

Therapeutic misconception in early phase gene transfer trials

Gail E. Henderson*, Michele M. Easter, Catherine Zimmer, Nancy M.P. King, Arlene M. Davis, Barbra Bluestone Rothschild, Larry R. Churchill, Benjamin S. Wilfond, Daniel K. Nelson

School of Medicine, University of NC at Chapel Hill, CB# 7240, Chapel Hill, NC 27599 7240, USA

Available online 5 July 2005

Abstract

Many subjects in early phase clinical trials expect to benefit in some way from the research intervention. It is understandable that people hope for improvement in their condition, no matter what the evidence. Yet unreasonable expectation of medical benefit may reflect problems with informed consent: Investigators may not disclose clearly that direct medical benefit from an early phase experimental intervention is unlikely or impossible, or subjects may not appreciate the differences between treatment and research. This paper presents findings from recent interviews with researchers and subjects and analysis of consent forms in early phase gene transfer research, a cutting-edge technology often called ‘gene therapy’. We use three variables to construct a composite measure of therapeutic misconception TM, tapping misconceptions about the purposes of early phase research and the potential for direct medical benefit in these trials. Our multivariate model demonstrates the importance of both subject- and study-level factors as predictors of this TM index: education, disease type, and communication by study personnel about the likelihood of benefit. We hope that this work will deepen the discussion of how to define and measure TM, and refine the specification of factors that are related to subjects’ TM.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: USA; Clinical trials; Informed consent; Therapeutic misconception; Gene transfer research; Benefit

Introduction

Many subjects in early phase clinical trials are motivated by the expectation that they will benefit in some way from the research intervention (e.g., Ackerman, 1995). This may be especially true when there is a diagnosis of terminal illnesses for which standard treatments have been exhausted (Dresser, 2002), or when scientific and lay publications promote an exciting new technology (Churchill, Collins, King, Pemberton, & Wailoo, 1998). It is understandable that people hope for improvement in their condition, no matter what the

evidence. Yet unreasonable expectation of medical benefit from early phase trials may be the result of problems with informed consent. Investigators may not disclose clearly that direct medical benefit from an early phase experimental intervention is unlikely or impossible, or subjects may not appreciate the differences between treatment and research.

The term ‘early phase’ refers to small studies (phase I) of an experimental intervention that assess safety and side effects with increasing doses, and to slightly larger studies (phase II) designed to begin an evaluation of effectiveness at a dose level found to be safe, as well as to continue to test for safety and side effects. Effectiveness in phase II trials is often measured by changes in laboratory values that may be surrogates for clinically

*Corresponding author.

E-mail address: ghenders@med.unc.edu (G.E. Henderson).

meaningful measures of how a patient feels, functions or survives (Temple, 1995). Empirical evidence demonstrates that early phase studies hold far less potential for improved clinical outcomes for participants than phase III studies, which are designed to test the experimental intervention against standard treatment or placebo on a sample large enough to demonstrate whether the intervention is effective (Horstmann et al., 2005).

Subjects' misunderstandings of the nature of early phase trials have been an object of scholarly investigation, both conceptual and empirical. Appelbaum, Roth, and Lidz (1982) first defined 'therapeutic misconception' (TM) in subjects¹ as the mistaken belief "that the research, like the therapy [subjects] have received previously, is designed and will be executed in a manner of direct benefit to them". While the original definition was primarily applied to phase III trials, empirical studies have subsequently confirmed that subjects in early phase trials misunderstand both their purpose and their potential to provide direct medical benefit (e.g., Daugherty et al., 1995). Many of these studies have been conducted with subjects in phase I cancer trials, the majority of whom think they will "get medical benefit from the treatment in this study" (Daugherty, Banik, Janish, & Ratain, 2000) and are unaware of "the unproven nature of treatment and the uncertainty of benefits to self" (Joffe, Cook, Cleary, Clark, & Weeks, 2001). More recently, Horng and Grady (2003) argue that it is misunderstanding the nature and intent of research that is most ethically problematic, while misestimating the probability of direct benefit may be less worrisome.

