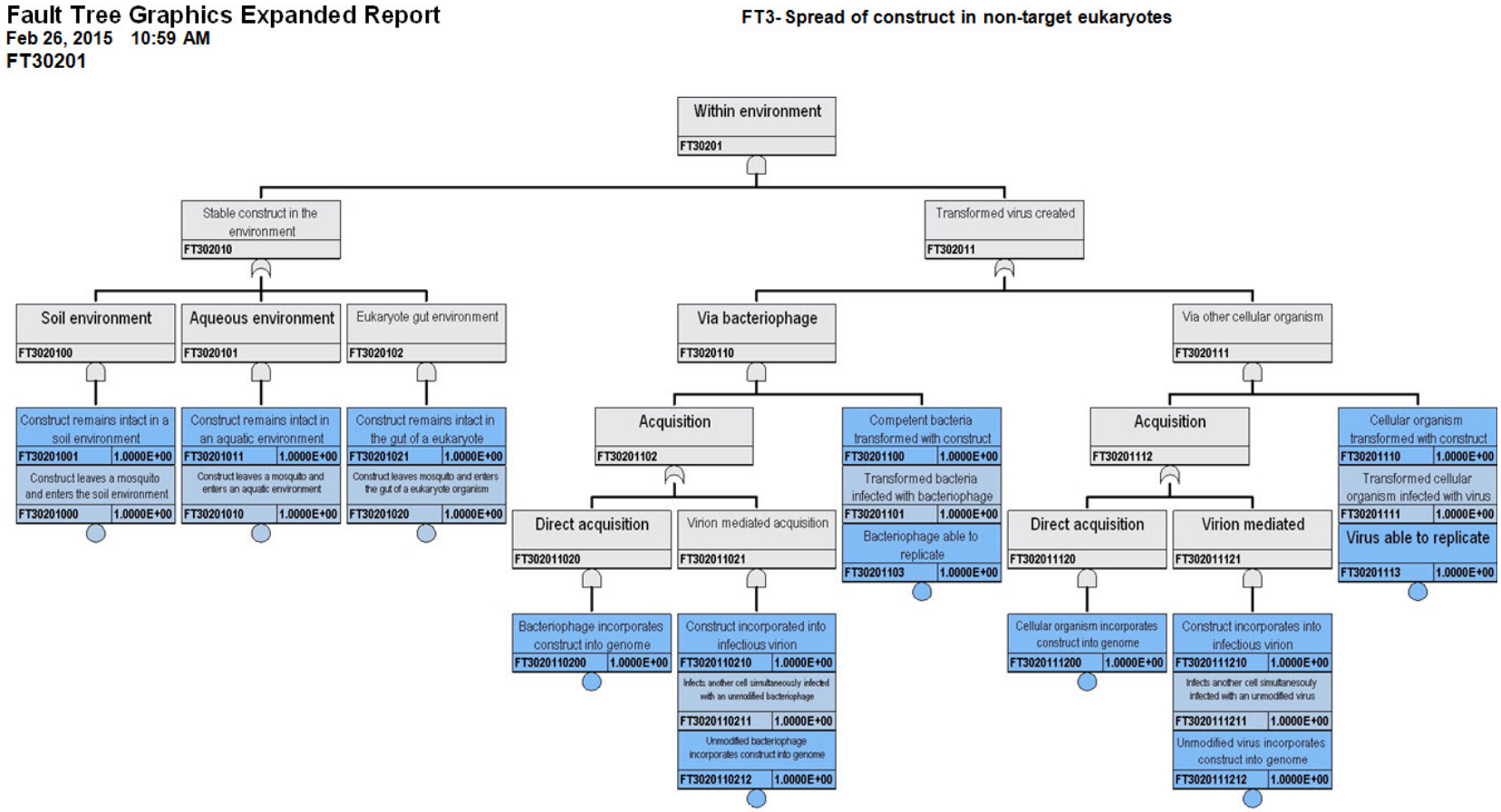


Figure 5.11: Page 5 of fault tree 3

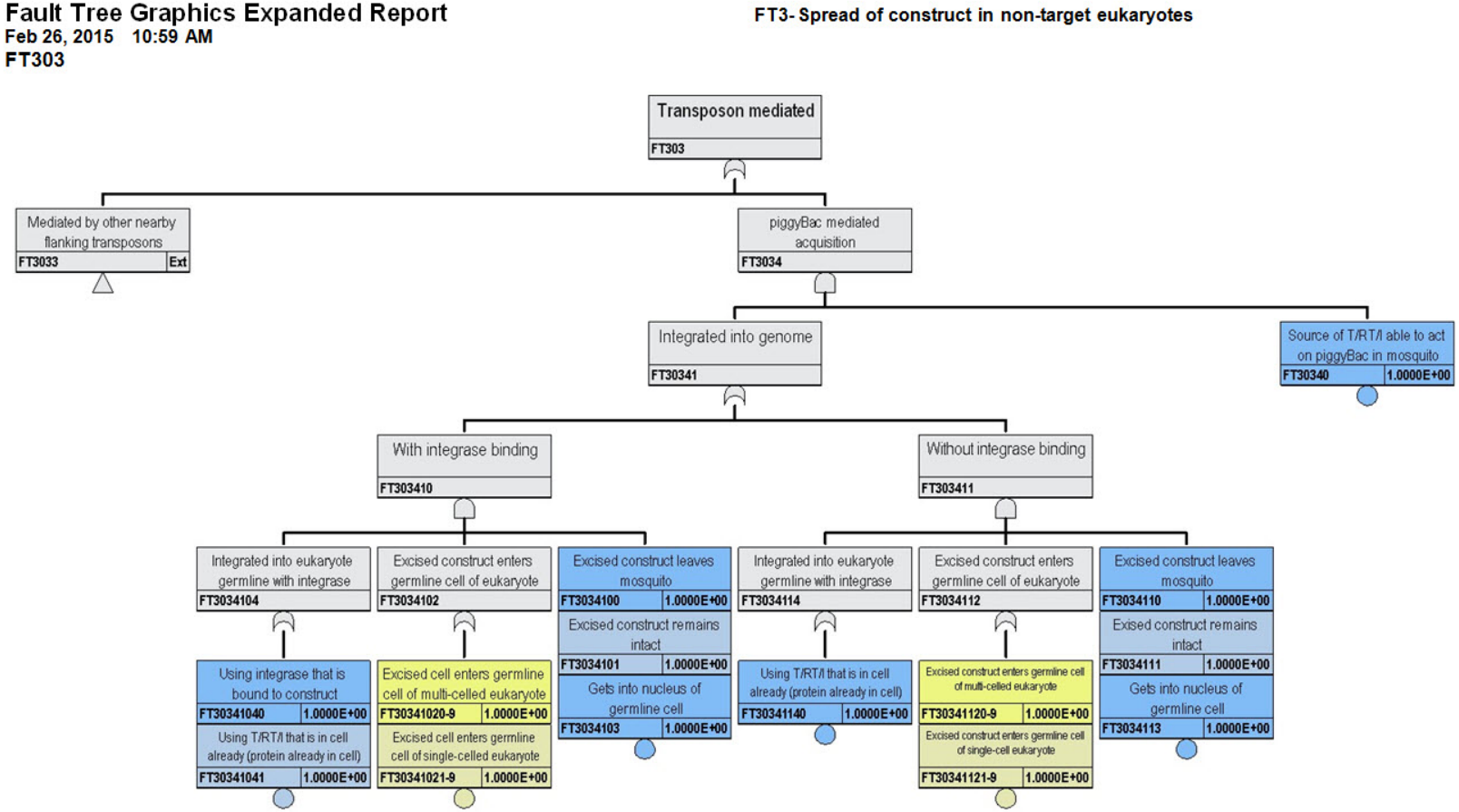


Figure 5.12: Page 6 of fault tree 3

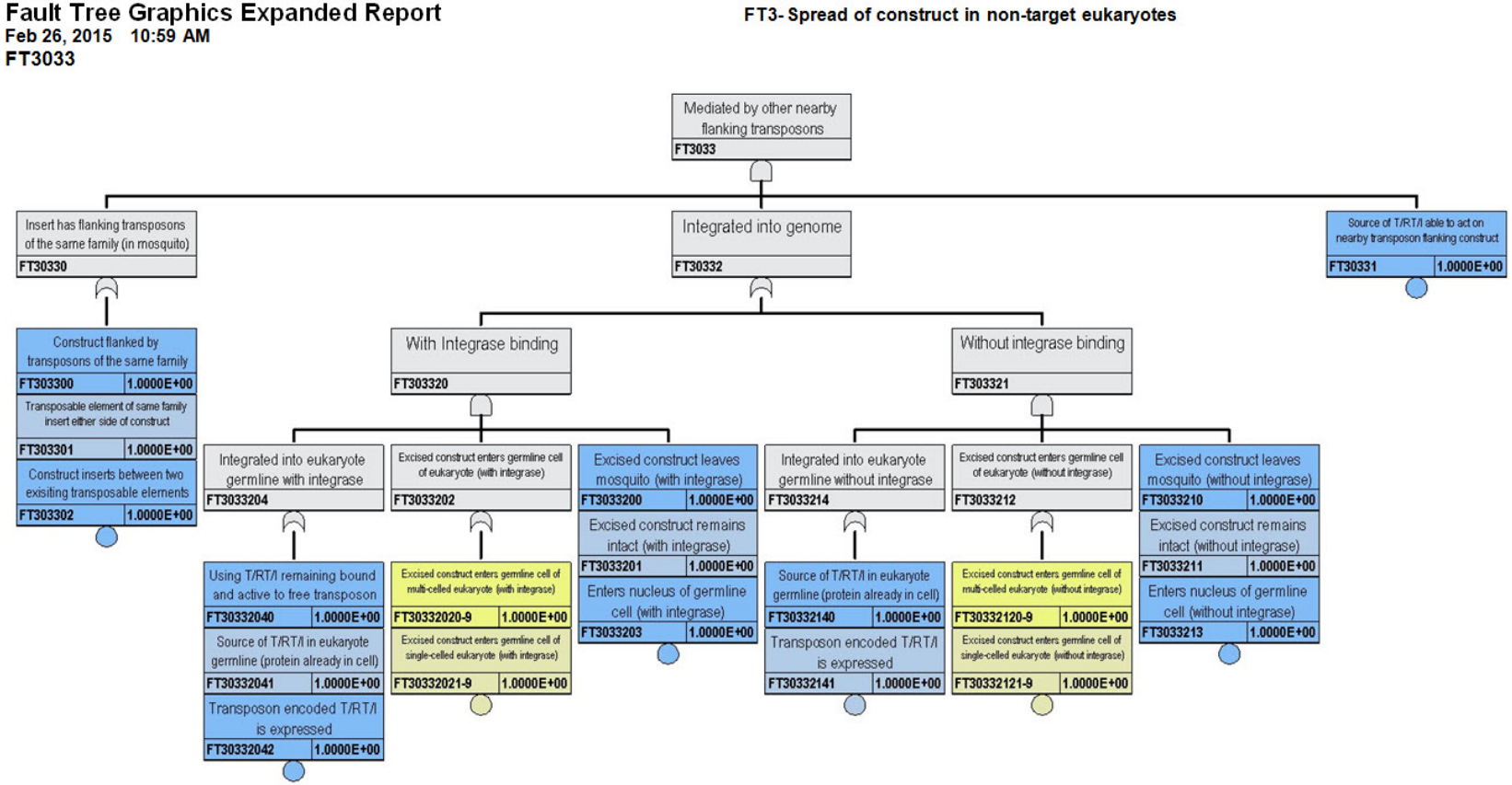


Figure 5.13: Page 7 of fault tree 3

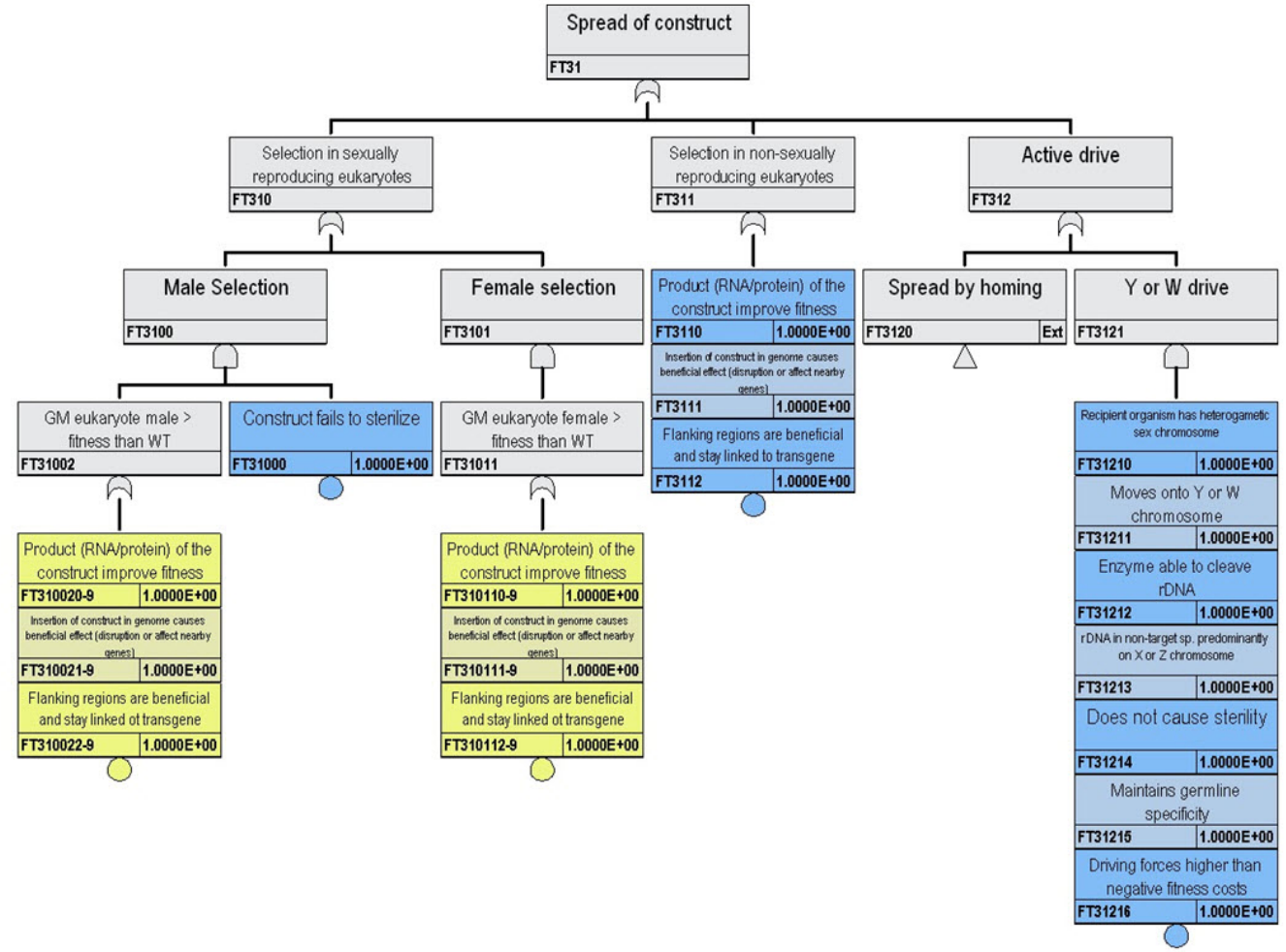


Figure 5.14: Page 8 of fault tree 3

Fault Tree Graphics Expanded Report
Feb 26, 2015 10:59 AM
FT3120

FT3- Spread of construct in non-target eukaryotes

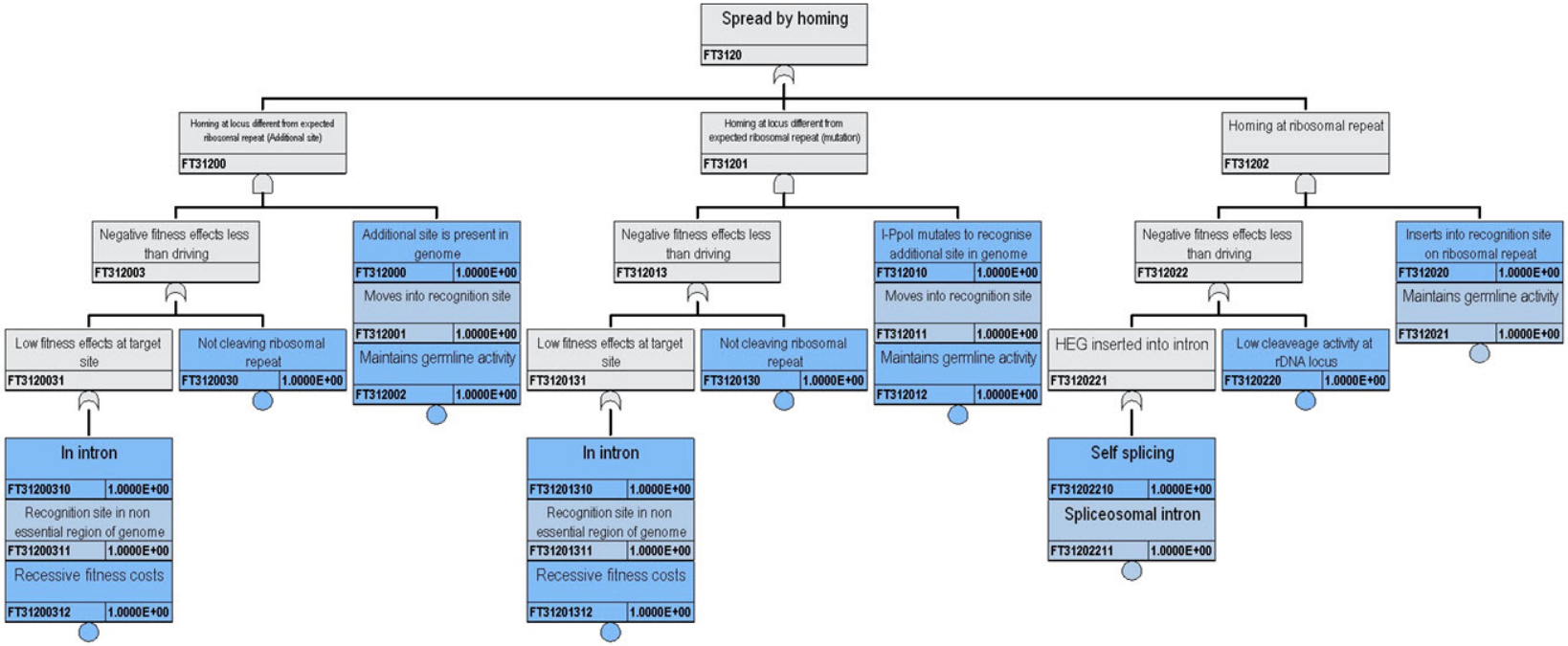


Figure 5.15: Page 9 of fault tree 3

5.2.2 Risk calculations

The results of the quantitative analysis for fault tree 3, using the AFTC and CFTA methods, are summarised in Figure 5.16. Table 5.2 shows the 50th, 90th and 99th percentiles of these results for a combined computational strategy – truncation of 10th order terms and sub-trees. This is not our preferred strategy because there will be a very small, but unknown, truncation error in the results, most likely in the 90th or 99th percentile. This tree, however, is large (it has 816 cut sets and the tree logic text file has 86 rows) and we were unable to complete the probability calculations with the other computational strategies discussed in Section 5.1.2.

Method B10 & C: Squared terms of repeats + truncation of >10 terms + sub-trees			
	50%	**90%**	**99%**
AFTC	1.2×10^{-10}	3.1×10^{-6}	0.02
CFTA_LP	2.4×10^{-10}	7.7×10^{-7}	0.016

Table 5.2: Median and upper tail percentiles for the probability of the top event in fault tree 3, for two analysis methods under the best available computation strategy.

The percentiles of the probability of the top event for fault tree 3 are substantially lower (approximately three orders of magnitude lower at the median) than fault tree 2. The patterns of response noted in fault tree 2, however, are broadly similar: (i) there is little difference between the cumulative distribution functions for AFTC and CFTA (with linear pool for missing values); (ii) there is no evidence for motivational bias on behalf of the Target Malaria consortium; but, (iii) one member of the Target Malaria consortium, and one independent expert, are strong outliers relative to the responses of other experts (Figure 5.16).

The outlying distribution of the independent expert reflects beliefs about the probability of piggyBac mediated transposition (FT3034). Here the expert noted that transposition via piggyBac has occurred “many times” in evolutionary history but acknowledged that the absolute frequency of horizontal gene transfer at short time scales is unknown. The percentiles of their distribution are nonetheless extremely high compared to the other experts interviewed here, they significantly increase the 99th percentile of FT3 (Table 5.2), and are difficult to reconcile with recent literature that suggests only tens to hundreds of horizontal gene transfer events have occurred within populations of vertebrates and invertebrates over hundreds of millions of years (Crisp et al., 2015).

5.2.3 Sensitivity analysis

The sensitivity analysis for FT3 was performed assuming no repeat events due to the very high memory requirements of the algorithms. The results are summarised in Figure 5.17 for the top 15 events. With the exception of FT3110, FT3112 and FT3111 (under the AFTC method) all of these events appear to have very similar influence. Fault tree 3 is very large. The smallest cut sets have six elements, and there are 15 of these in the tree. Fault tree theory shows that the elements in these sets should exert a strong influence on the top event probability. In terms of frequency within these sets FT3001, FT30340 and FT3002 (12 sets) rank highest, but only one of these events are shown here. The next most common events are FT3034100, FT3034101, FT3034103 (8 sets) none of which are shown here, followed by FT31000, FT3110, FT3111, FT3112 (6 sets) which all feature prominently here.

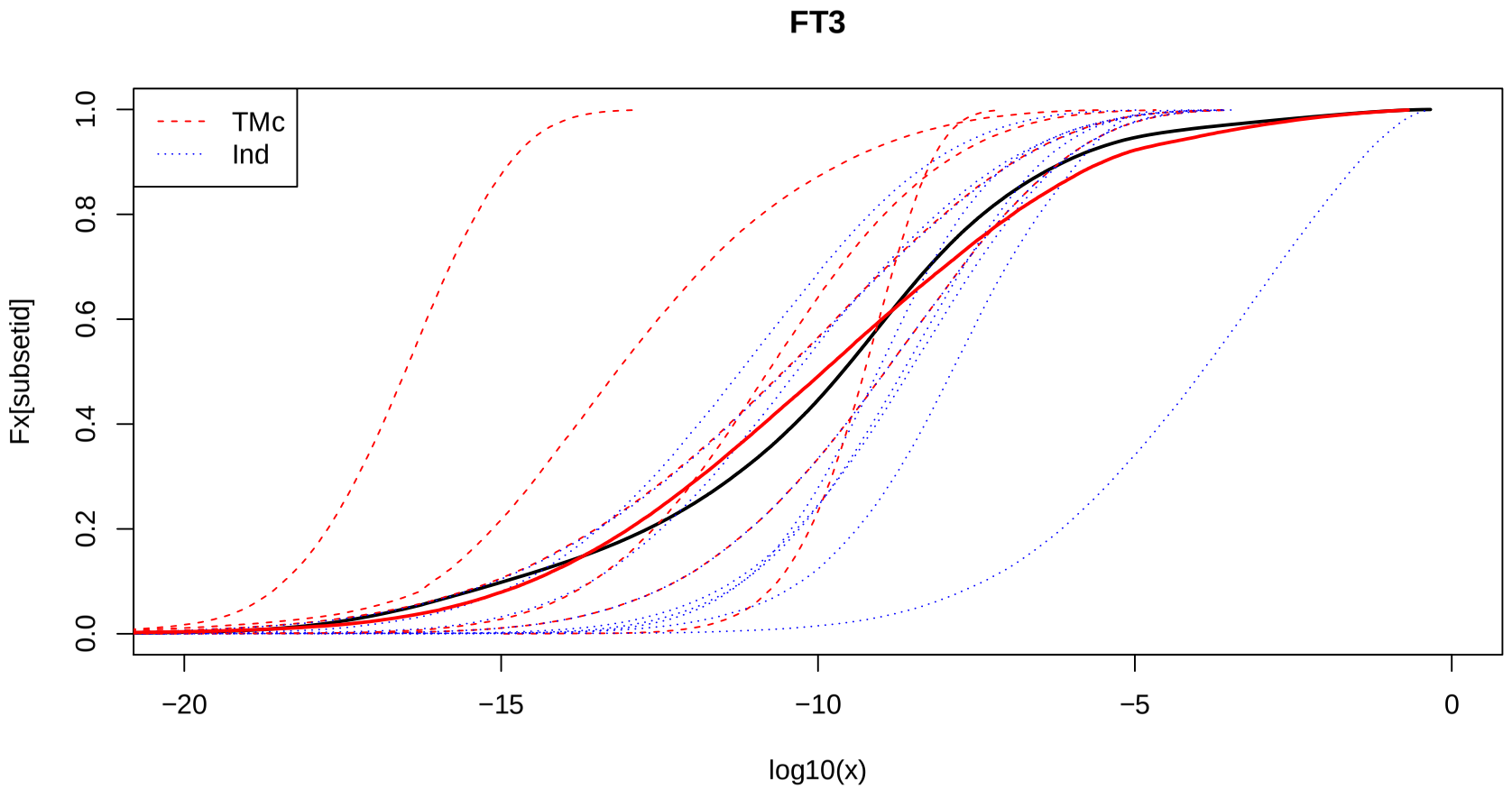


Figure 5.16: Cumulative distribution functions (CDF) of the probability of the top event in fault tree 3. Red solid curve shows the result for the AFTC method. Solid black curve shows the CFTA method using the linear pool of expert opinion for missing values at the basic events. Dashed curves show result for each expert (TMc = Target Malaria consortium, Ind = Independent) under the CFTA strategy using a linear pool for missing values.

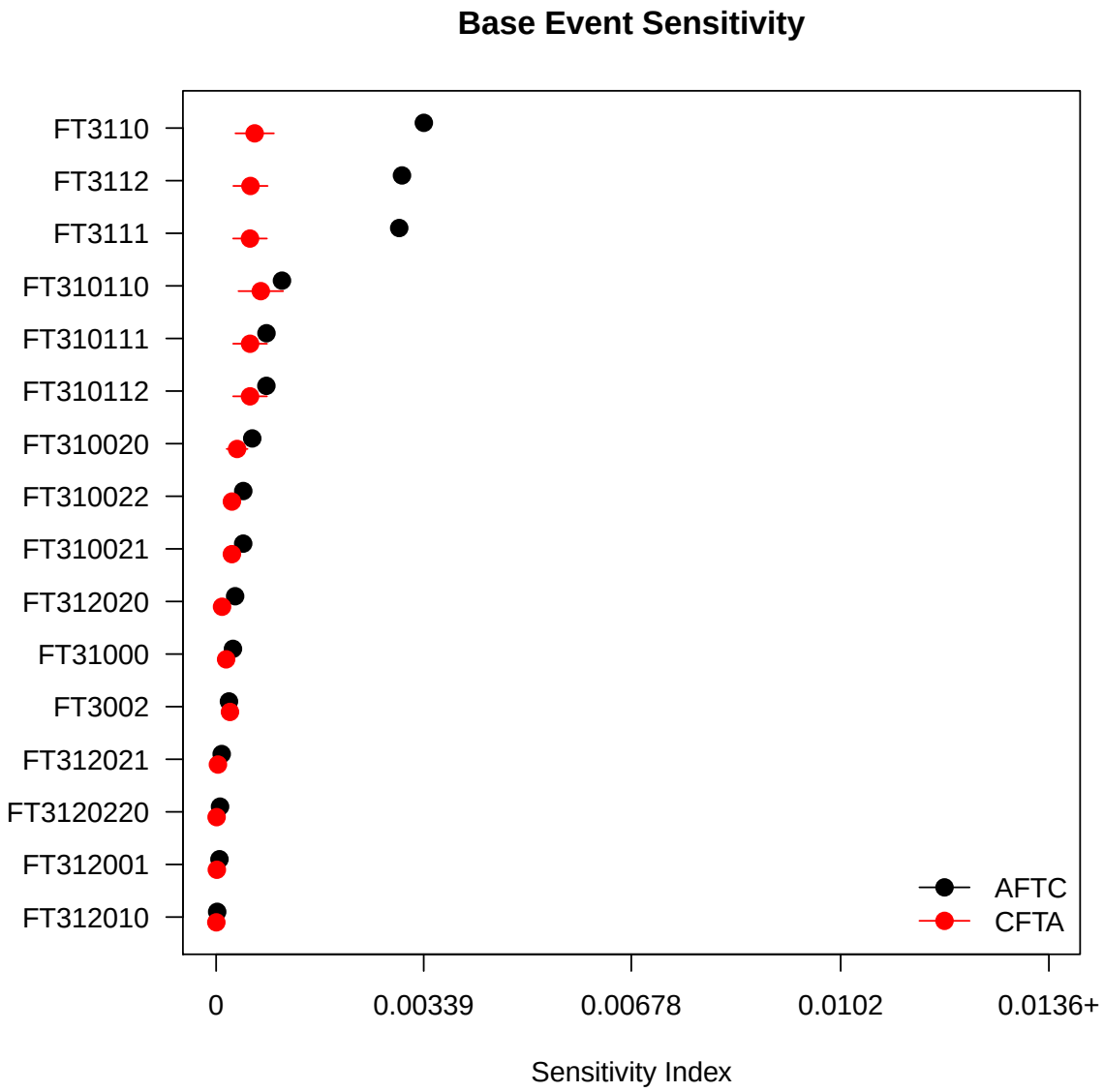


Figure 5.17: Result of the sensitivity analysis for FT3 for the AFTC and CFTA with linear pool analysis methods. Solid circles show the median of the importance measure (Equation 4.6) for the top 15 most important events

5.3 FT4: Spread of the construct in non-eukaryotes

5.3.1 Fault tree structure

The top event in fault tree 4 is defined as *"The probability that the construct spreads in non-eukaryotes over a year following an escape of 10,000 genetically modified mosquitoes"*. This analysis complements fault tree 3 wherein prokaryotes and viruses were intermediaries in the spread of the construct. This tree addresses the probability of spread in these groups (Figure 5.18).

As before, spread of the construct in any non-target population requires acquisition and spread, either by selection or homing. In prokaryotes acquisition can occur in a soil or aquatic environment or in the gut of a eukaryote, as previously outlined . Spread via selection is possible via one of three positive selection mechanisms: (i) HE protein is expressed and somehow confers a selective advantage to the prokaryote; (ii) the insertion of the HEG into the genome creates a positive fitness effect by disrupting nearby genes or the sequences flanking the construct stay linked during transformation and these somehow improve fitness; or, (iii) the construct is linked to one of the prokaryotes mobile genetic elements and increases in frequency by "hitching-hiking" (Figure 5.19).

Homing is possible but only in a small group of prokaryotes whose genetic systems allows for the possibility of HEG- and HEG+ alleles. In this group the process is very similar to that previously described under fault tree 3: the I-PpoI recognition site must either already be present, or the HEG must mutate to recognize a site in the genome, the HEG must move into this site, be expressed but then impose a relatively low fitness cost.

The mechanisms by which Viruses might acquire the I-PpoI construct are identical to those discussed under fault tree 3 (Figures 5.20 and 5.21). The mechanisms for spread of the construct in Viruses are identical to Prokaryotes discussed above, and spread by homing is only possible in DNA viruses with HEG- and HEG+ co-infection.

