**The value of further research: the added value of individual-participant level data**

**Pedro Saramago**, PhD\*, Centre for Health Economics, University of York, York, U.K. (pedro.saramago@york.ac.uk) ORCID: 0000-0001-9063-8590

**Manuel A. Espinoza**, PhD, Departamento de Salud Pública, Pontificia Universidad Católica de Chile, Chile. (manuel.espinoza@uc.cl) ORCID:

**Alex J. Sutton**, PhD, Department of Health Sciences, University of Leicester, UK. (ajs22@leicester.ac.uk) ORCID:

**Andrea Manca**, PhD, Centre for Health Economics, University of York, York, U.K. (andrea.manca@york.ac.uk) ORCID: 0000-0001-8342-8421

**Karl Claxton**, PhD, Centre for Health Economics, University of York, York, U.K. (karl.claxton@york.ac.uk) ORCID:

\* Corresponding author.

Centre for Health Economics, Team for Economic Evaluation and Health Technology Assessment, The University of York, Alcuin College, A Block, York YO10 5DD United Kingdom, Tel: +44 (0)1904 321428; Fax: +44 (0)1904 321402; e-mail: pedro.saramago@york.ac.uk

**Total words: 5000**

**Running title: Value of further research and the added value of IPD**

**Abstract**

Judgements based on average cost-effectiveness estimates may disguise significant heterogeneity in net health outcomes. Decisions about coverage of new interventions are often more efficient when they consider between-patient heterogeneity, which is usually operationalized as different selections for different subgroups. While most model-based cost-effectiveness studies are populated with aggregated-level sub-group estimates, individual-level data is recognized as the best source of evidence to produce unbiased and efficient estimates to explore this heterogeneity. This paper extends a previously published framework to assesses the added value of having access to individual-level data, compared to using aggregate-level data only, in the absence/presence of mutually exclusive population subgroups. Supported by a case study on the cost-effectiveness of interventions to increase uptake of smoke-alarms, the extended framework provided a quantification of the benefits foregone of not using individual-level data, pointed to the optimal number of subgroups and where should further research be undertaken. Although not indicating changes in reimbursement decisions, results showed that irrespective of using aggregate or individual-level data, no substantial additional gains are obtained if more than two subgroups are taken into account. However, depending on the evidence type used, different subgroups are revealed as warranting larger research funds. The use of individual-level data, rather than aggregate, may however influence not only the extent to which an appropriate understanding of existing heterogeneity is attained, but, more importantly, it may shape approval decisions for particular population subgroups and judgements of future research.

**Key points for decision makers**

* Compared to the use of aggregate data, the use of individual-level data in economic evaluations has the potential to provide a better understanding of the differences between patients characteristics of the target population.
* With the use of individual-level data decision-makers may experience further gains in efficiency since different reimbursement decisions for different subsets of the population may be made. In addition, it may enable determining the value of performing further research on the population subgroups where this is most needed.
* However, relative to aggregate data, the magnitude of the benefits of considering individual-level data might not always outweigh the costs of obtaining, analysing, exploring and disentangling patient differences, together with assessing if new research is required for particular subgroups.

## Introduction

In healthcare, decisions are inevitably made under uncertainty. In this context, a decision maker should not only consider *i)* whether the provision of healthcare given the available information is consistent with a good use of the limited resources, but also *ii)* whether it is worthwhile to fund supplementary research to resolve existing uncertainty and improve decisions in the future.[[1](#_ENREF_1), [2](#_ENREF_2)]

In estimating the value of conducting further research [item *ii)* listed above], characterising and quantifying decision uncertainty and its consequences are key components. One of the main reasons for the existence of decision uncertainty is that decisions are based on sampled data, so that true decision model parameter values cannot be known with certainty.[[3](#_ENREF_3)] An important consequence of decision uncertainty is that a wrong decision may be made, determining suboptimal resource allocation and subsequent health losses at population level. This supports the idea of undertaking further research to produce more precise estimates and reduce the likelihood of decision error and the subsequent health losses.[[4](#_ENREF_4), [1](#_ENREF_1)]