Empirical studies have clearly documented the presence of TM but, reflecting lack of agreement in the conceptual literature, they have not defined or measured it in the same way, nor have they differentiated studies of subjects in early phase from later phase trials. Most studies have used simple "yes-no" or forced-choice questions about whether subjects expect medical benefit, how likely they think it is, or what motivates them to join a trial. Two studies have employed measures that attempt to tap more than one dimension of TM, thus acknowledging that it is a complex phenomenon. Joffe et al. (2001) developed a knowledge score for subjects that combines questions about the basic elements of informed consent and trial-specific questions about the purpose of clinical research and the difference between research and medical treatment. Most recently, Appelbaum, Lidz, and Grisso (2004) created a qualitative measure of TM that gives equal weight to two components: inaccurate beliefs about individualized

treatment and unrealistic appraisals of the likelihood of medical benefit.

Studies also differ in the choice of variables theorized to predict TM. Most have focused on individual characteristics such as older age, lower education, illness severity, and being recruited by one's physician (e.g., ACHRE, 1995). Others have examined trial characteristics (e.g., Schaeffer et al., 1996) and various aspects of consent forms and the consent process, including who is present (e.g., Cheng et al., 2000). There is no consistency across these studies regarding which variables are studied or how they are defined; moreover, there are few attempts to include both subject- and study-level variables, and only one published study attempts to jointly assess their contribution through multivariate analysis. In that study, Joffe et al. (2001) found that "improved knowledge" among cancer trial subjects was significantly related to subjects' age and education, and to four factors that measured aspects of the consent process and the quality of consent forms.

In this paper, we use a unique set of semi-structured interviews from early phase gene transfer trials to examine predictors of TM, broadly defined as misunderstanding the meaning of research. Gene transfer research (GTR) has been applied to a wide range of disorders. The majority are cancer trials, thus providing a natural comparison to the literature on TM in phase I oncology trials, although inherited disease, infectious disease, and peripheral and coronary artery disease are also represented. Since the first trial in 1989, GTR—often known as "gene therapy"—has inspired scientists and the public alike with the possibility of a genetic treatment for medical conditions (Churchill et al., 1998; Friedmann, 1996; King, 1999). As a "cutting edge" technology, GTR is an appropriate place to investigate the presence of TM among study subjects.

We employ three variables from our interview questions as measures of the concept of TM. While each has strengths and weaknesses, in combination they tap fundamental misconceptions about the purposes of early phase research and the potential for direct medical benefit in these trials. Our multivariate model demonstrates the importance of both subject- and study-level factors as predictors of this TM index. While our approach is exploratory, we hope that this work will deepen the discussion of how to define and measure TM, and refine the specification of factors that are related to subjects' TM.

Methods

Sample

As part of a broad study of benefit in early phase gene transfer research (GTR), we contacted principal inves-

¹We also recognize the potential for therapeutic misconception in investigators, as do Appelbaum and colleagues, but do not address it in this paper (see Henderson et al., 2004).

Table 1
Distribution of eligible trials, and study sample by disease and phase

Study characteristics	A	B	C	D	E	F	G
Total GTR protocols 1990–2001 ($N = 457$) frequency (%)		Eligible GTR studies 12/98–12/00 ($N = 78$) frequency (%)	Studies sampled ($N = 19$) frequency (%)	Studies not sampled ($N = 59$) frequency (%)	Subjects interviewed in C ($N = 68$) frequency (%)	Subjects not interviewed in C ($N = 70$) frequency (%)	Number of subjects interviewed per study: mean (range)
Total	457 (100)	78 (100)	19 (100)	59 (100)	68 (100)	70 (100)	
<i>Disease</i>							
Cancer	314 (69)	57 (73)	9 (47)	48 (81)	44 (65)	54 (77)	4.9 (1–19)
Vascular	52 (11)	9 (12)	3 (16)	6 (10)	10 (15)	8 (11)	3.3 (1–5)
Inherited	54 (12)	7 (9)	5 (26)	2 (3)	8 (12)	4 (6)	1.6 (1–3)
Infectious	37 (8)	5 (6)	2 (11)	3 (5)	6 (9)	4 (6)	3.0 (2–4)
<i>Column comparisons</i>	A–B	B–C B–D	C–D*		E–F		
<i>Phase</i>							
I	—	44 (56)	10 (53)	34 (58)	22 (32)	26 (37)	2.6 (1–8)
I/II	—	13 (17)	4 (21)	9 (15)	26 (38)	22 (31)	6.5 (1–19)
II	—	21 (27)	5 (26)	16 (27)	20 (29)	22 (31)	3.2 (1–7)
<i>Column comparisons</i>		B–C B–D	C–D		E–F		

