INCLUDED

Promoting INCLUsivity through improving the practice anD utility of Ethnicity Data collection in trials

Recommendations

"Ethnicity is a multi-faceted quality that refers to the group to which people belong, and/or are perceived to belong, as a result of certain shared characteristics, including geographical and ancestral origins, but particularly cultural traditions and languages. The characteristics that define ethnicity are not fixed or easily measured, so ethnicity is imprecise and fluid. Ethnicity differs from race, nationality, religion, and migrant status, sometimes in subtle ways, but may include facets of these other concepts."

Raj Bhopal, 2004[1]

Introduction

This document presents recommendations based on the INCLUDED study. The aim of the INCLUDED study was to produce initial recommendations for researchers on collecting, handling and reporting ethnicity data in clinical trials. During the study, we considered the definition of ethnicity. Dictionary definitions are not consistent. The quote above is a health-related explanation of ethnicity that the research team felt reflected our thinking about what ethnicity is.

The INCLUDED study comprised (i) a survey to 24 UK Clinical Research Collaboration registered Clinical Trials Units, collection of 41 case studies (mostly randomised trials) from these units and interviews with survey respondents; (ii) opinion gathering from a number of trials-related networks; (iii) focus groups and interviews with individuals from diverse ethnic communities; and (iv) a consensus meeting.

We identified individuals for focus groups and interviews through four community groups (Social Action for Health, South Asian Health Action, Caribbean and African Health Network, Latin American Group), three of which were identified by Egality Health, a community organisation linking researchers and community groups. The work was underpinned throughout by a public advisory group.

A key element of the INCLUDED study was interaction with diverse ethnic communities at all stages of the research cycle. We identified three distinct types of interaction. These three types of interaction are referred to in our recommendations:

- Interaction with **organisations or networks** that connect relevant diverse ethnic communities with the research sector
- Interaction with individuals who are part of specific ethnic communities and who, because of their role and/or standing in that community can bridge the gap between their own community and researchers. IN INCLUDED these individuals requested to be called "**community connectors**" and this is the terminology we use in this document.
- Interaction with individual public/patient members from diverse ethnic communities, as part of **patient and public involvement and engagement** (PPIE).

This document lists 13 recommendations that have come from the work of the INCLUDED study. In the main text, the recommendations are divided into five sections reflecting different stages of the research cycle: prior to grant application (section 1), grant application (section 2), trial set-up (sections 3 and 4), reporting (section 5).

Executive Summary

This document contains 13 recommendations covering early planning for ethnicity data collection and reporting; material to be included in a grant application; further planning and collecting data once the trial is funded; training and oversight required; and reporting (figure 1).

Four recommendations focus on the planning stage of a trial, prior to submitting the grant application. The recommendations acknowledge the importance of meaningful involvement of diverse ethnic communities at this stage of the research. They are highly interrelated and most likely to be followed concurrently. We recommend that researchers identify the diverse ethnic communities from which participants in the trial need to be drawn (recommendation 1); clearly articulate the reason(s) for ethnicity data collection in the trial (recommendation 2); if possible, engage with organisations or networks that connect relevant diverse ethnic communities with the research sector (recommendation 3) and engage individual patient/public members in communities identified in recommendation 1 (recommendation 4).

Recognising the need to put the recording of ethnicity on a similar footing to other data collected in trials, we recommend that investigators communicate to funders in the grant application, why and how ethnicity data will be collected and disseminated (recommendation 5).
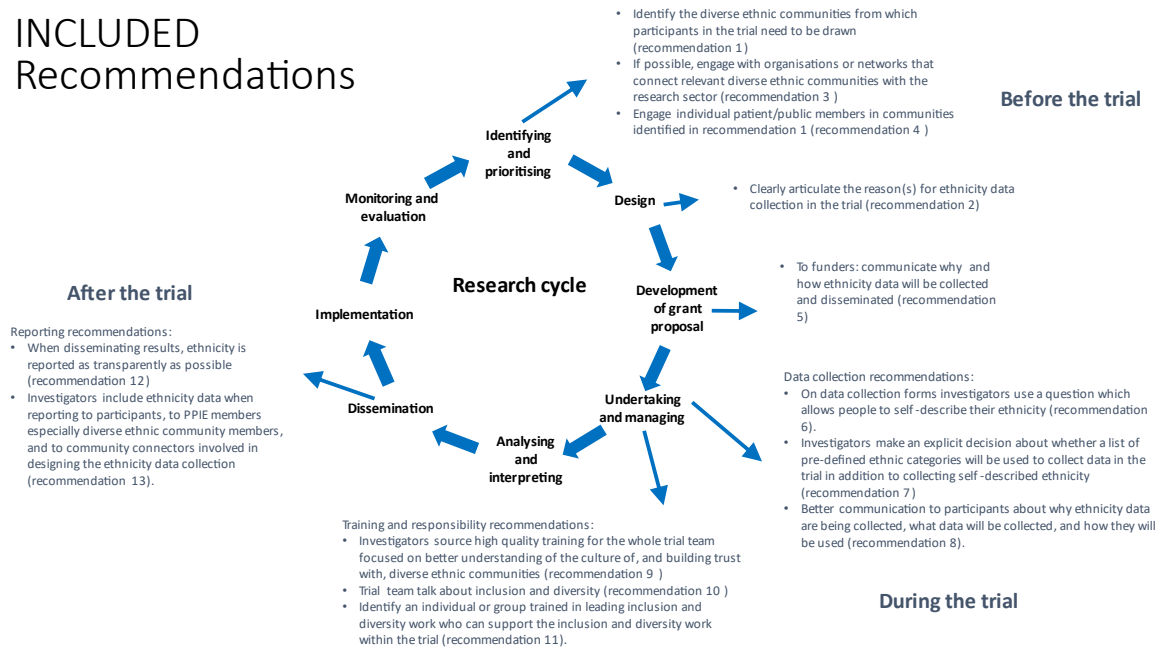
Recommendations relevant to the trial set up stage focused on two distinct areas: data collection, and training and responsibility. In relation to data collection, we recommend that on data collection forms investigators use a question which allows people to self-describe their ethnicity (recommendation 6). At the moment it is more common to use a pre-specified list of ethnic categories. However, during the INCLUDED study we became aware of the substantial issues around lack of inclusivity that this raises alongside arguments about consistency and ease of analysis that are often voiced. We therefore recommend that investigators make an explicit decision about whether a list of pre-defined ethnic categories will be used to collect data in the trial in addition to collecting self-described ethnicity (recommendation 7). We also recommend better communication to participants about why ethnicity data are being collected, what data will be collected, and how they will be used (recommendation 8).