Fault Tree Graphics Expanded Report
Feb 26, 2015 10:57 AM

FT4 - Spread of construct in non-eukaryotes

FT4

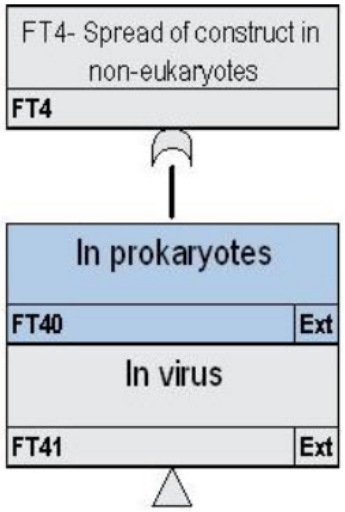


Figure 5.18: Page 1 of fault tree 4

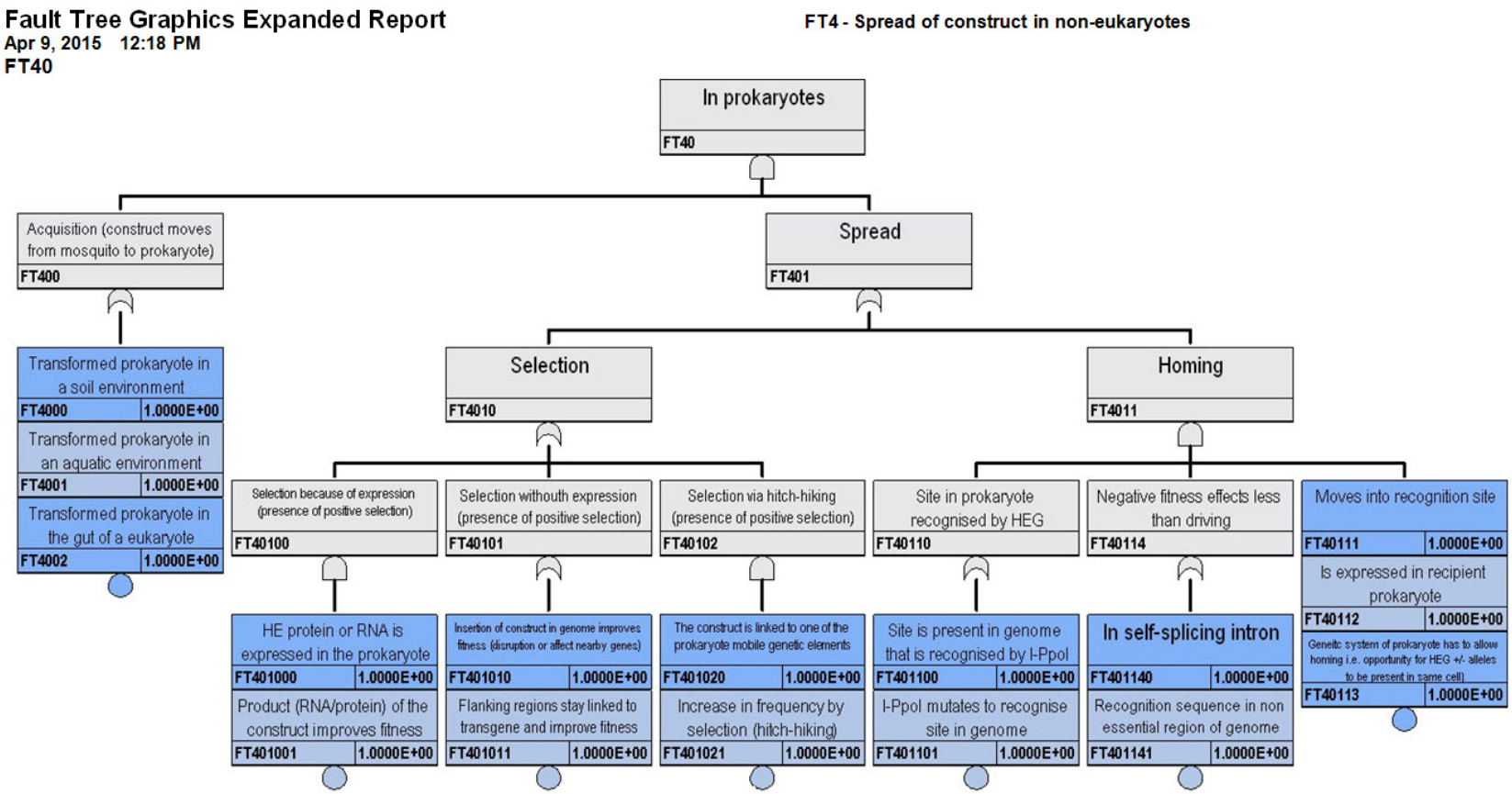


Figure 5.19: Page 2 of fault tree 4

Fault Tree Graphics Expanded Report
Feb 26, 2015 10:57 AM
FT41

FT4 - Spread of construct in non-eukaryotes

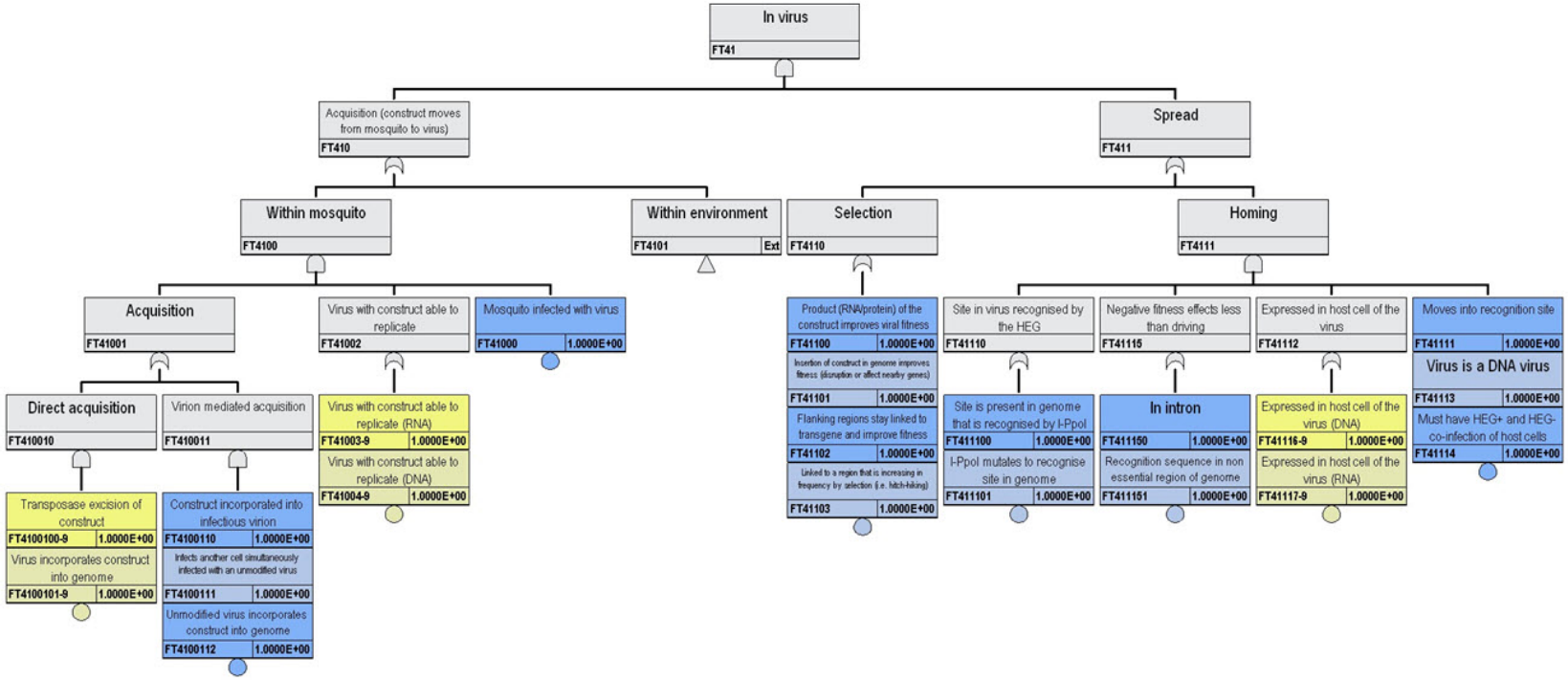


Figure 5.20: Page 3 of fault tree 4

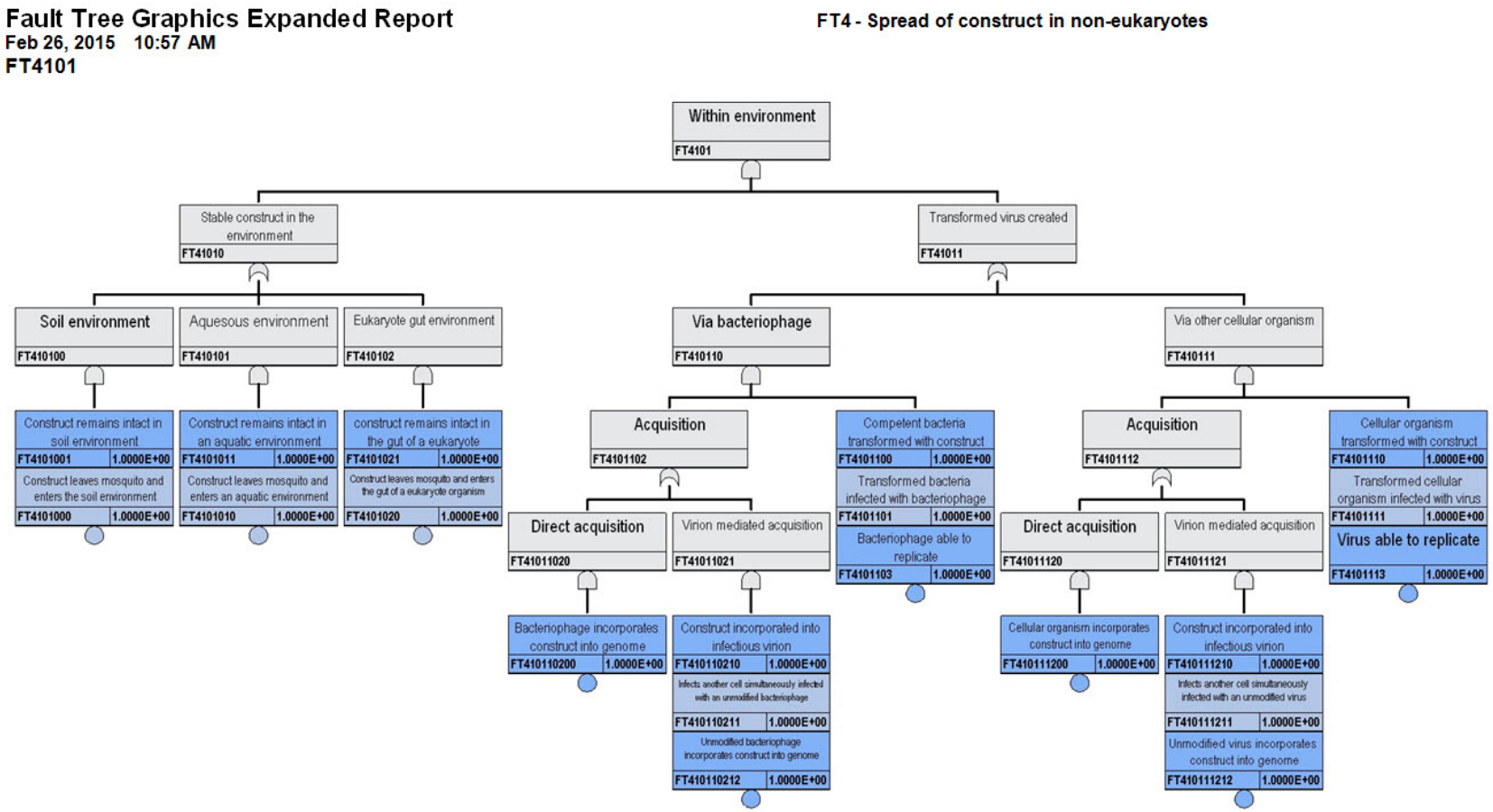


Figure 5.21: Page 5 of fault tree 4

5.3.2 Risk calculations

The results of the quantitative analysis for fault tree 4, using the AFTC and CFTA methods, are summarised in Figure 5.22. Table 5.3 shows the 50th, 90th and 99th percentiles of these results. In this tree there are no repeat events hence computational strategy E is sufficient, and there is no truncation error.

Method D No repeat events			
	50%	**90%**	**99%**
AFTC	6.7×10^{-7}	7.8×10^{-4}	0.057
CFTA_LP	1.3×10^{-7}	2.3×10^{-4}	0.002

Table 5.3: Median and upper tail percentiles for the probability of the top event in fault tree 4, for two analysis methods under the best available computation strategy.

The percentiles of the top event probability of fault tree 4 are (perhaps unsurprisingly) higher than fault tree 3 – fundamentally there are fewer steps in the causal chain of spread in non-eukaryotes than spread in non-target eukaryotes. The risk estimates are similar in magnitude to fault tree 2 but there are two important differences in the pattern of expert response. Firstly, the uncertainty in the responses of individual experts (indicated by the slope of their distribution functions) and the difference of opinions across the experts (indicated by the distance between their distribution functions) is higher than fault tree 2. Secondly, the difference between the AFTC method and the CFTA with linear pool method is higher than in fault tree 2 or 3. This may be due to the higher levels of uncertainty noted earlier (Figure 5.22).

Fewer experts responded to the conditional probability questions in fault tree 4 than fault tree 2, but again with the possible exception of a single “outlier”, there is no obvious difference in the spread pattern between independent experts and experts from the Target Malaria consortium. In this context it should be noted that many of the basic events in fault tree 4 are identical to the events in fault tree 3 (see Appendix D), and hence the CFTA results are very similar between the two trees.

5.3.3 Sensitivity analysis

The results of the sensitivity analysis for FT4 suggest that two basic events – FT401011 and FT401020 – have the highest influence on the probability of the top event for FT4 followed by FT401010 and FT41003. In this instance the structure of the tree appears to be a strong contributor to this effect. FT4 has 6 minimum cut sets with only two elements, all of the first three include FT401010, and all of the last three contain FT401011. The other members are FT4000, FT4001 and FT4002, one of which also appear in the top 15 here. FT401010, FT401011 and FT401020 also have significant levels of disagreement between experts. These results highlight how the importance measure is confounded by inter-expert uncertainty and position of the event in the tree.

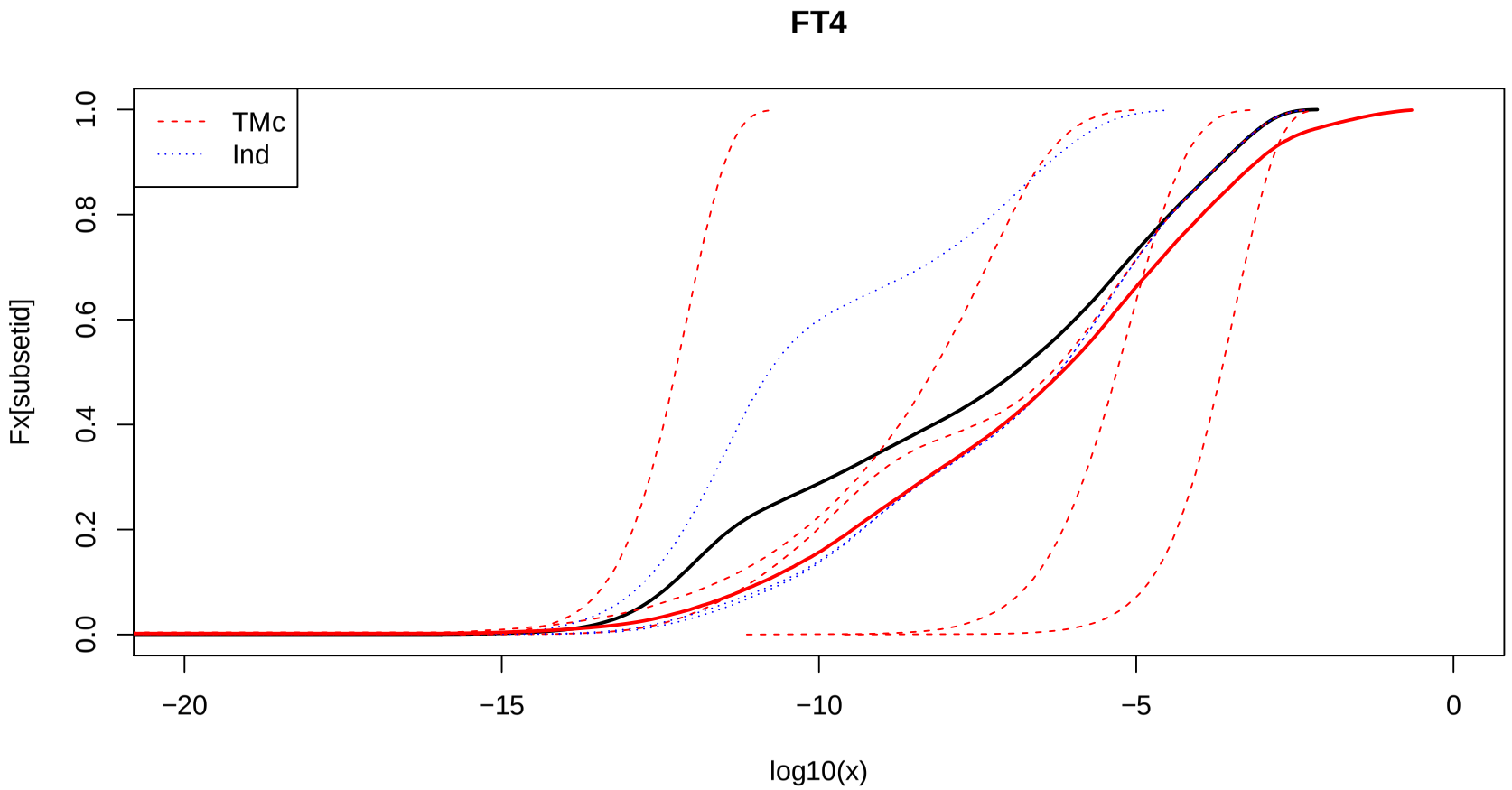


Figure 5.22: Cumulative distribution functions (CDF) of the probability of the top event in fault tree 4. Red solid curve shows the result for the AFTC method. Solid black curve shows the CFTA method using the linear pool of expert opinion for missing values at the basic events. Dashed curves show result for each expert (TMc = Target Malaria consortium, Ind = Independent) under the CFTA strategy using a linear pool for missing values.

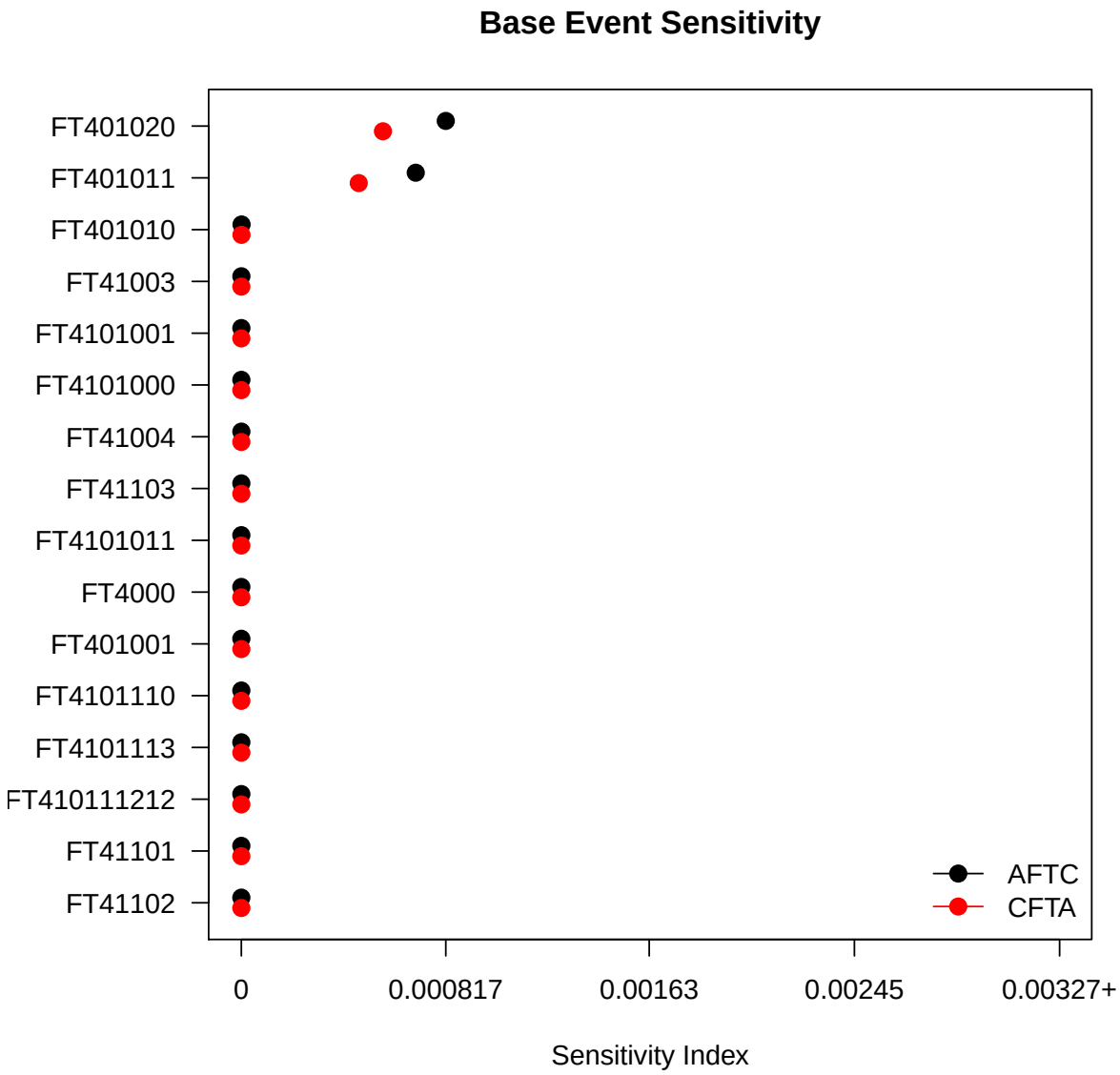


Figure 5.23: Result of the sensitivity analysis for FT4 for the AFTC and CFTA with linear pool analysis methods. Solid circles show the median of the importance measure (Equation 4.6) for the top 15 most important events

5.4 FT50: Spread of the construct in *An. gambiae*

5.4.1 Fault tree structure

The top event in fault tree 50 is defined as *"The probability that the I-PpoI construct will spread in An. gambiae over a year following an escape of 10,000 genetically modified mosquitoes"*. The fault tree structure suggests two fundamentally different causal pathways. The first is mediated by *Wolbachia* infections of genetically modified mosquitoes. The second occurs via (non-*Wolbachia*) vertical gene transfer mechanisms (Figure 5.24).

Wolbachia is a genus of maternally transmitted intracellular bacteria that spread by infecting the testes and ovaries of their arthropod hosts and manipulating their reproduction. These manipulations favour the birth rate of infected females in one of four ways, most notably by killing infected males during larval development, feminization of infected males or by cytoplasmic incompatibility – a process whereby infected males are only able to reproduce with uninfected females or females infected with a different strain of *Wolbachia*.

The reproductive advantages that cytoplasmic incompatibility confers on infected females has made *Wolbachia* a target for controlling many arthropod transmitted diseases, including mosquito vectored malaria and dengue fever (Alphey, 2014). The discovery of populations of *An. gambiae* in Burkina Faso, naturally infected with *Wolbachia* (Baldini et al., 2014), however, provides two theoretical avenues for the I-PpoI construct to spread through uninfected populations. The first occurs via transformation of the *Wolbachia* bacteria with the construct, although we note here that transformation of *Wolbachia* has not been demonstrated in the laboratory despite decades of effort (McGraw and O'Neill, 2013). The second is via the construct becoming inserted in mitochondrion that then spreads through the population by "hitch-hiking" with *Wolbachia*. In both cases spread occurs through the usual reproductive advantage that *Wolbachia* confers on infected females.

The causal pathways of spread via vertical gene transfer are essentially identical to those previously described for spread of the construct in sexually reproducing eukaryotes in Figure 5.14 and Figure 5.15 (but clearly the probability of the basic events are quite different). In particular, spread can occur either via selection or active drive, where the former requires some males to be fertile, and genetically modified males or females to have a higher fitness than wild types, or the construct somehow mutates to increase the fitness of genetically modified females enough to compensate for the fitness costs of the sterile males (Figure 5.24).

Again active drive can occur through homing or movement of the I-PpoI construct onto the Y chromosome, and homing can occur at the ribosomal repeat (unmutated construct) or some other locus (mutated or unmutated construct) (Figures 5.25 and 5.26).

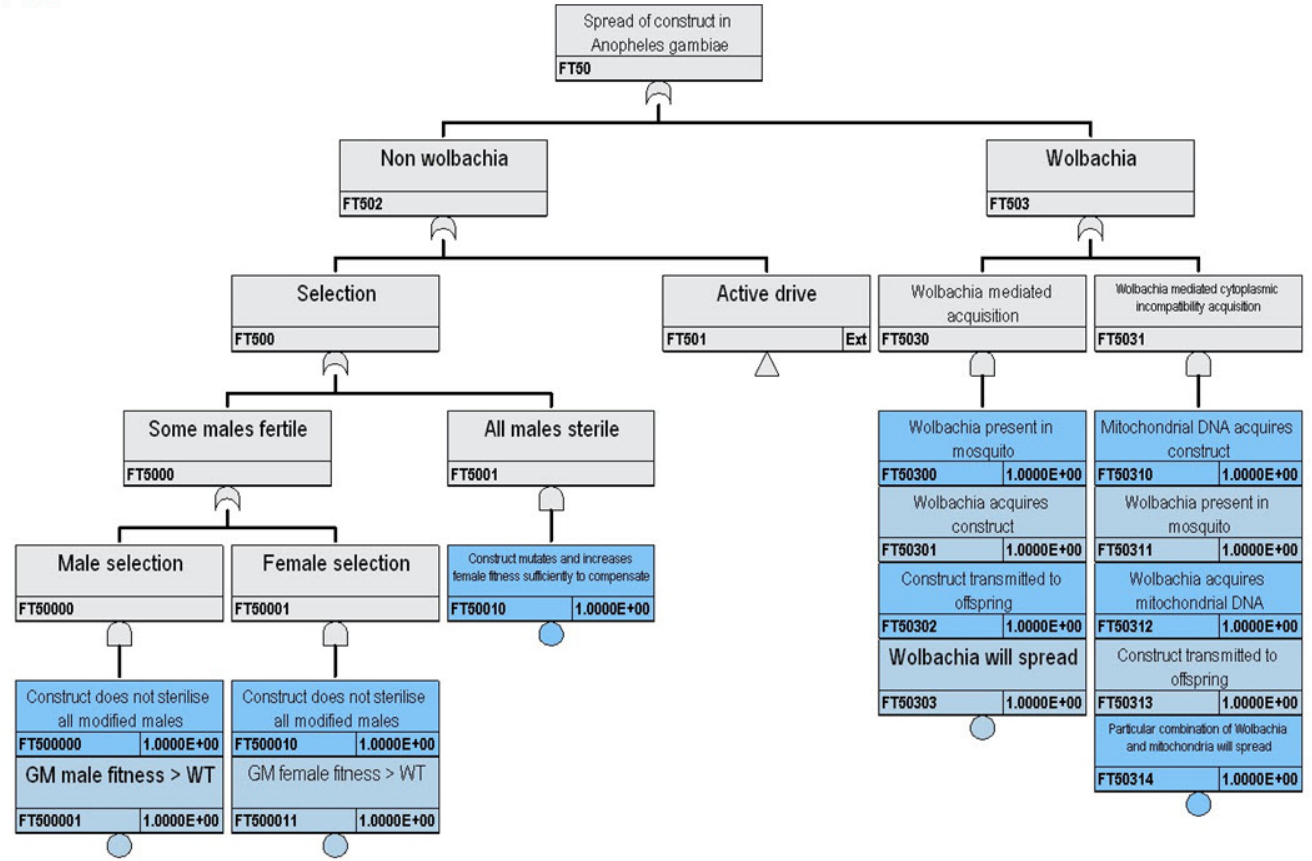


Figure 5.24: Page 1 of fault tree 50

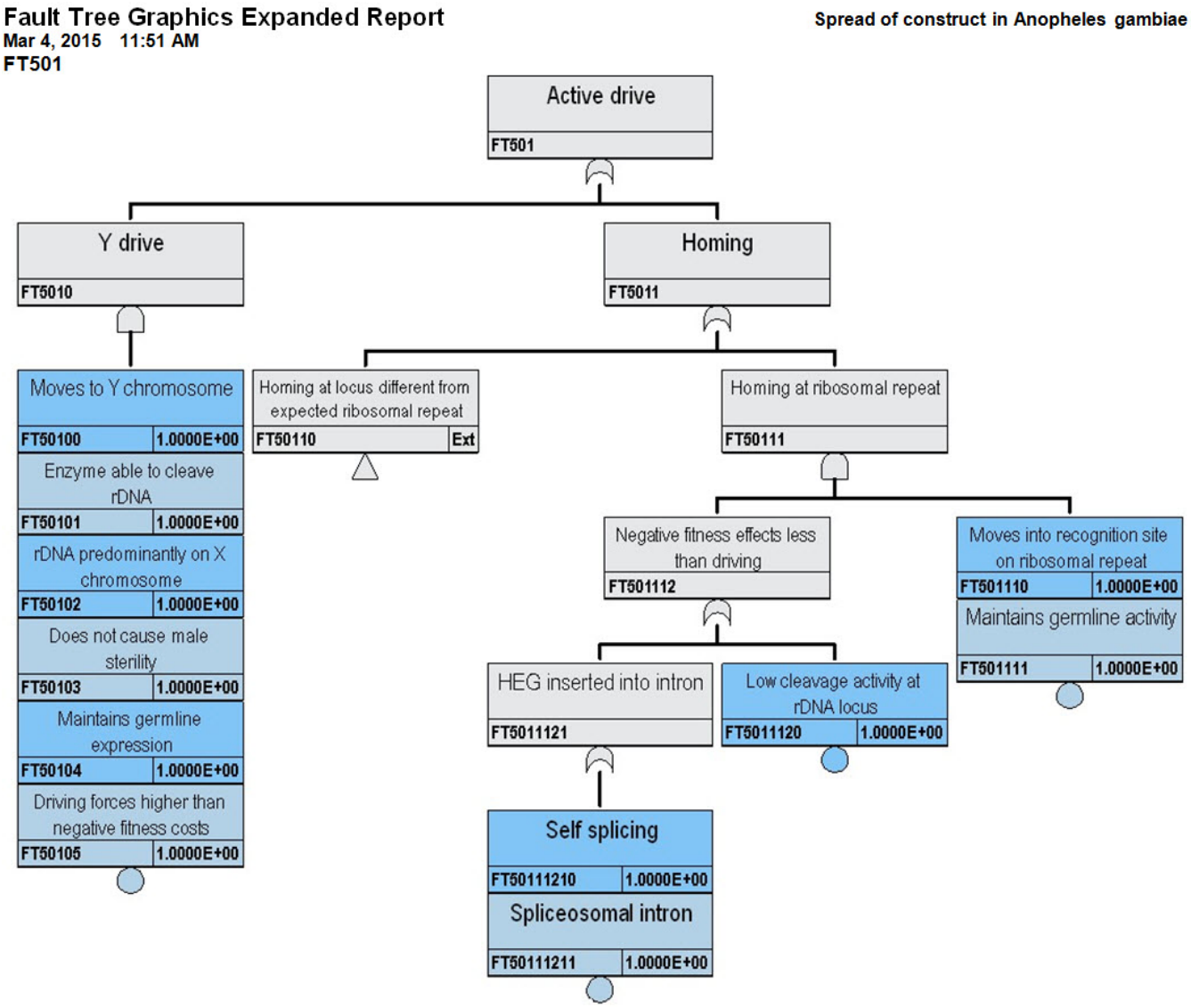


Figure 5.25: Page 2 of fault tree 50

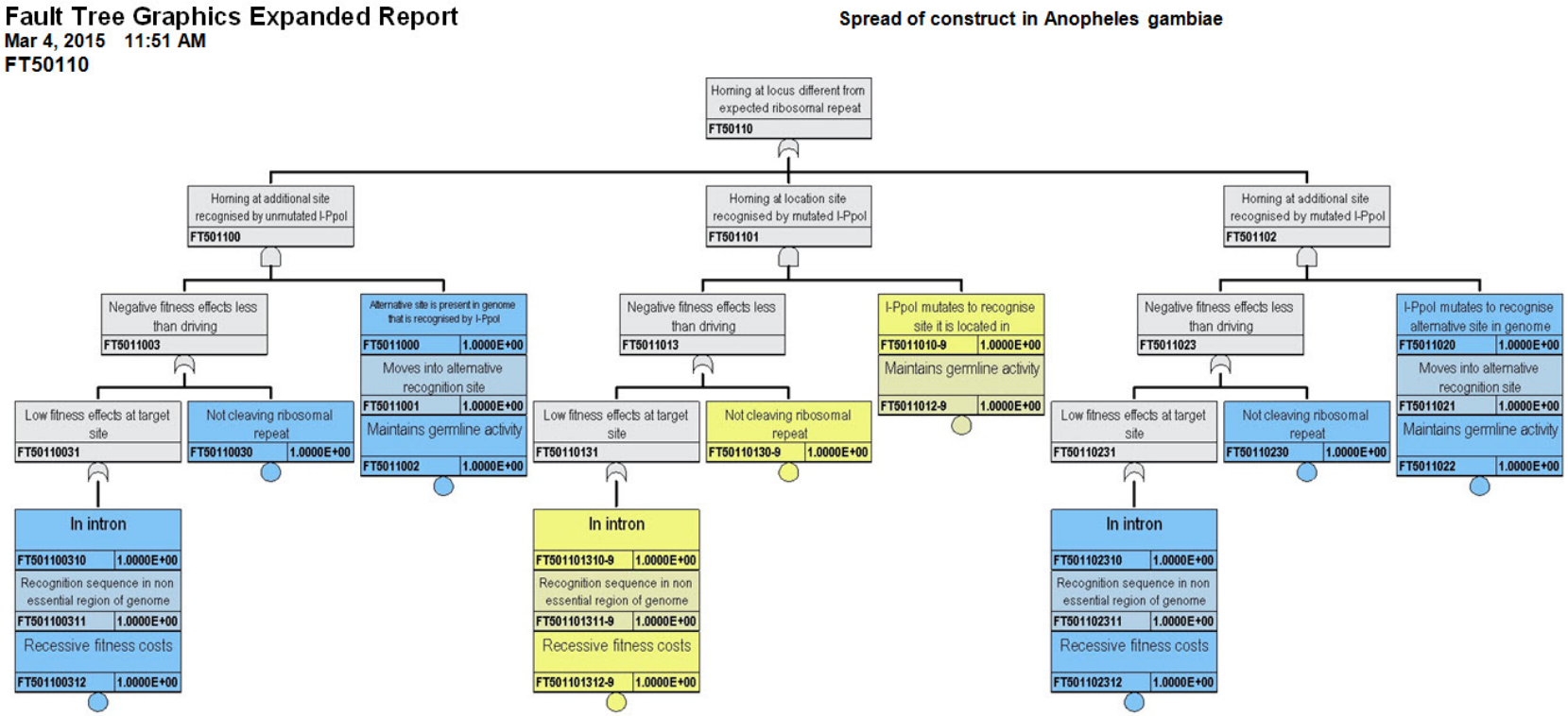


Figure 5.26: Page 3 of fault tree 50

5.4.2 Risk calculations

The results of the quantitative analysis for fault tree 50, using the AFTC and CFTA methods, are summarised in Figure 5.27. Table 5.4 shows the 50th, 90th and 99th percentiles of these results for two comparable computation strategies – in this case accounting for squared terms of repeat event with 12th order term truncation, and allowing these terms (i.e. ignoring the error they introduce). There is only one repeated event in FT50 so, as expected, the results are almost identical. This confirms that there are unlikely to be any errors in the R code for the handling squared repeat events with truncation strategies.

Method B12 Squared terms of repeats handled + truncation of order $>$12 terms			
	50%	**90%**	**99%**
AFTC	0.0014	0.25	0.88
CFTA_LP	3.7×10^{-4}	0.06	0.78
Method E. Squared terms of repeats allowed			
	50%	**90%**	**99%**
AFTC	0.0021	0.26	0.88
CFTA_LP	4×10^{-4}	0.06	0.78

Table 5.4: Median and upper tail percentiles for the probability of the top event in fault tree 50, for two analysis methods under two comparable computation strategies.

The percentiles of the cumulative distribution function for the top event of FT50 are the highest of all the results reported so far. This is not surprising and the reasons are clear from the structure of page 1 of the fault tree (Figure 5.24). The selection route via FT500 is a very short causal pathway. In a cut set analysis, short causal pathways lead to cut sets with only one or two members, which is qualitatively indicative of events that are likely to dominate the probability of the top event.

In our analysis the probability of the product of FT500000 and FT500001, or the product of FT500010 and FT500011, or the probability FT50010 are likely to dominate the probability of the top event. In this context it is relevant to note that FT50010 is an event with significant uncertainty between experts (significant defined as a difference greater than five orders of magnitude between the enveloping distribution functions) which may indicate one or more experts has misinterpreted the elicitation question.

The pattern of individual expert assessments (Figure 5.27) shows a reasonably high degree of uncertainty across the experts, again with no strong differences in the pattern of spread between the two groups of experts.

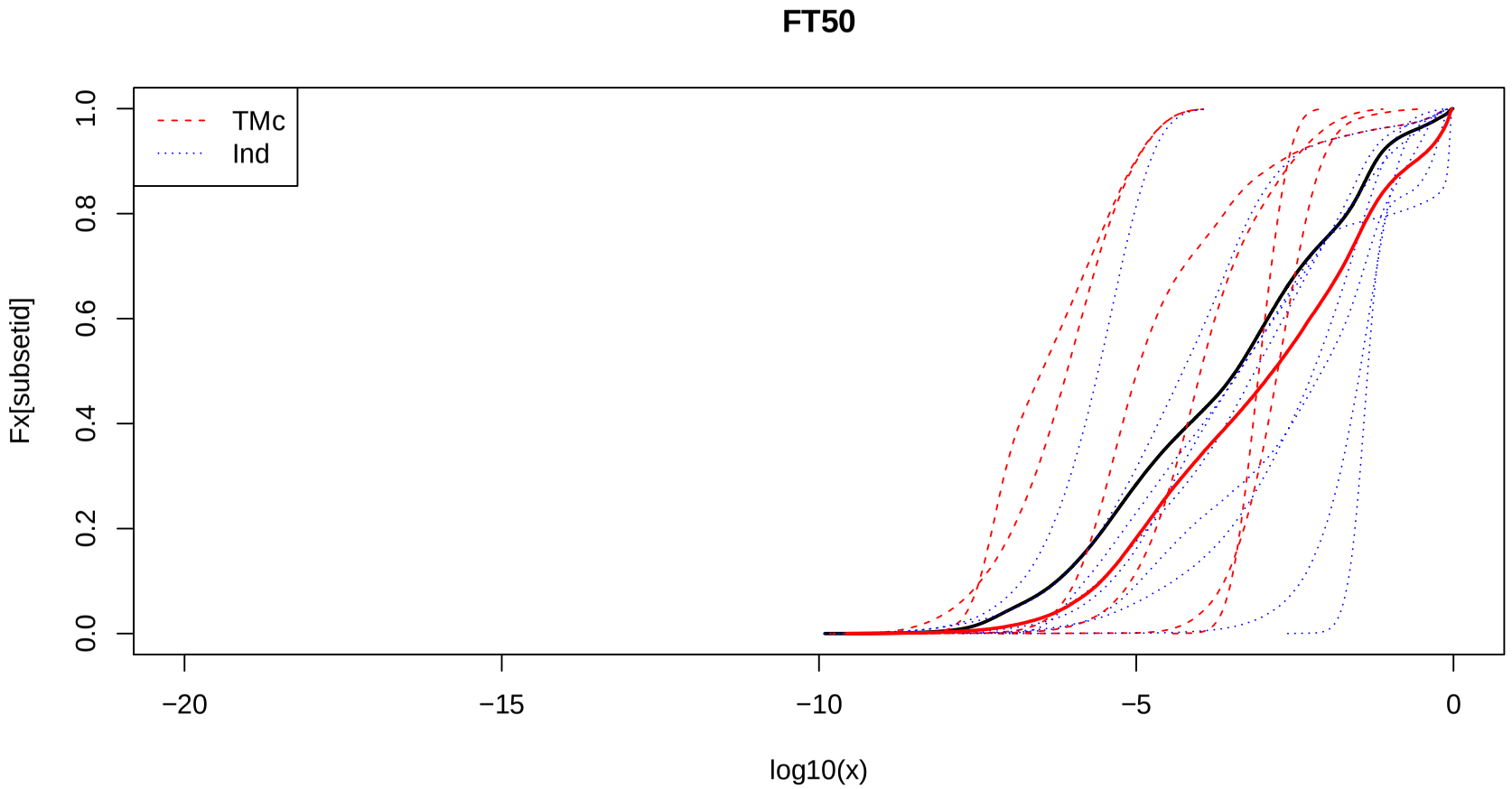


Figure 5.27: Cumulative distribution functions (CDF) of the probability of the top event in fault tree 50. Red solid curve shows the result for the AFTC method. Solid black curve shows the CFTA method using the linear pool of expert opinion for missing values at the basic events. Dashed curves show result for each expert (TMc = Target Malaria consortium, Ind = Independent) under the CFTA strategy using a linear pool for missing values.