In addition, whilst a particular health technology may be on average, and for the population as a whole, cost-effective, our confidence in this assessment may be low for individuals or groups of individuals within that population that depart from the average. This has led several authors and national and international institutions to recommend exploring heterogeneity, mostly as subgroup analyses, as usual part of the health technology assessment process.[[5-10](#_ENREF_5)] This examination of existing heterogeneity should include, for instance, exploring heterogeneity around costs, baseline risk, treatment effects and preferences underlying utilities.[[11](#_ENREF_11)] Ultimately, if heterogeneity is identified and acknowledged, different reimbursement decisions may be performed for different subgroups.[[8](#_ENREF_8), [7](#_ENREF_7)] This, however, may bring ethical and equity issues, particularly when subgroup analyses are focused on certain characteristics (e.g. age) that may be viewed as unacceptable to determine which individuals should or should not have access to the health technology.[[5](#_ENREF_5)] Although, examining variability and exploring heterogeneity may be considered an important analytical task; it is not the only one. Quantifying parameter uncertainty for each population strata can and should be performed if heterogeneity is identified.[[5](#_ENREF_5), [6](#_ENREF_6), [12](#_ENREF_12)] In other words, it is not only the possibility of making different decisions for different subsets of the population that matters, the value of performing further research may vary between the subgroups identified.

It is in these circumstances where the role of (and the benefits from) using individual participant-level data (IPD), compared with aggregate-level data (AD), becomes more evident. It has been shown that, relative to AD, IPD may provide unbiasedness and precision improvements. Additionally, the quantification of uncertainty in subgroups-specific estimates will, generally, be reduced when IPD is available. In evidence synthesis for health care decision-making the usefulness of IPD in exploring heterogeneity has been widely documented.[[13-20](#_ENREF_13)] When multiple studies are available in IPD format, there is the opportunity to standardise the statistical analysis across the evidence base. Nevertheless, IPD’s main advantage is that it enhances the exploration of heterogeneity in treatment effects (which may be performed by using treatment by covariate interactions) across patient groups, further explaining between-study heterogeneity and reducing potential evidence inconsistencies.

The paper starts by considering the value of additional research in the absence of subgroups when IPD is available. Making use of and extending previously published frameworks [[21](#_ENREF_21), [22](#_ENREF_22), [12](#_ENREF_12)], this paper further explores and estimates the magnitude of the additional value of using IPD, when performing subgroup value of information (VoI) analysis and how it may affect decision making. Both sections are supported by a case study on the cost-effectiveness of programmes intended to increase the uptake of functioning smoke alarm equipment by households with pre-school children.[[23-26](#_ENREF_23)] Finally, the last section summarizes the main points of the article, discusses limitations of the extended framework and includes suggestions for future research.

## Value of additional research in the absence of subgroups

We begin by examining the scenario of having access to relevant evidence at aggregate-level and/or at individual-level without considering subgroups. Given the presence of uncertainty surrounding the expected net benefits (NBs) associated with the use of alternative technologies, decision recommendations based on current information may change if these uncertainties are resolved. The probability and the consequences of making a wrong decision can be assessed and quantified [[27](#_ENREF_27)] through the Expected Value of Perfect Information (EVPI). For the optimal strategy among *y* mutually exclusive alternatives and given *θ*, an uncertain vector of parameters, the EVPI can be defined as the difference between the estimated payoffs of having perfect information from those obtained under current information [[1](#_ENREF_1)], that is:

|  |  |
| --- | --- |
|  | (1) |

The above expression refers to decisions made for one individual that can easily be translated to population values.[[28](#_ENREF_28)]

###  The added value of individual-level data to decisions on further research

Although in many instances access to IPD is preferred, the current state of art of decision analysis considers summary level data as satisfactory. Nonetheless, the type of evidence available may influence one’s certainty about the adoption decision. We now focus on the impact the type of evidence may have on the estimation of the need for further research.