In relation to training and responsibility, it was clear that both participants in the INCLUDED focus groups and interviews, and those from CTUs who were interviewed, felt that researchers could be better equipped to understand how to interact with diverse ethnic community members involved in a trial. Thus, we recommend that, if possible, investigators source high quality training for the whole trial team focused on better understanding of the culture of, and building trust with, diverse ethnic communities (recommendation 9), that the trial team talk about inclusion and diversity (recommendation 10), and that investigators identify an individual or group trained in leading inclusion and diversity work who can support the inclusion and diversity work within the trial (recommendation 11).

The INCLUDED study highlighted the need for better reporting of ethnicity in trials and we therefore recommend that when disseminating results, ethnicity is reported as transparently as possible (recommendation 12), and that investigators include ethnicity data when reporting to participants, to PPIE members especially diverse ethnic community members, and to community connectors involved in designing the ethnicity data collection (recommendation 13).

Figure 1: INCLUDED recommendations within the research cycle

INCLUDED
Recommendations

Before the trial

Identifying and prioritising
- Identify the diverse ethnic communities from which participants in the trial need to be drawn (recommendation 1)
- If possible, engage with organisations or networks that connect relevant diverse ethnic communities with the research sector (recommendation 3)
- Engage individual patient/public members in communities identified in recommendation 1 (recommendation 4)

Design
- Clearly articulate the reason(s) for ethnicity data collection in the trial (recommendation 2)

Development of grant proposal
- To funders: communicate why and how ethnicity data will be collected and disseminated (recommendation 5)

Undertaking and managing

Data collection recommendations:
- On data collection forms investigators use a question which allows people to self-describe their ethnicity (recommendation 6).
- Investigators make an explicit decision about whether a list of pre-defined ethnic categories will be used to collect data in the trial in addition to collecting self-described ethnicity (recommendation 7)
- Better communication to participants about why ethnicity data are being collected, what data will be collected, and how they will be used (recommendation 8).

During the trial

Research cycle

Monitoring and evaluation

Implementation

After the trial

Dissemination

Reporting recommendations:
- When disseminating results, ethnicity is reported as transparently as possible (recommendation 12)
- Investigators include ethnicity data when reporting to participants, to PPIE members especially diverse ethnic community members, and to community connectors involved in designing the ethnicity data collection (recommendation 13).

Analysing and interpreting

Training and responsibility recommendations:
- Investigators source high quality training for the whole trial team focused on better understanding of the culture of, and building trust with, diverse ethnic communities (recommendation 9)
- Trial team talk about inclusion and diversity (recommendation 10)
- Identify an individual or group trained in leading inclusion and diversity work who can support the inclusion and diversity work within the trial (recommendation 11).

INCLUDED recommendations 20/03/2023

Recommendations

Section 1: Prior to grant application - planning data collection with diverse ethnic communities (Recommendations 1-4)

All four recommendations in this section apply to the period of trial development prior to grant application submission. The recommendations are highly interrelated and more likely to be followed concurrently than in any specific order. The process of following these recommendations will differ depending on trial and trial team and should be tailored accordingly.

Our research in the INCLUDED study has highlighted that communities feel it is important to be meaningfully involved in developing plans for ethnicity data collection at an early stage, to ensure inclusivity and best practice and that the trial fulfils its purposes. Our recommendations reflect this. However, currently, interaction with diverse ethnic communities in planning ethnicity data collection at the pre-grant submission stage is minimal with only two out of 41 trial teams in the INCLUDED survey[2] of recently completed trials doing so. Thus, compared to current, less explicitly inclusive, practice there are likely to be additional resource implications of following these recommendations both in terms of the time needed to do this work, and in terms of finance to ensure those involved are fairly reimbursed for their time and expertise.

1. **Identify the diverse ethnic communities from which participants in the trial need to be drawn**

    1.1 Why? Ultimately, we want trial participants to be representative of those who could potentially benefit from the intervention(s) being tested. We therefore need to correctly identify those who need to participate. In relation specifically to ethnicity data collection, without correctly identifying those who need to participate, data collection instruments may inappropriately exclude relevant communities, or cause unnecessary harm or further exclusion (Box 1). Only eight out of 41 trial teams that responded as part of the INCLUDED survey of recently completed trials had had discussions about the diverse ethnic communities from which the participants in their trials should be drawn. Thus, we know discussion about this aspect of the research needs improving.