In this analysis we found that the effect of the different missing data strategies for the CFTA analysis methods led to errors in the probability calculations. In large fault trees, a high proportion of missing events within an individual expert's response is possible. If these events are assumed to have probability 1, then high order terms may not be negligible, hence truncating them can lead to probability estimates greater than 1. This effect is identical to the error that occurs if the rare event approximation $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ is applied to events that are not in fact rare. We do not therefore report these results here.

5.4.3 Sensitivity analysis

FT50 is not a large tree and it has a number of small cut sets including one with just a single event (FT50010). Based on the cuts we would predict the following events (in descending order) to have the greatest influence on the overall probability of FT50: FT50010, FT500000, FT500001, FT500010, FT500011, because these events are in cut sets with cardinality of one and two, followed by FT5011010, FT5011012, FT50110130, FT501101310, FT501101311, FT501110, FT501111, FT5011120, FT50111210 and FT50111211, because these events are in cut sets with cardinality three.

The sensitivity analysis for FT50 identifies FT50010, FT5011012, FT5011010, FT501110 and FT50303 as the top five events influencing the overall probability of the tree (Figure 5.28). Only four of these events are in the low cardinality cut sets identified above, although some of the others are identified by the sensitivity analysis as within the top 15 events. Given the small size of the tree we might expect to see more fidelity between our importance measure and the events in the smallest cut sets. Importantly, however, many of the low cardinality cut set events identified here are also events that exhibit significant inter-expert uncertainty. The effect of the high level of variation between experts appears to mask the importance that their location in the tree would confer.

5.5 FT51: Spread of the construct in *An. coluzzii* or *An. arabiensis*

5.5.1 Fault tree structure

FT51 was originally part of a fault tree 3 under a vertical gene transfer arm of the tree. It was moved and merged with FT5 late in the analysis because the scope of fault tree 5 was changed from spread of the construct in *An. gambiae* to spread of the construct in the *An. gambiae* complex, to more accurately reflect the target species for the Target Malaria consortium's proposal. Fault tree 5 was subsequently separated to 50 and 51 to clarify that only two other species from the complex are addressed in this analysis – *An. coluzzii* and *An. arabiensis*. It should be noted, however, that during the elicitations, one expert recommended that the consortium consider the risk of spread in the outdoor resting Goundry subgroup recently reported in Burkina Faso (Gneme et al., 2013).

The structure of FT51 is very similar to FT50 except that acquisition of the construct, and spread via *Wolbachia*, entails additional hybridisation steps in the causal chain (Figures 5.29 and 5.33), and spread via non-*Wolbachia* related mechanisms requires that hybrids are not sterile (see for example Figure 5.30). These additional steps occur at multiple points in the tree.

It is important to note that the tree structure shows many "OR" gates with event inputs from *An. coluzzii* and *An. arabiensis*. Typically these gates are also coloured yellow as many expert chose to answer at the gate rather than distinguish the probabilities of events for these two species.

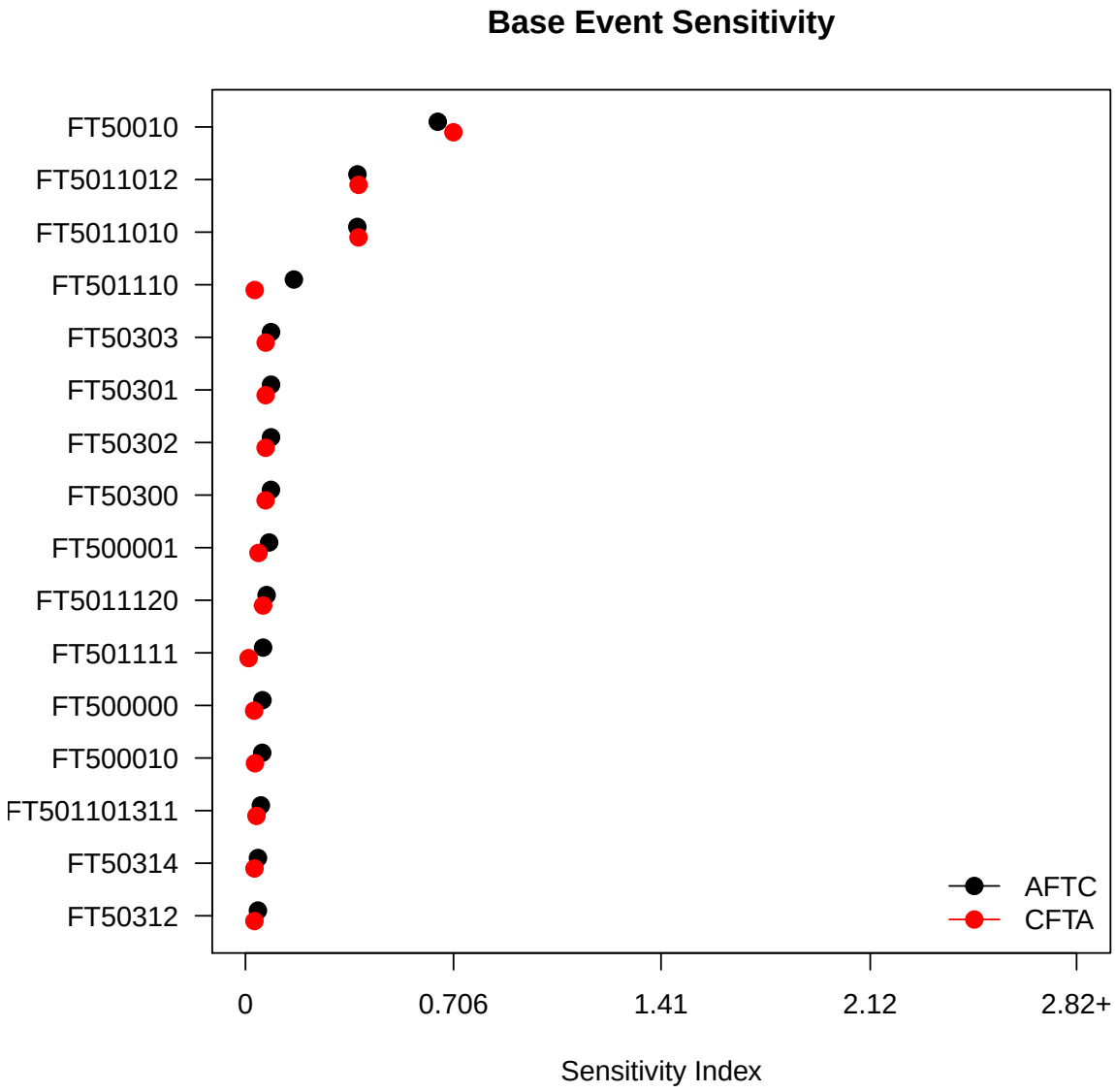


Figure 5.28: Result of the sensitivity analysis for FT50 for the AFTC and CFTA with linear pool analysis methods. Solid circles show the median of the importance measure (Equation 4.6) for the top 15 most important events

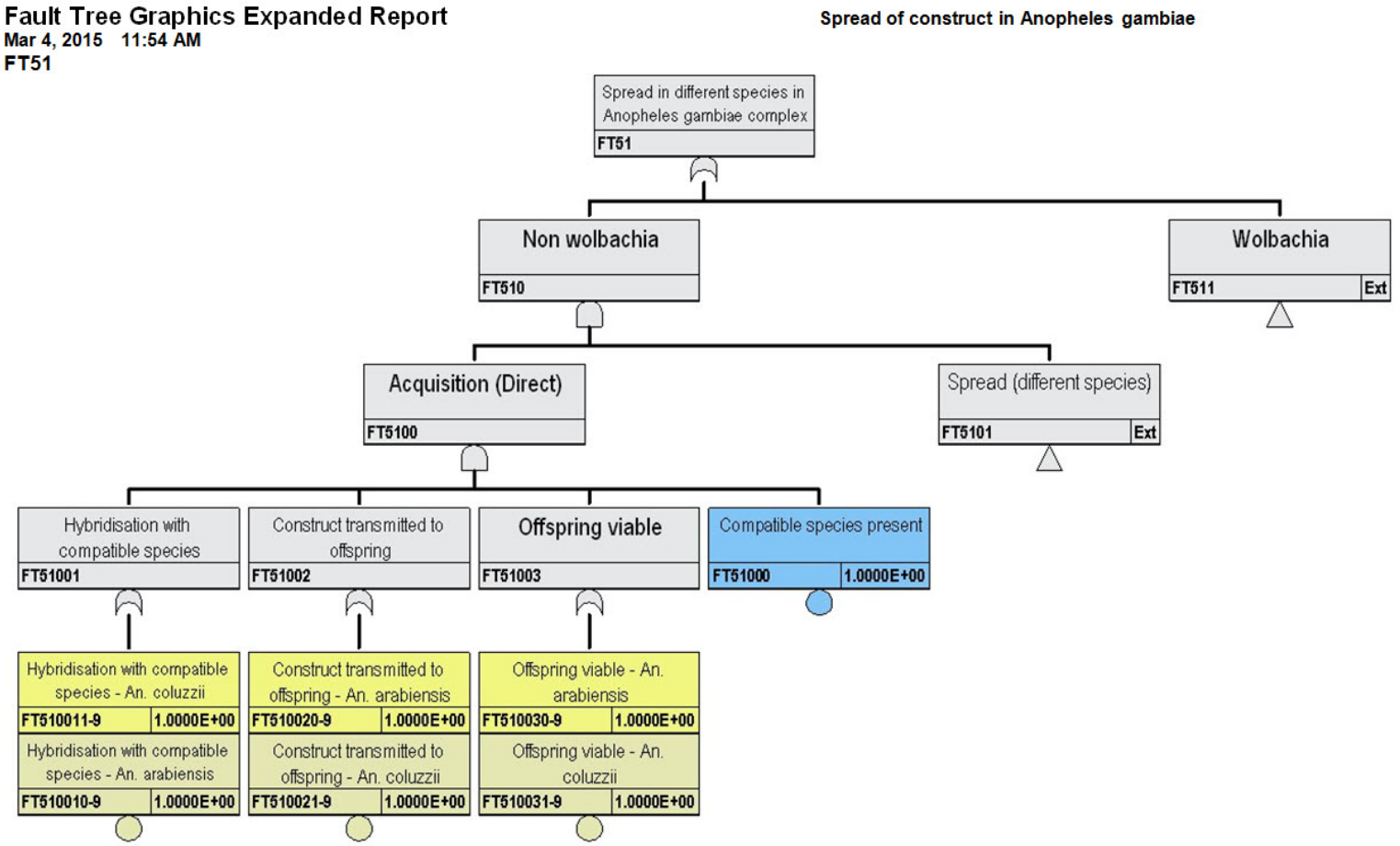


Figure 5.29: Page 1 of fault tree 51

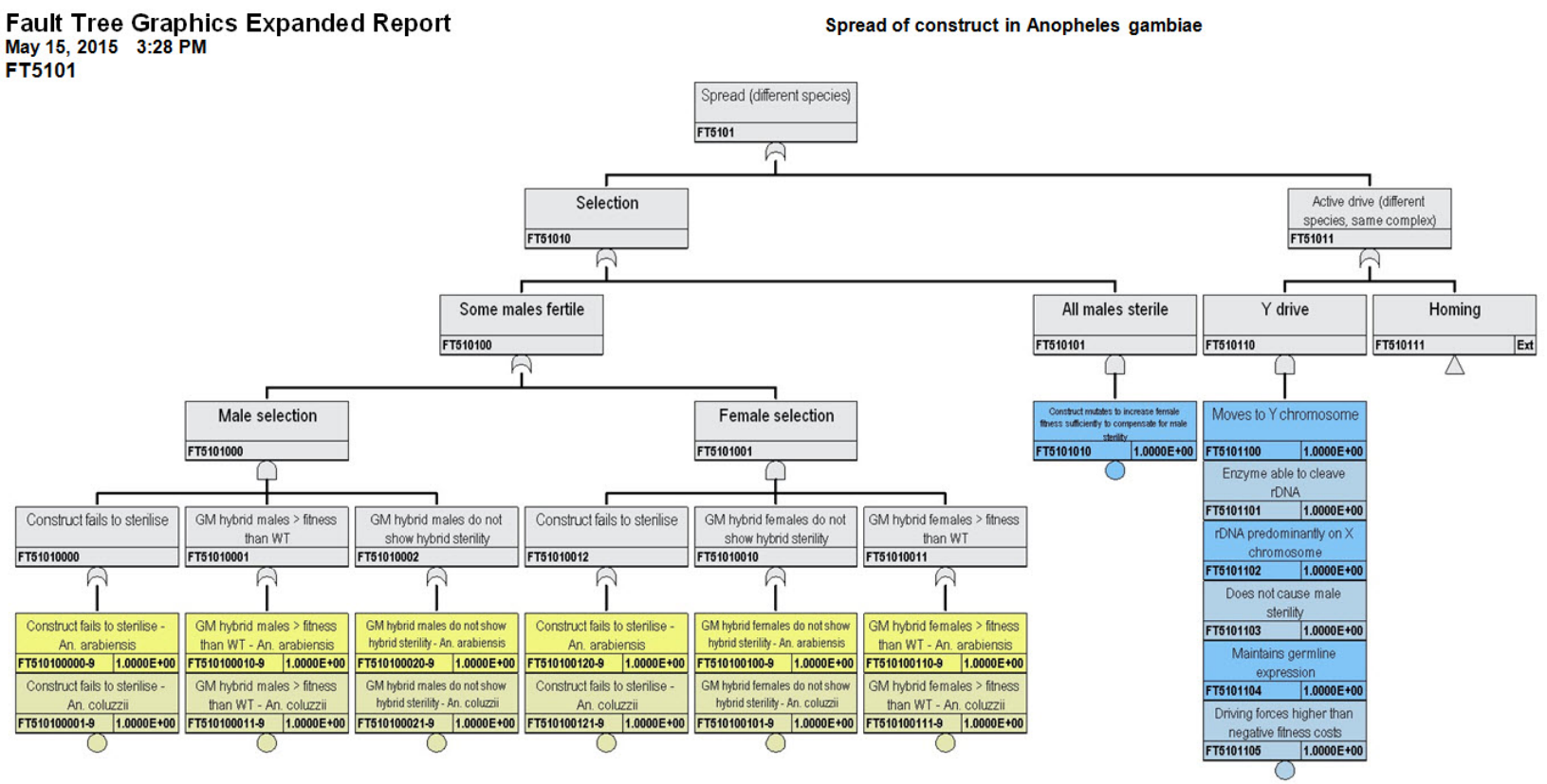


Figure 5.30: Page 2 of fault tree 51

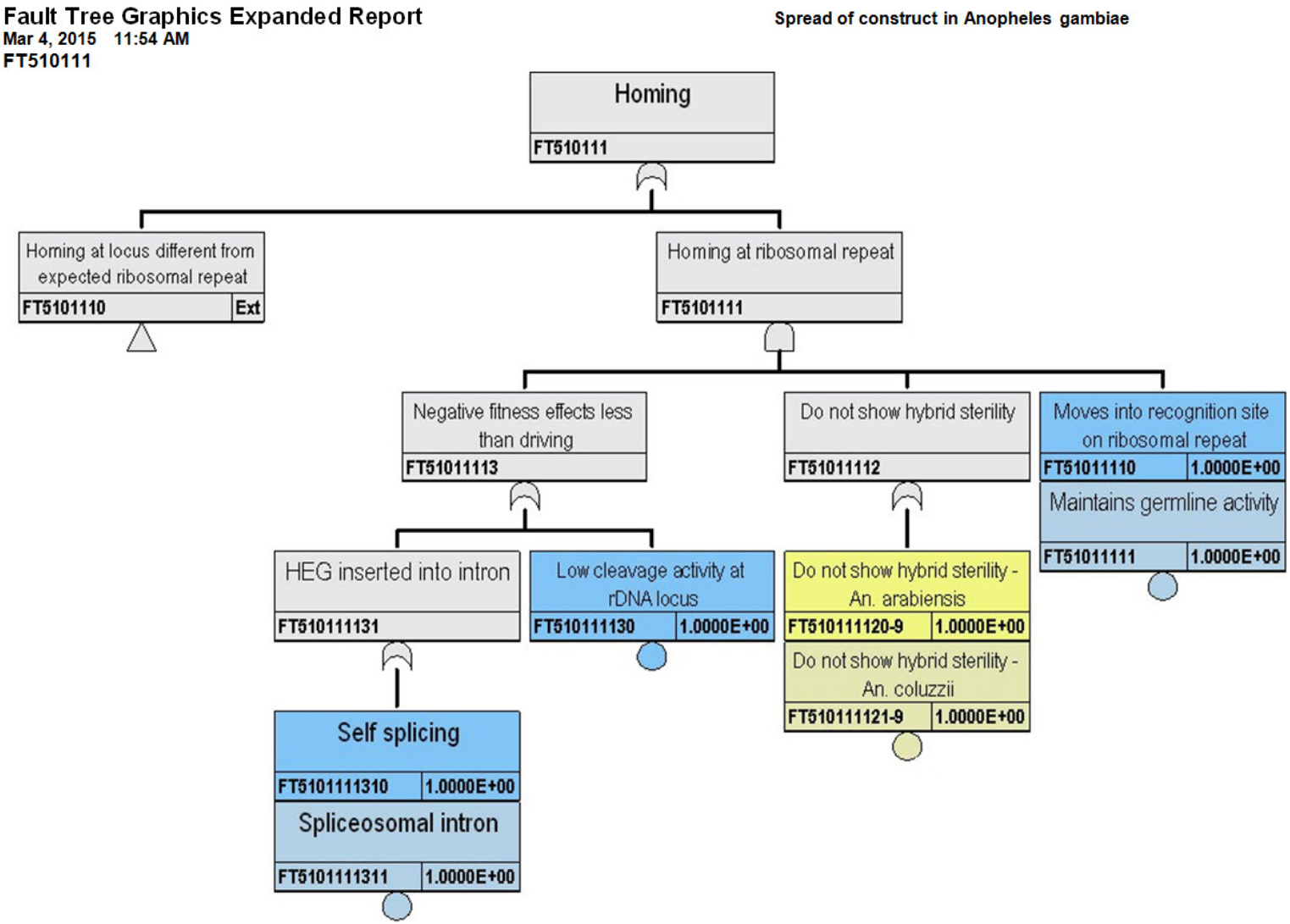


Figure 5.31: Page 3 of fault tree 51

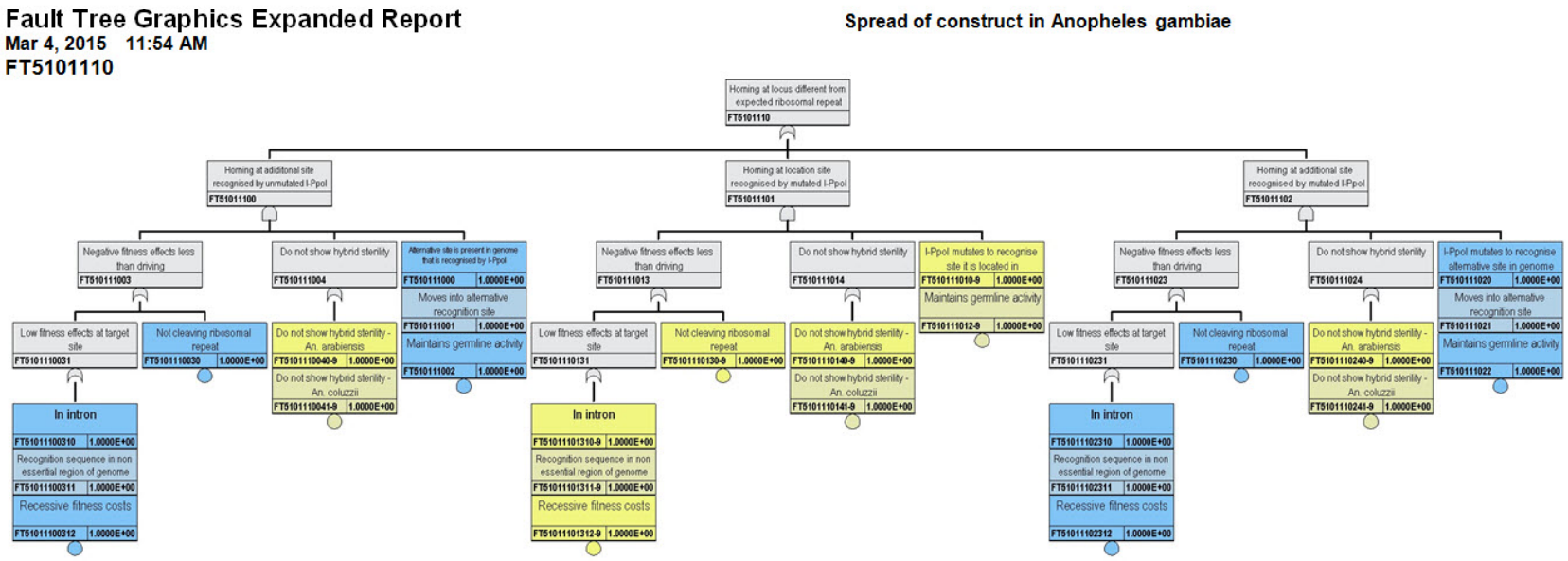


Figure 5.32: Page 4 of fault tree 51

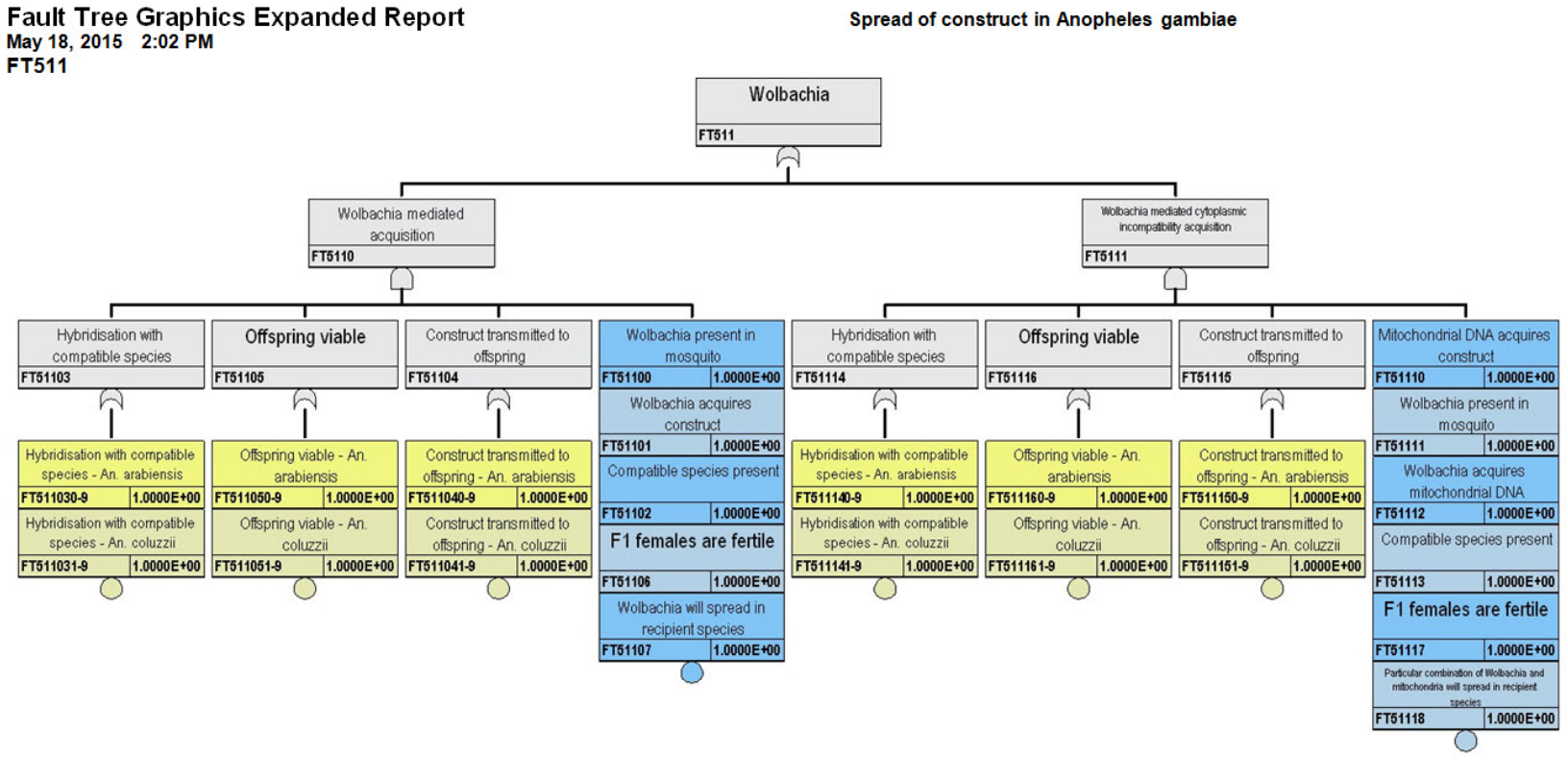


Figure 5.33: Page 5 of fault tree 51

5.5.2 Risk calculations

The results of the quantitative analysis for fault tree 51, using the AFTC and CFTA methods, are summarised in Figure 5.34. Table 5.5 shows the 50th, 90th and 99th percentiles of these results for the best available computation method – in this case accounting for squared terms of repeat events with subtrees and 10th order term truncation.

Method B10 & C Squared terms of repeats handled + truncation of >10 terms + sub-trees			
	50%	**90%**	**99%**
AFTC	1.1×10^{-6}	0.0024	0.095
CFTA_LP	1.5×10^{-5}	0.01	0.21

Table 5.5: Median and upper tail percentiles for the probability of the top event in fault tree 51, for two analysis methods under the best available computation strategy.

The probability of the top event for FT51 under the AFTC analysis method is lower than the equivalent probability for FT50. This is to be expected and reflects the additional hybridisation and hybrid fertility steps in the causal chain. The results for the CFTA analysis method are also lower than the equivalent probability for FT50, but these results are potentially unreliable because of the high degree of repeat events within FT51, and large proportions of missing data under the CFTA method.

The pattern of responses between the Target Malaria consortium experts and the independent experts is similar to previous cases. There is no obvious sign of motivational bias, but there is a potential “outlier” among the group of responses.

5.5.3 Sensitivity analysis

The results of the sensitivity analysis for FT51 are shown in Figure 5.35. FT51 is the second largest tree in the analysis with 400 minimum cut sets. The cardinality of the smallest cut sets is five, and there are eight of these sets. These sets contain seven unique events from the non-*Wolbachia* acquisition side of the tree (FT51000, FT510010, FT510011, FT510020, FT510021, FT510030, FT510031) and one from the spread side of the tree (FT5101010). All of these events feature in the top 15 most influential events. So here it appears as if the structure of the fault tree is again having a strong influence on the overall importance of the basic events in the tree. The remaining events in the top 15 are all represented in the collection of the second smallest cut sets (with cardinality seven).

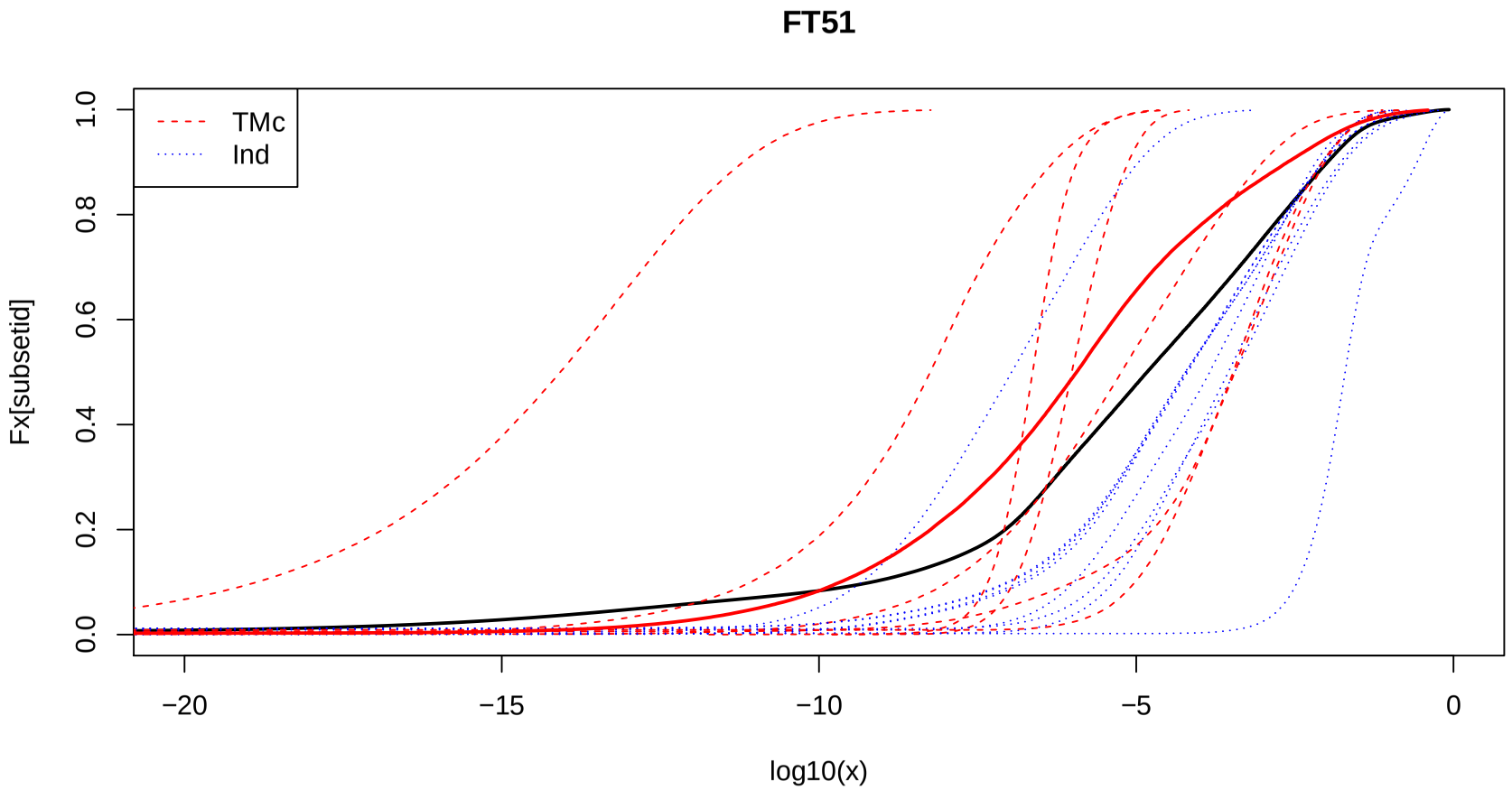


Figure 5.34: Cumulative distribution functions (CDF) of the probability of the top event in fault tree 51. Red solid curve shows the result for the AFTC method. Solid black curve shows the CFTA method using the linear pool of expert opinion for missing values at the basic events. Dashed curves show result for each expert (TMc = Target Malaria consortium, Ind = Independent) under the CFTA strategy using a linear pool for missing values.

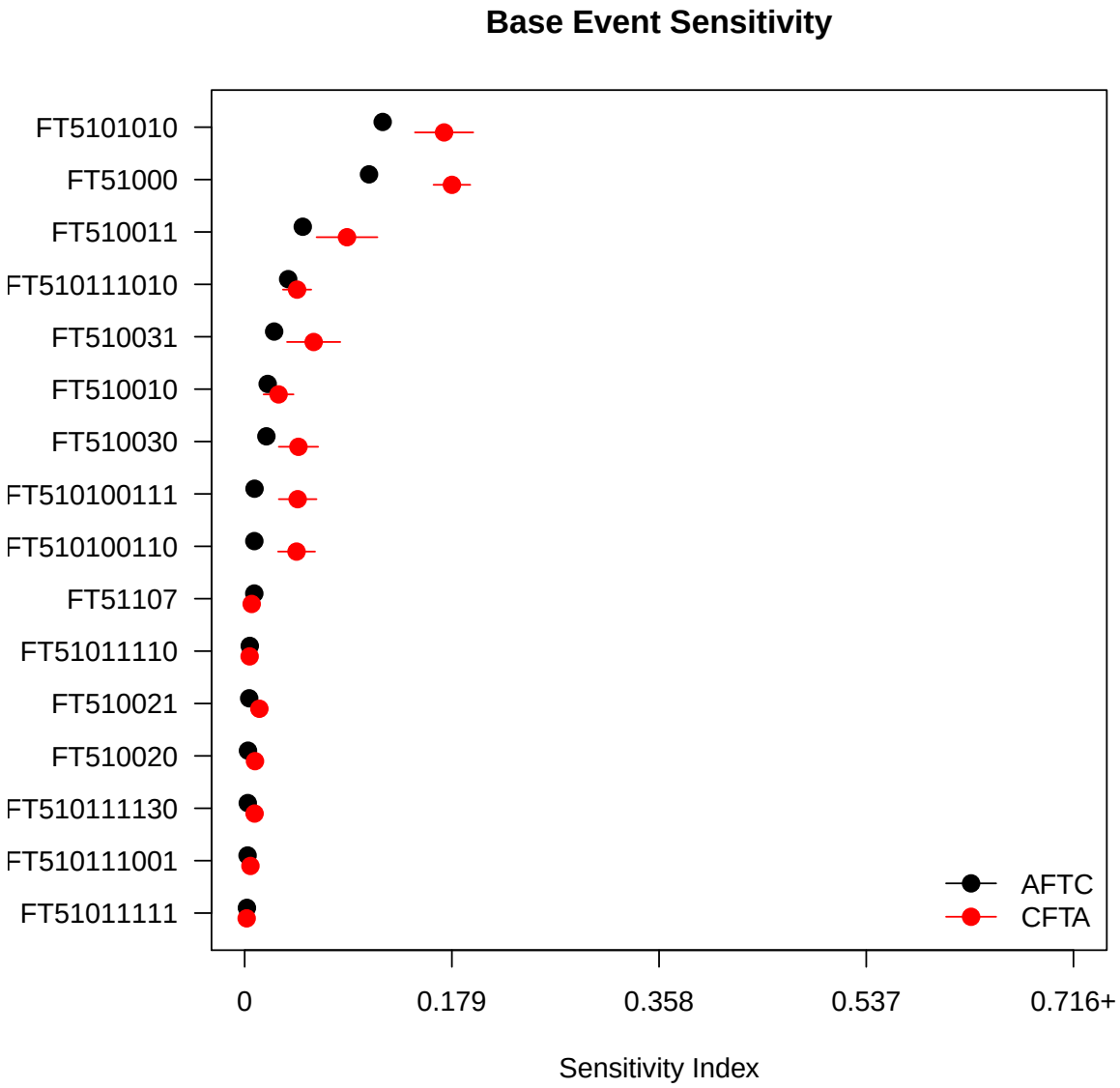


Figure 5.35: Result of the sensitivity analysis for FT51 for the AFTC and CFTA with linear pool analysis methods. Solid circles show the median of the importance measure (Equation 4.6) for the top 15 most important events

KEY POINTS: INFERENCE WITH NULL EVENTS

- The risk assessment results reported in the preceding sections rely entirely on expert opinion, and our confidence in these estimates would be improved if they were supported, at least in part, by results from carefully controlled experiments
- A challenge in this context is that many of the events being considered here are so rare that we know *a priori* that we will probably observe no outcomes in these experiments.
- There are a large number of methods in the literature to estimate the probability p of rare events given n independent experiments that observe no outcomes ($x = 0$).
- By focusing on methods that consider the special case of $x = 0$, p close to zero and large n, and by comparing this case with other solutions, we recommend that if a cautious approach is required the Exact (Clopper and Pearson, 1934) or Wilson's method (Wilson, 1927) are used to estimate the upper 95% confidence interval for p.
- If we consider the example of basic event FT500000 (the probability p that the construct fails to sterilize all males), treat the experiments conducted to date by the Target Malaria consortium as independent with n number of crosses between transgenic males and wild type females of the order 500, and $x = 0$ number of fertile events, then the exact method gives the upper value of the 95% confidence interval for the true value of p to be 8.7$\times 10^{-3}$, whereas Wilson's method gives 9.9$\times 10^{-3}$.

6 NULL EVENT INFERENCE

The risk assessment results reported in the preceding sections rely entirely on expert opinion. The confidence in these estimates would be improved if they were supported, at least in part, by results from carefully controlled experiments. A challenge in this context, however, is that many of the events being considered here are so rare that we know *a priori* that we will probably observe no outcomes in these experiments.