Assuming the general case of a statistical synthesis model aiming to obtain a pooled effect size statistic based on any observed data *xi* (*i=1, …, m* studies). The analyst may have access to *m* studies at a summary level – an AD evidence base, or, we may envisage a situation in which access to IPD from each of these *m* studies is attained – an IPD evidence base. The possible benefits that the synthesis of IPD may bring, over and above the synthesis of AD, may be translated in terms of EVPI. In the absence of subgroups, estimated EVPI is not expected to differ when using AD or IPD to describe the evidence base, assuming AD is a sufficient statistic, i.e. no loss of information in comparison to IPD is expected.[[17](#_ENREF_17)] Three ways can be envisaged in which access to IPD can influence results in the absence of subgroups. The following scenarios are intended to be as generic as possible.

*Scenario 1.1*: If the use of IPD yields an increase in parameter(s) precision (efficiency), the distribution of relative treatment effects obtained, portrayed in Figure 1a by a probability density function, will have lower or equal variance, i.e. $Var\left[\hat{θ}^{AD}\right]\geq Var\left[\hat{θ}^{IPD}\right]$. Therefore, across a range of possible threshold values, EVPI values are expected to decrease due to the precision gained on this particular decision model parameter, all things being equal – Figure 1b.

*Scenario 1.2*: If the use of IPD is associated to a reduction or elimination of bias in the effect estimates, then this may imply a shift in the distribution of effects (for simplicity it is here assumed an impact on the scale parameter, and not on the shape and/or dispersion parameters). Figure 1c depicts a hypothetical situation where the estimates derived from AD evidence are biased upwards, i.e. $E\left[\hat{θ}^{AD}\right]>E\left[\hat{θ}^{IPD}\right]$. This may be translated in shifts in the mean ICER and consequently (left or right) shifts in EVPI curves as depicted in Figure 1d, *ceteris paribus*. Whilst in scenario 1.1 it can be predicted that EVPI is lower when using IPD, in the current scenario it cannot.

*Scenario 1.3*: This situation entails both an increase in precision and a reduction in bias from using individual-level evidence relative to AD. The graphical representations by Figure 1e and 1f illustrate this scenario and interpretations can be inferred from the two previous examples. In this scenario it is unpredictable how EVPI curves, derived from using IPD, may shift relative to the one derived using AD, though a decrease is expected due to the precision gained.

<< Figure 1 here >>

###  Application using the smoke alarms case-study and AD and AD+IPD in the absence of subgroups

Figure 2 summarizes the main cost-effectiveness results of the analysis over interventions to increase presence of smoke alarms in households. Results presented relate to when AD is synthesised using network meta-analysis (NMA) methods and pooled results are used to inform the smoke alarms decision model taken from Saramago *et al* [[25](#_ENREF_25), [26](#_ENREF_26)].[[1]](#footnote-2) Here, no subgroups are taken into account (i.e. it is assumed that no covariate information is available) and; therefore, they relate to population average estimates and, consequently, decisions are based on those. Usual care, identified as *(1)* in the figure, has the highest expected NBs and is also associated with the highest probability of being cost effective at a cost per QALY gained of £30,000.[[30](#_ENREF_30)]

<< Figure 2 here >>

Since the outcome variable analysed in the NMA synthesis model is binary (i.e. working smoke alarm present or not) [[23](#_ENREF_23)], population average (cost-) effectiveness results obtained using AD only versus AD + IPD where available (essentially a mix of aggregate and individual-level data since IPD was not available from all studies) are expected to be similar in the absence of covariates (no subgroups). Figure 3 shows that population EVPI estimates are almost equivalent at different threshold values when using AD and AD + IPD, respectively. Population EVPI estimate peaks at approximately £35,000 per QALY gained, when it equals the expected ICER of intervention that entails providing education and free or sponsored smoke alarm equipment, E + FE.[[25](#_ENREF_25), [26](#_ENREF_26), [31](#_ENREF_31)]

<< Figure 3 here >>

## Value of additional research in the presence of mutually exclusive subgroups

Following from equation 1, Espinoza *et al* [[12](#_ENREF_12)] show that when the decision maker is interested in subgroup specific results, the maximum expected net benefits for each subgroup *s* (out of a total of *K* mutually exclusive subgroups), the expected opportunity cost of uncertainty for subgroup *s* can be expressed as:

|  |  |
| --- | --- |
|  | (2) |

The EVPI considering subgroups (EVPI*K*, for *K* number of subgroups) is simply the weighted average across subgroups considering the proportion of patients in each subgroup (*ws*). In addition, population EVPI (PEVPI*K*) considering subgroups can also easily be derived [[12](#_ENREF_12)].