    Box 1: Quotes from participants in INCLUDED focus groups and interviews about exclusion from ethnicity data collection instruments

    > *"……the moment you open that I saw just the beginning, I was like okay, I know what this is, I've seen this plenty of times and I ignore it every single time. Because it is just so limiting. You're either… It's almost like you're either English or others. And then if others, there's a couple of options like other…white plus Caribbean. That's the only option if you're not British. That's the only other combination you can have. So I always just tick other and move on. So I could be other from Brazil or I could be other from China. It could be anywhere in the world. There's nothing really in there."* (Latin American)
    >
    > *"We make up a lot of communities spread out across the world and we still don't have our own box. (Latin American)*
    >
    > *"As we are different, both being white or being black, there is quite a difference between us…Sometimes they may not be interested because they're just asking, then they put what they like down."'* (Caribbean African Health Network)
    >
    > *"'Even when you have Latin American options, when we do, it's like white European. What? No. [it is] White Latin American.'"* (Latin American)

1.2 How? Ideally such an exercise will involve the chief investigator and those who might be part of the eventual trial team including statisticians, trial managers, those likely to collect data, and data managers; and relevant community connectors. Community connectors will have an important perspective to bring. For example, they may highlight that a community that researchers think of as homogeneous e.g. South Asian, African, or Caribbean is better thought of as several different communities, having different barriers and facilitators to research participation. The INCLUDE Ethnicity Framework[3] could be used to help with this. By using the Framework, trial teams can identify the diverse ethnic communities from which trial participants need to be drawn and then identify features of the disease, intervention and trial delivery that may make the involvement of these diverse ethnic groups harder or easier. A recent evaluation confirms that identification of these factors works better if researchers and community connectors use this framework together.[4]

2. **Clearly articulate the reason(s) for ethnicity data collection in the trial**

2.1 Why? Information gathered in the INCLUDED study suggested considerable uncertainty amongst researchers and potential trial participants about why such data are collected. Clear articulation of the reason(s) for collecting ethnicity data can motivate individuals from diverse ethnic groups to provide such data (Box 2). It also helps trialists to comply with best practice advice and only collect necessary data in trials. Understanding the purpose of collecting data is also essential for ensuring the way data are collected and analysed matches the purpose of collecting the data.

Box 2: Quotes from participants in INCLUDED focus groups and interviews about the purpose of data collection

> *"Yeah, it's quite often like disease, we are not included as part of any study. So how can they treat us if they don't know anything about us or our conditions" (South Asian Health Action)*
>
> *"..So I just feel it's very upsetting and it confuses me even more when taking part in research that they don't explain the purpose of the research" (Latin American)*

The most common reasons for collecting ethnicity data given by the 41 trials in the INCLUDED survey were generalisability, representativeness and to describe the population (Figure 2); three-quarters of the trials presented ethnicity data in tables designed to show the make-up of the population at baseline. In some trials, the collection of ethnicity data may be thought useful for analysis purposes, for example, in forming groups that can be compared in exploratory subgroup analyses. For example, in the ELECTRA study,[5] an exploratory subgroup analysis highlighted potential health inequalities that were explored further in the OEDIPUS study.[6] Only 15% of the trials in the INCLUDED survey used ethnicity data for such analyses, several using only two subgroups (white and non-white/other). Addressing health inequalities is a major reason for collecting ethnicity data in many studies. As authors of the MBRRACE study (not a trial) on maternal death rates highlight, "ethnic inequality endures…… [and] warrant[s] a continuing focus" even when data shows reductions in such inequality.[7] However, sometimes collecting data on a different aspect of culture and ancestry - for example, beliefs or skin colour - may be more suited to the analyses required. For example, during the COVID pandemic researchers highlighted that pulse oximeters did not work well for those with darker skin.[8] While this disadvantaged

those from certain ethnic communities, it was not primarily an issue related to ethnicity, and further studies on pulse oximeters would do better to focus data collection on skin colour (ideally measured as a range of skin tones) rather than ethnicity.