Consider for example, FT500000 – the probability that the construct fails to sterilize all males (Figure 5.24). To date the Target Malaria consortium have detected no fertile eggs ($x = 0$) among more than 50,000 eggs laid be females mated to I-PpoI male mosquitoes (*pers. comm.* Mark Benedict). What does this tell us about the probability of FT500000?

Observing no failures (no fertile eggs) after n trials (independent matings between wild type females and transgenic males) does provide some information about p (FT500000). For example consider if 10,000 trials were completed and no failures were observed, we can be confident that p is small and the likelihood of p being for example 0.5 is practically impossible. A standard way to approach this idea statistically is to make inferences on the upper confidence interval of p based on n.

There are a large number of methods in the literature to estimate confidence intervals of the binomial probability p (see for example Wilson, 1927; Clopper and Pearson, 1934; Berry and Armitage, 1995; Jovanovic and Levy, 1997; Newcombe, 1998; Agresti and Coull, 1998) and a number of studies comparing methods to estimate binomial confidence intervals have been published. Table 6.1 gives a summary of the findings of a selection of papers in the literature.

The results of these reviews at first appear to be contradictory. The results are more consistent, however once you consider that some of the papers are not specifically concerned with the case of $x = 0$ but rather generally estimating confidence intervals for p in a binomial experiment. Furthermore, how well a method performs depends on the magnitude of n and the true p. We are interested in a rather special case of $x = 0$, very small p, and large n.

The difference between these methods is best visualised by plotting the results of each method in terms of what the estimated upper confidence interval of p would be for a given n (Figure 6.1). Also we can plot, for a specified estimated upper confidence interval of p what n would be required (Figure 6.2).

Focusing on the review papers that consider the special case of $x = 0$, p close to zero and large n, as well as the comparisons in Figures 6.1 and 6.2, we recommend that if a cautious approach is required the Exact (Clopper and Pearson, 1934) or Wilson's method (Wilson, 1927) are used. Otherwise cbox (Balch, 2012) or Laplace seem to provide a middle ground and reduces the sample size required.

Returning to the example of FT500000, expert opinion exhibited significant disagreement on this event, with the expected value of p ranging from 0.59 to 1.0×10^{-7}. For the purposes of illustration, if we assume that the number of eggs laid by wild type females mated with transgenic males is similar to FT1-5a – i.e. the number of eggs laid by wild type females mated with wild type males with a mean value of 93 – and that all mating experiments conducted by the consortium to date have been independent trials, then the number of trials conducted to date would be in the order of $n = 500$. The exact method gives the upper value of the 95% confidence interval for the true value of p to be 8.7×10^{-3}, whereas Wilson's method gives 9.9×10^{-3}.

Reference	Conditions	Recommendation
Robinson et al. (2008) and references therein		Use Jeffrey Interval rather than exact
Agresti and Coull (1998) Ross (2003)		Exact not as good as approximate methods
Looney (2013)	$x = 0$ (i.e. small p)	Clopper-Pearson (Exact)
Winkler et al. (2002)	Compares rule of 3 versus Bayesian ($x = 0$)	
He and Wu (2009)		Wald's method
Newcombe (1998)	Various p and n	Suggests multiple methods, finds asymptotic pretty bad, Exact, mid-P, and likelihood based reasonable. In terms of minimum coverage Exact performs the best.
Tobi et al. (2005)	$p \leq 0.01$	$p \leq 0.01$ use the Exact method or the Continuity-Corrected Wilson method
Dunnigan (2008)		States that Wald's method is most commonly used but is flawed. So some statisticians resort to older Clopper-Pearson (Exact), however if a conservative approach is needed they recommend Wilson

Table 6.1: Review of recommendations from papers comparing methods for null event inference

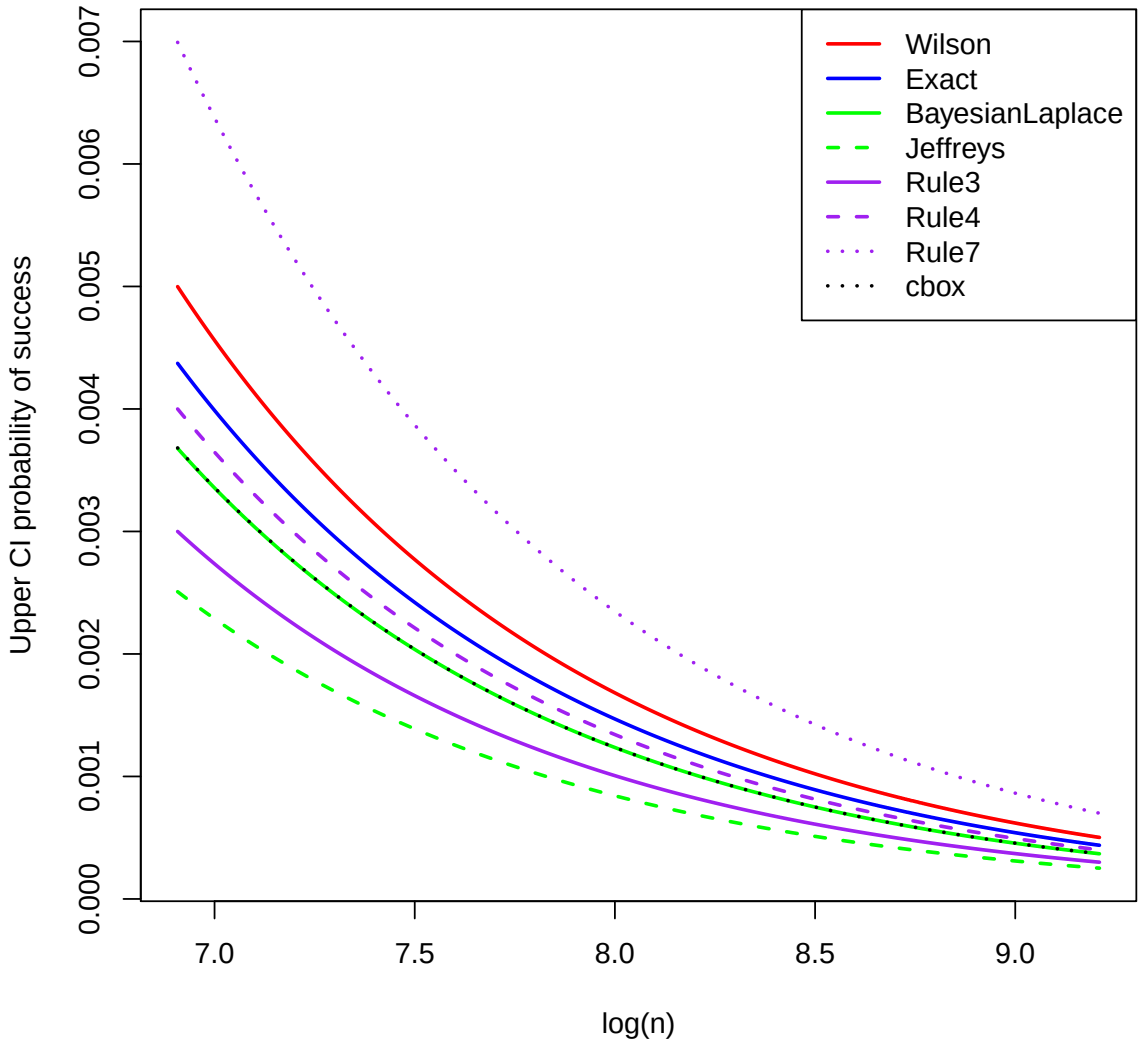


Figure 6.1: Comparison of estimates for upper confidence interval from various methods.

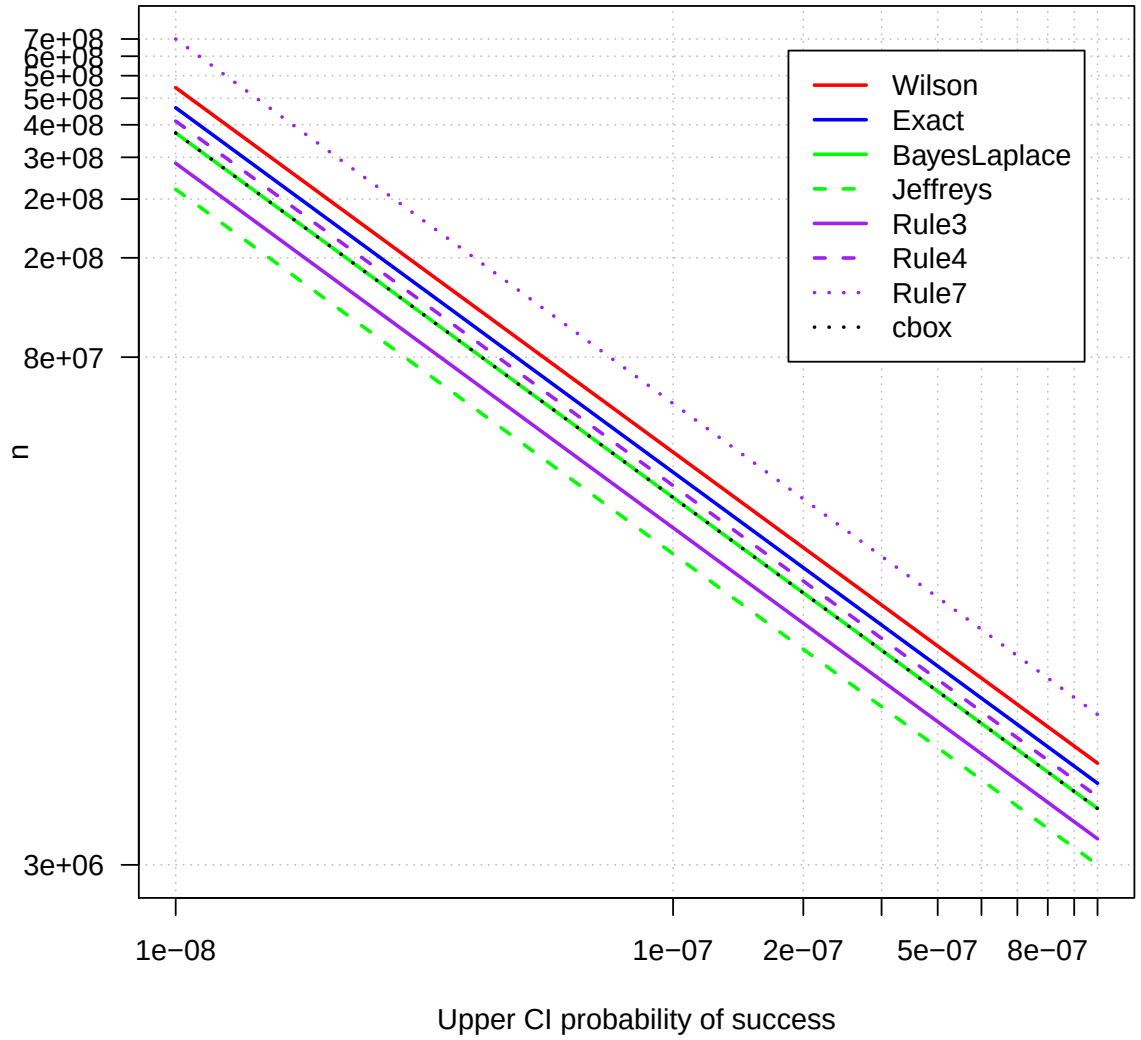


Figure 6.2: Log-log plot of n required to obtain a specific upper confidence interval for each of the selected methods.

KEY POINTS: DISCUSSION

- A scientific risk assessment should be transparent, repeatable and wherever possible make predictions that are measurable and falsifiable. This assessment maintains transparency, and coherently propagates uncertainty through its calculations, by quantifying risks, adhering to the laws of probability theory and avoiding qualitative risk assertions.
- Quantitative risk estimates in our opinion are essential to meet science quality criteria, and we achieved this by using direct elicitation to fit subjective probability density functions to expert beliefs about the probability of the events in fault trees, and for a set of well established vectorial capacity parameters.
- In quantifying the possible risks associated with the accidental release of I-PpoI mosquitoes we encountered a variety of technical and computational challenges, some of which are common to all risk assessments, and some that were particular to this analysis.
- The assessment of risk of an increase in vectorial capacity was based on a direct elicitation of parameters that are widely accepted to be the most relevant in this context. If we make the assumption of strong positive dependence between strains, then the mean intrinsic transmission risk index for G3 and I-PpoI mosquitoes is -0.41 and -0.23 respectively.
- The results of the analysis indicate that the median value of the risk of G3 strain mosquitoes transmitting a novel blood-based pathogen in a year following a complete escape of 10,000 mosquitoes is 5.2×10^{-7}. Comments provided by the experts during the elicitation suggest that a linear pool estimate of the risk of the I-PpoI mosquitoes vectoring a novel pathogen would be the same as, or lower than, these values.
- The analysis indicates that the median risk of the HEG spreading in non-target eukaryotes or non-eukaryotes is 1.2×10^{-10} and 6.7×10^{-7} respectively. We infer that the probability of incidental impacts on populations of non-target species, over the course of a year, will be no higher than these values, and would be lower if: (i) the probability of any of the events in the causal chain between spread of the construct and any specific type of detrimental impact was less than one; and/or (ii) only a sub-set of the spread pathways quantified here could lead to a specific impact.
- The median value of the risk of the construct spreading in local populations of the related mosquito species *An. coluzzii* or *An. arabiensis* in a year, following the complete loss of 10,000 I-PpoI modified mosquitoes, was estimated to be 1.1×10^{-6}. The median value for the *An. coluzzii* or *An. arabiensis* risk, however, is sensitive to the analysis method used in the fault tree, and this rises to 1.5×10^{-5} under an alternative strategy because the beliefs of one expert have a stronger influence on the risk estimate under this alternative strategy.
- The median value of the risk of the construct spreading in local populations of *An. gambiae* in a year following the complete release of 10,000 I-PpoI mosquitoes was estimated to be 0.0014.

7 DISCUSSION

7.1 Rationale for the risk assessment method

A scientific risk assessment should be transparent, repeatable and wherever possible make predictions that are measurable and falsifiable. The scientific quality of a risk assessment should be determined by the extent to which it meets these criteria. Risk assessments should also be faithful to any assumptions about the source and types of uncertainty they address, they should carry these uncertainties through the analysis, and they should represent and communicate them clearly and reliably. These are the hallmarks of an "honest" risk assessment (Burgman, 2005).

This risk assessment was designed to be scientific, honest and conservative. The assessment quantifies the risk of a set of endpoints that were carefully chosen to minimise the complexity of the assessment without compromising its relevance to decision makers and stakeholders. The assessment maintains transparency, and coherently propagates uncertainty through its calculations, by quantifying risks, adhering to the laws of probability theory and avoiding qualitative risk assertions.

All risk assessments, qualitative and quantitative, are based on a conceptual model of how things go wrong. For four of the five endpoints addressed in this analysis we used fault trees to make the conceptual model explicit, transparent and amenable to quantification. We were also able to accommodate uncertainty in this conceptual model when individual experts expressed a different conception. On the whole, however, different experts tended to modify sub-sections of the initial fault trees rather than suggest an entirely different risk model. For the first endpoint – change in vectorial capacity – we used a well accepted model that distils the complexity of malaria transmission down to a few key parameters.

Risk assessments for biotechnology products are typically qualitative. Quantitative risk estimates are harder to obtain, but in our opinion they are essential to meet the science quality criteria described above. We achieved this by using direct elicitation to fit subjective probability density functions to expert beliefs about the probability of the events in the fault trees, and the vectorial capacity parameters. We consulted experts that are independent from, and part of, the Target Malaria consortium. The use of independent experts is important to guard against the potential for motivational bias. We found no evidence of this in our analysis.

The construction of a fault tree is a heuristic exercise that helps identify ways things can go wrong – i.e. identify hazards. Fault tree analysis, however, is not designed to identify all potential hazards – it is a deductive "top-down" approach that focuses on causal pathways leading to a pre-defined event. We have addressed the possibility of additional hazards by complementing the deductive fault tree approach with an inductive, "bottom-up" analysis.

By virtue of its whole-of-system approach to hazard analysis, HHM analysis is more likely to identify, or at least suggest, unexpected interactions that may lead to additional hazards. Taken together, HHM and FTA enable the analyst to postulate certain hazards and then investigate in more detail how they might occur. There is no guarantee, however, that these processes together will identify all hazards. There are no such guarantees in any form of hazard analysis or risk assessment (hence the need to continually compare the predictions of a risk assessment with reality). The logical and systematic structure of HHM and fault tree analysis, however, helps minimise the probability of missing important causal pathways and we believe that it performs much better in this regard than unstructured brainstorming techniques.

7.2 Quantitative risk estimates

In quantifying the possible risks associated with the accidental release of I-PpoI mosquitoes we encountered a variety of technical and computational challenges, some of which are common to all risk assessments, and some that were particular to this analysis. The effect of dependence on risk calculations is a common challenge. It influences the results of all arithmetic operations involving random variables, and hence complicates our fault tree calculations and the assessment of the relative malaria transmission risk.

The dependence induced by repeat events and mutually exclusive events in a fault tree analysis is usually handled via the method of minimum cut sets. This method, however, assumes all information occurs at the basic events but in our analysis some experts elected to provide information at the gates – a challenge particular to this analysis that in effect results in a slightly different fault tree structure.

To include information elicited at a gate, we developed our own analysis functions in the R computing language that symbolically implements the probability laws for union and intersection in a step-wise, gate-to-gate, fashion through the fault tree. Doing this symbolically enables us to retain information about the identity of the gate and associated information such as an experts probability density function. Computationally, however, this is not an efficient approach and for the large fault trees in this analysis, we were forced to adopt a mixture of computational strategies – identifying independent sub-trees and truncating terms in the expanded fault tree equation with more than 10 or 12 elements.

Our computational strategies were successful for the AFTC analysis method, but they failed during the CFTA strategy when we implemented a prior that assumed the event would occur in the face of missing data at the gate. In this case truncating terms led to errors that are equivalent to the errors that occur when the rare event approximation is used to calculate the probability of the sum of two random variables that are not in fact rare. For these reasons our preferred approach is the AFTC analysis method.

The CFTA method is an attractive alternative when experts provide subjective probability judgements for all the basic events in a fault tree. In complex biological systems, however, this is a daunting task for any single expert, and in light of the multi-disciplinary nature of the events within the tree, is arguably an undesirable goal. Unfortunately with “missing data” in the fault tree none of the work arounds needed to implement the CFTA are very satisfactory. We discovered, however, that (in this analysis at least) the difference in the probability of the top event between the CFTA with a linear pool for missing data, and the AFTC method, was generally less than an order of magnitude, except for FT51 which displays a large difference in the 90th and 99th percentiles because the beliefs of a single expert have a weaker influence.

On these grounds we recommend that the results of the fault tree analysis are presented for the AFTC method, for the median and upper tail percentiles of the distribution function of the probability of the top event, but that the discrepancy for FT51 is noted. Table 7.1 summarizes the results of the fault tree analysis in this manner. This table, together with the results of the relative transmission risk analysis - which suggests the difference in the mean intrinsic transmission risk between G3 and wild type is -0.41, and between I-PpoI and wild type is -0.23 – completes the quantitative risk assessment component of this project.

FT2 Vector novel blood-based pathogen			
	50%	**90%**	**99%**
AFTC estimate	5.2×10^{-7}	10^{-4}	0.0046
Standard error	7.2×10^{-9}	2.0×10^{-6}	1.6×10^{-4}
FT3 Construct spread in non-target Eukaryotes			
	50%	**90%**	**99%**
AFTC estimate	1.2×10^{-10}	3.1×10^{-6}	0.02
Standard error	7.2×10^{-9}	2.0×10^{-6}	1.6×10^{-4}
FT4 Construct spread in non-Eukaryotes			
	50%	**90%**	**99%**
AFTC estimate	6.7×10^{-7}	7.8×10^{-4}	0.057
Standard error	1.9×10^{-8}	1.6×10^{-5}	1.6×10^{-3}
FT50 Construct spread in *An. gambiae*			
	50%	**90%**	**99%**
AFTC estimate	0.0014	0.25	0.88
Standard error	3.2×10^{-5}	6.6×10^{-3}	1.3×10^{-3}
FT51 Spread of construct in *An. Coluzzii* or *An. arabiensis*			
	50%	**90%**	**99%**
AFTC estimate	1.1×10^{-6}	0.0024	0.095
Standard error	3.0×10^{-7}	2.8×10^{-4}	4.2×10^{-3}
CFTA_LP	1.5×10^{-5}	0.01	0.21
Standard error	1.4×10^{-5}	3.3×10^{-4}	2.5×10^{-3}

Table 7.1: Risk and standard error estimates for five assessment endpoints following the complete loss of 10,000 G3 strain mosquitoes (FT2) and 10,000 mosquitoes genetically modified with the I-PpoI construct (FT3, FT4, FT50 and FT51) from African insectaries in Mali, Burkina Faso and Kenya, over the period of a year. Risk calculations were performed via fault tree analysis using a range of computational strategies designed to account for dependency (repeated events) within the trees

.

7.3 Key events

One of the human-health risks addressed by this analysis is the potential for additional malaria cases following a complete escape of all insectary mosquitoes, and the extent to which this risk is enhanced or diminished with G3 strain mosquitoes or I-PpoI modified mosquitoes when compared to wild type. The issue of dependence between the parameters that capture the main dynamics of malaria transmission proved to be important in this context. If we impose positive dependence within these parameters between strains then we are 95% sure that the malaria transmission index of I-PpoI mosquitoes is lower than WT. We believe on biological grounds that positive dependence is a more realistic situation but we did not test this assumption with the experts we consulted. We also did not explore the potential for positive dependency between parameters but within strains.

The results in Table 7.1 suggest that the highest risk posed by a complete escape of 10,000 I-PpoI modified mosquitoes is the spread of the construct in *An. gambiae* either through *Wolbachia*-mediated mechanisms, or through vertical gene transfer. The next highest risk is the spread of the construct through the closely related species *An. coluzzii* and *An. arabiensis* via essentially identical mechanisms. As noted in Section 2.3, these end points are not harmful per se, but were included in the risk analysis because they are unexpected for this particular construct.The sensitivity analysis conducted as part of the fault tree analysis, together with a comparison of the minimum cut-set membership, suggests the following:

- ***Wolbachia* mediated spread mechanisms**. Our sensitivity analysis highlighted the importance of all of the events under FT5030 – that is *Wolbachia* mediated acquisition and spread. *Wolbachia* related events were also highlighted in FT51 for *An. arabiensis* and *An. coluzzii*, but not to the same extent. An additional analysis of the median probabilities at each basic event and gate, however, indicates that the vast majority of the risk in FT50 and FT51 occurs via the non-*Wolbachia* mediated side of the the fault trees. The median probabilities of the *Wolbachia* gates – FT503 and FT511 – are of the order of 10^{-7} and 10^{-8} (Figures 7.1 and 7.2).
- **Uncertainty around mutation, male sterility and fitness costs**. Comparison between our analysis and the size of the minimum cut sets highlighted significant uncertainty around events that would otherwise have an important influence on the probability of this endpoint. These events are FT50010 – the probability that the construct mutates and increases female fitness sufficiently to compensate for the sterile males fitness cost; FT500000 and FT500010 – the probability that the construct does not sterilize all males; and, FT500001 and FT500011 – the probability GM male or GM female fitness is higher than wild type. The apparent disagreement between experts over these issues warrants further attention.
- **Homing at unexpected locus**. The sensitivity analysis for FT51 also highlighted the importance of five basic events under FT5101110 - that is homing by the I-PpoI construct at a locus different from the expected ribosomal repeat. None of these events occur in minimum cut sets with relatively (for this tree) low cardinality, but two of these events have high uncertainty.

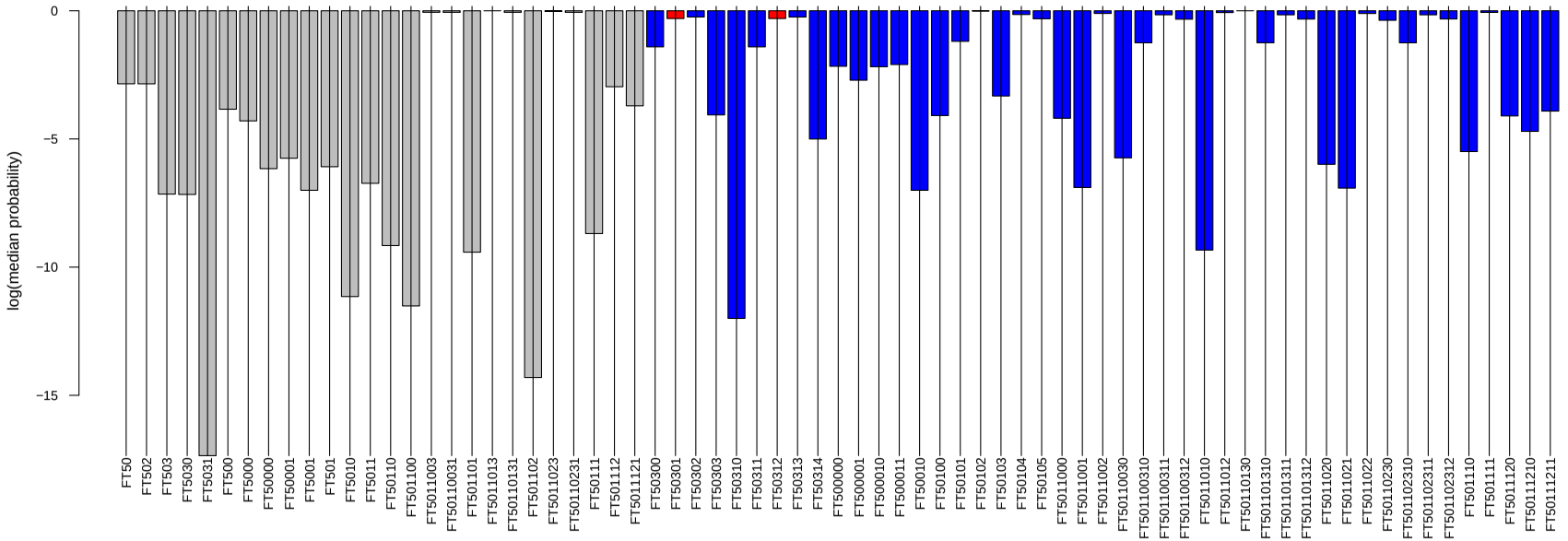


Figure 7.1: Bar chart showing the median probability for all basic events and gates within fault tree 50. Blue bars represent basic events, grey bars are gates. Red bars are basic events that were not addressed by any expert during the elicitation. These events are assigned a uniform distribution on the range [0,1].

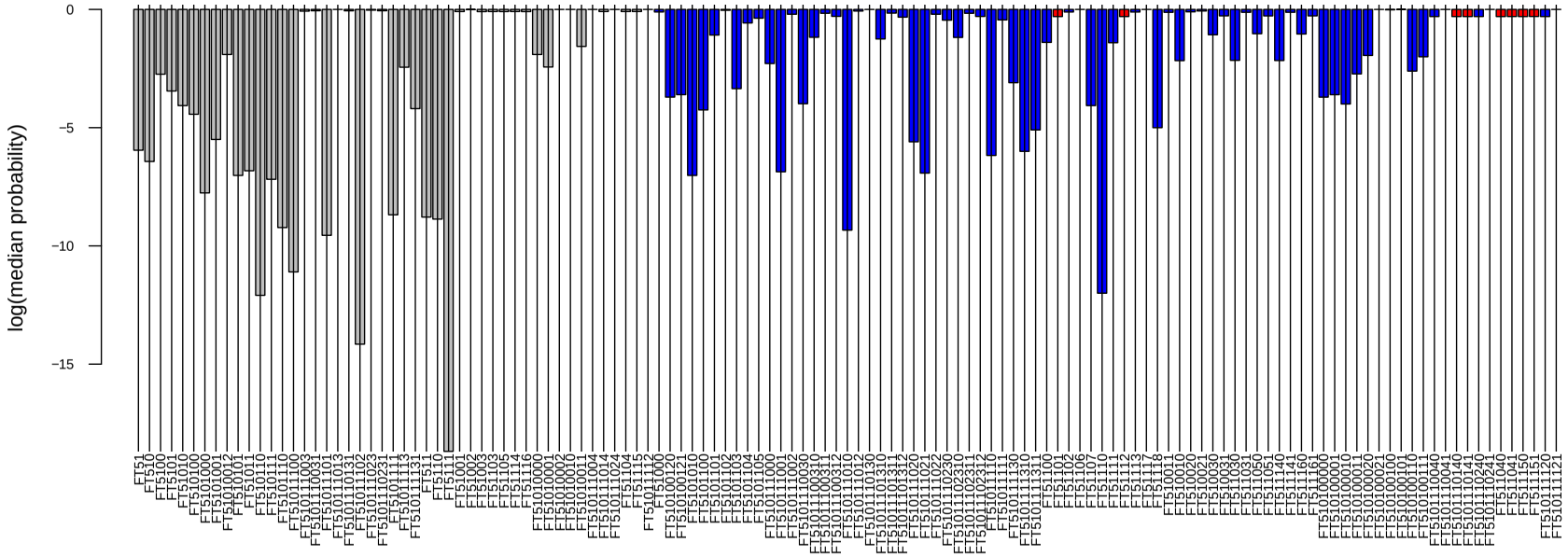


Figure 7.2: Bar chart showing the median probability for all basic events and gates within fault tree 51. Blue bars represent basic events, grey bars are gates. Red bars are basic events that were not addressed by any expert during the elicitation. These events are assigned a uniform distribution on the range [0,1].

The risk of the G3 strain mosquitoes vectoring a novel blood-borne pathogen and the risk of the construct spreading through populations of non-eukaryotes are broadly similar. The analysis also notes that most of the experts we consulted believed that the risk of I-PpoI modified mosquitoes vectoring a novel-blood based pathogen would be lower than the value reported here for G3 strain mosquitoes, on the grounds that the genetically modified mosquitoes are likely to have a higher mortality rate than G3 strain. The sensitivity analysis for fault tree 2 proved inconclusive whereas that for FT4 suggests that the following events may have a strong influence on its results:

- **Uncertainty about selection in prokaryotes without expression**. The two basic events under FT40101 (positive selection without expression) appear in the smallest minimum cut sets of FT4. They are also two (of only five) events in fault tree 4 around which experts significantly disagreed. It is therefore important to confirm that this disagreement is genuine and not caused by experts misinterpreting the question. This issue is important generally but particularly so in this case.

The spread of the construct to non-target eukaryotes via horizontal gene transfer is the lowest of the risk estimates completed here. The large number of steps in the causal chain of acquisition and spread, together with the very low probability associated with some of these steps, strongly mitigates against the probability of this endpoint. The extreme percentiles of this endpoint, however, have been significantly influenced by the beliefs of one independent expert, whose views are strongly at odds with the other experts we interviewed under FT3. Again we believe that going forward it would be useful to further examine the reasons behind this disagreement and if possible provide experts an opportunity to discuss their respective beliefs with each other in light of these results.

8 REFERENCES

Agresti, A. and Coull, B. A. (1998). Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126.

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., and Petit, E.and Gabaldón, T. (2014). High variability of mitochondrial gene order among fungi. *Genome Biology and Evolution*, 6(2):451–465.

Alonso, P. L. and Tanner, M. (2013). Public health challenges and prospects for malaria control and elimination. *Nature Medicine*, 19:150–155.

Alphey, L. (2014). Genetic control of mosquitoes. *Annual Review of Entomology*, 59(1):205–224. PMID: 24160434.

American Committee of Medical Entomology and the American Society of Tropical Medicine and Hygiene (2003). Arthropod containment guidelines. A project of the American Committee of Medical Entomology and American Society of Tropical Medicine and Hygiene. *Vector Borne and Zoonotic Diseases*, 3:61–98.

Baeshan, R., Ekechukwu, N. E., Toure, M., Paton, D., Coulibaly, M., Traoré, S., and Tripet, F. (2014). Differential effects of inbreeding and selection on male reproductive phenotype associated with the colonization and laboratory maintenance of *Anopheles gambiae*. *Malaria Journal*, 13:19. DOI: 10.1186/1475–2875–13–19.

Bailey, N. (1982). *The Biomathematics of Malaria*. Charles Griffin and Company, High Wycombe, UK.

Balch, M. S. (2012). Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning*, 53(7):1003–1019.

Baldini, F., Segata, N., Pompon, J., Marcenac, P., Robert Shaw, W., Dabiré, R. K., Diabaté, A., Levashina, E. A., and Catteruccia, F. (2014). Evidence of natural Wolbachia infections in field populations of *Anopheles gambiae*. *Nature Com*, 5:doi:10.1038/ncomms4985.

Barnthouse, L. W., Suter, G. W., Bartell, S. M., Beauchamp, J. J., Gardner, R. H., Linder, E., O'Neill, R. V., and Rosen, A. E. (1986). User's manual for ecological risk assessment. Technical report, Environmental Sciences Division Publication No. 2679, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, 207 pp.

Berry, G. and Armitage, P. (1995). Mid-p confidence intervals: a brief review. *The Statistician*, pages 417–423.

Birnbaum, Z. (1969). On the importance of different components in a multicomponent system. In Krishnaiah, P., editor, *Multivariate Analysis II*, pages 591–592. Academic Press.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley and Sons, New York, USA.

Burgman, M. (2005). *Risks and Decisions for Conservation and Environmental Management*. Cambridge University Press, Cambridge, England.

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London, Series B*, 270:921–928.

Burt, A. and Koufopanou, V. (2004). Homing endonuclease genes: the risk and fall and rise again of a selfish element. *Current Opinion in Genetics and Development*, 14:609–615.

Clopper, C. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, pages 404–413.

Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A., and Micklern, G. (2015). Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol*, 16:DOI 10.1186/s13059–015–0607–3.

Deredec, A., Godfray, H. C. J., and Burt, A. (2011). Requirements for effective malaria control with homing endonuclease genes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(43):E874–E880.

Desquesnes, M. and Dia, M. L. (2003). Mechanical transmission of *Trypanosoma congolense* in cattle by the African tabanid *Atylotus agrestis*. *Experimental Pa*, 105:226–231.

Diekmann, O. and Heesterbeek, J. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. Wiley, New York.

Dietz, K., Molineaux, L., and Thomas, A. (1974). A malaria model tested in the African savannah. *Bulletin of the World Health Organization*, 50(3-4):347.

Dunnigan, K. (2008). Confidence interval calculation for binomial proportions. In *Midwest SAS user group*.

Dye, C. (1986). Vectorial capacity: must we measure all its components? *Parasitology Today*, 2(8):203–209.

Ericson, C. A. (2011). *Fault Tree Analysis Primer*. CreateSpace Inc., Charleston, USA, 136pp.

Esvelt, K. M., Smidler, A. L., Catteruccia, F., and Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *eLife*, 3:e03401.

Ferson, S. (1995). Quality Assurance for Monte Carlo Risk Assessment. In *Uncertainty Modeling and Analysis, 1995, and Annual Conference of the North American Fuzzy Information Processing Society. Proceedings of ISUMA - NAFIPS '95., Third International Symposium*.

Ferson, S. (1996). What monte carlo methods cannot do. *Human and Ecological Risk Assessment*, 2:990–1007.

Fournet, F., Cussac, M., Ouari, A., Meyer, P. E., Toè, H. K., Gouagna, L. C., , and Dabirè, R. K. (2010). Diversity in anopheline larval habitats and adult composition during the dry and wet seasons in Ouagadougou (Burkina Faso). *Malaria Journal*, 9:78.

Garrett-Jones, C. (1964a). The human blood index of malaria vectors in relation to epidemiological assessment. *Bull. World Health Org*, 30:241–261.

Garrett-Jones, C. (1964b). Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. *Nature*, 204:1173–1175.

Garthwaite, P. H. and O'Hagan, A. (2000). Quantifying expert opinion in the uk water industry: an experimental study. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(4):455–477.

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–135.

Gneme, A., Guelbeogo, W. M., Riehle, M. M., Sanou, A., Traore, A., Zongo, S., Eiglmeier, K., Kabre, G. B., N'Fale, S., and Vernick, K. D. (2013). Equivalent susceptibility of Anopheles gambiae M and S molecular forms and Anopheles arabiensis to Plasmodium falciparum infection in Burkina Faso. *Malaria*, 12:204.

Goddard, M. R. and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences*, 96:13880–13885.

Gould, F., Magori, K., and Huang, Y. (2006). Genetic strategies for controlling mosquito-borne diseases. *American Scientist*, 94:238–246.

Haimes, Y. Y. (1981). Hierarchical holographic modelling. *IEEE Transactions on Systems, Man and Cybernetics*, 11:606–617.

Haimes, Y. Y. (1998). *Risk Modeling, Assessment and Management*. John Wiley and Sons Ltd, New York, USA, 726 pp.