### Available evidence and the value of further research in the presence of subgroups

Returning to the general scenarios discussed in section 2 and taking into account the need to consider the presence of subgroups, the use of IPD, compared to the use of AD only, may correct for bias and/or increase precision in relation to population average estimates. This section will analyse the impact of these effects now in the context of subgroup analysis. Scenarios 2.1 to 2.3 (listed below) will consider potential gains in precision.

*Scenario 2.1*: The use of IPD may facilitate the use of formal modelling of treatment-covariate associations, which in particular circumstances, may not be attainable when using AD (e.g. study manuscript does not provide subgroup specific effect estimates). In this scenario, we consider the case where EVPI for subgroups can be estimated only when IPD is available, allowing the appropriate quantification of uncertainty in subgroup related effect estimates. A possible graphical representation may be found in Figure 4a, where a distribution of treatment effects for the overall population (i.e. 1 subgroup) was obtained from both evidence formats. In this case, as with scenario 1.1, the EVPI estimated with AD is expected to be higher or equal than the EVPI estimated with IPD (empty circle on or above full circle in Figure 4b). Given that IPD, compared to AD, may enable further exploration of existing between-patient heterogeneity when stratifying, an estimate of EVPI for two subgroups is only obtained when using IPD. With two subgroups, different adoption decisions may be made for each subgroup, which implies distinct Net Monetary Benefit (NMB) distributions for each. By exploring existing heterogeneity, subgroups may be specified that facilitate the characterisation of uncertainty (i.e. using informative specifications), the variance between observations (i.e. patients) decreases, reducing subgroup related parameter uncertainty. Consequently, the EVPI is here expected to be lower when compared to the overall population (i.e. , with *K* number of subgroups – as represented in Figure 4b). As the use of IPD allows the characterisation of heterogeneity, in extending this from two to *K* subgroups, and to (*I*) decisions made at individual level, the situation illustrated in Figure 4b may be obtained.[[32](#_ENREF_32)]

<< Figure 4 about here >>

*Scenario 2.2*: If AD allows the estimation of *q* subgroup effects (through, for instance, meta-regression analysis – assuming the required number of degrees of freedom exist), and the use of IPD facilitates resolving existing heterogeneity for *q* and *l* subgroups (where *q* < *l < K*, with *K* representing all possible subgroups), a situation as represented in Figure 4c may be obtained. This diagram reinforces the idea that, if the interest is to perform analysis on *l* subgroups, the type of evidence available may restrict this (e.g. AD). If IPD is available, it may meet the necessary conditions to perform the analysis, with the additional potential benefit of obtaining estimates with greater precision (for *q* and *l* subgroup effect estimates) in comparison with the distribution of effects derived from AD. Replicating this scenario in real world decisions, where these are made on a continuum, EVPI is expected to be lower or equal when IPD are available and used in the model compared to estimated EVPI when AD is used – Figure 4d.

*Scenario 2.3*: Assuming that AD also enables estimation of all required subgroup effects, a situation as represented in Figure 4e may exist. However, less precision is expected in interaction estimates derived from AD when compared to IPD. This fact is translated in a higher capacity of IPD to explore heterogeneity (and reduce the level of uncertainty within each subgroup) – Figure 4f.

The expected phenomena described in scenarios 2.1 to 2.3 will occur up to a point where, the size of the subgroup sub-sample implies an increase in subgroup-related variances, and, consequently, an increase in subgroup-related parameter uncertainty. In addition, we are assuming that we are working with subgroup specifications on the efficiency frontier, that is, the best subgroup specifications to characterise the different heterogeneity layers are being used.