Figure 2: Reasons for collecting ethnicity data
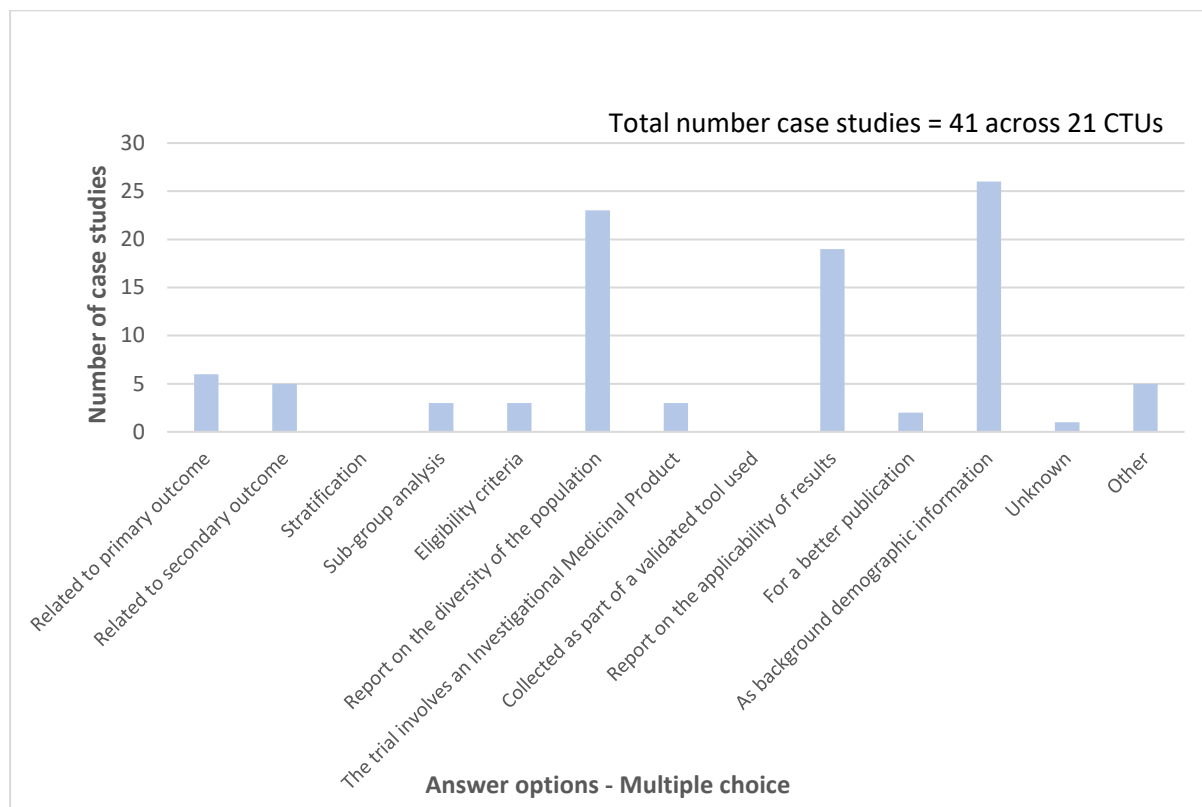
2.2 *How?* A focused discussion about reasons for collecting ethnicity data provides a forum for debating the issues raised above. Having this discussion prior to submitting a grant application means that it is at an early enough stage to allow the most appropriate domains of culture and ancestry (which may or may not include ethnicity) to be collected to match the research question(s), thus ensuring the best design for data collection. Currently it is most often only chief investigators, trial manager and statisticians who are involved in discussions about why and how ethnicity data are collected in trials. Of the 27 INCLUDED case study trials that had such discussions, the Chief Investigator was involved in 25, trial manager/coordinator and statistician in 22 each, trial methodologist and PPI member in ten each, clinicians in five, health economist and data manager in two each, and one trial included the whole project group. However, none of these individuals are routinely involved in providing or collecting trial data. Involving community connectors will make sure that the reasons articulated are clear to participants and other community stakeholders, and any particular aspects of culture and ethnicity relevant to the study can be taken on board in doing so. Involving a wider range of researchers, such as research nurses and research assistants who are more familiar/experienced with data collection processes will provide first hand insight in to the practicalities and challenges of collecting this data. Although not specifically related to ethnicity data collection, the ORINOCO study highlighted the usefulness of talking to collectors of data about the way the data are collected.[9] If some of the individuals included in these discussions end up being part of the trial team, these earlier discussions will help them to explain to trial participants the reasons for collecting the data. The results of the discussions could usefully be communicated to the rest of the trial team (e.g. as part of training), if and when the grant is successful.

INCLUDED recommendations 20/03/2023

3. **If possible, engage with organisations or networks that connect relevant diverse ethnic communities with the research sector**

3.1 *Why?* Researchers often lack knowledge on how to connect with diverse ethnic communities, or have some knowledge but nevertheless struggle to actually connect with these communities. Community organisations and networks can facilitate this connection, often for multiple communities. They can also support and advocate for such communities ensuring good mutual understanding between researchers and communities. Engaging with organisations and networks in this way supports increased inclusivity by facilitating the following of recommendations such as those outlined above. This facilitation is likely to be primarily in the specific trial under consideration. However, it could also lead to longer-term relationships which promote inclusivity in other trials by providing the continuity in relationships with diverse ethnic communities that researchers often find difficult due to time and resource constraints.

3.2 *How?* The conditionality in this recommendation recognises the lack of existing relationships between researchers and such organisations and networks, lack of funding for pre-grant stages to support developing relationships, and grant timelines that make it difficult to allow time for such discussion. Many such organisations exist, sometimes national (eg Egality Health https://egality.health/) but more often regional (eg Grampian Regional Equality Council https://grec.co.uk). Up until now relationships between such organisations and researchers have been limited. NIHR's People in Research aims to put individuals interested in being involved in research and those researchers looking for such individuals in touch with each other.[10] However, a system specifically designed to link researchers and community organisations does not yet exist. We suggest researchers look for organisations in their areas. NIHR provides guidance on how to involve the public in research.[11] Though the focus is largely on involving individuals rather than organisations, some of the principles are useful in thinking about making links.

4. **Engage individual patient/public members in communities identified in recommendation 1**

4.1 *Why?* From focus groups and interviews with diverse ethnic communities involved in the INCLUDED study, we heard that such communities wanted to be, and felt it was important to be, meaningfully involved at an early stage in developing plans for ethnicity data collection. Community connectors involved in the INCLUDED study also pointed out that their perspectives may not be representative of all of those in their communities and so it is important to talk to other individuals in these communities.

4.2 How? Community connectors and organisations and networks mentioned in recommendations 1-3 may be able to identify suitable individuals to engage with. INCLUDED focus group and interview participants suggested that researchers meet members of their communities in places used by communities rather than always using settings such as NHS premises. These settings should be those that community members feel are appropriate, where they feel safe, and which they feel represents their identity (Box 3). This could include, for example, the family home, places of worship, or other community hubs. For researchers this may be a change from expecting members of the public to come to academic or clinical settings, but it is likely to have ongoing benefits in terms of inclusivity in research. Many of the participants in INCLUDED focus groups and interviews felt that they and members of their community knew very little about opportunities to participate in research (Box 3). Engaging with communities in this way may also help to increase knowledge of research opportunities and improve trust and engagement long-term.