Hayes (2003a). *Bioinvasions: Pathways, Vectors, and Management Strategies.*, chapter Biosecurity and the role of risk-assessment, pages 382–414. Island Press, Washington, D.C., USA.

Hayes, K. R. (2002). Identifying hazards in complex ecological systems, Part 1: Fault tree analysis for biological invasions. *Biological Invasions*, 4:235–249.

Hayes, K. R. (2003b). KRA Project 1: Robust methodologies for GMO risk assessment. Inductive hazard analysis for GMOs. Technical report, Final report for the Australian Government Department of Environment and Heritage, CSIRO Division of Marine Research, Hobart, Australia, 59 pp.

Hayes, K. R., Greg, g. P. C., Gupta, V., Jessop, R., Lonsdale, W. M., Sindel, B., Stanley, J., and Williams, C. K. (2004). Identifying hazards in complex ecological systems. part 3: Hierarchical holographic model for herbicide tolerant oilseed rape. *Environmental Biosa*, 3:109–128.

He, X. and Wu, S.-J. (2009). Confidence intervals for the binomial proportion with zero frequency. In *Pharmaceutical Industry SAS Users Group*.

Hoch, A. L., Gargan, T. P., and Bailey, C. L. (1985). Mechanical transmission of Rift Valley fever virus by hematophagous Diptera. *American Journal of Tropical Medicine and Hygiene*, 34:188–193.

Hosack, G. R., Rossignol, P. A., and van den Driessche, P. (2008). The control of vector-borne disease epidemics. *Journal of Theoretical Biology*, 255:16–25.

Joardar, V., Abrams, N. F., Hostetler, J., Paukstelis, P. J., Pakala, S., Pakala, S. B., Zafar, N., Abolude, O. O., Payne, G., Andrianopoulos, A., Denning, D. W., and Nierman, W. C. (2012). Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. *BMC Genomics*, 13(1):698.

Jovanovic, B. D. and Levy, P. S. (1997). A look at the rule of three. *The American Statistician*, 51(2):137–139.

Keese, P. (2008). Risks from GMOs due to Horizontal Gene Transfer. *Environmental Biosafety Research*, 7:123–149.

Khan, S., Guo, L., Maimaiti, Y., Mijit, M., and Qiu, D. (2012). Entomopathogenic fungi as microbial biocontrol agent. *Molecular Plant Breeding*, 3(1):63–79.

Klein, T. A., Windbichler, N., Deredec, A., Burt, A., and Benedict, M. Q. (2012). Infertility resulting from transgenic i-ppoi male anopheles gambiae in large cage trials. *Pathogens and Global Health*, 106(1):20–31.

Lavitrano, M., Busnelli, M., Cerrito, M. G., Giovanonni, R., Manzini, S., and Vargiolu, A. (2006). Sperm-mediated gene transfer. *Reproduction, Fertility and Development*, 18:19–23.

Looney, S. (2013). Much ado about almost nothing: Methods for dealing with limited data.

Macdonald, G. (1952). The analysis of equilibrium in malaria. *Trop. Dis. Bull.*, 49:813–828.

Macdonald, G. (1957). *The Epidemiology and Control of Malaria*. Oxford University Press, London.

Marshall, J. (2011). Commentary: The Cartagena Protocol in the context of recent releases of transgenic and *Wolbachia*-infected mosquitoes. *Asia Pacific Journal of Molecular Biology and Biotechnology*, 19(3):93–100.

Marshall, J. M. (2010). The Cartagena Protocol and genetically modified mosquitoes. *Nature Biotechnology*, 28(9):896–897.

McGraw, E. A. and O'Neill, S. L. (2013). Beyond insecticides: new thinking on an ancient problem. *Nature Reviews Microbiology*, 11:181–193.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, England.

Mori, A., Chadee, D. D., Graham, D. H., and Severson, D. W. (2004). Reinvestigation of an Endogenous Meiotic Drive System in the Mosquito,*Aedes aegypti* (Diptera: Culicidae). *Journal of Medical Entomology*, 41:1027–1033.

Murray, C., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., and Lopez, A. D. (2012). Global malaria mortality between 1980 and 2010: A systematic analysis. *Lancet*, 379:413–431.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17(8):857–872.

North, A., Burt, A., and Godfray, H. C. J. (2013). Modelling the spatial spread of a homing endonuclease gene in a mosquito population. *Journal of Applied Ecology*, 50:1216–1225.

O'Hagan, A., Buck, C. E., Daneskhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts Probabilities*. John Wiley and Sons, Ltd., Chichester, England.

Oye, K. A., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., Lightfoot, S. B.-Y., McNamara, J., Smidler, A., and Collins, J. P. (2014). Regulating gene drives. *Science*, 345:626–628.

Pedroni, N. and Zio (2013). Uncertainty analysis in fault tree models with dependent basic events. *Risk*, 33:1146–1173.

Reiner, R. C., Perkins, T. A., Barker, C. M., Niu, T., Chaves, L. F., Ellis, A. M., George, D. B., Le Menach, A., Pulliam, J. R. C., Bisanzio, D., Buckee, C., Chiyaka, C., Cummings, D. A. T., Garcia, A. J., Gatton, M. L., Gething, P. W., Hartley, D. M., Johnston, G., Klein, E. Y., Michael, E., Lindsay, S. W., Lloyd, A. L., Pigott, D. M., Reisen, W. K., Ruktanonchai, N., Singh, B. K., Tatem, A. J., Kitron, U., Hay, S. I., Scott, T. W., and Smith, D. L. (2013). A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of The Royal Society Interface*, 10(81):20120921.

Robert, C. and Casella, G. (2005). *Monte Carlo statistical methods*. Springer, New York, USA, second edition.

Robinson, A., Burgman, M., Atkinson, W., Cannon, R., Miller, C., and Immonen, H. (2008). Aqis import clearance data framework. Technical report, Australian Centre of Excellence for Risk Analysis, University of Melbourne, Parkville, Victoria, Australia, Technical Report 0804.

Ross, R. (1911). *The Prevention of Malaria*. Murray, London, 2nd edition.

Ross, T. D. (2003). Accurate confidence intervals for binomial proportion and poisson rate estimation. *Computers in biology and medicine*, 33(6):509–531.

Scholte, E.-J., Knols, B. G. J., Samson, R. A., and Takken, W. (2004). Entomopathogenic fungi for mosquito control: A review. *Journal of Insect Science*, 4:19.

Sherwin, D. J. and Bossche, A. (1993). *The Reliability, Availability and Productiveness of Systems*. Chapman and Hall, London, England, 279 pp.

Smith, D. and McKenzie, F. (2004). Statics and dynamics of malaria infection in *Anopheles* mosquitoes. *Malaria Journal*, 3:13.

Smith, D. L., Drakeley, C. J., Chiyaka, C., and Hay, S. I. (2010). A quantitative analysis of transmission efficiency versus intensity for malaria. *Nature communications*, 1:108.

Suter, G. W. (1990). Endpoints for regional ecological risk assessment. *Environmental Management*, 14:9–23.

The Royal Society (1983). *Risk Assessment*. The Royal Society of Great Britain, London, England, 198 pp.

Tobi, H., van den Berg, P. B., de Jong-van den Berg, L., et al. (2005). Small proportions: what to report for confidence intervals? *Pharmacoepidemiology and drug safety*, 14(4):239–247.

UNEP (2010). Final report of the *ad hoc* technical expert group on risk assessment and risk management under the Cartegna Protocol on BioSafety. Technical report, United Nations Environment Program, Convention on Biological Diversity, 40 pp.

van den Driessche, P. and Watmough, J. (2002). Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180:29–48.

Vesely, W. E., Goldberg, F. F., Roberts, N. H., and Haasl, D. F. (1981). Fault Tree Handbook.

Technical report, Systems and Reliability Research, Office of Nuclear Regulatory Research, US Nuclear Regulatory Commission, Washington, USA, 209 pp.

Vickerman, K. (1973). The mode of attachment of *Trypanosoma vivax* in the proboscis of the tsetse fly *Glossina fuscipes*: an ultrastructural study of the epimastigote stage of the trypanosome". *Journal of Protozoology*, 20:394–404.

Vohra, P. and Blakely, G. W. (2013). Easing the global burden of diarrhoeal disease: can synthetic biology help? *Systems and Synthetic Biology*, 7(3):73–78.

Wang, W., Loman, J., and Vassiliou, P. (2004). Reliability importance of components in a complex system. In *Reliability and Maintainability, 2004 Annual Symposium - RAMS*, pages 6–11.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.

Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, A., Burt, A., and Crisanti, A. (2007). Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Research*, 35:5922–5933.

Windbichler, N., Papathanos, P. A., and Crisanti, A. (2008). Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. *PLoS Genetics*, 4:e1000291.

Winkler, R. L., Smith, J. E., and Fryback, D. G. (2002). The role of informative priors in zero-numerator problems: being conservative versus being candid. *The American Statistician*, 56(1):1–4.

Wolt, J., Keese, P., Raybould, A., Fitzpatrick, J. W., Burachik, M., Gray, A., Olin, S. S., Schiemann, J., Sears, M., , and Wu, F. (2010). Problem formulation in the environmental risk assessment for genetically modified plants. *Transgenic Research*, 19:425436.

World Health Organisation (2011). World malaria report.

Zio, E. (2009). *Computational Methods for Reliability and Risk Analysis*. World Scientific Publishing, Singapore.

Appendix A Results of the Boolean literature search

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., and Gabaldo n, T. (2014). High variability of mitochondrial gene order among fungi. Genome Biology and Evolution, 6(2):451-465. Elements of the fungus mitochondrial genome would facilitate up-take of exogenous DNA. This could be an avenue of escape of the construct into the environment if interacting with other control strategies based on pathogenic fungus. However, it would still need to be beneficial to the fungus to increase in frequency, and data suggest that uptake of HEG is largely detrimental (not addressed in this article).

Akbari, O. S., Chen, C.-H., Marshall, J. M., Huang, H., Antoshechkin, I., and Hay, B. A. (2012). Novel synthetic Medea selfish genetic elements drive population replacement in Drosophila; a theoretical exploration of Medea-dependent population suppression. ACS Synthetic Biology. Points out the importance of mutation leading to failure of the construct (reduced fitness of GMM will bring about strong selection to silence the construct). However, this paper seems more relevant for the second and third generation constructs, when a self-sustaining option will be the goal.

Akbari, O. S., Papathanos, P. A., Sandler, J. E., Kennedy, K., and Hay, B. A. (2014). Identification of germline transcriptional regulatory elements in Aedes aegypti. Scientific Reports, 4. Identifies genes expressed in the germ line to identify potential promoters for constructs. No risks identified.

Alphey, L. (2014). Genetic control of mosquitoes. Annual Review of Entomology, 59(1):205. Points out that environment and genetic background can affect expression of the construct, and its overall efficacy. No new risks identified.

Alphey, L., McKemey, A., Nimmo, D., Neira Oviedo, M., Lacroix, R., Matzen, K., and Beech, C. (2013). Genetic control of Aedes mosquitoes. Pathogens and Global Health, 107(4):170-179. Argues that one of the challenges is separating the sexes in mass before release, as it is highly undesirable to release females (as they transmit disease).

Alphey, L.S. and Beech, C. J. (2012). Genetically engineered insects-regulatory progress and challenges. In Regulation of Agricultural Biotechnology: The United States and Canada, pages 281-299. Springer. Provides a summary of regulatory frameworks. Will be more important in the second and third generation constructs where there is still much to learn. Points out one example of accidental release of mozzies (del Valle 2003 - screw-worm in Mexico).

Alphey, N. and Bonsall, M. B. (2014). Interplay of population genetics and dynamics in the genetic control of mosquitoes. Journal of the Royal Society Interface, 11(93):20131071. Mathematical model of ecological interactions affecting the efficacy of the construct depending on when the construct is expressed during the development of the mozzie. Important for second and third generation constructs.

Andersson, K. (2008). Ribonucleotide reductase. Nova Publishers. Incidental reference.

Appels, R., Adelson, D. L., Moolhuijzen, P., Webster, H., Barrero, R., and Bellgard, M. (2011). Genome studies at the PAG 2011 conference. Functional and Integrative Genomics, 11(1):1-11. Incidental reference.

Arnould, S., Bruneau, S., Cabaniols, J.-P., Chames, P., Choulika, A., Duchateau, P., Epinat, J.-C., Gouble, A., Lacroix, E., Paques, F., et al. (2010). Use of meganucleases for inducing homologous recombination ex vivo and in toto in vertebrate somatic tissues and application thereof. US Patent 7,842,489. Patents proposing the use of HEGs in human genetic therapy. Point out that integration into genomes is easier in germ lines (in eggs and embryos) than in somatic cells. This is mainly because these cells are still largely undifferentiated. Relevant to the potential hazard scenario identified already of eDNA interacting with eggs/embryos/larvae of free-living aquatic organisms.

Arnould, S., Chames, P., Duchateau, P., Epinat, J.-C., Lacroix, E., and Paques, F. (2013). Use of meganucleases for inducing homologous recombination ex vivo and in toto in vertebrate somatic tissues and application thereof. US Patent 8,530,214. As above.

Arnould, S., Chames, P., Duchateau, P., Epinat, J.-C., Lacroix, E., and Paques, F. (2014). Use of meganucleases for inducing homologous recombination ex vivo and in toto in vertebrate somatic tissues and application thereof. US Patent 8,624,000. As above.

Aryan, A. (2013). Gene Editing in Aedes aegypti. PhD thesis, Virginia Polytechnic Institute and State University. Results are more relevant to the second and third generation constructs. Demonstrates that when the HEG copies itself into the insertion site through homologous recombination, it can often do so with error (high mutation rate).

This can often silence or inactivate the construct.

Baldini, F., Gabrieli, P., Rogers, D. W., and Catteruccia, F. (2012). Function and composition of male accessory gland secretions in Anopheles gambiae: a comparison with other insect vectors of infectious diseases. Pathogens and Global Health, 106(2):82-93. Review of the reproductive biology of Anophele gambiae. Suggests that reproductive isolation between WT and GMM could happen if the construct affects male swarming behaviour or how it beats its wings. Also, it suggests that WT females that mate with GM male could seek other matings if the seminal fluid is modified in any way by the construct, and obstructing or precluding the physiological queues that lead to egg laying and egg production.

Beech, C. J., Koukidou, M., Morrison, N. I., and Alphey, L. (2012). Genetically modified insects: science, use, status and regulation. Collection of Biosafety Reviews, 6:66-124. Interesting comment about a piggyBac transposon inserted gene into A. stephensi escaping, even though the auxiliary transposase was not available. Suggests that piggyBac inserts might not be the most stable in Anophelenes as suspected.

Bell, G. (1993). The sexual nature of the eukaryote genome. Journal of Heredity, 84(5):351-359. Excellent summary of ways genes can move between organisms, although a little dated now. Brings up the information that gene transfer from eukaryotes to prokaryotes is substantially less common than other gene transfer events.

Bertolini, L., Bunting, M., Madden, K., Maga, E., and Murray, J. (2006). Methods for increasing efficiency of homologous recombination. US Patent App. 11/490,474. No additional risks identified.

Bertolotti, R. and Ozawa, K. (2008). Autologous and Cancer Stem Cell Gene Therapy, volume 3. World Scientific. A patent and a book. Notes that random insertions of genes into eukaryotic genomes can lead to unintended problems, and that the same construct inserted into different locations can have different expression patterns. This raises the possibility that the two lineages are very different in their expression levels and genome-wide consequences.

UNEP (2010). Secretariat of the Convention on Biological Diversity. A discussion thread under the banner of the Biosafety Clearing House, that discusses briefly some elements of RA associated with GM mosquitoes, and the various ways in which they could be used to reduce mosquito impacts. Essential background reading. Alludes to a lot of APHIS work on developing RA strategies for insects engineered for agricultural practices. The thread itself outlines no new risks, but very clearly relevant, and especially related threads (i.e., formal guidelines on risk assessment of Living Modified Organisms, on monitoring LMOs, etc.).

Blackbourn, D. Landschaft und umwelt in der deutschen geschichte. Akademievorlesungen der Interdisziplina ren Arbeitsgruppe Funktionen des Bewusstseins. No additional risks identified.

Boete, C. and Beisel, U. (2013). Transgenic mosquitoes for malaria control: from the bench to the public opinion survey. No additional risks identified.

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1518):921-928. No additional risks identified. HEG site has natural variation/resistance alleles.

Burt, A. (2014). Heritable strategies for controlling insect vectors of disease. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1645). No additional risks identified.

Caragata, E. P. and Walker, T. (2012). Using bacteria to treat diseases. Expert Opinion on Biological Therapy, 12(6):701-712. No additional risks identified.

Carninci, P. and Hayashizaki, Y. (2002). Cloning vectors and method for molecular cloning. US Patent App. 10/469,508. No additional risks identified.

Carroll, D. (2014). Genome engineering with targetable nucleases. Annual Review of Biochem- istry, 83(1):409-439. PMID: 24606144. No additional risks identified.

Catteruccia, F., Crisanti, A., Wimmer, E. A., et al. (2009). Transgenic technologies to induce sterility. Malaria Journal, 8(Suppl 2):S7. No additional risks identified.

Chan, Y.-S., Takeuchi, R., Jarjour, J., Huen, D. S., Stoddard, B. L., and Russell, S. (2013). The design and In Vivo evaluation of engineered I-OnuI -based enzymes for HEG gene drive. PloS One, 8(9). No additional risks identified.

Cost, G. J. (2013). Methods and compositions for delivery of biologics. US Patent App. 13/939,633. No additional risks identified.

Cost, G. J., Gregory, P. D., Guschin, D., Holmes, M. C., Miller, J. C., Paschon, D., Rebar, E. J., Reik, A., Urnov, F., Zhang, L., et al. (2013). Methods and compositions for treatment of a genetic condition. US Patent App. 14/013,250. No additional risks identified.

Craig, N. L. (1997). Target site selection in transposition. Annual Review of Biochemistry, 66(1):437-474. No additional risks identified.

Darbani, B., Eimanifar, A., Stewart, C. N., and Camargo, W. N. (2007). Methods to produce marker-free transgenic plants. Biotechnology Journal, 2(1):83-90. No additional risks identified.

David, A. S., Kaser, J. M., Morey, A. C., Roth, A. M., and Andow, D. A. (2013). Release of genetically engineered insects: a framework to identify potential ecological effects. Ecology and Evolution, 3(11):4000-4015. Worth reading. All suggested risks already covered. Could HEG select and move insecticide resistance allele? Discusses evolutionary consequence of removal of An. gambiae to other species such as An. Arabiensis. Introgression with An. arabiensis and removal of local populations could result in losses in herd immunity (transient reduction in immunity in human populations can facilitate explosive malaria epidemic). Effect on virulence of the pathogen (Plasmodium) transmission by mosquitoes of lower vector capacity may lead to more virulent pathogen. Evolution of resistance through reproductive isolation pre/post zygotic.

De Visser, A., Nijhuis, E., van Elsas, J., Dueck, T., and Inventory, A. (2000). Crops of uncertain nature. Controversies and Knowledge Gaps concerning Genetically Modified Crops. No additional risks identified.

Deredec, A., Godfray, H. C. J., and Burt, A. (2011). Requirements for effective malaria control with homing endonuclease genes. Proceedings of the National Academy of Sciences of the United States of America, 108(43):E874-E880. HEG cleavage events being repaired in other ways leading to some change at the target site that would make it no longer recognized by the enzyme. Resistance to cleavage may lead to fitness effects (fit may go to fixation). Decreased HEG resistance by targeting multicopied rDNA but this marker can evolve.

Dorscht, J. (2007). Comparative genomics of Listeria bacteriophages. PhD thesis, Dissertation. Technische Universita t Mu nchen, Weihenstephan-Freising, Germany. No additional risks identified.

Egan, R. S. (2007). Recent literature on lichens-206. The Bryologist, 110(3):577-593. Approximately 15 pages of citations covering literature on LichensEverett, K. D. (2002). Chlamydiae. eLS. No additional risks identified.

Fahrenkrug, S. C. and Hackett, P. B. Precision editing of large animal genomes. No additional risks identified.

Fernandes, N. D. (2011). Molecular studies on the role of bacteria in a marine algal disease. PhD thesis, The University of New South Wales. No additional risks identified.

Fishman, L. and Jaenike, J. (2013). Selfish genetic elements and genetic conflict. The Princeton Guide to Evolution, page 347. Can HEGs lead to antagonistic co-evolution with other components of the genome? (DNA methylation, RNAi, small RNA regulatory pathways).

Folcher, M. and Fussenegger, M. (2012). Synthetic biology advancing clinical applications. Current Opinion in Chemical Biology, 16(3):345-354. No additional risks identified.

Franz, A. W., Sanchez-Vargas, I., Adelman, Z. N., Blair, C. D., Beaty, B. J., James, A. A., and Olson, K. E. (2006). Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America, 103(11):4198-4203. No additional risks identified.

Fraser Jr, M. J. (2012). Insect transgenesis: current applications and future prospects. Annual Review of Entomology, 57:267-289. No additional risks identified.

Gaardbo Kuhn, K., Campbell-Lendrum, D. H., and Davies, C. R. (2002). A continental risk map for malaria mosquito (Diptera: Culicidae) vectors in Europe. Journal of Medical Entomology, 39(4):621-630. No additional risks identified.

Gaj, T., Gersbach, C. A., and Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology, 31(7):397-405. No additional risks identified.

Gaj, T., Sirk, S. J., Tingle, R. D., Mercer, A. C., Wallen, M. C., and Barbas III, C. F. (2014). Enhancing the specificity of recombinase-mediated genome engineering through dimer interface redesign. Journal of the American Chemical Society. No additional risks identified.

Garcia, P., Martinez, B., Obeso, J. M., Lavigne, R., Lurz, R., and Rodriguez, A. (2009). Functional genomic analysis

of two Staphylococcus aureus phages isolated from the dairy environ- ment. Applied and Environmental Microbiology, 75(24):7663-7673. No additional risks identified.

Grube, M. and Berg, G. (2009). Microbial consortia of bacteria and fungi with focus on the lichen symbiosis. Fungal Biology Reviews, 23(3):72-85. No additional risks identified.

Hagen, D. E. (2007). Identification and characterization of germline-specific promoters for re- mobilization of transgenes in the mosquitoes, Aedes aegypti and Anopheles gambiae. PhD thesis, Texas A&M University. No additional risks identified.

Haugen, P., Bhattacharya, D., Palmer, J. D., Turner, S., Lewis, L. A., and Pryer, K. M. (2007). Cyanobacterial ribosomal RNA genes with multiple, endonuclease-encoding group I introns. BMC Evolutionary Biology, 7(1):159. No additional risks identified.

Hellinga, H. W., Jantz, D., and SMITH, J. J. (2012). Method of cleaving DNA with rationally designed meganucleases. US Patent 8,124,369. No additional risks identified.

Heng, B. C. and Fussenegger, M. (2013). Design and application of synthetic biology devices. Synthetic Biology: Tools and Applications. No additional risks identified.

Horton, N. C. and Perona, J. J. (2004). DNA cleavage by Eco RV endonuclease: Two metal ions in three metal ion binding sites. Biochemistry, 43(22):6841-6857. No additional risks identified.

Hranueli, D. (2005). Industrial applications of genomics, proteomics and bioinformatics. NATO Science Series Sub Series I Life and Behavioural Sciences, 368:176. No additional risks identified.

Hsu, P. D. and Zhang, F. (2012). Dissecting neural function using targeted genome engineering technologies. ACS Chemical Neuroscience, 3(8):603-610. No additional risks identified.

Humbert, O., Davis, L., and Maizels, N. (2012). Targeted gene therapies: tools, applications, optimization. Critical Reviews in Biochemistry and Molecular Biology, 47(3):264-281. Reviews a number of technologies, including HEGs. Of relevance, mentions that target-specificity might be dose dependent, with higher doses making the HE toxic. This is probably why the PpoI construct only works well in low doses. Thus, it suggests it can hit off-target. However, the effect seems to cause death. So, it seems unlikely that increase dosage of the HE would be problematic.

Jain, K. (2012). Synthetic biology and personalized medicine. Medical Principles and Practice, 22(3):209-219. A good, semi-technical overview of synthetic biology (building genes, chromosomes, etc. from basic molecular building blocks), with a discussion of likely applications ranging from pharmacology to personalized gene therapy. Nothing novel relevant to hazard analysis.

Joardar, V., Abrams, N. F., Hostetler, J., Paukstelis, P. J., Pakala, S., Pakala, S. B., Zafar, N., Abolude, O. O., Payne, G., Andrianopoulos, A., et al. (2012). Sequencing of mitochondrial genomes of nine Aspergillus and Penicillium species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics, 13(1):698. Describes the mitochondrial genome of nine fugal species from two genera (Aspergillus and Penicilium). A number of HEG elements are described. Significant diversity in auxiliary elements and HEG type elements among species is thought to conform with the notion of significant intra-species HGT. The idea of interspecies HGT is not considered. Unclear what this could mean for risks. HEGs occur naturally, and can be shared among organisms, yet there is no record of it causing damage beyond that.

Kang, H.-W., Kim, J.-W., Jung, T.-S., and Woo, G.-J. (2013). wksl3, a new biocontrol agent for Salmonella enterica serovars enteritidis and typhimurium in foods: Characterization, application, sequence analysis, and oral acute toxicity study. Applied and Environmental Microbiol- ogy, 79:1956-1968. Describes the isolation of a phage that is toxic to Salmonella. Of relevance is the diversity of phages found associated with bacteria. This implies some potential for genetic material to transfer in both directions. Although, this is not addressed in this paper.

Kapuscinski, A. R. and Patronski, T. J. (2005). Genetic methods for biological control of non- native fish in the Gila River Basin: Final report to the US Fish and Wildlife Service. Brief and slightly dated overview of daughterless and sterile male technologies as they might be applied to managing sunfish and other non-native species in Arizona, with an extensive discussion of the many levels of regulatory issues and jurisdictions that would need top be worked through if the technology was to be used. Risk issues raised in this report are well covered by our hazard analyses.

Khan, S., Guo, L., Maimaiti, Y., Mijit, M., Qiu, D., et al. (2012). Entomopathogenic fungi as microbial biocontrol agent. Molecular Plant Breeding, 3(1).

Konig, H., Frank, D., Heil, R., and Coenen, C. (2013). Synthetic genomics and synthetic biology applications between hopes and concerns. Current Genomics, 14(1):11. Reviews a number of uses of synthetic genomics and synthetic biology. Discusses some of the risks involved with release of GM organisms into the wild, including horizontal gene transfer, 'empty niche', human health concerns (e.g., allergic reactions). No new risks are raised.

Lee, T.-W., Verhey, T. B., Antiperovitch, P. A., Atamanyuk, D., Desroy, N., Oliveira, C., De- nis, A., Gerusz, V., Drocourt, E., Loutet, S. A., et al. (2013). Structural-functional studies of Burkholderia cenocepacia d-glycero-d-manno-heptose 7-phosphate kinase (HldA) and characterization of inhibitors with antibiotic adjuvant and antivirulence properties. Journal of Medicinal Chemistry, 56(4):1405-1417. A study describing the structure and function of a specific enzyme potentially involved in antibiotic resistance in medically important bacteria. There is concern that it is transmitted to other bacteria via HGT. No relevant risks are discussed.

Legros, M., Xu, C., Morrison, A., Scott, T. W., Lloyd, A. L., and Gould, F. (2013). Modeling the dynamics of a non-limited and a self-limited gene drive system in structured Aedes aegypti populations. PloS One, 8(12):e83354. A model based analysis of different release strategies for Dengue-resistance genes being driven into an urban Aedes population using two different gene drive systems (Medea and Killer-Rescue). Brief discussion of social and risk elements associated with the approach, but nothing new identified. Refers to a mosquito (Aedes) population dynamics model (Skeeter Buster) that might be useful (freely available) with Anopheles specific parameters.

Liu, W., Yuan, J. S., and Stewart Jr, C. N. (2013). Advanced genetic tools for plant biotechnology. Nature Reviews Genetics, 14(11):781-793. A review of current methods to modify organisms. Proposes mechanisms for removing the transgenic genes or of limiting its spread to address government concerns. At the end, suggests that regulations are potentially outdated, and need to be revised not to 'quash' innovation.

Lobocka, M., Hejnowicz, M. S., Dabrowski, K., Gozdek, A., Kosakowski, J., Witkowska, M., Ulatowska, M. I., Weber-Dabrowska, B., Kwiatek, M., Parasion, S., et al. (2012). Genomics of Staphylococcal Twort-like phages-potential therapeutics of the post-antibiotic era. Advances in Virus Research, 83(83):143-216. While phages might be a vehicle for transport of foreign DNA between bacteria, and perhaps between host and bacteria, there are a number of cellular mechanisms designed to stop the foreign DNA from integrating. Plus, as suggested in Bell (1993), transfer of genetic material from eukaryote to prokaryote is least common of the horizontal transfer events. The safety of these phages is thought such that the medical community is considering using them to fight resistant strains of bacteria.

Ma, L., Dong, J., Jin, Y., Chen, M., Shen, X., and Wang, T. (2011). RMDAP: A versatile, ready- to-use toolbox for multigene genetic transformation. PloS One, 6(5):e19883. Reports a new method to introduce multiple genes into plant genomes. No risks are reported or discussed.

Marshall, J. (2011). Commentary: The Cartagena protocol in the context of recent releases of transgenic and Wolbachia-infected mosquitoes. Asia Pacific Journal of Molecular Biology and Biotechnology, 19(3):93-100. Review of several issues relating to international movement of GM mosquitoes that are, basically, not well covered by the CP. Main issues are that importing countries cannot require exporters to undertake a risk assessment if the GMs are to be used in lab studies initially (even if subsequently used in fields releases, which would be subject only to national regulations), problems in dealing with non-parties to the CP (such as Australia) which are not obliged to undertake risks associated with international spread. Suggested that in most (all?) cases risks assessments would be done anyway, but may not be obligatory. Paper does not raise new specific risks/hazards, but does highlight that the CP does not require TMc to undertake a risk analysis prior to import if the GMs are to be used in laboratory trials.

Marshall, J. M. (2010). The Cartagena Protocol and genetically modified mosquitoes. Nature Biotechnology, 28(9):896-897. A look at the potential failings of the Cartagena Protocol to regulate trans-boundary effects of gene drive systems. Marshall is concerned that there are no mechanisms to control the spread of a gene through a gene drive systems into countries other than those directly involved in the release. He is also concerned about that the Cartagena Protocol may allow countries to develop and test transgenic mosquitoes in captivity without proper risk assessment.

Marshall, J. M. and Hay, B. A. (2012a). Confinement of gene drive systems to local populations: a comparative analysis. Journal of Theoretical Biology, 294:153-171. The authors undertake to model the ability to spatially confine a release of different transgenic technologies. The scenario is an intentional release, but with the caveat that it should not escape from the restricted spatial locations surrounding the release site. As stated by the authors, the models are simple and provide a basis for comparisons among the technologies, but should not be considered predictive. In comparison with other technologies the HEG system shows little chance of confinement.

Marshall, J. M. and Hay, B. A. (2012b). General principles of single-construct chromosomal gene drive. Evolution, 66(7):2150-2166. Models the Medea system, an alternative to the HEG, it carries the advantage of being easy to spatially confine. The authors mode, several scenarios of the system (e.g., sex-linked, autosomal), and examine the population dynamics of the genes. No risks beyond the spatial confinement issue are discussed.

Marshall, J. M., Pittman, G. W., Buchman, A. B., and Hay, B. A. (2011). Semele: a killer-male, rescue-female system for suppression and replacement of insect disease vector populations. Genetics, 187(2):535-551. A mathematical (stochastic) model describing the dynamics of a hypothetical male-carried female lethal system coupled with a female-carried immunity gene. Shows that under a range of conditions, localised gene drive can occur, which could be used to convey, e.g., pathogen resistance. Refers to X-drive systems, noting that the dynamics are similar to those in which the female resistance is carried by an autosome. No new specific hazards raised.

Mertens, M. (2008). Assessment of environmental impacts of genetically modified plants. Implementation of the Biosafety Protocol Development of Assessmant Bases FKZ, 20167430(07). A comprehensive review of potential impacts of GM plants, listing numerous case studies, examples of VGT, etc. Of primary interest is discussion of bacterially mediated HGT at evolutionary time scales, with suggestions that even very rare events could have long-term implications, but would be difficult to detect using conventional methods. Hazards raised are covered by our analyses, but warns of underestimating the risks imposed by bacterial/soil mediated HGT.

Minnick, M. F. and Raghavan, R. (2011). Genetics of Coxiella burnetii: On the path of specialization. Future Microbiology, 6(11):1297-1314. Coxiella is an obligate intracellular parasite that infects widely, and can be found everywhere except NZ. It is a bacterium, and has selfish genetic elements like the PpoI endonuclease. The endonucleases could have originated via HGT from a eukaryote, or through VGT from an ancestor. In spite of this observation, the authors say that HGT between an intracellular parasite and its host should be rare, although they don't specify why that should be the expectation. The following reference might be useful to follow up on: Moran NA, Plague GR. Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. 2004; 14:627???"633. [PubMed: 15531157]

Mlgaard, L., Patil, K. R., Thykr, J., Mortensen, U. H., and Eliasson Lantz, A. (2012). Engineering the Polyketide Cell Factory. PhD thesis, Technical University of DenmarkDanmarks Tekniske Universitet, Department of Systems Biology, Institute for Systems biology, Center for Microbial Biotechnology. PhD thesis demonstrating how to modify yeast cells to produce polyketides. No relevant risks are discussed.

Nguyen, M. T., Liu, M., and Thomas, T. (2013). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Molecular Ecology. Example of horizontal gene transfer from a eukaryote to a bacteria.

Olson,K. E. and Blair, C. D. (2012). Flavivirus-vector interactions. Molecular Virology and Control of Flaviviruses, page 297. No additional risks identified.

Palazzoli, F., Carnus, E., Wells, D. J., and Bigot, Y. (2008). Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. Current Gene Therapy, 8(5):367-390. No additional risks identified.

Papathanos, P. A., Windbichler, N., Menichelli, M., Burt, A., and Crisanti, A. (2009). The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies. BMC Molecular Biology, 10(1):65. No additional risks identified.

Pearlman, A. L. and Stern, B. S. (2011). Long lasting drug formulations. US Patent App.13/160,632. No additional risks identified.

Persad, D. L., Quach, U., Thorsteinsdo ttir, H., Salamanca-Buentello, F., Singer, P., and Daar, A. (2006). Enabling knowledge societies in developing countries: the example of genomics. International Journal of Biotechnology, 8(1):4-22. No additional risks identified.

Prax, M., Lee, C. Y., and Bertram, R. (2013). An update on the molecular genetics toolbox for Staphylococci. Microbiology, 159(Pt 3):421-435. No additional risks identified.

Pribat, A., Boureau, L., Mortain-Bertrand, A., Bert, L. S., Rolin, D., Teyssier, E., and Gallusci, P. (2013). Metabolic engineering of isoprenoid biosynthesis. In Natural Products, pages 2813-2851. Springer. No additional risks identified.

Protocol on Biosafety (2012a). Final Report of the ad hoc Technical Expert Group on Risk Assessment and Risk Management Under the Cartagena Protocol on Biosafety. No additional risks identified.

Protocol on Biosafety (2012b). Guidance on Risk Assessment of Living Modified Organisms I. Introduction. No additional risks identified.

Que, Q., Chilton, M.-D. M., de Fontes, C. M., He, C., Nuccio, M., Zhu, T., Wu, Y., Chen, J. S., and Shi, L. (2010). Trait stacking in transgenic crops. Pat, 24236:5. No additional risks identified.

Raghavan, R. (2008). Mobile genetic elements in Coxiella burnetii: Friends, foes or just indifferent? ProQuest. No additional risks identified.

Reumer, A., Van Loy, T., Clynen, E., and Schoofs, L. (2008). How functional genomics and genetics complements insect endocrinology. General and Comparative Endocrinology, 155(1):22-30. No additional risks identified.

Riccombeni, A., Vidanes, G., Proux-We ra, E., Wolfe, K. H., and Butler, G. (2012). Sequence and analysis of the genome of the pathogenic yeast Candida orthopsilosis. PLoS One,7(4):e35750. No additional risks identified

Rosas, C. T., Goodman, L. B., von Einem, J., and Osterrieder, N. (2006). Equine herpesvirus type 1 modified live virus vaccines: quo vaditis? No additional risks identified.

Saha, J. (2011). Development of Defined Biological Models for Complex Radiation-Induced DNA Lesions Using Homing Endonucleases and Transposon Technology: Feasibility and Initial Characterization. PhD thesis. No additional risks identified.