###  Application using the smoke alarms case-study and AD and AD+IPD in the context of subgroups

Three alternative subgroup specifications based on the number of parents in the household and/or the parents’ employment status were considered. NMB and EVPI estimates were obtained when using AD only and when using both, AD and IPD, to inform key effectiveness parameters. The alternative *specifications* analysed comprise:

1. the average population (no subgroups) – presented in section 2.3;
2. two-subgroups (*K* = 2) specification 1, defined by the number of parents in a family per household. The proportion of patients in each of the two subgroups is estimated from the available trial evidence (average proportion across trials) and will be used further as the weights, *ws* – 71.0% of households have two parents and 29.0% have single parents.
3. two-subgroups (*K* = 2) specification 2, defined by the parents’ employment status in the household. In the samples of the trials used to evaluate effectiveness, 45.7% are employed and 54.3% have at least one parent unemployed in the family;
4. four-subgroups (*K* = 4) considering both binary variables, with following weights: two employed parents, 38.0%; two parents, at least one of them unemployed, 33.0%; single parent employed, 7.7%; and single unemployed parent, 21.3%.

#### *Further research: for which population?*

Table 1 shows the estimated population EVPI for a £20,000 threshold for each of the specifications *a)* to *d)* above. The same decision is being made when the population is split (i.e. rejection of any of the ‘active’ alternatives). Overall, results show that two parents strata in specification 1 (and that strata split into employed and at least one unemployed parent in the four subgroups setting) is, in relative terms, the one of highest population EVPI estimates and, thus, justifying further research, irrespective of the type of evidence used.

Furthermore, estimated (weighted) population EVPI is lower when it is split into two subgroups than when it is considered as a whole – no subgroups (PEVPI*K*=1 > PEVPI*K*=2). In addition, as the number of subgroups increases, population EVPI further decreases (PEVPI*K*=2 > PEVPI*K*=4). This is true when using AD to inform the decision model, but not verified by a small margin when AD + IPD is used. When four subgroups are being considered and AD + IPD is being used, population EVPI estimates are marginally higher than when using AD estimates (i.e. ). This result is inconsistent with the expected outcome of IPD resolving some of the existing decision uncertainty, lowering the EVPI. Reasons underlying this phenomenon may be related to: *(i)* specifications used are not the best to characterise existing heterogeneity and, thus, there is another (unknown) combination of specifications whose NMB estimate under current information is higher and EVPI smaller; *(ii)* the fact that different synthesis models are analysing different subgroup specifications, as described in section 1.2; *(iii)* confounding factors and/or ecological bias affecting the AD estimates; and *(iv)* due to simulation error.

Also, different subgroups contribute differently to the (total weighted) population EVPI. For the same subgroup, this estimated contribution is different if IPD is being considered or if it is not. While the use of AD indicates a small imbalance between the contributions of these subgroups (e.g. £5,405 and £10,954 for subgroups in specification 1), the use of AD + IPD points to a stronger imbalance between these two strata, with the single parent family subgroup having an estimated value of further research of £4,486. When considering four subgroups, and for AD + IPD, the subgroup for which the expected cost of uncertainty is higher is the ‘two employed parents’ subgroup, contributing £5,943 to the total of £15,442. The same is not true if AD estimates are used. In this case, the ‘two parents with at least one unemployed’ subgroup is the one that would require a larger amount of research funds (£5,374 to the total of £15,392).

<< Table 1 about here >>

#### *Further research: optimal number of subgroups?*

The following exercise is intended to understand the optimal number of subgroups to be defined for intervention(s) approval and subsequent further research development. Figure 5 depicts the population NMBs obtained with current and perfect information for different subgroup specifications (i.e. situations *a)* to *d)* described above) and for when AD and AD + IPD synthesis estimates are used to inform the modelling. The reader is reminded that in the current assessment, the choice of the number of subgroups is limited to the definition of a maximum of four subgroups and that the available subgroup specifications may not be at the efficiency frontier. This relates to the concept of static value of heterogeneity, that is, the additional benefits that could be gained if different decisions were made for different subgroups, under current information. In the ideal world, where all information is available and multiple subgroups can be explored, the efficiency frontier (i.e. the set of specifications that at different levels of disaggregation maximise benefits) can be revealed. Thus, different decisions can be taken for different subgroups according to the specification that produces the maximum health, at a level of disaggregation that can be implemented in practice. However, in most cases strata related information is restricted, and only a limited number of specifications can be explored.