Box 3: Quotes from participants in INCLUDED focus groups and interviews about settings and opportunities to participate in research

"if you're talking about Brazilian people there are groups, almost like pockets or even like scattered all over. And there are so many... And I think it's almost finding places where they usually concentrate in, in how you can utilise that. So from a more traditional community hub, for example, there's just this newly opened London LGBTQ centre, which is really great and they already do many things, just trying to bring the community together and help people. So that's a pretty obvious example." (Latin American)

"And I came into contact with one of the researchers and I said have you ever thought about going to BME communities and thinking about how they feel? They said oh... The word she used then, she said oh, they are sometimes very hard to reach. I said I don't think they're really very hard to reach, but have you got the right facilities or the right recruitment to reach out to these communities and bring them?" (South Asian Health Action)

"I didn't get a chance to do (take part in research) here in the UK, because I came, I started the job and as a social worker, after that I got a job in the NHS. ... I love to do, but I didn't get platform" (Social Action for Health)

Section 2: Grant application - communicating about ethnicity data collection to funders (recommendation 5)

During the INCLUDED study we discussed the importance of interactions with funders. This recommendation is concerned with communication to the funders by investigators at the grant submission stage.

5. Communicate to funders in the grant application, why and how ethnicity data will be collected and disseminated

5.1 *Why?* Participants in INCLUDED focus groups and interviews felt that communication with funders was important for transparency. It also puts the description of ethnicity data collection on the same level as other information that needs to be articulated in the funding application, and makes it clearer what finance is likely to be needed to do the data collection appropriately.

5.2 *How?* The Chief Investigator is responsible for ensuring that there is clarity amongst grant applicants on why collecting data on ethnicity is important, how these data will be collected, and how they will be presented in the results. Discussions described in section 1 should provide a good basis for communicating this information in the grant application. Depending on funder and type of application some discussion of recommendations 6, 7, and 8 below may also be needed prior to grant application submission and decisions may need to be communicated to funders in the application.

Section 3: Trial set-up - designing data collection (recommendations 6-8)

The recommendations in this section are focused on the design of documents including the participant information sheet and data collection instruments. These discussions may take place before grant application submission or in the set-up phase of the trial, or both.

6. On data collection forms, use a question which allows people to self-describe their ethnicity

6.1 *Why?* Asking individuals to self-describe avoids the problem of providing lists that inevitably exclude some individuals (see box 1). Many participants in the INCLUDED focus groups and interviews felt that their ethnic identity was either not represented or not accurately represented in current lists used for ethnicity data collection. For example, in the UK Office for National Statistics classification used in the 2021 census, there is no category for Latin American; only a few Asian countries are specified; and no African countries are specified because Africa is specified as a continent.[11] Those who designed this classification had reasons for doing so. Although it is never the intention of researchers who use such lists, the lists often feel exclusionary and discriminatory to those in diverse ethnic communities who are asked to complete them and cannot find themselves reflected in the categories they are presented with (Box 4). The more inclusive approach of allowing self-description may make people more likely to complete the question on ethnicity and improve data accuracy.

Box 4: Quotes from participants in INCLUDED focus groups and interviews about how collect ethnicity data collection instruments can feel exclusive and discriminatory

> *Discrimination, probably, the first thing as a foreigner coming in my mind. Now I know, so views .. have definitely changed, I know the procedure now. But as a foreigner, because that's not even explained why. So that seems like - okay, why are they questioning me like this"* *(Social Action for Health)*
>
> *It is 'fascist' and 'imperialist' to tell people they can only be 'this, this, this, this''. (Latin American)*

6.2 *How?* Without restricting what individuals can enter as their self-described ethnicity, in some instances it may be useful to provide some pointers as to what participants might want to think about when completing this question. An explanation of why collecting ethnicity information is important for the trial and for people like themselves is likely to be useful. Given that dictionary definitions of ethnicity are not consistent it may also be useful to say something about what the researchers mean by ethnicity (see quote at start of this document for one example). Involve those from diverse ethnic communities when writing this text to ensure that the question and any explanation about ethnicity will be clear to the communities they come from.

7. Make an explicit decision about whether a pre-specified list of ethnic categories will be used in the trial for data collection in addition to collecting self-described ethnicity

7.1 *Why?* The strong recommendation of the INCLUDED team is to use self-described ethnicity (recommendation 6). However, it may occasionally be necessary to use a pre-specified list of ethnicity categories in addition.

There are arguments for and against using a pre-specified list for collecting ethnicity data in addition to self-description. Both views were expressed by participants in the INCLUDED study, and the INCLUDED research team recognised the validity of both views. We are therefore suggesting that the decision about using a pre-specified list for data collection or not needs to be made for each trial, though for most trials such a list will not be necessary (note that this does not preclude grouping self-described responses for analysis if necessary and appropriate) and such a discussion may therefore be short.

Arguments for using a pre-specified list are that it can potentially provide immediate categories to be used in analysis and this may be an important consideration depending whether ethnicity data are to be included in an analysis, and the type of analysis. If the same categories are used consistently across different trials this will facilitate comparisons and synthesis. However, currently the advantage of consistency is not being realised: in the INCLUDED survey of recently completed trials we found no consistency in the use of lists for ethnicity data collection. Rather, there was substantial variation in terminology and length of lists, with almost every study surveyed using a different list.

Arguments against using a pre-specified list are that existing lists are generally disliked by those from diverse ethnic communities because they are not felt to represent the UK population and can feel exclusive and sometimes discriminatory (see section 6.1 and Boxes 1 and 4). Those in our focus groups and interviews were, however, realistic about the seemingly impossible task of constructing a list covering all possible ethnicities of those who might be asked to complete them because of practical constraints on the length of a list. We presented these participants with a newly devised list focused on continent and country with no mention of skin colour and which allowed multiple boxes to be ticked. This list had been carefully designed to be fully consistent so that every respondent was able to identify both a continent and country that they felt reflected their ethnicity (although not every country was listed new countries could be added). The list was felt by some to be somewhat better than, for example, the standard Office of National Statistics list, but far from ideal as some groups of individuals (eg Arab and Latin American individuals) may not see their primary ethnic identity as related to the major continents or country. Thus, it appears that constructing a 'good' list is not straightforward, and any list that is used risks alienating some participants from diverse ethnic communities. Further research is needed in this area.