Samuels, M. A. (2011). Structural and Biochemical Studies of UvrA, the Bacterial NER DNA Damage Sensor, and the Biochemical Characterization of a Bacterial MCM Protein. Harvard University. No additional risks identified.

Sarker, S. A., McCallin, S., Barretto, C., Berger, B., Pittet, A.-C., Sultana, S., Krause, L., Huq, S., Bibiloni, R., Bruttin, A., et al. (2012). Oral T4-like phage cocktail application to healthy adult volunteers from Bangladesh. Virology, 434(2):222-232. No additional risks identified.

Schleef, M., Blaesen, M., Schmeer, M., Baier, R., Marie, C., Dickson, G., and Scherman, D. (2010). Production of non-viral DNA vectors. Current Gene Therapy, 10(6):487-507. No additional risks identified.

Scolari, F., Siciliano, P., Gabrieli, P., Gomulski, L., Bonomi, A., Gasperi, G., and Malacrida, A. (2011). Safe and fit genetically modified insects for pest control: from lab to field applications. Genetica, 139(1):41-52. No additional risks identified

Segal, D. J. and Meckler, J. F. (2013). Genome engineering at the dawn of the golden age. Annual Review of Genomics and Human Genetics, 14:135-158. No additional risks identified.

Sinkins, S. P. and Gould, F. (2006). Gene drive systems for insect disease vectors. Nature Reviews Genetics, 7(6):427-435. No additional risks identified.

Sinkovics, J. G. (2011). Horizontal gene transfers with or without cell fusions in all categories of the living matter. In Cell Fusion in Health and Disease, pages 5-89. Springer. No additional risks identified.

Smith, J. J. and Jantz, D. (2013). Rationally-designed single-chain meganucleases with non-palindromic recognition sequences. US Patent 8,445,251. No additional risks identified.

Smith, J. J., Jantz, D., and Hellinga, H. W. (2012). Methods of cleaving DNA with rationally designed meganucleases. US Patent 8,143,016. No additional risks identified.

Smith, J. J., Jantz, D., and Hellinga, H. W. (2013). Method for producing genetically-modified cells with rationally designed meganucleases with altered sequence specificity. US Patent 8,377,674.

A set of 3 stacked patents that discuss the design features of synthetic HEGs (= meganucleases) and describes improved methods for producing them that improves their targeting efficiency. Raises the issues of “pseudo” and completely palindromic sequences that would be targets for naturally occurring HEGs, notes that such palindromes would be rare, and describes design specifications that improve targeting on non-palindromic sequences.

Sourdive, D. (2011). Method for targeted genomic events in algae. US Patent App. 13/813,705. A patent that describes design and build of algae-targeting HEGs. Raises issue of “large set of genes [in chromophytic algae diatoms] that arise from lateral transfer from bacteria”, and a reference thereto (Bolwler et al., 2008, Nature 456:239-244) that is not in the Boolean search list. A read of that paper indicated that most were of ancient origin, so frequent but only on evolutionary time scales.Diatoms are apparently diverse and common in African freshwater systems.

Subramanian, R. A. (2008). The behavior and evolution of Class II transposable elements in the malarial mosquito,

Anopheles gambiae. ProQuest. Detailed study of a class II (large and common class) of transposable elements. Notes several features in A gambiae – site deletion is expected to be common in transposable elements, though apparently not in this case; TE expected to increase in frequency in genome during “active” phase (though also not noted in this case). HAT element thought to derive from a “recent” HGT, but again recent is in an evolutionary, not ecological, context. Raises issues of gradual increase and site deletion (= mutation and loss of function, but not re-targeting?) in TEs.

Sun, N., Abil, Z., and Zhao, H. (2012). Recent advances in targeted genome engineering in mammalian systems. Biotechnology Journal, 7(9):1074-1087. Review of HE and zinc finger gene insertion/deletion mechanisms in human medical applications.

Sutherland, W. J., Aveling, R., Brooks, T. M., Clout, M., Dicks, L. V., Fellman, L., Fleishman, E., Gibbons, D. W., Keim, B., Lickorish, F., et al. (2014). A horizon scan of global conservation issues for 2014. Trends in Ecology and Evolution, 29(1):15-22. Notes use of GM technology is one of ten emerging issues in conservation biology. Highlights some (already noted) environmental effects and further notes they are not well investigated.

Tan, W., Carlson, D. F., Walton, M. W., Fahrenkrug, S. C., and Hackett, P. B. (2011). Precision editing of large animal genomes. Advances in Genetics, 80:37-97. An extensive review of GM methods for applications in animals. Notes virally mediated integration tends to occur into and/or proximal to resident genes and thereby influence normal cellular function. Something to be considered in relation to viral mediated HGT.

Thresher, R. E., Hayes, K., Bax, N. J., Teem, J., Benfey, T. J., and Gould, F. (2013). Genetic control of invasive fish: technological options and its role in integrated pest management. Biological Invasions, pages 1-16. A review of some GM applications and options for managing invasive fish.

Tolman, J. S. and Valvano, M. A. (2012). Global changes in gene expression by the opportunistic pathogen Burkholderia cenocepacia in response to internalization by murine macrophages. BMC Genomics, 13(1):63. Comments on issues relating to intracellular invasion by bacteria, and notes the prevalence of very high diversity of habitats and hosts (amoebae, fungus, insects, plants and animals) in this single opportunistic group of related species (17 apparently). Further check of references indicates that single species can survive as free living, in plants or in amoeba and, in the latter case, can act as a “trojan horse” to get the bacteria into human respiratory tissues. Possible route of human impact of horizontally transferred DNA? HGT common in the genus over evolutionary time scales.

Trevors, J. and Masson, L. (2010). DNA technologies: what’s next applied to microbiology research? Antonie van Leeuwenhoek, 98(3):249-262. Notes HEG approach to pest control. Otherwise a lengthy article about future applications of microbial DNA applications.

Uriz, M. J., Turon, X., et al. (2012). Sponge ecology in the molecular era. Advances in Marine Biology, 61:345-410. Detailed review of phylogenetics, etc. of sponges. Of limited relevance.

Voght, S. P. (2007). Establishment of a *Drosophila* model of intestinal sterol absorption and trafficking. PhD thesis, University of Washington. No additional risks identified.

Vohra, P. and Blakely, G. W. (2013). Easing the global burden of diarrhoeal disease: can synthetic biology help? Systems and Synthetic Biology, 7(3):73-78. Good review of some of the key issues regarding governance and social acceptability of synthetic organisms when used in a disease management context. Specifically notes the need for a genetic “off switch’ of some kind that can be integrated into the GMO to prevent its spread or having an effect outside of the target organism, citing examples of temperature-sensitivity as an exogenously triggered natural barrier or the Jurassic Park scenario (organism requires something artificially supplied to reproduce or survive). Notes fecal-oral uptake in poor quality drinking water major cause of death of under 5 yo’s in developing nations. Implications for up-take of bacteria (in particular E coli) that might have taken up construct in urban water bodies?

Weber, W. and Fussenegger, M. (2012). Emerging biomedical applications of synthetic biology. Nature Reviews Genetics, 13(1):21-35. Generally positive review of GM approaches to disease management. Notes HEG/RIDL efforts at mosquito control.

White, M. A. (2010). Replication, recombination and chromosome segregation in Escherichia coli. Experimental and observation studies of chromosomal structure and sequencing in E coli. No direct RA implications.

Wilke, A. B. B., Nimmo, D. D., St John, O., Kojin, B. B., Capurro, M. L., Marrelli, M. T., et al. (2009). Mini-review: Genetic enhancements to the sterile insect technique to control mosquito populations. Asia Pacific Journal of Molecular Biology and Biotechnology, 17(3):65-74. Short review of SIT in insects, and specifically GM approaches to

improve SIT. Notes HEG and Wolbachia efforts, and summarises current efforts involving Oitec RIDL. No discussion of risks.

Yau, Y.-Y. and Stewart, C. N. (2013). Less is more: Strategies to remove marker genes from transgenic plants. BMC Biotechnology, 13(1):36. Discusses laboratory options for excision of a marker gene in the production of a GM plant line, on the basis that the marker gene is only of value during GM production and disadvantageous for a wide range of reasons in production lines. Could be useful in the context of ways to limit penetration of the HEG outside of the target species, pointing to approaches that could be applicable or adapted to a field situation.

Yin, L.-F., Wang, F., Zhang, Y., Kuang, H., Schnabel, G., Li, G.-Q., and Luo, C.-X. (2014). Evolutionary analysis revealed the horizontal transfer of the Cyt b gene from fungi to chromista. Molecular Phylogenetics and Evolution. HGT across phyla, but ancient/of evolutionary significance.

Appendix B HHM hazard scenarios

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
1. GM mosquitoes incorrectly marked (fluoro) or not marked at all. Or fluorescence is no longer visible due to mutation. With successive generations change in fitness that is not noticed, but male fertility appears. Colonies not sufficiently separated.	1	yes	1.1	1.1	1
2. Possible fitness change but largest effect from total number released. Large release might have strong local effect.	1	yes	1.1	1.1	1
3. Change in harvest or mortality rate of predators leads to change in abundance of mosquitoes.	1	yes	1.1	1.1	1
4. Any change to breeding sites or distribution of adults (bio-transport).	1	yes	1.1	1.1	1
5. Nursery changes background environment. Transport of plants could lead to dispersal of mosquitoes.	1	yes	1.1	1.1	1
6. Changes in mosquito survival and dispersal (wet versus dry season). In dry season relatively larger effect of a release with respect to wild population.	1	yes	1.1	1.1	1
7. Seasonal differences lead to different potential for pathogen transmission (likely greatest risk in wet season, but in dry season great relative effect of GM mosquitoes with respect to wild population).	1	yes	1.1	1.1	1
8. In Lake Victoria, larvae get carried in boats, but insectaries far from the boats. Issue of release. Also issue of proximity to international border.	1	yes	1.1	1.1	1
9. Beginning of wet season could differentially affect persistence.	1	yes	1.1	1.1	1
10. Interaction between trade and fire and chemicals in environment resulting in local knock down of predators/prey and local competition.	1	yes	1.1	1.1	1
11. Fire affects vegetation cover which can affect mosquito movement & survival (interpreted as an increase in vectorial capacity).	1	yes	1.1	1.1	1
12. Knock out of local predator or prey due to extreme events (including locust control) altering local ecology of receiving environment (interpreted as a release from predators leading to an increase in vectorial capacity).	1	yes	1.1	1.1	1

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
13. Semi-abandoned fishing boats great breeding sites for dispersal and establishment (BF). Use of and transport of clay pots (used for transport of water) in Mali.	1	yes	1.1	1.1	1
14. Human mediated long distance dispersal.	1	yes	1.1	1.1	1
15. Dusty environments less amenable to predators and resulting increase in persistence.	1	yes	1.1	1.1	1
16. Urban and physical environment affects persistence.	1	yes	1.1	1.1	1
17. In war, or in large shifts in market forces, increased level slaughter of cattle leads to increased biting.	1	yes	1.3	1.1	1
18. Any business that attracts infected people (i.e., clinics, schools, pharmacy).	1	yes	2	1.1	1
19. Change in cultivation practices leading to change in plant community which affects distribution, breeding sites, or resources. Could change human-mosquito interactions.	1	yes	7	1.1	1
20. Protestors or curious public leads to more people near facilities increasing their interactions with laboratory and escapees.	1	yes	7	1.1	1
21. Changes in mosquito survival and dispersal (wet versus dry season). In dry season relatively larger effect of a release with respect to wild population. Changes in human behaviour changes their exposure to escapee mosquitoes.	1	yes	7	1.1	1
22. In urban environment low mosquito density, men go to work in rural area, but don't transmit due to low numbers of mosquito in urban area, so now post release/escape malaria transmission due to increased numbers of mosquito. Increase in malaria transmission in urban environment due to increased number of mosquito in urban areas post release/escape. Immigrant populations also moving into cities and now increased population increases malaria transmission.	1	yes	7	1.1	1
23. Abundance seasonally affected. Laboratory reared mosquitoes might have different predator-prey responses, generally this would decrease their survival and diminish their risk, but there could be the opposite (i.e., fly lower less bat predation).	2	yes	1.1	1.1	2

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
24. Porous versus heavy or impermeable soils mediate breeding sites and access to breeding sites. Laboratory-reared mosquitoes may be better adapted to enriched or less clean water. Differences in water quality might have different levels of predation pressure for larval or adult mosquitoes.	2	yes	1.1	1.1	2
25. Change in competition due to potential differential affinity to dirty/clean water compared to WT mosquitoes.	2	yes	1.1	1.1	2
26. Laboratory colonies infected with plasmodium, not noticed or unable to respond properly or they have enhanced vectorial capacity. Workers in laboratory are infected.	2	no	2	1.1	2
27. Treatment of malaria carrying workers in laboratory. If colonies infected but not destroyed just prior to release.	2	yes	2	1.1	2
28. Infected worker does not disclose their infection.	2	yes	2	1.1	2
29. Antibiotics and hormones used in agricultural production could differentially affect wild-type and GM mosquito survival.	3	yes	1	1.1	3
30. If construct alters vectorial capacity, then change in seasonal levels of plasmodium could affect human health.	3	yes	1	1.1	3
31. Biological control via predation (i.e. Gambusia, fungus) is targeted around wild type, but GM mosquito has different predation risk.	3	yes	1	1.1	
32. Transgene results in improved vectorial capacity.	3	yes	1	1.1	3
33. Differential fitness (fecundity, movement, time to maturity etc.) so that less susceptible to predation leading to increased fitness. Competition between GM and WT mosquitoes for larval habitat - laboratory conditions select for phenotype or genotype that do better in certain conditions (i.e. dirty water) or in high density populations. Selection over time resulting in genotype persistence.	3	yes	1	1.1	3

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
34. Possible route to toxicity if the construct cuts ribosomal repeats in the egg. An off-target cut in mosquito could create mutation which destroys mosquito immune system and somehow increases vector capacity, or changes behaviour so that bites more. (Increase fitness of harmful and decrease fitness of beneficial).	3	yes	1	1.1	3
35. Changes in agriculture practices favour or disfavour GM mosquitoes (i.e., maze pollen falling into larval habitats alters success of GM larvae). Differential preference of wild type or GM type for agriculture lands. GxE interaction, e.g., perhaps in different environment there is a change in expression of gene in mosquito larvae.	3	yes	1.1	1.1	3
36. Abundance seasonally affected. GM mosquitoes might have different predator-prey responses, generally this would decrease their survival and diminish their risk, but there could be the opposite (i.e., fly lower less bat predation).	3	yes	1.1	1.1	3
37. Change in life expectancy increases vectorial capacity.	3	yes	1.2	1.1	3
38. Differential feeding or biting preference of GM mosquitoes and result on health of human sub-populations.	3	yes	1.3	1.1	3
39. GM mosquitoes have differential susceptibility to pathogens or vectors than WT mosquitoes.	3	yes	2	1.1	3
40. Outcomes of 38 and 42 might result in a governance response, a change in vector control program, or treatment of disease in human population leading to increase in human disease.	3	yes	9	1.1	3
41. Gene expression of I-Ppol or fluorescence changes some fitness component or disease transmission in GM mosquitoes. Fluorescence and interaction with other animals. Multiple mating behaviour; males have trouble emerging from pupae cases. Change in function of vision/fertility due to fluorescent marker. Increased effect of sexual maturity. Changes to immune system resulting in differential transmission.	3	yes	1.1, 2, 10	1.1	3
42. Different susceptibility in human sub-populations due to clothing, or preference of GM mosquitoes.	3	yes	1.3,7	1.1	3

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
43. Changes in abundance of plants associated with mosquitoes. If GM mosquitoes have different affinities for certain plants or if there is different levels of human activities or vector control around these plants, could lead to change in mosquito interactions with people and environment (i.e., shade, spraying programs, availability of flowers as food source, and breeding sites).	3	yes	7, 9	1.1	3
44. Seasonal weather patterns (i.e., wet versus dry season) leads to heightened transmission of malaria near insectaries.	4	no	1	1.1	4
45. Other insectaries in cities raising AG mosquitoes for other studies. Escapee mosquitoes from one of these could increase incidence of malaria near project insectaries.	4	no	1.1	1.1	4
46. Temperature may affect biting rates, wind movement disease transmission and spatial distribution.	4	no	1.3	1.1	4
47. False media report regarding mosquito-plant interaction changes agriculture or horticulture practices leading to change in exposure.	4	no	7	1.1	4
48. False reporting on mosquito predators changes human behaviour and exposure.	4	no	7	1.1	4
49. Atypical weather affecting human behaviour and thus change in exposure.	4	no	7	1.1	4
50. Seasonal change in people moving in or out of cities. Seasonal change in behaviour such as windows open/closed.	4	no	7	1.1	4
51. Changes in community population over week (i.e. market day). Behaviour of individuals or groups, which affects exposure (acquiring a TV, alcohol etc.).	4	no	7	1.1	4
52. People displaced by natural disasters - more vulnerable or exposed. Bring malaria into a new area. Ability to fight malaria diminished as a result of nutritional status following natural disaster.	4	no	7	1.1	4
53. Urban population have higher susceptibility to malaria due to lack of exposure or resistance.	4	no	7	1.1	4
54. False reporting leads to change in vector control or human behaviour.	4	no	9	1.1	4

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
55. Insectaries with different levels of exposure to public or road access. Fewer predators in urban environment. Buildings can be more mosquito proof in urban environment. Not sure that 5000 - 10000 AG would be noticeable in low background numbers in urban environment.	4	no	9	1.1	4
56. Change in crops leads to change in breeding habitat, which changes mosquito population abundance and species composition.	none	no	1.1	1.1	5
57. Fertilization increasing nutrient runoff, which enriches larval food resources.	none	no	1.1	1.1	5
58. Change in vegetative cover changes levels of competitive interactions.	none	no	1.1	1.1	5
59. Agriculture could change economic status and thus health of locals changing leading to change in exposure.	none	no	7	1.1	5
60. More exposure and vulnerability in periods of unrest.	none	no	7	1.1	5
61. When hot don't like using bed nets. May not be using bed nets as Culex doesnt transmit malaria.	none	no	7	1.1	5
62. Background change to environment alters population distribution and abundance. Change in lighting (i.e., street or site lighting at night) could alter insects (moths) or, which alters their predators (bats) and prey. Change in economic standards changes exposure and health status of people.	none	no	1, 7	1.1	5
63. Differential public response arising from an external event involving other nuisance or disease vectors leading to change in human behaviour that causes increased susceptibility to malaria.	none	no	7, 9	1.1	5
64. Importation of livestock introduces novel pathogens and exposes escapee mosquitoes to these novel pathogens.	1	yes	3	1.2	6
65. More/different animals present in urban environment which could result potential differential survival, and interaction with different animals and potential transmission of alternative pathogen (bacteria/virus).	1	yes	3	1.2	6

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
66. Differential exposure: fishermen or children that are playing or living around animals, or poor quality housing near insectary become co-infected with other pathogen and leading to increased exposure of escapee mosquitoes to novel pathogens. Also potential for a differential individual susceptibility to spraying post escape or release.	1	yes	7	1.2	6
67. Does transgene change preference to food source (i.e. biting habit, or feeding on different animals)? Change exposure to other pathogens, makes unpredictable.	3	yes	1.3, 3	1.2	7
68. GM mosquitoes have differential susceptibility to pathogens or vectors than WT mosquitoes.	3	yes	3	1.2	7
69. GM mosquitoes able to transmit previously novel cattle pathogen.	3	yes	3	1.2	7
70. HGT to gut micro-biota in humans if ingested.	3	yes	5	1.2	7
71. Local depletion of invertebrates from intensive spraying. This loss of invertebrates allows GM mosquitoes to bite other vertebrates and increases transmission of extant diseases.	3	yes	10, 3	1.2	7
72. Presence of transgene changes microbiome of mosquitoes. Mosquitoes pick up human, plant or animal pathogens in laboratory, become new (or locally novel) vectors inside or outside of laboratory.	3	yes/no	3	1.2	7
73. Change in immunogenicity of GM mosquitoes causes allergic reaction in humans (more than WT mosquitoes) from biting; toxic if ingested.	3	yes		1.3	8
74. In spraying for escape could impact bee colonies. Also spraying and larvicide impacts.	1	yes	8	1.4	9
75. Local economic loss due to perceived risk to potential new businesses and thus limited investing in region.	4	no		1.4	9
76. Local spraying reduces populations of pollinating insects, which reduces pollination of local floral.	2	yes	8	2.1	10

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
77. Wild type colonies are insecticide resistant, or change in behaviour (i.e., less anthropophilic, change in egg-laying, swarming, etc.). This causes humans to change their behaviour, which affects vector control practices. If there is an ensuing increase in the intensity of the local spraying program then this could lead to reduction in local invertebrate populations.	2	yes	7, 8, 9	2.1	10
78. If GM mosquitoes change their biting habit, then impact of vector control could be changed. If release causes dramatic suppression of wild mosquito populations, could have empty niche which has knock-on effects to ecosystem.	3	yes	1.3, 8	2.1	11
79. Change in use of building by humans leads to increased biting by GM mosquitoes and subsequently increased spraying intensity, which could suppress local invertebrate populations.	3	yes	1.3, 8	2.1	11
80. VGT eradicates other members of species complex.	3	yes	5	2.1	11
81. Possible fitness change from GM, but largest effect could follow from total number released. A large release might have strong local effect that leads to HGT.	1	yes	5	3.1	12
82. Potential for HGT with fungi.	2	no	5	3.1	12
83. HGT involving vertebrates and other invertebrates.	2	yes	5	3.1	12
84. Laboratory rearing practices may affect the fitness, which facilitates HGT to another organism.	2	yes	5	3.1	12
85. Humidity in soil could affect persistence of DNA. Leading to (for example) increased exposure of bacteria to transgene and increase in HGT risk.	3	no	5	3.1	12
86. Remobilisation and transposable element. Shown remobilisation of piggyback in Anopheles. Potential for remobilisation due to exogenous transposase. Construct migrate to different part of mosquito genome.	3	no	5	3.1	12
87. HGT issue from biological control agent consuming GM larvae.	3	yes	5	3.1	12

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
88. Seasonal differences lead to different potential for HGT (likely greatest risk in wet season, but in dry season great relative effect of GM mosquitoes with respect to wild population).	3	yes	5	3.1	12
89. Soil bacteria mediated HGT.	3	yes	5	3.1	12
90. Bodies of GM mosquitoes end up in soil, possible start of HGT.	3	yes	5	3.1	12
91. HGT leading to sterility, fitness costs to non-target eukaryotes and prokaryotes	3	yes	5	3.1	12
92. Predator switching leading to focus on consuming GM mosquito (including larvae), which increases chance for HGT.	3	yes	5	3.1	12
93. Topography may channel wind. Wind speeds different at different altitudes resulting in a change in dispersal or concentration, with flow on affect to monitoring and control. Exogenous DNA in soil into dust storms potentially dispersed over very large areas and wide range of habitats, thus increasing potential for HGT.	3	yes	5	3.1	12
94. Bio-accumulation influenced by shrinking water bodies increases potential for HGT.	3	yes	5	3.1	12
95. HGT - temperature might affect degradation rates.	3	yes	5	3.1	12
96. Bio-accumulation due to concentration of dead GM mosquitoes in urban environment, leads to increased potential for HGT.	3	yes	5	3.1	12
97. Food and feed production or other changes that alter local environment leading to HGT in microbial community.	4	no	5	3.1	12
98. Prolonged cycles of drought, locus swarms, bush fire. Change to environment resulting in change to persistence - impact on modelling.	1	yes	1	3.2	13
99. Civic, holidays or religious observances, adverse weather can influence abilities to respond to an escape. Also questions around circadian and lunar cycles affecting monitoring efficiency.	1	yes	7	3.2	13

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
100. Laboratory rearing process likely to select for a number of life history attributes such as fast emergence or assimilation of laboratory diet but not wild diet; these changes could alter their fitness in wild environments, and the change could be positive or negative (most likely negative from a wide range of examples such as fruit flies) depending on receiving environment (e.g.,. laboratory raised males are selected for rapid development, but are less competitive for wild females, or they have lower parasite load). These differences could decrease our ability to effectively monitor or interpret responses.	2	yes	1	3.2	14
101. Laboratory reared mosquitoes (especially those with constructs) change their behaviour with relationship to plants (i.e., male swarms or refuge). Olfactory or pollination behaviour changes make them difficult to monitor.	2	yes	4	3.2	14
102. HGT with respect to bee keeping. This or any source of fluorescent dyes in invertebrate populations would confound monitoring programs. Industrial sites that create breeding sites could affect background population dynamics of wild or escaped mosquitoes.	3	no	1, 5	3.2	15
103. Atypical weather affecting human behaviour and thus change in exposure and change in mitigation/monitoring approaches.	4	no	7	3.2	16
104. Vegetative cover limits access to sites for background monitoring.	4	no		3.2	16
105. Disruption of one access route due to criminal activity (etc.) overlapping with flood (etc.) precludes access to monitoring sites. Perhaps insectaries are more vulnerable after storms due to breakdown of services.	4	no		3.2	16
106. Monitoring compromised in urban environment due to inability to capture, or to use capture methods such as light traps, or to capture swarms when in built urban environment.	4	no		3.2	16
107. Change in background mosquito population changes our ability to monitor and control.	4	no	1.1	3.2, 3.3	17
108. Governance structure could affect nature or ability to respond.	4	yes	7	3.2, 3.3	17

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
109. War, criminal activity: might lead to release (or alleged release) in an area that we are not prepared to react to. Could also lead to harassment of project.	4	yes	9	3.2, 3.3	17
110. Vegetative cover affects access and monitoring efficiency and vector control following a release.	4	yes	9	3.2, 3.3	17
111. Compromised monitoring and control due to black market on GM mosquitoes, insecticide, monitoring equipment, traps etc. Biosecurity compromised by attempt/actual theft inside building.	4	yes	9	3.2, 3.3	17
112. Breakdown of local services might force shut down of laboratory, or change local levels. Breakdown or failure of health and safety procedures or biosecurity procedures, leads to release to unknown distant sites.	4	yes		3.2, 3.3	17
113. Insectaries not accessible due to emergency unable to execute emergency measures. Not present to detect escapes.	4	yes		3.2, 3.3	17
114. Differential access to areas in urban environment resulting in compromised monitoring or spraying.	4	yes		3.2, 3.3	17
115. Topography may channel wind. Wind speeds different at different altitudes resulting in changed dispersal or concentration with flow on affect to monitoring and control.	4	yes		3.2, 3.3	17
116. Seasonal change and signals in local resistance due to seasonal use of insecticides in agriculture leads decreased effectiveness of vector control efforts by project.	3	no	9	3.3	18
117. Use of insecticides in agriculture could cause local population to be resistance - if these local WT mosquitoes used as backcross could create insecticide resistance in GM mosquitoes.	3	no		3.3	18
118. Change in insecticide application changes local invertebrate populations. Could also lead to increased or decreased pesticide resistant invertebrates (including mosquitoes), leading to differential susceptibility of wild type and GM mosquitoes to insecticides.	3	yes	6	3.3	18

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
119. Lack of effective control may lead to bio-accumulation. Inappropriate storage of insecticide spray leads to ineffective control. Insecticide resistance due to backcross with WT mosquitoes. Cross between local WT mosquitoes and source population, from which there has been an interaction in local genetics. If there is such a change then backcross may change gene expression. From Baeshan et al. 2014: inbreeding and selection differentially affect laboratory strains over time and that heterotic 'super males' could be used to rescue some male reproductive characteristics.	3	yes	9	3.3	18
120. Wild type colonies are insecticide resistant, or change in behaviour (less anthropophilic, change in egg-laying, swarming, etc.). In response, humans change their behaviour, which decreases effectiveness of vector control program.	4	no	6	3.3	19
121. Stealing insecticide from insectaria for agricultural use, so not available if required. Compromised monitoring quality due to access and opinions within community. Government or community response to release being overreaction and spraying of insecticide (e.g., DDT).	4	no	9	3.3	19
122. Mali insectaries on a hillside, but all others are on flats. Buildings and city structures an issue for spraying and access.	4	no		3.3	19
123. Hard to implement control in adverse weather. Greater breeding sites after storms and in wet season.	4	no		3.3	19
124. Greater uncertainty about status of local conditions or population near insectaries compared to potential release sites, leading to less effective control programs.	4	no		3.3	19
125. Increased risk of non-native plant species or food source near transport hubs could result in persistence of released males.	1	yes	1	3.4	20
126. Change in background mosquito population changes spread of transgene.	2	no	1.1	3.4	21
127. Knock out mutation in fluorescent gene or in enzyme causing male sterility, thus leading to loss of male sterility.	2	no		3.4	21
128. Negligent/intentional actions from workers resulting in unknown status (sterile versus not sterile) of mosquito within insectaria.	2	no		3.4	21

Hazard Scenario	Project event	Release required	Mediating event or mechanism	Impact	Summary statement
129. Laboratory culture alters assortative mating and heterosis (hybrid vigour), leading to increased VGT.	2	no	5	3.4	21
130. As before, a change in reproductive behaviour leads to change in swarm behaviour leading to change in spread of transgene.	2	yes	10	3.4	21
131. More control near laboratory. Or locals believe that there is less need for vector control. In unauthorized release there would be local spraying, but if planned release there could be change or reduction in vector control(NB: no discernible impact recorded).	2	no	9	none	22
132. Could have fitness competition with target or non-target mosquitoes. Wild types are introduced into laboratory and introduce unknown genes or gene sequences (e.g., viral sequences) to colonies (NB: no discernible impact recorded).	2	no		none	22
133. Flooding influences larval dispersal. Storms could knock out electricity, and backup drains, which diminishes laboratory procedures (NB: no discernible impact recorded).	2	no		none	22
134. Chemical mutagens have differential effect on GM mosquitoes (NB: no discernible impact recorded).	3	no		none	22
135. Temperature change/extreme potentially destabilise protein. Degradation of construct - stage 2/3 construct. GxE interactions (NB: no discernible impact recorded).	3	no		none	22
136. Minute weight of 5000-10000 GM mosquitoes and associated protein (NB: no discernible impact recorded).	3	yes		none	22
137. GxE interactions arising effects of physical processes—i.e., wind & water movement, storms, weather, temperature, etc. (NB: no discernible impact recorded).	3	yes		none	22
138. Governance structure more/less responsive to complaints about mosquitoes due to introduction of a mutagen into the local urban environment (NB: no discernible impact recorded).	4	no	9	none	22
139. Earthquake result in potential short term increase in background radiation leading to greater mutation (NB: no discernible impact recorded).	4	no		none	22

Appendix C Fault Tree Analysis

C.1 Fault tree construction

This Appendix provides a brief overview of the methods of Fault Tree Analysis (FTA). It provides a basic introduction to the construction and quantification of a fault tree using a simple hypothetical example of missing a bus, and highlights the issues that arise when basic events within the tree are dependent.

A fault tree is a graphical model of the causal chain that leads to a system failure, defined by the top event. The causal chain is drawn as a tree with the branches connected by gates that represent one of a number of possible logical functions. Many logical functions are used in fault tree analysis, but in practise the two must important gates are the "AND" gate and "OR" gate.

The "OR" gate represents logical disjunction of the events below it. An OR gate can have any number of inputs (branches) running into it. The event above the gate is realised if any one of the inputs are true. The "AND" gate represents the logical conjunction of events below it. An "AND" gate can also have any number of inputs running into it, but the event above the gate is only realised if all the inputs are true. The tree terminates with a set of basic events. In engineering contexts these are Boolean events, taking one of only two values such as true/false, open/closed or success/failure, that are not developed further. If the probability of the basic events in a tree can be determined, then the probability of the top event can also be calculated using standard rules of probability, but taking care to account for the dependency and correct conditioning between the events within the tree.

Figure C.1 shows a fault tree constructed for the hypothetical event of missing a bus. Gates in the tree are marked as grey rectangular boxes, and basic events are shown in blue. "OR" gates are indicated by, crescent moon symbols below the grey boxes, whilst "AND" gates are indicated by dome shaped symbols.

Fault trees like this are typically constructed iteratively with domain experts. The trees are a graphic model of the way in which things can go wrong, and as with any model they do not represent the truth but rather, if constructed carefully, a useful abstraction of it – or more succinctly *"Essentially, all models are wrong, but some are useful"*(Box and Draper, 1987).

C.2 Quantification of the tree

Fault tree analysis can be used to quantify the probability of the top event because there is a one-to-one relationship between the logical operations "AND" and "OR" and the basic probability laws for union and intersection:

$$\begin{aligned} \text{UNION} \quad & \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \\ \text{INTERSECTION} \quad & \Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(B), \end{aligned} \tag{C.1}$$

where $\Pr(A|B)\Pr(B) = \Pr(A)\Pr(B)$ if events A and B are independent, and $\Pr(AB) = 0$ if A and B are mutually exclusive.

Fault tree calculations are trivial in the special case that all the events that lead into a gate are unique (there are no repeat events) and independent. Dependency between events in the tree, however, introduces additional complexity into these calculations.

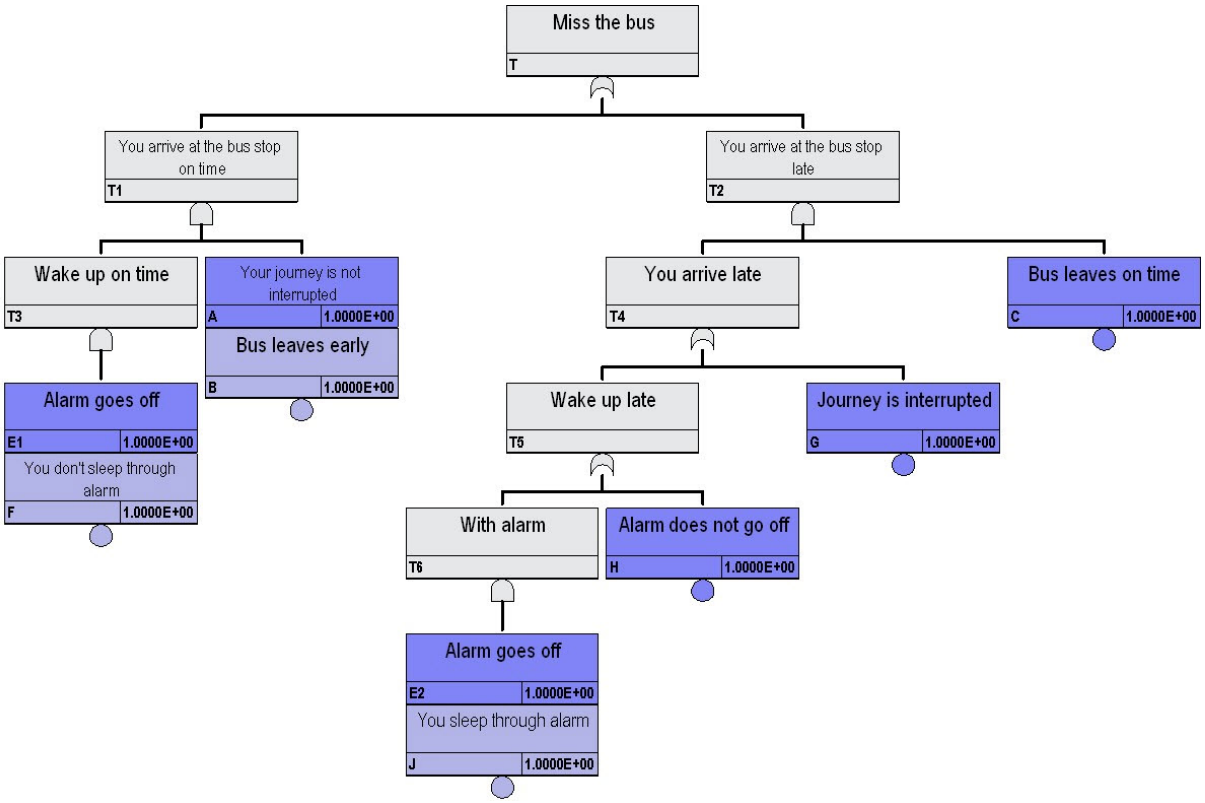


Figure C.1: Hypothetical fault tree for the top event “Missing the bus”. The tree contains a number of mutually exclusive events (events that cannot both be true at the same time), and one repeat event (the alarm goes off). Repeat events and mutually exclusive events are examples of dependent events

The dependency introduced by repeat events and mutually exclusive events is usually eliminated from the calculations by presenting the fault tree equation as the sum of minimum cut sets. Cuts sets are a unique set of basic events that together can cause the top event to occur. The members of a cut set are the basic events that in union will cause the top event. A minimum cut set is a cut set with no superfluous elements – i.e. if you remove one or more events from a minimum cut set, the set will no longer cause the top event to occur. Typically a fault tree will have several minimum cut sets, anyone of which will cause the top event, and large trees can have hundreds.