If four subgroups are considered, similar NMB gains are obtained, as in the two-subgroup specifications (i.e. ‘number of parents in the family’ or ‘parents’ employment status’). This means that four subgroups do not provide additional static value. With both current and perfect information, it can be said that no apparent substantial additional gains are obtained if more than two subgroups are taken into account. If no further research is undertaken, that is, judgements are made with existing current information; the same conclusion can be derived. This is equivalent to say that no dynamic value of heterogeneity is observed with the specifications explored, because no additional benefits are expected when comparing the expected health with two or four specifications.

<< Figure 5 about here >>

It can be observed that under perfect information, and compared to no or four subgroup’ settings, higher NMBs are attained if two subgroups are considered and employment specification used. Except for the four subgroups situation, perfect information estimates are smaller when using IPD compared to when not.

## Discussion

This paper examines the advantages of having access to IPD when quantifying decision uncertainty with the purpose of estimating an upper boundary for the value of additional research. The rationale behind this is that parameter uncertainty is translated into decision uncertainty. Applying our extended framework, which explicitly makes use of available IPD, to the case study on interventions for smoke alarms uptake provided interesting findings. Existing decision uncertainty in the form of population EVPI estimates were obtained for the population as a whole and broken down into subgroups using two subgroup specifications. From the tested specifications, results indicated that the two parents’ strata had relatively higher decision uncertainty than other subgroup specifications, thus, indicating where future research investments should be done. More surprisingly perhaps was to find only marginal differences between the use of AD and AD+IPD evidence in terms of population EVPI estimates across the different stratifications. Although the use of IPD may have offered greater precision in appropriately characterising decision uncertainty and at what number of subgroups highest net benefits would be provided, fundamentally it would not alter research funding decisions, as using only AD provided similar results. We note, however, that the case study used and the stratifications undertaken were conditional on data available on covariates.

Our analyses adds to existing frameworks by assessing the advantages and disadvantages of using the synthesis of AD and IPD in appropriately disentangling existing heterogeneity and informing decisions at the subgroup level.[[12](#_ENREF_12), [22](#_ENREF_22), [10](#_ENREF_10)] Different formats of the same evidence set (i.e. AD and IPD) may provide different distributional ‘scenarios’ about the same set of parameters. These will provide different estimations of existing decision uncertainty, which can be quantified in expected costs. The extended framework highlights the capabilities of available evidence in quantifying the value of further research in the presence of mutually exclusive subgroups, pointing to the magnitude of the benefits. It allows investigating whether requesting further information for particular population *strata* is justified, and, moreover, how those judgements depend on the type of evidence (i.e. AD and/or IPD) being used in the analysis. Conceptually, it is argued that, when subgroup effects exist, the use of IPD is translated in a more appropriate characterisation of decision uncertainty compared to using AD only [[23](#_ENREF_23), [16](#_ENREF_16)] – this is true when making decisions for the population as a whole and when splitting the population into subgroups.

This work in not without assumptions and limitations. The rationale behind each subgroup scenario analysed in the case study is that the covariate information available is the one analysed, that is, for instance, in the population average scenario no subgroups are considered because it is assumed that information on covariates was non-existent. Moreover, it was not the purpose of this work to seek determining suitable grounds for the stratification applied to the case study.[[10](#_ENREF_10)] Comparisons made across subgroup scenarios and between NMB estimates with current and perfect information obtained using AD and IPD (i.e. AD + IPD) were performed with caution and should not be seen as generalisable. This is because different synthesis models are used to analyse different subgroup specifications. The approach taken may have limited the comparisons’ interpretability and highlighted as the main factor contributing to some unpredictable results obtained for particular subgroup specifications.

Throughout the paper, EVPI estimates obtained by using AD and IPD have been compared and contrasted depending on a number of factors, the most important being the number of subgroups considered. Care must be taken in interpreting the differences between EVPI estimates coming from these two different data structures. To view this EVPI difference in terms of the value of acquiring additional evidence or of acquiring IPD rather than AD to resolve existing decision uncertainty is incorrect, and the temptation to do so may be considered a limitation of this framework and of the analyses presented.