7.2 *How?* Individuals from diverse ethnic communities should be involved in these discussions, as should those who are familiar with data collection, analysis and reporting. It appears that, in most trials, ethnicity data is used to describe the trial sample and not for subgroup analysis (see section 2.2). Therefore, in thinking about using a pre-specified list, the main considerations are likely to be whether adding such a list to data collection instruments will enhance inclusivity, or transparency, or the feasibility of describing the sample (see also Recommendation 12). If a pre-specified list is used, there may be considerations specific to the disease area that need to be taken into account in constructing an appropriate list. Individuals from diverse ethnic communities, including community connectors, will bring important perspectives on the consequences of using a list, the categories to be used if a list is decided on, and clarity of terminology. Those familiar with data collection will bring important perspectives on the practicalities of collecting the data.

If it is decided to use a list, it is important to:
- explain if the list has been used elsewhere and why it is being used in your trial
- explain how those with mixed ethnicity should respond, for example, being able to specify more than one category, making it clear that this is their assessment
- if respondents are not able to choose more than one category, avoid using categories that are not mutually exclusive and may therefore cause confusion, for example, "White" and "British"
- ensure that the tool is as inclusive as possible of all ethnic groups that could potentially be included in the trial. For example, if the trial participants come from a city, these tend

to be multicultural places that can include almost any ethnicities, and diversity in terms of ethnicity may be different in different trial sites.

- avoid the category "other" as this continues to marginalise those groups that are not explicitly listed.

Apart from trials that specifically target a small number of well-defined ethnic groups we expect it to be hard or impossible to create a list that is truly inclusive for the reasons given in Recommendation 6. We are also aware of the challenges that may arise from using a single question to collect ethnicity data, given the multiple attributes that can potentially comprise ethnicity. Investigators who use a single self-description question to collect ethnicity data could usefully contribute reflections to further research in this area.

8. Communicate to participants why ethnicity data are being collected, what will be collected, and how they will be used

8.1 *Why?* While focus group participants understood the importance of collecting ethnicity data, they said that researchers need to be much better at communicating what the data will be used for, how the data will be used and what makes the data useful. Doing this is likely to improve participation of diverse ethnic communities in specific trials. It may also increase engagement in research more generally as researchers become better at informing people about the potential benefits to their diverse communities of taking part in (and helping to design) research that values and acknowledges their needs and perspectives.

8.2 *How?* Any explanations should relate to the purpose of the trial and to the communities involved. An explanation of *why* ethnicity data are being collected could be included within the participant information sheet and/or during the recruitment discussion, as could an explanation of *what* ethnicity data will be collected if this is feasible. Communication about *what* ethnicity data will be collected and how it will be used is likely to be useful at the time of data collection. Videos may be useful for these sorts of communication. Decisions about when and how to do this communication will need to be made within each trial. Community connectors may be able to advise on the best methods to use for their own communities, and multiple methods may be required to ensure effective communication with diverse groups.

Section 4: Trial set-up - training and responsibility (recommendations 9-11)

9. If possible, source high quality training for the whole trial team focused on better understanding of the culture of, and building trust with, diverse ethnic communities

9.1 *Why?* Focus group and interview participants wanted to see more ethnic diversity in research. To achieve this, they suggested that researchers needed to (1) learn more about diverse ethnic communities, including cultural and religious sensitivities; (2) move away from tokenism in including ethnically diverse communities; (3) engage with communities by working with grass roots organisations and local community stakeholders; (4) build trust with communities; (5) build understanding about how to widen research participation. In the INCLUDED survey, some researchers on recently completed trials expressed surprise about those from diverse ethnic communities involved in their trials still referencing historical abusive events occurring in and outside health settings and health-related research that contribute to mistrust

of research amongst such communities. One example of a historical abusive event is the Tuskegee study, which finished in 1972. This study involved African-American men with syphilis. Investigators wanted to observe the effects of the disease when untreated. Medical advances meant that syphilis became entirely treatable during the course of the study but the men were not informed: more than 100 participants died as a result.[12] Participants in INCLUDED focus groups and interviews also specifically mentioned racism in health care (Box 5).

Box 5: Quotes from participants in INCLUDED focus groups and interviews about reasons for mistrust of research including providing ethnicity data

> *"To me, frankly, even if they explain to me I always am suspicious. I'm sorry to say. I have always this suspicion that maybe there's more to it than meets the eye. Do you understand? So that okay, they will say we are doing this because we want to know exactly. But because of the past racism and so on, I always have this feeling that I don't trust it."(Caribbean and African Health Network)*
>
> *"…….where we say we don't trust, the thing that we're not trusting with is our data or the information we might be providing as the question earlier said. I think what we're finding, a lot of the forms that we find is, for example, I'm onto a diabetes study.., but they want to know my address, my name, how long I've lived in the UK, whether I've got a driving licence and that. First of all, how is this information relevant, and then also on the questions…"(South Asian Health Action)*
>
> *"I contacted these weight management people... And she was asking my questions, what is your nationality? And I said I'm British. She again repeated the question because she wasn't happy with what I said, from my accent she was probably thinking she's not British. And she again asked me that question, I said British. What do you mean? What do you want to know from me? And she said no, no, no, where are you from? Where are you born? And I said that's what you want to know. And I said it's a need to know basis, I need to know why you are asking me that question. I don't mind telling you I'm British Indian. But the question you asked me specifically was your nationality. I'm British. It makes me so angry, it makes me so angry that time. I said you shouldn't be asking me this question. But the way some people don't know how to ask questions, how to explain that question. But this is how even the person who probably knows… And imagine if someone doesn't understand."(South Asian Health Action)*