Figure C.2 shows one way to identify the cut sets for the hypothetical "miss the bus" fault tree. Working from the top event down, each "OR" gate in the tree generates a new cut set, whilst each "AND" gate adds a new member to an existing set determined by the precedence imposed by the tree structure. In the "miss the bus" tree there are four cut sets of order 2, 2, 3, 4, where the order of the cut set refers to the number of elements within the set.

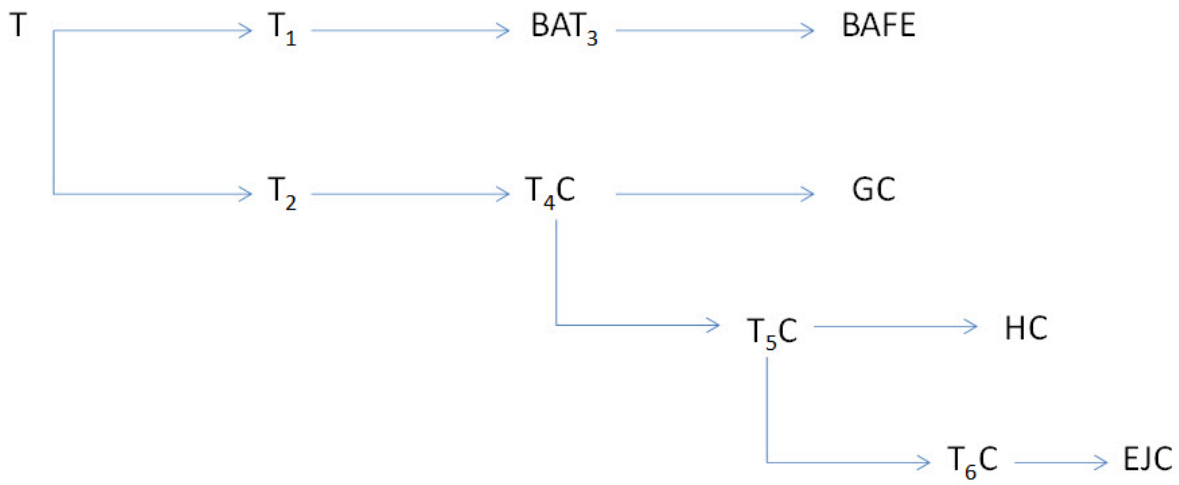


Figure C.2: Diagram showing how the cut sets of the "miss the bus" fault tree are constructed. Each "OR" gate starts a new cut set, each "AND" gate adds a member to an existing cut set.

The main motivation for identifying the cut sets of a fault tree is to present the fault tree equation in a sum-of-products form that can be simplified using the laws of Boolean Algebra. For Boolean basic events, the Law of Absorption reduces the cut sets down to the minimum cut sets. The Idempotent Law eliminates repeat events, and mutually exclusive events are eliminated by the complement $A.\bar{A} = 0$ where the dot operator . stands for logical union.

The number and order of the cut sets also provide important clues to the most likely ways the top event may occur. Low order cut sets, for example, identify the most likely failure pathways in a fault tree, and cut sets with order 6 or more are often dropped from the overall calculation because (in the context of engineering systems) their contribution to the top probability is typically negligible (Ericson, 2011).

Once the fault tree equation has been presented in a simplified sum-of-products form the probability of the top event can be computed using the exclusion-inclusion rule which for a sum-of-product expression with n terms is given by

$$
\begin{aligned}
\Pr\left(F(X_i)\right) = &+ \sum_{i=1}^{n} \Pr\left(X_i\right) && \text{single terms} \\
&- \sum_{\substack{i=2^{n-1} \\ i \neq j}} \Pr\left(X_i \cap X_j\right) && \text{first order intersections} \\
&\vdots \\
&+ (-1)^{n+1} \Pr\left(X_1 \cap X_2 \cap \cdots X_n\right) && \text{nth order intersections}
\end{aligned} \tag{C.2}
$$

The big advantage of the minimum cut set approach is that is reduces the number of terms in the fault tree equations whilst simultaneously accounting for the repeat events and mutually exclusive events. The disadvantage from our perspective, however, is that information at the gate is lost. For this reason we implemented an alternative calculation method.

In this analysis we used the fault tree analysis software SAPHIRE https://saphire.inl.gov/ to draw the fault trees. Once the tree is drawn in SAPHIRE, one of the fault tree reporting options is used to export the Fault Tree Logic as text. Table C.1, for example, shows the output for the miss the bus example.

In the case where the tree had no repeated events we converted this text file into a symbolic list objects using the Ryacas package http://cran.r-project.org/web/packages/Ryacas/index.html and then sequentially implemented the probability laws for union and intersection through the tree, using a list of repeat events (Appendix D) within each tree to identify, and eliminate squared terms of repeat events in the expanded form of the equations. Squared terms of repeat events are eliminated on the grounds that $\Pr\left(A\right)\Pr\left(A\right) = \Pr\left(A, A\right) = \Pr\left(A\right)$. Our analysis contains no mutually exclusive events, but if it did these could be handled in a similar fashion.

The big disadvantage of this approach is that the size of the fault tree equation grows dramatically through the tree. This effect can be seen in Table C.1, and this effect is accentuated when the equations are expanded which is necessary in order to identify squared terms of repeat events.

To handle this computational difficulty we used two approaches. Firstly, the code identified independent subtrees (that is tree sections that contained no repeated events) and left these in factored (unexpanded) form as the equation was built up the tree. Secondly, for very large trees, the standard practice of setting terms in the expansion that are the product of more than 10 probabilities to zero was also used.

T	OR	T_1, T_2	$1-(1-\Pr(T_1))(1-\Pr(T_2))$
T_1	AND	T_3, A, B	$1-(1-\Pr(T_3AB))(1-\Pr(T_2))$
T_3	AND	E, F	$1-(1-\Pr(EFAB))(1-\Pr(T_2))$
T_2	AND	C, T_4	$1-(1-\Pr(EFAB))(1-\Pr(CT_4))$
T_4	OR	T_5, G	$1-(1-\Pr(EFAB))(1-\Pr(C)[1-(1-\Pr(T_5))(1-\Pr(G))])$
T_5	OR	T_6, H	$1-(1-\Pr(EFAB))(1-\Pr(C)[1-(1-(1-(1-\Pr(T_6))(1-\Pr(H))))(1-\Pr(G))])$
T_6	AND	E, J	$1-(1-\Pr(EFAB))(1-\Pr(C)[1-(1-(1-(1-\Pr(EJ))(1-\Pr(H))))(1-\Pr(G))])$

Table C.1: Step wise compilation of the Miss the bus fault tree in factored (unexpanded) form

Appendix D Fault tree repeat events

Table D.1: Fault Tree dependencies

Question	Basic Event	Basic Event	Basic Event	Basic Event
What is the probability that a mosquito from the escaped population, in a year, will contact a human or other vertebrate infected with a pathogen not previously known to be vectored by Anopheles gambiae?	FT2000	FT2110	FT2100	
Given that a mosquito has contacted an infected human or other vertebrate, what is the probability that it acquires a novel pathogen through a blood meal?	FT2001	FT2111		
Given that a mosquito has survived the pathogen's incubation period what is the probability that it bites a subsequent human or vertebrate?	FT2022	FT21021		
Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters the soil environment?	FT300000	FT301000	FT30201000	FT4101000
Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in the soil?	FT300001	FT301001	FT30201001	FT4101001
Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters an aquatic environment?	FT300010	FT301010	FT30201010	FT4101010
Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in an aquatic environment?	FT300011	FT301011	FT30201011	FT4101011
Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct enters the gut of a eukaryote organism?	FT300020	FT301020	FT30201020	FT4101020

Given that the I-PpoI construct has entered the gut of a eukaryote organisms what is the probability that the construct DNA remains intact?	FT300021	FT301021	FT30201021	FT4101021
Given a complete release of all 10,000 GM insectary mosquitoes, how many already are or will be infected with a mosquito virus?	FT302000	FT41000		
Direct acquisition	FT3020010	FT410010		
Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the construct is excised from the mosquito genome by a transposase?	FT30200100	FT4100100		
Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, and the construct is excised from the genome via a transposase, what is the probability that the virus incorporates the construct into its own genome?	FT30200101	FT4100101		
Direct acquisition	FT302011120	FT41011120		
Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?	FT30200111	FT4100111		
Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?	FT3020111211	FT410111211		
Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?	FT30200112	FT4100112		
Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?	FT3020111212	FT410111212		
Given that a mosquito virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?	FT302002	FT41002		

Given that there is intact I-PpoI construct in the soil, aqueous or gut environment, what is the probability that a cellular organism is transformed with construct?	FT30201110	FT4101110
Given that a cellular organism has been transformed with the construct, what is the probability that it is infected with a virus?	FT30201111	FT4101111
Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct is stably acquired by a prokaryote organism by transformation in an aquatic environment?	FT4001	FT30101
Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?	FT30200110	FT4100110
Given that a mosquito RNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?	FT302003	FT41003
Given that a mosquito DNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?	FT302004	FT41004
Given that there is intact I-PpoI construct in either soil, aqueous or gut environment, what is the probability that a competent bacteria is transformed with the construct?	FT30201100	FT4101100
Given that a competent bacteria has been transformed with the construct, what is the probability that it is infected with a bacteriophage?	FT30201101	FT4101101
Given that transformed bacteria is infected with a bacteriophage, what is the probability that the bacteriophage incorporates the construct into its genome?	FT3020110200	FT410110200

Given that transformed bacteria is infected with a bacteriophage, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?	FT410110210	FT3020110210	
Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified bacteriophage?	FT410110211	FT3020110211	
Given that a bacteriophage has acquired the construct into its genome, what is the probability that it remains able to replicate (autonomously or non-autonomously)?	FT4101103	FT30201103	
Given that transformed cellular organism is infected with a virus, what is the probability that the virus incorporates the construct into its genome?	FT3020111200	FT410111200	
Given that transformed cellular organism is infected with a virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?	FT3020111210	FT410111210	
Given that a virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?	FT30201113	FT4101113	
Given that the I-PpoI construct has entered the germline cell of a eukaryote what is the probability that it gets into the nucleus?	FT3001	FT30112	
Given the I-PpoI construct has entered the nucleus of a germline cell of a eukaryote what is the probability it recombines into the eukaryote's genome?	FT3002	FT30113	
What is the probability that there are compatible species (i.e. species that could mate with GM insectary mosquitoes) in the vicinity of the insectary?	FT51000	FT51102	FT51113
Given that there are compatible species in the vicinity, what is the probability that hybrids will be formed, following a complete release of all 10,000 mosquitoes?	FT51001	FT51103	FT51114

Given that there are compatible species in the vicinity, what is the probability that Coluzzii hybrids will be formed, following a complete release of all 10,000 mosquitoes?	FT510011	FT511031	FT511141	
Given that there are compatible species in the vicinity, what is the probability that Arabiensis hybrids will be formed with these species, following a complete release of all 10,000 mosquitoes?	FT510010	FT511030	FT511140	
Given that the construct has been transmitted to the Coluzzii F1 offspring, what is the probability that they are viable?	FT510031	FT511051	FT511161	
Given that the construct has been transmitted to the Arabiensis F1 offspring, what is the probability that they are viable?	FT510030	FT511050	FT511160	
Given that GM males carrying the construct are viable, what is the probability that the construct allows some fertile males (not all are sterilised)?	FT500000	FT500010		
Given that mitochondrial DNA has acquired the construct, what is the probability that Wolbachia is present in the mosquito?	FT51111	FT51100	FT50311	FT50300
Given that the construct has been transmitted to the F1 offspring, what is the probability that they are viable?	FT51003	FT51105	FT51116	
Given that hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?	FT51104	FT51115	FT50302	FT50313
Given that Arabiensis hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?	FT511040	FT511150		

Given that Coluzzii hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?	FT511041	FT511151
Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?	FT501100312	FT501102312
Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?	FT51011100312	FT51011102312
Given that Wolbachia is present, what is the probability that Wolbachia acquires the construct?	FT51101	FT50301
Given that the F1 females carrying the modified Wolbachia are fertile, what is the probability that the modified Wolbachia will spread in the recipient species?	FT51107	FT50303
Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that mitochondrial DNA acquires the construct?	FT51110	FT50310
Given that Wolbachia is present, what is the probability that Wolbachia acquires the mitochondrial DNA with the construct?	FT51112	FT50312
Given that the modified F1 females carrying the Wolbachia and the transformed mitochondria are fertile, what is the probability that the particular combination of Wolbachia and construct bearing mitochondria will spread in the recipient species?	FT51118	FT50314
Given that hybrids carrying the construct are viable, what is the probability that I-PpoI mutates and recognises the site where it is already located?	FT510111010	FT5011010

Given that I-PpoI has mutated to recognise the site that it is in, what is the probability that it maintains its germline expression?	FT510111012	FT5011012	
Given that I-PpoI has mutated to recognise the site that it is in and maintains germline expression, what is the probability that its cleavage rate at the rDNA repeat does not preclude spread through the population?	FT5101110130	FT50110130	
Low fitness effects at original target site	FT5101110131	FT50110131	
What is the probability that the recognition site is in an Intron?	FT51011101310	FT501101310	
What is the probability that the recognition site is in a non-essential region of the genome?	FT51011101311	FT501101311	
What is the probability that the negative fitness effects are recessive?	FT51011101312	FT501101312	
Given that all GM males carrying the construct are sterile, what is the probability that the construct mutates to increases female fitness sufficiently to compensate for the male sterility?	FT50010	FT5101010	
Given that An. arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?	FT510111120	FT5101110040	FT5101110240
Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?	FT510111121	FT5101110041	FT5101110241
Given that hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?	FT51010000	FT51010012	
Given that Arabiensis hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?	FT510100000	FT510100120	

Given that Coluzzii hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?	FT510100001	FT510100121

Appendix E Fault tree basic events and gates

Table E.1: Fault Tree Questions

Basic Event or Question Gate number	
FT1-1a	Number of bites per day on humans by female mosquitoes (WT)
FT1-1b	Number of bites per day on humans by female mosquitoes (G3)
FT1-1b-factor	Number of bites per day on humans by female mosquitoes (G3)
FT1-1c	Number of bites per day on humans by female mosquitoes (I-PpoI)
FT1-1c-factor	Number of bites per day on humans by female mosquitoes (I-PpoI)
FT1-2a	Proportion of bites from infectious human that transmit infectious agent to mosquito (WT)
FT1-2b	Proportion of bites from infectious human that transmit infectious agent to mosquito (G3)
FT1-2b-factor	Proportion of bites from infectious human that transmit infectious agent to mosquito (G3)
FT1-2c	Proportion of bites from infectious human that transmit infectious agent to mosquito (I-PpoI)
FT1-2c-factor	Proportion of bites from infectious human that transmit infectious agent to mosquito (I-PpoI)
FT1-3a	Proportion of bites from infectious mosquitoes that transmit infectious agent to humans (WT)
FT1-3b	Proportion of bites from infectious mosquitoes that transmit infectious agent to humans (G3)
FT1-3b-factor	Proportion of bites from infectious mosquitoes that transmit infectious agent to humans (G3)

FT1-3c	Proportion of bites from infectious mosquitoes that transmit infectious agent to humans (I-PpoI)
FT1-3c-factor	Proportion of bites from infectious mosquitoes that transmit infectious agent to humans (I-PpoI)
FT1-4a-lifespan	Mortality rate of mosquitoes once outside of insectary (WT_male)
FT1-4a-mortality	Mortality rate of mosquitoes once outside of insectary (WT_male)
FT1-4a-survival	Mortality rate of mosquitoes once outside of insectary (WT_male)
FT1-4b-lifespan	Mortality rate of mosquitoes once outside of insectary (WT_female)
FT1-4b-mortality	Mortality rate of mosquitoes once outside of insectary (WT_female)
FT1-4b-survival	Mortality rate of mosquitoes once outside of insectary (WT_female)
FT1-4c-factor	Mortality rate of mosquitoes once outside of insectary (G3_male)
FT1-4c-lifespan	Mortality rate of mosquitoes once outside of insectary (G3_male)
FT1-4c-mortality	Mortality rate of mosquitoes once outside of insectary (G3_male)
FT1-4c-survival	Mortality rate of mosquitoes once outside of insectary (G3_male)
FT1-4d-factor	Mortality rate of mosquitoes once outside of insectary (G3_female)
FT1-4d-lifespan	Mortality rate of mosquitoes once outside of insectary (G3_female)
FT1-4d-mortality	Mortality rate of mosquitoes once outside of insectary (G3_female)
FT1-4d-survival	Mortality rate of mosquitoes once outside of insectary (G3_female)
FT1-4e-factor	Mortality rate of mosquitoes once outside of insectary (I-PpoI_male)

FT1-4e-lifespan	Mortality rate of mosquitoes once outside of insectary (I-PpoI_male)
FT1-4e-mortality	Mortality rate of mosquitoes once outside of insectary (I-PpoI_male)
FT1-4e-survival	Mortality rate of mosquitoes once outside of insectary (I-PpoI_male)
FT1-4f-factor	Mortality rate of mosquitoes once outside of insectary (I-PpoI_female)
FT1-4f-lifespan	Mortality rate of mosquitoes once outside of insectary (I-PpoI_female)
FT1-4f-mortality	Mortality rate of mosquitoes once outside of insectary (I-PpoI_female)
FT1-4f-survival	Mortality rate of mosquitoes once outside of insectary (I-PpoI_female)
FT1-5a	Number of fertile eggs laid by blood fed female (following WT male mating) (WT_female)
FT1-5b	Number of fertile eggs laid by blood fed female (following WT male mating) (G3_female)
FT1-5b-factor	Number of fertile eggs laid by blood fed female (following WT male mating) (G3_female)
FT1-5c	Number of fertile eggs laid by blood fed female (following WT male mating) (I-PpoI_female)
FT1-5c-factor	Number of fertile eggs laid by blood fed female (following WT male mating) (I-PpoI_female)
FT1-6a	Average (over a month) daily dispersal distance (WT_male)
FT1-6b	Average (over a month) daily dispersal distance (WT_females)
FT1-6c	Average (over a month) daily dispersal distance (G3_male)
FT1-6d	Average (over a month) daily dispersal distance (G3_female)
FT1-6e	Average (over a month) daily dispersal distance (I-PpoI_male)

FT1-6f	Average (over a month) daily dispersal distance (I-PpoI_female)
FT1-6f-factor	Average (over a month) daily dispersal distance (I-PpoI_female)
FT1-7a	Probability of biting for two blood meals after release (WT_female)
FT1-7b	Probability of biting for two blood meals after release (G3_female)
FT1-7b-factor	Probability of biting for two blood meals after release (G3_female)
FT1-7c	Probability of biting for two blood meals after release (I-PpoI_female)
FT1-8a	Extrinsic incubation period (number of days from mosquito acquiring infection to becoming infectious) (WT_female)
FT1-8b	Extrinsic incubation period (number of days from mosquito acquiring infection to becoming infectious) (G3_female)
FT1-8c	Extrinsic incubation period (number of days from mosquito acquiring infection to becoming infectious) (I-PpoI_female)
FT1-9a	Percentage of blood meals taken on humans (WT_female)
FT1-9aEast	Percentage of blood meals taken on humans in East Africa (WT_female)
FT1-9aWest	Percentage of blood meals taken on humans in West Africa (WT_female)
FT1-9b	Percentage of blood meals taken on humans (G3_female)
FT1-9c	Percentage of blood meals taken on humans (I-PpoI_female)
FT2	Probability that (non-GM) insectary mosquitoes transmits a novel (not previously known to be vectored by A. gambiae) blood-based pathogen in a year following an escape of all 10,000 mosquitoes
FT20	Biological transmision (via standard transmission route mid gut/salivary gland)

FT200	Mosquito acquires pathogen
FT2000	What is the probability that a mosquito from the escaped population, in a year, will contact a human or other vertebrate infected with a pathogen not previously known to be vectored by Anopheles gambiae?
FT2001	Given that a mosquito has contacted an infected human or other vertebrate, what is the probability that it acquires a novel pathogen through a blood meal?
FT201	Pathogen reaches infectious load
FT2010	Given that a mosquito has acquired a novel pathogen through a blood meal, what is the probability that the pathogen survives all of the mosquito's immune systems (cellular, humoral and RNA interference)?
FT20100-9	Given that a mosquito has acquired a novel pathogen through a blood meal (biological contamination), what is the probability that it survives the digestive enzymes
FT20101-9	Given that a mosquito has acquired a novel pathogen through a blood meal (biological contamination), and has survived the digestive enzymes, what is the probability that the novel pathogen infects (or enters) a gut cell (which is the first barrier)
FT2010GM	How would you change your probabilities for FT2-010 if the 10,000 insectary mosquitoes were all genetically modified with the I-PpoI construct?
FT2011	Given that the novel pathogen has survived the mosquito's immune system, what is the probability that the pathogen replicates in the mosquito?
FT202	Transmits to second human or vertebrate
FT2020	Given that the pathogen has replicated in the mosquito, what is the probability that it travels to the salivary gland? (point of inoculation)
FT2021	Given that a mosquito has acquired a novel replicating pathogen what is the probability that the mosquito survives the pathogen's incubation period?

FT2021GM	How would you change your probabilities for FT2-021 if the 10,000 insectary mosquitoes were all genetically modified with the I-PpoI construct?
FT2022	Given that a mosquito has survived the pathogen's incubation period what is the probability that it bites a subsequent human or vertebrate?
FT2023	Given that a mosquito with a novel replicating pathogen has bitten a second human or vertebrate what is the probability that it transmits an infectious load to this individual?
FT21	Mechanical transmission (via non-standard transmission route)
FT210	Via proboscis (mechanical)
FT2100	What is the probability that a mosquito escaped from an insectary, in a year, will contact a human or vertebrate infected with a pathogen not previously known to be vectored by Anopheles gambiae?
FT2101	Given that an infected human or vertebrate is contacted, what is the probability that a novel pathogen sticks to the proboscis?
FT2102	Transmits to second human or vertebrate
FT21020	Given that the novel pathogen has adhered to the proboscis, what is the probability that the pathogen remains viable between bites?
FT21021	Given that the pathogen has survived on the proboscis, what is the probability that the mosquito bites a second human or vertebrate?
FT21022	Given that a mosquito with a novel pathogen on its proboscis has bitten a second individual, what is the probability that it transmits an infectious load from its proboscis to this individual's bloodstream?
FT211	Via blood transfer
FT2110	What is the probability that a mosquito escaped from an insectary, in a year, will contact an individual infected with a pathogen not previously known to be vectored by Anopheles gambiae?

FT2111	Given that a mosquito has contacted an infected human or other vertebrate, what is the probability that it acquires a novel pathogen through a blood meal?
FT2112	Transmits infectious dose to second human or vertebrate
FT21120	Given a mosquito that has acquired a novel pathogen through a blood meal, what is the probability that the pathogen remains viable between infection and contact with a subsequent individual?
FT21120GM	How would you change your probabilities for FT2-1120 if the 10,000 insectary mosquitoes were all genetically modified with the I-PpoI construct?
FT21121	Blood fed or partially fed mosquito contacts second human or vertebrate
FT211210	Exposure through ingestion
FT2112100	Given that a mosquito has acquired a novel pathogen through a blood meal that remains viable between contacts, what is the probability that it is ingested by a second human or vertebrate?
FT2112101	Given the mosquito has been ingested, what is the probability that the pathogen survives the human or vertebrate digestion system and infects the subsequent individual?
FT211211	Exposure through wound
FT2112110	Given that a mosquito has acquired a novel pathogen through a blood meal that has survived between contacts, what is the probability that the mosquito lands on an open wound of a subsequent human or vertebrate?
FT2112111	Given that the mosquito has landed on the open wound, what is the probability that an infectious dose is delivered to the subsequent individual's bloodstream (e.g., the mosquito is squashed releasing the blood in its gut onto the wound).
FT211212	Exposure through inhalation
FT2112120	Given that a mosquito has acquired a novel pathogen through a blood meal that has survived between contacts, what is the probability that the pathogen is passed through mosquito faeces?

FT2112121	Given that the pathogen has been passed through mosquito faeces, what is the probability that an infectious dose is delivered to the subsequent individual's bloodstream through inhalation or self inoculation through mucosa.
FT22245GM	How would you change your probabilities for FT2-2245 if the 10,000 insectary mosquitoes were all genetically modified with the I-PpoI construct?
FT2-GM-difference	For each of the questions above how would you change your probabilities if the 10,000 insectary mosquitoes were all genetically modified with the I-PpoI construct?
FT3	Spread of construct in non-target eukaryotes over a year following an escape of all 10,000 genetically modified mosquitoes
FT30	Acquisition of construct
FT300	Unmediated acquisition
FT3000	Acquisition environment
FT30000	Soil environment
FT300000	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters the soil environment?
FT300001	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in the soil?
FT300002	Given that there is intact I-PpoI construct in the soil environment, what is the probability that it will enter a eukaryote?
FT300003	Given that there is intact I-PpoI construct in the soil environment, what is the probability that it will enter the germline cell of a eukaryote?
FT300003	Given that the I-PpoI construct has entered a non-target eukaryote, what is the probability that it will enter the germline cell of the eukaryote?

FT300004	Given that the I-PpoI construct has entered a non-target eukaryote, what is the probability that it will enter the germline cell of a multi-cellular eukaryote?
FT300005	Given that the I-PpoI construct has entered a non-target eukaryote, what is the probability that it will enter the germline cell of a single-celled eukaryote?
FT30001	Aqueous environment
FT300010	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters an aquatic environment?
FT300011	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in an aquatic environment?
FT300012	Given that there is intact I-PpoI construct in an aquatic environment, what is the probability that it will enter a eukaryote?
FT300013	Given that there is intact I-PpoI construct in an aquatic environment, what is the probability that it will enter the germline cell of a eukaryote?
FT300013	Given that there is intact I-PpoI construct in an aquatic environment, what is the probability that it will enter the germline cell of a eukaryote?
FT300014	Given that there is intact I-PpoI construct in a eukaryote, what is the probability that it will enter the germline cell of a multi-cellular eukaryote?
FT300015	Given that there is intact I-PpoI construct in a eukaryote, what is the probability that it will enter the germline cell of a single-celled eukaryote?
FT30002	Eukaryote gut environment
FT300020	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct enters the gut of a eukaryote organism?

FT300021	Given that the I-PpoI construct has entered the gut of a eukaryote organisms what is the probability that the construct DNA remains intact?
FT300022	Given that there is intact I-PpoI construct in the gut of a eukaryote organisms, what is the probability that it will enter the germline cell of the eukaryote?
FT3001	Given that the I-PpoI construct has entered the germline cell of a eukaryote what is the probability that it gets into the nucleus?
FT3002	Given the I-PpoI construct has entered the nucleus of a germline cell of a eukaryote what is the probability it recombines into the eukaryote's genome?
FT301	Prokaryote mediated acquisition
FT3010	Construct moves from mosquito to prokaryote
FT30100	In a soil environment
FT301000	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters a soil environment?
FT301001	Given that the I-PpoI construct has left the mosquito and entered a soil environment, what is the probability that the construct DNA remains intact?
FT301002	Given that there is intact I-PpoI construct in the soil environment, what is the probability that there is direct contact with competent bacteria?
FT301003	Given that there is direct contact between the I-PpoI construct and competent bacteria, what is the probability that it is taken up and survives the restriction enzymes?
FT301004	Given that the I-PpoI construct has been taken up and survived the restriction enzymes, what is the probability that it is recombined into the DNA and the microbe has sufficient energy to replicate?

FT30101	In an aqueous environment
FT301010	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters an aqueous environment?
FT301011	Given that the I-PpoI construct has left the mosquito and entered an aqueous environment, what is the probability that the construct DNA remains intact?
FT301012	Given that there is intact I-PpoI construct in an aqueous environment, what is the probability that there is direct contact with competent bacteria?
FT301013	Given that there is direct contact between the I-PpoI construct and competent bacteria, what is the probability that it is taken up and survives the restriction enzymes?
FT301014	Given that the I-PpoI construct has been taken up and survived the restriction enzymes, what is the probability that it is recombined into the DNA and the microbe has sufficient energy to replicate?
FT30102	In the gut of a eukaryote
FT301020	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters the gut of a eukaryote?
FT301021	Given that the I-PpoI construct entered the gut of a eukaryote what is the probability that the construct DNA remains intact?
FT301022	Given that there is intact I-PpoI construct in the gut, what is the probability that there is direct contact with competent bacteria?
FT301023	Given that there is direct contact between the I-PpoI construct and competent bacteria, what is the probability that it is taken up and survives the restriction enzymes?
FT301024	Given that the I-PpoI construct has been taken up and survived the restriction enzymes, what is the probability that it is recombined into the DNA and the microbe has sufficient energy to replicate?
FT30103	Within mosquito

FT301030	Given that there is intact I-PpoI construct within the mosquito, what is the probability that there is direct contact with competent bacteria?
FT301031	Given that there is direct contact between the I-PpoI construct and competent bacteria, what is the probability that it is taken up and survives the restriction enzymes?
FT301032	Given that the I-PpoI construct has been taken up and survived the restriction enzymes, what is the probability that it is recombined into the DNA and the microbe has sufficient energy to replicate?
FT3011	Construct moves from prokaryote to eukaryote germline
FT30110	Given a viable I-PpoI transformed prokaryote organism has been produced, what is the probability that it comes into contact with a eukaryote organism?
FT30111	Given that the I-PpoI construct has entered a non-target eukaryote, what is the probability that it will enter the germline cell of the eukaryote?
FT30112	Given that the I-PpoI construct has entered the germline cell of a eukaryote what is the probability that it gets into the nucleus?
FT30113	Given the I-PpoI construct has entered the nucleus of a germline cell of a eukaryote what is the probability it recombines into the eukaryote's genome? (Given that contact between a transformed I-PpoI prokaryote and a eukaryote has been established, what is the probability it recombines into the eukaryote's genome?)
FT302	Viral mediated acquisition
FT3020	Construct moves from mosquito to virus
FT30200	Within mosquito
FT302000	Given a complete release of all 10,000 GM insectary mosquitoes, how many already are or will be infected with a mosquito virus?

FT302001	Acquisition
FT3020010	Direct acquisition
FT3020010	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the virus incorporates the construct into its own genome?
FT30200100	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the construct is excised from the mosquito genome by a transposase?
FT30200101	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, and the construct is excised from the genome via a transposase, what is the probability that the virus incorporates the construct into its own genome?
FT3020011	Virion mediated acquisition
FT30200110	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?
FT30200111	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?
FT30200112	Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?
FT302002	Given that a mosquito virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT302003-9	Given that a mosquito RNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT302004-9	Given that a mosquito DNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT30201	Within environment

FT302010	Stable construct in the environment
FT3020100	Soil environment
FT30201000	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters the soil environment?
FT30201001	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in the soil?
FT3020101	Aqueous environment
FT30201010	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters an aquatic environment?
FT30201011	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in an aquatic environment?
FT3020102	Eukaryote gut environment
FT30201020	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct enters the gut of a eukaryote organism?
FT30201021	Given that the I-PpoI construct has entered the gut of a eukaryote organisms what is the probability that the construct DNA remains intact?
FT302011	Transformed virus created
FT3020110	Via Bacteria/Bacteriophage combination
FT30201100	Given that there is intact I-PpoI construct in either soil, aqueous or gut environment, what is the probability that a competent bacteria is transformed with the construct?

FT30201101	Given that a competent bacteria has been transformed with the construct, what is the probability that it is infected with a bacteriophage?
FT30201102	Acquisition
FT302011020	Direct acquisition
FT3020110200	Given that transformed bacteria is infected with a bacteriophage, what is the probability that the bacteriophage incorporates the construct into its genome?
FT302011021	Virion mediated acquisition
FT3020110210	Given that transformed bacteria is infected with a bacteriophage, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?
FT3020110211	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified bacteriophage?
FT3020110212	Given simultaneous infection, what is the probability that the unmodified bacteriophage incorporates the construct into it own genome?
FT30201103	Given that a bacteriophage has acquired the construct into its genome, what is the probability that it remains able to replicate (autonomously or non-autonomously)?
FT3020111	Via other organism/virus combination
FT30201110	Given that there is intact I-PpoI construct in the soil, aqueous or gut environment, what is the probability that a cellular organism is transformed with construct?
FT30201111	Given that a cellular organism has been transformed with the construct, what is the probability that it is infected with a virus?
FT30201112	Acquisition

FT302011120	Direct acquisition
FT3020111200	Given that transformed cellular organism is infected with a virus, what is the probability that the virus incorporates the construct into its genome?
FT302011121	Virion mediated acquisition
FT3020111210	Given that transformed cellular organism is infected with a virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?
FT3020111211	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?
FT3020111212	Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?
FT30201113	Given that a virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT3021	Construct moves from virus into eukaryote germline
FT30210	Given that a transformed viable virus has been created, what is the probability that there is a eukaryote organism that is a suitable host for this virus?
FT30211	Given that there is a suitable eukaryote host for the transformed virus, what is the probability that there is a suitable transmission mechanism to transmit the virus to the new host?
FT30212	Virus in germline nucleus
FT302120	Given that there is suitable host and transmission mechanism, what is the probability that the transformed virus infects the early zygote of the eukaryote (first few cell division before cell differentiation)?
FT302121	Given that there is suitable host and transmission mechanism, what is the probability that the transformed virus infects a later stage of the eukaryote's life-cycle but then moves to the ovaries or testes?

FT30213	Given that a transformed virus has infected the germline of the eukaryote, what is the probability of integration of the I-PpoI construct into the host's genome?
FT303	Transposon mediated acquisition
FT3033	Mediated by other nearby flanking transposons
FT30330	Insert has flanking compatible transposon ends (in the mosquito)
FT303300	What is the probability that the existing I-PpoI construct is flanked by transposons of the same family?
FT303301	What is the probability that transposable elements of the same family insert either side of the I-PpoI construct (in the mosquito)?
FT303302	What is the probability that the I-PpoI construct moves within the mosquito genome into another transposable element (or between two elements of the same family)?
FT30331	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that there is a source of transposase / reverse transcriptase / integrase (T/RT/I) that acts on nearby transposons that are flanking the construct?
FT3033200	Given that there is a source of T/RT/I that acts on nearby flanking transposons, what is the probability that the excised I-PpoI construct will leave the mosquito?
FT3033201	Given that the excised I-PpoI construct has left the mosquito, what is the probability that it remains intact?
FT3033201	Given that the excised I-PpoI construct has left the mosquito with the integrase still bound to the construct, what is the probability that it remains intact?
FT3033202	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a eukaryote?
FT30332020-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a multi-cellular eukaryote?

FT30332021-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a single-celled eukaryote?
FT3033203	Given that the excised I-PpoI construct enters the germline cell of a eukaryote, what is the probability that it gets into the nucleus?
FT3033204	Integrated into eukaryote germline
FT30332040	Given that the excised I-PpoI construct has inserted into the nucleus of a germline cell with flanking transposons of the same family, what is the probability that it integrates into the genome because the T/RT/I remains bound and active to the free transposon?
FT30332040	Given that the excised I-PpoI construct together with its bound and active tranposase (RT/T/I) has inserted into the nuclues of a germline cell, what is the probability that it integrates into the genome?
FT30332041	Given that the excised I-PpoI construct has inserted with flanking transposons of the same family, what is the probability that there is a source of T/RT/I in the eukaryote germline cell?
FT30332041	Given that the excised I-PpoI construct with flanking transposons of the same family has inserted into the nucleus of a germline cell that contains endogenous transposons of the same family, what is the probability that it integrates into the eukaryotes germline because there is a source of T/RT/I in the eukaryote germline cell?
FT30332042	Given that the excised I-PpoI construct with flanking transposons of the same family has inserted into the nucleus of a germline cell that does not contain endogenous transposons of the same family, what is the probability that it integrates into the eukaryotes germline because transposon encoded T/RT/I is expressed in the eukaryote germline cell?
FT3033210	Given that there is a source of T/RT/I that acts on nearby flanking transposons, what is the probability that the excised I-PpoI construct will leave the mosquito?
FT3033211	Given that the excised I-PpoI construct has left the mosquito, what is the probability that it remains intact?
FT3033211	Given that the excised I-PpoI construct has left the mosquito without the integrase still bound to the construct, what is the probability that it remains intact?