It is worth highlighting also that a single study available in IPD format (or even multiple) is not an ‘unlimited’ source of evidence as it is likely not to have been collected for the purpose of exploring subgroups. As more and more stratifications of the dataset(s) are performed (and hence, expected population health gains potentially increased), subgroup sample size decreases (smaller amounts of evidence available) and subgroup-specific uncertainty is expected to increase. Thus, the saturation of a particular dataset may be achieved at different stratification levels. All these (transaction) costs should be considered when assessing the impact of using IPD when quantifying subgroup level decision uncertainty. These were not under consideration here.

Important elements of the analyses are the transaction costs related to exploring available evidence, the understanding of information about existing heterogeneity and also the costs attached to implementing treatment decisions for each subgroup. These, as highlighted by Espinoza *et al* [[12](#_ENREF_12)], are expected to increase with the number of population subgroups being considered. In the case study, and relative to AD, the use of IPD did not demonstrate an impact on policy and/or practice, indicating that circumstances exist where the use of AD will provide similar results to the use of IPD, i.e. the full potential benefits of using IPD are not attained. In the framework discussed throughout this paper, and in the case study used, transaction costs are assumed negligible or not outweighing the gains, which may not always be true. The reality may be that the availability of IPD adds an extra layer of complexity to the transaction costs equation, as the task of obtaining and exploring the IPD is in itself a burden. In addition, to fully understand existing heterogeneity, new research may be needed in order to attempt to resolve it. Thus, in this context, a difficult balance exits between obtaining and using IPD and understanding and/or predicting its benefits relative to using AD alone. The reality is that decisions made at subgroup-level, or even at patient-level, may depend on other factors beyond efficiency, including the decision-making context, ethical and moral considerations, and the clarity and acceptability of stratification criteria.

This paper emphasises the position that standard economic analyses of health technologies are currently being performed to a non-optimal choice, as usual mean cost-effectiveness estimates may not take into account underlying population heterogeneous factors. However, exploring and accounting for existing between-patient heterogeneity may be conditional on the type and characteristics of the evidence available to the analysis. The use of IPD rather than AD, or in combination with AD, when performing subgroup cost-effectiveness analysis may offer better decisions with respect to the optimisation of (limited) available health resources and maximize population health, potentially culminating in different approval decisions for ethically sound population subgroups. Additionally, the use of IPD may enable a better understanding of which subgroup(s) justify investment of additional funds in order to increase certainty about approval recommendations.

## Acknowledgments

We would like to thank members of the Keeping Children Safe study [[26](#_ENREF_26)] for providing the case study data. In addition, we acknowledge Professor Mark Sculpher and Cynthia Iglesias for useful comments and suggestions as Thesis Advisory Group member of Dr. Pedro Saramago’s PhD.

## Compliance with Ethical Standards

Pedro Saramago was funded through a Medical Research Council Capacity Building Grant (grant number G0800139) and the Portuguese Fundacao para a Ciencia e a Tecnologia (grant number SFRH/BD/61448/2009).

## Declaration of Conflicting Interests

Pedro Saramago, Manuel A. Espinoza, Alex J Sutton, Andrea Manca and Karl Claxton: None to declare.

## Authors contributions

Pedro Saramago, Manuel A. Espinoza and Karl Claxton conceptually developed the framework. Pedro Saramago and Manuel A. Espinoza were responsible for the analysis, interpretation of the findings and writing of the manuscript. Alex J Sutton and Andrea Manca contributed to the interpretation of results and commented on drafts of the manuscript. Karl Claxton oversaw the work and commented on drafts of the manuscript. All authors contributed to the editing of this manuscript. All authors read and approved the final manuscript.

## References

1. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341-64. doi:S0167-6296(98)00039-3 [pii].

2. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. Pharmacoeconomics. 2006;24(11):1055-68. doi:24113 [pii].

3. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Mak. 2012;32(5):722-32.

4. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ. 1996;5(6):513-24. doi:10.1002/(SICI)1099-1050(199611)5:6<513::AID-HEC237>3.0.CO;2-9 [pii].

5. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics. 2008;26(9):799-806. doi:2699 [pii].

6. NICE. Guide to the Methods of Thechnology Appraisal. London: NICE, NHS2013.

7. Grutters JP, Sculpher M, Briggs AH, Severens JL, Candel MJ, Stahl JE et al. Acknowledging patient heterogeneity in economic evaluation : a systematic literature review. Pharmacoeconomics. 2013;31(2):111-23. doi:10.1007/s40273-012-0015-4.