9.2 *How?* Trial teams will need to liaise with those who might be providing such training. This can include universities, trials networks, funders and other researchers. Cultural competence training has become popular in recent years. However, there has been some criticism that although such training tries to help trainees to recognize how culture, race and identity define health among diverse ethnic communities, it fails to acknowledge how members of dominant cultural groups are similarly influenced by a culture of their own.[13] We recommend that as well as covering the five points above, there is also time within training for members of the trial team to reflect on their own culture in order to understand how this may influence the research. One good example of such training is run by The Centre for Ethnic Health Research.[14] Guidance from Trial Forge contains a similar recommendation to the one discussed here, and further resources.[15]

10. Talk as a trial team about inclusion and diversity

10.1 *Why?* Responsibility for inclusion and diversity should lie with the whole trial team. In relation specifically to ethnicity recording, such conversations are likely to increase the understanding of the whole team about how to improve this. In addition, the research community is less diverse than the UK population.[16] From the focus groups and interviews in INCLUDED we learnt that many of those from diverse ethnic communities would feel more comfortable if they could see themselves reflected in the make-up of research teams. It is clearly impossible for all ethnicities to be represented within a single research team. Nevertheless, a team is likely to be better able to widen participation and include individuals from diverse ethnic communities appropriately if the team reflects on the effect their own ethnic make-up might have on engagement, as well as specific aspects of data collection. For example, language can be a barrier and was specifically mentioned as an issue in the INCLUDED focus groups (Box 6).

Box 6: Quotes from participants in INCLUDED focus groups and interviews about language being a barrier

> *'..because language is not only like a community skill. Of course, it's a basic. But it's also part of our culture. So by maybe knowing the language you can better understand the culture.'(Latin American)*
>
> *'And I think probably the language barrier is one of the main things in ethnicity, in ethnic groups. So whoever is being given these forms, they should be made clear whoever gives them, why. Even though it's written, it should be written on the forms. But people pay less attention on reading. It's more on listening I guess." (Social Action for Health)*
>
> *I think they (in the Office of National Statistics categorisation) do have that bit where it says any other Asian. They do give you the option of… But my only thing is we know that, but maybe someone older might not know or someone with a language barrier might not actually know where they fit in, kind of. So I'm thinking…I don't know, they might be put off by that to be honest." (Social Action for Health)*

10.2 *How?* Involve those from diverse ethnic communities in these discussions. Discussions could include whether inclusion and diversity need adding to site set-up meeting agendas, whether standard operating procedures, checklists or reminder documents would be useful. Those from organisations or networks that link researchers and community groups might have experience to bring to running such discussions.

11. Identify an individual or group trained in leading inclusion and diversity work who can support the inclusion and diversity work within the trial

11.1 *Why?* The work in the INCLUDED study has identified a number of ways that trials could improve the design of ethnicity data collection, data collection itself, and the analysis and reporting of these data. Much of this improvement relates to the wider issues of inclusion and diversity in trials, and indeed it was one of the aims of the INCLUDED study to improve inclusion and diversity by considering how ethnicity data handling in trials could be improved. However, the INCLUDED research team recognises that research schedules are usually governed by tight timetables and deadlines. It is therefore likely to be helpful to identify someone or a group of

people who have a focus and responsibility for the inclusion and diversity work to make sure it does not get overlooked.

11.2 *How?* It may be that an individual or a group of people who have a focus and responsibility for the inclusion and diversity work sit outside the trial team, for example within a clinical trials unit, or a wider department. If this is not the case, an individual or small group within the trial team could take responsibility. Either way, there may be resourcing issues, particularly if there is staff turnover. The Chief Investigator would need to be responsible for ensuring such individuals or groups have an overview of the inclusion and diversity work, are clear about their responsibilities, and have time to contribute to meetings as and when appropriate.

Section 5: Reporting (recommendations 12 and 13)

12. When disseminating results, report ethnicity as transparently as possible

12.1 *Why?* Trialists should report on ethnicity data collected. This is in keeping with good practice to use and report data provided by participants. However, of the recently completed trials in the INCLUDED survey, 18% did not report ethnicity at all even though they had collected ethnicity data. Many more did not report ethnicity as they had collected it, often just reporting white and "other" even though they had collected data using categories that were more detailed than this. There was no explanation at the time of dissemination as to why ethnicity categories were collapsed into 'other' although some explained during the INCLUDED study that this was because there were very few participants in their trial who were not white. Thus, both reporting and transparency in reporting needs improving.

12.2 *How?* The way that ethnicity data are reported should be related to the purpose for which they are collected (see section 2). It is particularly important that ethnicity data are reported in the journal paper or report that contains the main results, preferably in the baseline table that contains information on other participant characteristics. Authors should state that self-described ethnicity was collected.

We recommend that investigators report categories *as collected* to maximise transparency and inclusivity. However, self-description may mean a much longer list of categories than currently used in baseline tables of most trial reports. If a longer list in a baseline table proves an obstacle to publication, investigators should consider longer tables in supplementary material, or appropriate explanations in footnotes. In the INCLUDED study we did not discuss these various options for publication with journal editors and publishers, and this requires further work.