FT3033212	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a eukaryote?
FT30332120-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a multi-cellular eukaryote?
FT30332121-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a single-celled eukaryote?
FT3033213	Given that the excised I-PpoI construct enters the germline cell of a eukaryote, what is the probability that it gets into the nucleus?
FT3033214	Integrated into eukaryote germline
FT30332140	Given that the excised I-PpoI construct has inserted with flanking transposons of the same family, what is the probability that there is a source of T/RT/I in the eukaryote germline cell?
FT30332140	Given that the excised I-PpoI construct with flanking transposons of the same family has inserted into the nucleus of a germline cell that contains endogenous transposons of the same family, what is the probability that it integrates into the eukaryotes germline because there is a source of T/RT/I in the eukaryote germline cell?
FT30332141	Given that the excised I-PpoI construct without the integrase still bound to it, has inserted with flanking transposons of the same family, what is the probability that there is a source of T/RT/I in the eukaryote germline cell?
FT30332141	Given that the excised I-PpoI construct with flanking transposons of the same family has inserted into the nucleus of a germline cell that does not contain endogenous transposons of the same family, what is the probability that it integrates into the eukaryotes germline because transposon encoded T/RT/I is expressed in the eukaryote germline cell?
FT3034	piggyBac mediated acquisition
FT30340	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that there is a source of T/RT/I that acts on piggyBac in the GM mosquitoes?

FT3034100	Given that there is a source of T/RT/I that acts on piggyBac in the GM mosquitoes, what is the probability that the excised I-PpoI construct will leave the mosquito?
FT3034101	Given that the excised I-PpoI construct has left the mosquito, what is the probability that the DNA remains intact?
FT3034101	Given that the excised I-PpoI construct with the integrase still bound to it, has left the mosquito, what is the probability that the DNA remains intact?
FT3034102	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a eukaryote?
FT30341020-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a multi-cellular eukaryote?
FT30341021-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a single-celled eukaryote?
FT3034103	Given that the excised I-PpoI construct enters the germline cell of a eukaryote, what is the probability that it gets into the nucleus?
FT3034104	Integrates into eukaryote germline
FT30341040	Given that the excised I-PpoI construct has inserted with flanking transposons of the same family, what is the probability that the T/RT/I remains bound and active to the free transposon?
FT30341041	Given that the excised I-PpoI construct has inserted with flanking transposons of the same family, what is the probability that there is a source of T/RT/I in the eukaryote germline cell?
FT3034110	Given that there is a source of T/RT/I that acts on piggyBac in the GM mosquitoes, what is the probability that the excised I-PpoI construct will leave the mosquito?
FT3034111	Given that the excised I-PpoI construct has left the mosquito, what is the probability that the DNA remains intact?

FT3034111	Given that the excised I-PpoI construct without the integrase still bound to it, has left the mosquito, what is the probability that the DNA remains intact?
FT3034112	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a eukaryote?
FT30341120-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a multi-cellular eukaryote?
FT30341121-9	Given that there is intact excised I-PpoI construct in the environment, what is the probability that it gets into the germline cell of a single-celled eukaryote?
FT3034113	Given that the excised I-PpoI construct enters the germline cell of a eukaryote, what is the probability that it gets into the nucleus?
FT3034114	Integrates into eukaryote germline
FT30341140	Given that the excised I-PpoI construct has inserted with flanking transposons of the same family, what is the probability that there is a source of T/RT/I in the eukaryote germline cell?
FT31	Spread of construct
FT310	Spread of construct: Selection in sexually reproducing eukaryotes
FT3100	Male selection
FT31000	Given that eukaryote males carrying the construct are viable, what is the probability that the construct fails to sterilise them?
FT31002	Given that the construct has introgressed into a wild type population, what is the probability that GM eukaryote male has higher fitness than a wild type male?
FT31002	GM Eukaryote male has fitness than WT

FT310020-9	Given that the construct has introgressed into a wild type population, what is the probability that a product (RNA/protein) of the construct improves fitness?
FT310021-9	Given that the construct has introgressed into a wild type population, what is the probability that the insertion has a beneficial effect by disrupting or affecting nearby genes?
FT310022-9	Given that the construct has introgressed into a wild type population, what is the probability that any flanking regions are beneficial and stay linked to the transgene?
FT3101	Female selection
FT31011	Given that the construct has introgressed into a wild type population, what is the probability that a GM eukaryote female has higher fitness than a wild type female?
FT31011	GM Eukaryote female $>$ fitness than WT
FT310110-9	Given that the construct has introgressed into a wild type population, what is the probability that a product (RNA/protein) of the construct improves fitness?
FT310111-9	Given that the construct has introgressed into a wild type population, what is the probability that the insertion has a beneficial effect by disrupting or affecting nearby genes?
FT310112-9	Given that the construct has introgressed into a wild type population, what is the probability that any flanking regions are beneficial and stay linked to the transgene?
FT311	Selection in non-sexually reproducing eukaryotes
FT3110	Given that the I-PpoI construct is successfully inserted into the genome of a eukaryote's germline, what is the probability that a product (RNA/protein) of the construct improves fitness?
FT3111	Given that the I-PpoI construct is successfully inserted into the genome of a eukaryote's germline, what is the probability that the insertion has a beneficial effect by disrupting or affecting nearby genes?

FT3112	Given that the I-PpoI construct is successfully inserted into the genome of a eukaryote's germline, what is the probability that any flanking regions are beneficial and stay linked to the transgene?
FT312	Active drive
FT3120	Spread by homing
FT31200	Homing at locus different from ribosomal repeat (additional site)
FT312000	Given that eukaryotes carrying the construct are viable, what is the probability that their genome contains an additional site outside the X-linked ribosomal repeat array that is recognised by I-PpoI?
FT312001	Given that I-PpoI recognises an additional site what is the probability that it moves into this site?
FT312002	Given that I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT312003	Negative fitness effects less than driving (additional site)
FT3120030	Given that unmutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT3120031	Low fitness effects at target site
FT31200310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT31200311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT31200312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT31201	Homing at locus different from ribosomal repeat (mutation)

FT312010	Given that eukaryotes carrying the construct are viable, what is the probability that I-PpoI mutates and recognises an additional site outside the X-linked ribosomal repeat array?
FT312011	Given that I-PpoI recognises an additional site what is the probability that it moves into this site?
FT312012	Given that I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT312013	Negative fitness effects less than driving (mutation)
FT3120130	Given that mutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT3120131	Low fitness effects at target site
FT31201310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT31201311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT31201312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT31202	Homing at ribosomal repeat
FT312020	Given that eukaryotes carrying the construct are viable, what is the probability that the I-PpoI construct inserts into a recognition site on a ribosomal repeat?
FT312021	Given that I-PpoI inserted into a recognition site on a ribosomal repeat, what is the probability that it maintains its germline expression
FT312022	Negative fitness effects less than driving

FT3120220	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it has low cleavage rate at the rDNA locus?
FT3120221	HEG inserted in intron
FT31202210	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a self-splicing intron?
FT31202211	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a spliceosomal intron?
FT3121	Y or W drive
FT31210	Given that the I-PpoI construct is successfully inserted into the genome of a eukaryote's germline, what is the probability that the eukaryote has heterogametic sex determination (e.g., XY sex determination)?
FT31211	Given that the I-PpoI construct is successfully inserted into the germline genome of a heterogametic eukaryote, what is the probability that the construct moves to the sex-determining chromosome (Y or W)?
FT31212	Given that the I-PpoI construct inserted on the Y or W chromosome, what is the probability that it produces an enzyme that is able to cleave rDNA?
FT31213	Given that the I-PpoI construct inserted on the Y or W chromosome and produces an enzyme that is able to cleave rDNA, what is the probability that this rDNA is predominantly on the X or Z chromosome?
FT31214	Given that the I-PpoI construct inserted on the Y or W chromosome and produces an enzyme that is able to cleave rDNA that is predominantly on the X or Z chromosome, what is the probability that it does not cause sterility?
FT31215	Given that the I-PpoI construct inserted on the Y or W chromosome, what is the probability that it maintain germline specificity?
FT31216	Given an I-PpoI construct inserted on the Y or W chromosome that is able to cleave rDNA exclusively on the X or Z chromosome without causing sterility, what is the probability that the driving forces are higher than the fitness costs?

FT3-Y-Drive	For each of the questions would you change any of your probabilities if the I-PpoI gene was in the driving Y construct?
FT4	Spread of the construct in non-eukaryotes over a year following an escape of all 10,000 genetically modified mosquitoes
FT40	In prokaryotes
FT400	Acquisition (construct moves from mosquito to prokaryote)
FT4000	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct is stably acquired by a prokaryote organism by transformation in the soil?
FT4001	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct is stably acquired by a prokaryote organism by transformation in an aquatic environment?
FT4002	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct is stably acquired by a prokaryote organism by transformation in the gut of a eukaryote organism?
FT401	Spread of construct in prokaryote
FT4010	Selection
FT40100	Selection because of expression (presence of positive selection)
FT401000	Given that a viable transformed prokaryote organism has been created, what is the probability that the HE protein or RNA is expressed in the prokaryote?
FT401001	Given that the HE protein is expressed in the prokaryote, what is the probability that the product (RNA or protein) of the construct per se improves the prokaryote's fitness?
FT40101	Selection without expression (presence of positive selection)
FT401010	Given that a viable transformed prokaryote organism has been created, what is the probability that the insertion of the construct in the genome improves the prokaryote's fitness by disrupting or affecting nearby genes?

FT401011	Given that a viable transformed prokaryote organism has been created, what is the probability that regions of the mosquito genome flanking the construct stay linked to the transgene and improve the prokaryote's fitness?
FT40102	Selection via hitch-hiking (presence of positive selection)
FT40102	Given that a viable transformed prokaryote organism has been created, what is the probability that the construct is increasing in frequency because of selection (i.e. hitch hiking)
FT401020	Given that a viable transformed prokaryote organism has been created, what is the probability that the construct is linked to one of the prokaryotes Mobile Genetic Elements (MGE)?
FT401021	Given that the construct has integrated into one of the prokaryote's Mobiles Genetic Elements, what is the probability that it increases in frequency by selection (i.e., hitch-hiking)?
FT4011	Homing
FT40110	Site in prokaryote recognised by the HEG
FT401100	Given that a viable transformed prokaryote organism has been created, what is the probability that the prokaryote genome contains an I-PpoI recognition site?
FT401101	Given that a viable transformed prokaryote organism has been created, what is the probability that the I-PpoI construct mutates to recognise a site in the prokaryote genome?
FT40111	Given that the viable transformed prokaryote's genome contains an I-PpoI recognition site or I-PpoI mutates to recognise a site, what is the probability that the construct inserts into this site?
FT40112	Given that the I-PpoI construct inserts into a recognition site in the prokaryote's genome, what is the probability that the I-PpoI protein is expressed in the prokaryote?
FT40113	Given that the I-PpoI protein is expressed, what is the probability the prokaryote's genetic system allows for HEG + and HEG - alleles to be present in the same cell to allow homing?

FT40114	Negative fitness effects less than driving
FT401140	Given that the viable transformed prokaryote's genome contains an I-PpoI recognition site or I-PpoI mutates to recognise a site, what is the probability that this site is in a self-splicing intron?
FT401141	Given that the viable transformed prokaryote's genome contains an I-PpoI recognition site or it mutates to recognise a site, what is the probability that this recognition sequence is in a non-essential region of the genome?
FT41	Spread of construct in virus
FT410	Acquisition of construct (construct moves from mosquito to virus)
FT4100	Within mosquito
FT41000	Given a complete release of all 10,000 GM insectary mosquitoes, how many already are or will be infected with a mosquito virus?
FT41001	Acquisition
FT410010	Direct acquisition
FT410010	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the virus incorporates the construct into its own genome?
FT4100100	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the construct is excised from the mosquito genome by a transposase?
FT4100101	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, and the construct is excised from the genome via a transposase, what is the probability that the virus incorporates the construct into its own genome?
FT410011	Virion mediated acquisition

FT4100110	Given a mosquito modified with the I-PpoI construct is infected with a mosquito virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?
FT4100111	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?
FT4100112	Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?
FT41002	Given that a mosquito virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT41003-9	Given that a mosquito RNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT41004-9	Given that a mosquito DNA virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT4101	Within environment
FT41010	Stable construct in the environment
FT410100	Soil environment
FT4101000	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters the soil environment?
FT4101001	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in the soil?
FT410101	Aqueous environment
FT4101010	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct leaves a mosquito and enters an aquatic environment?

FT4101011	Given that the I-PpoI construct has left the mosquito what is the probability that the construct DNA remains intact in an aquatic environment?
FT410102	Eukaryote gut environment
FT4101020	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that the I-PpoI construct enters the gut of a eukaryote organism?
FT4101021	Given that the I-PpoI construct has entered the gut of a eukaryote organisms what is the probability that the construct DNA remains intact?
FT41011	Transformed virus created
FT410110	Via Bacteria/Bacteriophage combination
FT4101100	Given that there is intact I-PpoI construct in either soil, aqueous or gut environment, what is the probability that a competent bacteria is transformed with the construct?
FT4101101	Given that a competent bacteria has been transformed with the construct, what is the probability that it is infected with a bacteriophage?
FT4101102	Acquisition
FT41011020	Direct acquisition
FT410110200	Given that transformed bacteria is infected with a bacteriophage, what is the probability that the bacteriophage incorporates the construct into its genome?
FT41011021	Virion mediated acquisition
FT410110210	Given that transformed bacteria is infected with a bacteriophage, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?

FT410110211	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified bacteriophage?
FT410110212	Given simultaneous infection, what is the probability that the unmodified bacteriophage incorporates the construct into it own genome?
FT4101103	Given that a bacteriophage has acquired the construct into its genome, what is the probability that it remains able to replicate (autonomously or non-autonomously)?
FT410111	Via other organism/virus combination
FT4101110	Given that there is intact I-PpoI construct in the soil, aqueous or gut environment, what is the probability that a cellular organism is transformed with construct?
FT4101111	Given that a cellular organism has been transformed with the construct, what is the probability that it is infected with a virus?
FT4101112	Acquisition
FT41011120	Direct acquisition
FT410111200	Given that transformed cellular organism is infected with a virus, what is the probability that the virus incorporates the construct into its genome?
FT41011121	Virion mediated acquisition
FT410111210	Given that transformed cellular organism is infected with a virus, what is the probability that the I-PpoI construct is incorporated into an infectious virion (which may be non-autonomous)?
FT410111211	Given that the virion has acquired the construct what is the probability that it infects another cell simultaneously infected with an unmodified virus?
FT410111212	Given simultaneous infection, what is the probability that the unmodified virus incorporates the construct into it own genome?

FT4101113	Given that a mosquito virus has acquired the construct into its genome, what is the probability that it is able to replicate (autonomously or non-autonomously)?
FT411	Spread
FT4110	Selection
FT41100	Given that a replicating transformed virus has been created, what is the probability that a product (RNA or protein) of the construct improves the virus' fitness?
FT41101	Given that a replicating transformed virus has been created, what is the probability that the insertion of the construct in the genome improves the virus' fitness by disrupting or affecting nearby genes?
FT41102	Given that a replicating transformed virus has been created, what is the probability that regions of the mosquito genome flanking the construct stay linked to the transgene and improve the virus' fitness?
FT41103	Given that a replicating transformed virus has been created, what is the probability that the construct spreads because it is linked to a region of the viral genome that is increasing in frequency by selection (i.e., hitch-hiking)?
FT4111	Homing
FT41110	Site in virus recognised by the HEG
FT411100	Given that a replicating transformed virus has been created, what is the probability that the viral genome contains an I-PpoI recognition site?
FT411101	Given that a replicating transformed virus has been created, what is the probability that the I-PpoI construct mutates to recognise a site in the viral genome?
FT41111	Given that the replicating transformed virus' genome contains the I-PpoI recognition site or I-PpoI mutates to recognise a site, what is the probability that the construct inserts in this site?

FT41112	Given that the I-PpoI construct has inserted into a recognition site in the viral genome, what is the probability that I-PpoI protein is expressed in the virus' host cell?
FT41113	Given that the I-PpoI protein is being expressed, what is the probability that the virus is a DNA virus?
FT41114	Given that the viral-borne I-PpoI protein is being expressed in the virus' host cell, what is the probability that HEG+ and HEG- viral genomes co-infect the same host cell?
FT41115	Negative fitness effects less than driving
FT411150	Given that the replicating transformed virus genome contains the I-PpoI recognition site or I-PpoI mutates to recognise a site, what is the probability that this site is in an intron?
FT411151	Given that the replicating transformed virus' genome contains the I-PpoI recognition site or I-PpoI mutates to recognise a site, what is the probability that this recognition sequence is in a non-essential region of the virus' genome?
FT41116-9	Given that the I-PpoI construct has inserted into a recognition site in the viral genome, what is the probability that I-PpoI protein is expressed in the virus' host cell? (DNA)
FT41117-9	Given that the I-PpoI construct has inserted into a recognition site in the viral genome, what is the probability that I-PpoI protein is expressed in the virus' host cell? (RNA)
FT4-Y-Drive	For each of the questions would you change any of your probabilities if the I-PpoI gene was in the driving Y construct?
FT5	Spread of construct in A. gambiae complex over a year following an escape of all 10,000 genetically modified mosquitoes
FT50	Spread in same species
FT500	Selection
FT5000	Some males fertile
FT50000	Male selection

FT500000	Given that GM males carrying the construct are viable, what is the probability that the construct allows some fertile males (not all are sterilised)?
FT500001	Given that some GM males carrying the construct are viable and fertile, what is the probability that the GM male mosquito has higher fitness than a wild type male?
FT50001	Female selection
FT500010	Given that GM males carrying the construct are viable, what is the probability that the construct allows some fertile males (not all are sterilised)?
FT500011	Given that GM females carrying the construct are viable and fertile, what is the probability that the GM female mosquito has higher fitness than wild type female?
FT5001	All males sterile
FT50010	Given that all GM males carrying the construct are sterile, what is the probability that the construct mutates to increases female fitness sufficiently to compensate for the male sterility?
FT501	Active drive
FT5010	Y Drive
FT50100	Given that offspring carrying the autosomal construct are viable, what is the probability that I-PpoI construct moves to the Y chromosome?
FT50101	Given that the I-PpoI construct inserted on the Y chromosome, what is the probability that it produces sufficient enzyme that is able to cleave rDNA?
FT50102	Given that the I-PpoI construct inserted on the Y chromosome and produces an enzyme that is able to cleave rDNA, what is the probability that this rDNA is predominantly on the X chromosome?

FT50103	Given that the I-PpoI construct inserted on the Y chromosome and produces an enzyme that is able to cleave rDNA that is predominantly on the X chromosome, what is the probability that it does not cause male sterility?
FT50104	Given that the I-PpoI construct inserted on the Y chromosome, what is the probability that it maintain germline expression
FT50105	Given an I-PpoI construct inserted on the Y chromosome that is able to cleave rDNA predominantly on the X chromosome without causing sterility, what is the probability that the driving forces are higher than the fitness costs?
FT5011	Homing
FT50110	Homing at locus different from ribosomal repeat
FT501100	Homing at additional site recognised by unmutated I-PpoI
FT5011000	Given that GM mosquitoes carrying the construct are viable, what is the probability that their genome contains an alternative site that is recognised by I-PpoI?
FT5011001	Given that I-PpoI recognises an alternative site in the genome, what is the probability that it moves into this site?
FT5011002	Given that I-PpoI has moved into an alternative site, what is the probability that it maintains its germline expression?
FT5011003	Negative fitness effects less than driving
FT50110030	Given that I-PpoI has moved into an alternative recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT50110030	Given that unmutated I-PpoI has moved into an alternative recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT50110031	Low fitness effects at target site
FT501100310	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?

FT501100311	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT501100312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT501101	Homing at location site recognised by mutated I-PpoI
FT5011010-9	Given that GM mosquitoes carrying the construct are viable, what is the probability that I-PpoI mutates to recognise site it is located in?
FT5011012-9	Given that I-PpoI has moved into an alternative site, what is the probability that it maintains its germline expression?
FT5011013	Negative fitness effects less than driving
FT50110130-9	Given that mutated I-PpoI has recognises site it is located and moved into this site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT50110131	Low fitness effects at target site
FT501101310-9	Given that mutated I-PpoI has recognises site it is located and moved into this site and maintains germline expression, what is the probability that the recognition site is in an Intron?
FT501101311-9	Given that mutated I-PpoI has recognises site it is located and moved into this site and maintains germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT501101312-9	Given that mutated I-PpoI has recognises site it is located and moved into this site and maintains germline expression, what is the probability that the negative fitness effects are recessive?
FT501102	Homing at additional site recognised by mutated I-PpoI
FT5011020	Given that GM mosquitoes carrying the construct are viable, what is the probability that I-PpoI mutates and then recognises an alternative site in the genome?

FT5011021	Given that I-PpoI recognises an alternative site in the genome, what is the probability that it moves into this site?
FT5011022	Given that I-PpoI has moved into an alternative site, what is the probability that it maintains its germline expression?
FT5011023	Negative fitness effects less than driving
FT50110230	Given that mutated I-PpoI has moved into an alternative recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT50110231	Low fitness effects at target site
FT501102310	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT501102311	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT501102312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT50111	Homing at ribosomal repeat
FT501110	Given that GM mosquitoes carrying the construct are viable, what is the probability that the I-PpoI construct inserts into a recognition site in a ribosomal repeat?
FT501111	Given that I-PpoI inserted into a recognition site on a ribosomal repeat, what is the probability that it maintains its germline expression
FT501112	Negative fitness effects less than driving
FT5011120	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that its cleavage rate at the rDNA locus is sufficiently low that reduced male fitness or reduced X chromosome transmission do not preclude spread through the population, but still is active enough for homing to occur?

FT5011121	HEG inserted in intron
FT50111210	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a self-splicing intron?
FT50111211	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a spliceosomal intron?
FT502	Non wolbachia
FT503	Wolbachia
FT5030	Wolbachia mediated acquisition
FT50300	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that Wolbachia is present in the mosquito?
FT50301	Given that Wolbachia is present, what is the probability that Wolbachia acquires the construct?
FT50302	Given that the construct has been transmitted to the F1 offspring, what is the probability that they are viable?
FT50303	Given that the offspring are carrying the modified Wolbachia, what is the probability that the modified Wolbachia will spread?
FT5031	Wolbachia mediated cytoplasmic incompatibility acquisition
FT50310	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that mitochondrial DNA acquires the construct?
FT50311	Given that mitochondrial DNA has acquired the construct, what is the probability that Wolbachia is present in the mosquito?
FT50312	Given that Wolbachia is present, what is the probability that Wolbachia acquires the mitochondrial DNA with the construct?
FT50313	What is the probability that the construct will be transmitted to the F1 offspring?

FT50314	Given that the modified F1 offspring are carrying the Wolbachia and the transformed mitochondria, what is the probability that the particular combination of Wolbachia and construct bearing mitochondria will spread in the recipient species?
FT51	Spread in different species from same complex
FT510	Non wolbachia
FT5100	Acquisition of construct
FT51000	What is the probability that there are compatible species (i.e. species that could mate with GM insectary mosquitoes) in the vicinity of the insectary?
FT51001	Given that there are compatible species in the vicinity, what is the probability that hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT510010-9	Given that there are compatible species in the vicinity, what is the probability that Arabiensis hybrids will be formed with these species, following a complete release of all 10,000 mosquitoes?
FT510011-9	Given that there are compatible species in the vicinity, what is the probability that Coluzzii hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT51002	Given that hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT510020-9	Given that Arabiensis hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT510021-9	Given that Coluzzii hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT51003	Given that the construct has been transmitted to the F1 offspring, what is the probability that they are viable?
FT510030-9	Given that the construct has been transmitted to the Arabiensis F1 offspring, what is the probability that they are viable?

FT510031-9	Given that the construct has been transmitted to the Coluzzii F1 offspring, what is the probability that they are viable?
FT5101	Spread (different species same complex)
FT51010	Selection
FT510100	Some males fertile
FT5101000	Male selection
FT51010000	Given that hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?
FT510100000-9	Given that Arabiensis hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?
FT510100001-9	Given that Coluzzii hybrid males carrying the construct are viable, what is the probability that the construct fails to sterilise them?
FT51010001	Given that the F1 hybrid males are fertile, what is the probability that the F2 and beyond GM hybrid males have higher fitness than wild type males?
FT510100010-9	Given that the Arabiensis F1 hybrid males are fertile, what is the probability that the F2 and beyond GM Arabiensis hybrid males have higher fitness than wild type males?
FT510100011-9	Given that the F1 Coluzzii hybrid males are fertile, what is the probability that the F2 and beyond GM Coluzzii hybrid males have higher fitness than wild type males?
FT51010002	Given that hybrid males carrying the construct are viable and the construct fails to sterilise them, what is the probability that the F1 hybrid males do not show hybrid sterility?
FT510100020-9	Given that Arabiensis hybrid males carrying the construct are viable and the construct fails to sterilise them, what is the probability that the F1 Arabiensis hybrid males do not show hybrid sterility?

FT510100021-9	Given that Coluzzii hybrid males carrying the construct are viable and the construct fails to sterilise them, what is the probability that the F1 Coluzzii hybrid males do not show hybrid sterility?
FT5101001	Female selection
FT51010010	Given that hybrid females carrying the construct are viable, what is the probability that the F1 hybrid females do not show hybrid sterility?
FT510100100-9	Given that hybrid Arabiensis females carrying the construct are viable, what is the probability that the F1 hybrid Arabiensis females do not show hybrid sterility?
FT510100101-9	Given that hybrid Coluzzii females carrying the construct are viable, what is the probability that the F1 hybrid Coluzzii females do not show hybrid sterility?
FT51010011	Given that the F1 GM hybrid female is fertile, what is the probability that F2 and beyond GM hybrid females have higher fitness than wild type females?
FT510100110-9	Given that the F1 GM Arabiensis hybrid female is fertile, what is the probability that F2 and beyond GM Arabiensis hybrid females have higher fitness than wild type females?
FT510100111-9	Given that the F1 GM Coluzzii hybrid female is fertile, what is the probability that F2 and beyond GM Coluzzii hybrid females have higher fitness than wild type females?
FT510101	All males sterile
FT5101010	Given that all GM males carrying the construct are sterile, what is the probability that the construct mutates to increases female fitness sufficiently to compensate for the male sterility?
FT51011	Active drive (different species same complex)
FT510110	Y Drive

FT5101100	Given that offspring carrying the autosomal construct are viable, what is the probability that I-PpoI construct moves to the Y chromosome?
FT5101101	Given that the I-PpoI construct inserted on the Y chromosome, what is the probability that it produces sufficient enzyme that is able to cleave rDNA?
FT5101102	Given that the I-PpoI construct inserted on the Y chromosome and produces an enzyme that is able to cleave rDNA, what is the probability that this rDNA is predominantly on the X chromosome?
FT5101103	Given that the I-PpoI construct inserted on the Y chromosome and produces an enzyme that is able to cleave rDNA that is predominantly on the X chromosome, what is the probability that it does not cause male sterility?
FT5101104	Given an I-PpoI construct inserted on the Y chromosome, is able to cleave rDNA predominantly on the X chromosome without causing sterility, what is the probability that it maintains germline expression?
FT5101105	Given an I-PpoI construct inserted on the Y chromosome, is able to cleave rDNA predominantly on the X chromosome without causing sterility and maintains germline expression, what is the probability that the driving forces are higher than fitness costs?
FT510111	Spread by homing
FT5101110	Homing at locus different from expected ribosomal repeat
FT51011100	Homing at locus different from expected ribosomal repeat (unmutated I-PpoI recognises site elsewhere)
FT510111000	Given that hybrids carrying the construct are viable, what is the probability that their genome contains an additional site outside the X-linked ribosomal repeat array that is recognised by I-PpoI?
FT510111000	Given that hybrids carrying the construct are viable, what is the probability that their genome contains an additional site outside the X-linked ribosomal repeat array that is recognised by I-PpoI?
FT510111001	Given that I-PpoI recognises an additional site what is the probability that it moves into this site?

FT510111001	Given that I-PpoI recognises an additional site what is the probability that it moves into this site?
FT510111002	Given that I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT510111002	Given that I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT510111003	Negative fitness effects less than driving
FT510111003	Negative fitness effects less than driving
FT5101110030	Given that unmutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT5101110030	Given that the unmutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT5101110031	Low fitness effects at target site
FT5101110031	Low fitness effects at target site
FT51011100310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT51011100310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT51011100311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT51011100311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?

FT51011100312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT51011100312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT510111004	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT510111004	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110040-9	Given that GM Arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110040-9	Given that GM Arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110041-9	Given that GM Coluzzii hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110041-9	Given that GM Coluzzii hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT51011101	Homing at locus different from expected ribosomal repeat (mutated I-PpoI recognises the site it is in)
FT510111010-9	Given that hybrids carrying the construct are viable, what is the probability that I-PpoI mutates and recognises the site where it is already located?
FT510111011-9	Given that I-PpoI has mutated to recognise the site that it is iwhat is the probability that it moves into this site?
FT510111012-9	Given that I-PpoI has mutated to recognise the site that it is in, what is the probability that it maintains its germline expression?
FT510111013	Negative fitness effects less than driving
FT5101110130-9	Given that I-PpoI has mutated to recognise the site that it is in and maintains germline expression, what is the probability that its cleavage rate at the rDNA repeat does not preclude spread through the population?

FT5101110131	Low fitness effects at original target site
FT51011101310-9	What is the probability that the recognition site is in an Intron?
FT51011101311-9	What is the probability that the recognition site is in a non-essential region of the genome?
FT51011101312-9	What is the probability that the negative fitness effects are recessive?
FT510111014	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110140-9	Given that Arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110141-9	Given that Coluzzii hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT51011102	Homing at locus different from expected ribosomal repeat (mutated I-PpoI recognises site elsewhere)
FT510111020	Given that hybrids carrying the construct are viable, what is the probability that I-PpoI mutates and recognises an additional site outside the X-linked ribosomal repeat array?
FT510111020	Given that hybrids carrying the construct are viable, what is the probability that I-PpoI mutates and recognises an additional site outside the X-linked ribosomal repeat array?
FT510111021	Given that I-PpoI recognises an additional site what is the probability that it moves into this site?
FT510111021	Given that the mutated I-PpoI recognises an additional site what is the probability that it moves into this site?
FT510111022	Given that I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT510111022	Given that the mutated I-PpoI has moved into an additional site what is the probability that it maintains its germline expression?
FT510111023	Negative fitness effects less than driving

FT5101110230	Given that the mutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that its cleavage rate at the rDNA repeat does not preclude spread through the population?
FT5101110230	Given that the mutated I-PpoI has moved into an additional recognition site and maintains germline expression, what is the probability that it does not cleave ribosomal repeats?
FT5101110231	Low fitness effects at target site
FT5101110231	Low fitness effects at target site
FT51011102310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT51011102310	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in an Intron?
FT51011102311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT51011102311	Given that I-PpoI has moved into an additional recognition site and maintains its germline expression, what is the probability that the recognition site is in a non-essential region of the genome?
FT51011102312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT51011102312	Given that I-PpoI has moved into an alternative recognition site and maintains its germline expression, what is the probability that the negative fitness effects are recessive?
FT510111024	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101110240-9	Given that GM Arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?

FT5101110241-9	Given that GM Coluzzii hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT5101111	Homing at ribosomal repeat
FT51011110	Given that hybrids carrying the construct are viable, what is the probability that the I-PpoI construct inserts into its recognition site on a ribosomal repeat?
FT51011111	Given that I-PpoI inserted into a recognition site on a ribosomal repeat, what is the probability that it maintains its germline expression
FT51011112	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT510111120	Given that An. arabiensis hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT510111121	Given that hybrids carrying the construct are viable, what is the probability that they do not show hybrid sterility?
FT51011113	Negative fitness effects less than driving
FT510111130	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that its cleavage rate at the rDNA locus is sufficiently low that reduced male fitness or reduced X chromosome transmission do not preclude spread through the population, but still is active enough for homing to occur?
FT510111131	HEG inserted in intron
FT5101111310	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a self-splicing intron?
FT5101111311	Given that I-PpoI inserted into a ribosomal repeat and maintains its germline expression, what is the probability that it inserted in a spliceosomal intron?
FT511	Wolbachia

FT5110	Wolbachia mediated acquisition
FT51100	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that Wolbachia is present in the mosquito?
FT51101	Given that Wolbachia is present, what is the probability that Wolbachia acquires the construct?
FT51102	Given that Wolbachia has acquired the construct, what is the probability that there are compatible species (i.e. species that could mate with GM insectary mosquitoes) in the vicinity of the insectary?
FT51103	Given that there are compatible species in the vicinity, what is the probability that hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT511030	Given that there are compatible species in the vicinity, what is the probability that Arabiensis hybrids will be formed with these species, following a complete release of all 10,000 mosquitoes?
FT511031	Given that there are compatible species in the vicinity, what is the probability that Coluzzii hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT51104	Given that hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT511040	Given that Arabiensis hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT511041	Given that Coluzzii hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT51105	Given that the construct has been transmitted to the F1 offspring, what is the probability that they are viable?
FT511050	Given that the construct has been transmitted to the Arabiensis F1 offspring, what is the probability that they are viable?
FT511051	Given that the construct has been transmitted to the Coluzzii F1 offspring, what is the probability that they are viable?

FT51106	Given that the construct has been transmitted to F1 offspring and they are viable, what is the probability that the F1 females are fertile?
FT51107	Given that the F1 females carrying the modified Wolbachia are fertile, what is the probability that the modified Wolbachia will spread in the recipient species?
FT5111	Wolbachia mediated cytoplasmic incompatibility acquisition
FT51110	Given a complete release of all 10,000 GM insectary mosquitoes, what is the probability that mitochondrial DNA acquires the construct?
FT51111	Given that mitochondrial DNA has acquired the construct, what is the probability that Wolbachia is present in the mosquito?
FT51112	Given that Wolbachia is present, what is the probability that Wolbachia acquires the mitochondrial DNA with the construct?
FT51113	Given that Wolbachia has become associated with the transformed mitochondrial DNA, what is the probability that there are compatible species (i.e. species that could mate with GM insectary mosquitoes) in the vicinity of the insectary?
FT51114	Given that there are compatible species in the vicinity, what is the probability that hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT511140	Given that there are compatible species in the vicinity, what is the probability that Arabiensis hybrids will be formed with these species, following a complete release of all 10,000 mosquitoes?
FT511141	Given that there are compatible species in the vicinity, what is the probability that Coluzzii hybrids will be formed, following a complete release of all 10,000 mosquitoes?
FT51115	Given that hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT511150	Given that Arabiensis hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?

FT511151	Given that Coluzzii hybrids between the GM insectary mosquitoes and compatible species have been formed, what is the probability that the construct will be transmitted to the F1 offspring?
FT51116	Given that the construct has been transmitted to the F1 offspring, what is the probability that they are viable?
FT511160	Given that the construct has been transmitted to the Arabiensis F1 offspring, what is the probability that they are viable?
FT511161	Given that the construct has been transmitted to the Coluzzii F1 offspring, what is the probability that they are viable?
FT51117	Given that the construct has been transmitted to F1 offspring and they are viable, what is the probability that the F1 females are fertile?
FT51118	Given that the modified F1 females carrying the Wolbachia and the transformed mitochondria are fertile, what is the probability that the particular combination of Wolbachia and construct bearing mitochondria will spread in the recipient species?

Appendix F Standard error of Monte Carlo estimates

FT2 Method A Standard error			
	50%	**90%**	**99%**
AFTC	7.2e-09	2.0e-06	1.6e-04
CFTA_LP	9.8e-09	1.0e-05	5.4e-04
FT3 Method B10 Standard error			
	50%	**90%**	**99%**
AFTC	7.2e-09	2.0e-06	1.6e-04
CFTA_LP	9.8e-09	1.0e-05	5.4e-04
FT4 Method E Standard error			
	50%	**90%**	**99%**
AFTC	1.9e-08	1.6e-05	1.6e-03
CFTA_LP	4.7e-09	3.7e-06	3.0e-05
FT50 Method B12 Standard error			
	50%	**90%**	**99%**
AFTC	3.2e-05	6.6e-03	1.3e-03
CFTA_LP	7.6e-06	8.0e-04	8.3e-03
FT51 Method B10 Standard error			
	50%	**90%**	**99%**
AFTC	3.0e-07	2.8e-04	4.2e-03
CFTA_LP	1.4e-05	3.3e-04	2.5e-03

Table F.1: Standard error of Monte Carlo estimates of the probability of the top event for each fault tree and preferred computation strategy

CONTACT US

t 1300 363 400

+61 3 9545 2176

e enquiries@csiro.au

w www.csiro.au

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

CSIRO Biosecurity Flagship
Keith Hayes

t +61 3 6232 5260

e Keith.Hayes@csiro.au

w http://people-my.csiro.au/H/K/Keith-Hayes.aspx