8. Ramaekers BL, Joore MA, Grutters JP. How should we deal with patient heterogeneity in economic evaluation: a systematic review of national pharmacoeconomic guidelines. Value Health. 2013;16(5):855-62. doi:10.1016/j.jval.2013.02.013.

9. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. Med Decis Making. 2003;23(5):379-96.

10. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. Health Econ. 2003;12(5):421-7. doi:10.1002/hec.788.

11. Espinoza MA, Sculpher MJ, Manca A, Basu A. Analysing Heterogeneity to Support Decision Making A2 - Culyer, Anthony J. Encyclopedia of Health Economics. San Diego: Elsevier; 2014. p. 71-6.

12. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. Med Decis Mak. 2014;34(8):951-64.

13. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof. 2002;25(1):76-97.

14. Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. Journal of Evaluation in Clinical Practice. 2005;11(5):468-78.

15. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209-17.

16. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. Value in Health. 2012;15(5):639-49.

17. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Med Decis Mak. 2013;33(5):618-40.

18. Debray TP, Schuit E, Efthimiou O, Reitsma JB, Ioannidis JP, Salanti G et al. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? Stat Methods Med Res. 2016:962280216660741. doi:10.1177/0962280216660741.

19. van Walraven C. Individual patient meta-analysis--rewards and challenges. J Clin Epidemiol. 2010;63(3):235-7. doi:10.1016/j.jclinepi.2009.04.001.

20. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221. doi:10.1136/bmj.c221.

21. Basu A, Meltzer D. Quantitative Methods for Valuing Comparative Effectiveness Information. Biopharmaceutical Report. 2010;17(2):2-10.

22. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. Med Decis Making. 2007;27(2):112-27. doi:27/2/112 [pii] 10.1177/0272989X06297393.

23. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Statistics in Medicine. 2012;31(28):3516-36.

24. Cooper N, Kendrick D, Achana F, Dhiman P, He Z, Wynn P et al. Network meta-analysis to evaluate the effectiveness of interventions to increase the uptake of smoke alarms. Epidemiol Rev. 2012;34(1):32-45. doi:<http://dx.doi.org/10.1093/epirev/mxr015>.

25. Saramago P, Cooper NJ, Sutton AJ, Hayes M, Dunn K, Manca A et al. Cost-effectiveness of interventions for increasing the possession of functioning smoke alarms in households with pre-school children: a modelling study. BMC Public Health. 2014;14:459.

26. Kendrick D, Ablewhite J, Achana F, Benford P, Clacy R, Coffey F et al. Keeping Children Safe: a multicentre programme of research to increase the evidence base for preventing unintentional injuries in the home in the under-fives. Keeping Children Safe: a multicentre programme of research to increase the evidence base for preventing unintentional injuries in the home in the under-fives. Programme Grants for Applied Research. Southampton (UK)2017.

27. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). Lancet. 2002;360(9334):711-5. doi:S0140-6736(02)09832-X [pii] 10.1016/S0140-6736(02)09832-X.

28. Philips Z, Claxton K, Palmer S. The half-life of truth: What are appropriate time horizons for research decisions? Med Decis Mak. 2008;28(3):287-99. doi:<http://dx.doi.org/10.1177/0272989X07312724>.

29. ONS. Office for National Statistics, Household Projections, 2008 to 2033, England: Office for National Statistics - Communities and Local Governement2010.

30. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ. 2004;13(5):405-15. doi:10.1002/hec.903.

31. Claxton K, editor. Heterogeneity in the Cost-effectiveness of Medical Interventions: The Challenge of Matching Patients to Appropriate Care. ISPOR; 2011; Madrid.

32. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol. 2002;55(1):86-94. doi:S0895435601004140 [pii].

1. The analyses performed on the case study assumed a public health perspective. A 10 years expected lifetime of the programmes and an annual effective population (i.e. expected number of new households with dependent/s under 5 years old per year in the UK) of 31,000 households [[29](#_ENREF_29)] were assumed for the value of information analysis calculations. [↑](#footnote-ref-2)