Furthermore, reporting self-described ethnicity as collected could also mean that some categories have very few individuals in. If this is the case, discussions could take place with appropriate individuals to ensure that reporting is as transparent as possible while not breaching confidentiality or legislation such as the UK's General Data Protection Regulations.

If a pre-specified list has been used in addition to self-described ethnicity (see Recommendation 7), then the categories in the list should be described in the report. The work to ensure that categories in any list are appropriate to meet the trial purposes (including analysis requirements) should have been done at an earlier stage (Recommendation 7). Thus, categories used in lists should not be combined unless this is to ensure confidentiality. For transparency, an explanation for such combination should also be provided.

It may also be appropriate to follow other recommendations in terms of reporting ethnicity (CONSORT, journal recommendations e.g. JAMA (The Journal of the American Medical

Association). Currently work is being undertaken to extend CONSORT/SPIRIT for ethnicity data collection.
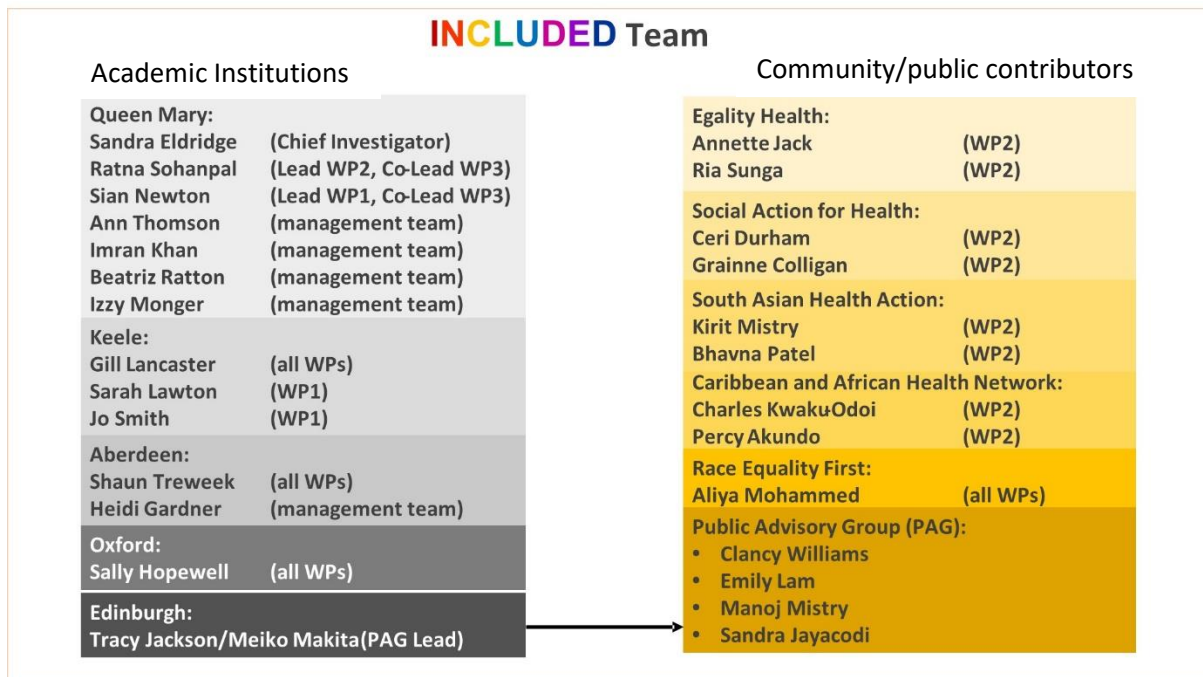
13. Include ethnicity data when reporting to participants, to PPIE members especially diverse ethnic community members, and to community connectors involved in designing the ethnicity data collection

13.1 *Why?* There are existing recommendations regarding the need to provide participants with the results of research they have contributed to.[17] To ensure that those who participated can understand how providing ethnicity data contributed to the purpose of the trial, Recommendation 13 extends existing recommendations to cover the inclusion of ethnicity data. Feeding back to PPIE members is recognised as part of good practice, and this recommendation extends this good practice to feeding back to all those involved in designing the ethnicity data collection. In addition, this dissemination will enable participants and community stakeholders to see the ethnicities of those involved in the research, which could potentially build trust.

13.2 *How?* In feeding back results to diverse ethnic communities that have participated in the research, individuals from these communities should be consulted over the reporting of the ethnicity data, including over the methods of dissemination which may include podcasts and videos as well as written documents and presentations. The NIHR Research Design Service Guidance also provides some general guidance about feeding back results.[17]

The INCLUDED team

The following diagram shows those who worked on the INCLUDED study and their involvement in it.

Acknowledgements

References

1. https://jech.bmj.com/content/jech/58/6/441.full.pdf
2. Report on INCLUDED survey available on request
3. https://www.trialforge.org/trial-forge-centre/include/
4. https://doi.org/10.1186/s13063-022-06747-2
5. https://www.nature.com/articles/pcrj200233
6. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158783
7. https://www.npeu.ox.ac.uk/mbrrace-uk/reports
8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9102088/
9. https://doi.org/10.1186/s13063-022-06973-8
10. https://www.peopleinresearch.org/
11. https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371
12. https://www.tuskegee.edu/about-us/centers-of-excellence/bioethics-center/about-the-usphs-syphilis-study
13. https://in-training.org/critique-cultural-competency-10456https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups#2021-census
14. https://ethnichealthresearch.org.uk/training-courses/cultural-competence/.
15. https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-022-06553-w
16. https://www.nihr.ac.uk/documents/about-us/nihr-diversity-data-report-2021-22.pdf
17. https://www.rdsresources.org.uk/dissemination-and-implementation

INCLUDED recommendations 20/03/2023