* $p < 0.05$ for Fisher's exact test.

tigators (PIs) of all trials registered with the Office of Biotechnology Activities (OBA) as early phase “therapeutic” GTR studies in adults between December 1998 and December 2000 ($N = 123$). 78 studies were eligible for inclusion, and 19 had one or more subjects willing and well enough to participate, producing our sample of 68 subject interviews. These 19 clustered datasets, including PI and study coordinator (SC) interviews² and at least one subject interview from the same study, make up the sample for the present analysis. The final, IRB-approved consent forms used in the 19 trials were obtained from the PIs, who gave permission to link the consent form analysis (King et al., 2005) to interview data. The project was approved by Institutional Review Boards (IRBs) at the University of North Carolina and the National Human Genome Research Institute. Interview and consent form instruments are available at <http://socialmedicine.med.unc.edu/scob/>.

Table 1 shows disease and phase distributions for the 78 eligible studies, the 19 sampled studies, and the 59 studies not sampled, as well as subjects from the 19 studies who were interviewed ($N = 68$) and not interviewed ($N = 70$). The distribution of phase is not statistically different for sampled and non-sampled

studies (compare column C–D), nor for subjects interviewed and not interviewed (E–F). However, compared to the 19 sampled studies, a significantly larger proportion of the 59 non-sampled studies were cancer trials (C–D) (Fisher's exact test, $p = 0.006$). Despite under-representation of cancer trials in the sampled studies,³ individuals in cancer trials are not underrepresented because sampled cancer trials had more eligible subjects than did other disease trials. The proportion of subjects in each disease category among those interviewed is similar to those who declined to be interviewed. Thus, our final sample was representative of the larger universe of GTR studies.

About half of the studies were multi-center trials, and half had corporate sponsors. With one exception, the trials were conducted at academic or governmental medical centers. All PIs in the 19 studies had medical degrees; 2 had Ph.D. degrees as well. Most PIs were male (13/19), while most SCs were female (15/18). Fourteen SCs had nursing degrees, and 2 had medical degrees. The average age was 47 for PIs and 45 for SCs. All PIs

³This may reflect the severity of illness experienced by subjects in cancer trials in which no one responded to our recruitment materials, or study subjects were judged by their researchers as too ill to recruit.

²One of the 19 trials had no study coordinator.

and most of the SCs had extensive clinical research experience; 8 of the PIs and 5 of the SCs had conducted at least one other GTR trial.

Interviews

Most of the 45-min telephone interviews were conducted by three of the authors (GH, AD, and ME) and completed between July 2000 and June 2002. The interviews focused on recruitment and enrollment of subjects in the particular trial, and the procedures involved in the conduct of research. All PIs, SCs, and subjects were asked about why subjects joined the study, and their expectations for medical and other benefits from participation. PIs and SCs were asked to describe their roles and relationships with subjects in the study; subjects were asked whether they saw the PI and SC as mostly taking care of them as patients or mostly conducting research, and about the overall intent of the study as care or research. Subjects were also asked questions about their participation in GTR in order to elicit reflection upon the meaning of this new genetic technology for them.

Quantitative interview data were entered directly into a computer-assisted telephone interview (CATI) software program. Interviews were taped and transcribed, then electronically coded using QSR N6 software (version 6.0, QSR International Pty. Ltd. 1999). Qualitative codes were developed and applied by teams of two investigators, with differences reconciled in discussion with four or more investigators (Miles & Huberman, 1994). Consent forms were coded by three of the authors (NK, GH, and AD) using a 94-item coding form developed for assessment of all GTR consent forms (King et al., 2005). Kappa scores, measuring coder agreement, were within the moderate range (Landis & Koch, 1977).

Measurement

Dependent variable: composite therapeutic misconception index

We constructed a composite measure of TM by combining quantitative and qualitative data from three interview questions. Each question addresses a different aspect of TM identified in the literature, and resembles measures found in prior studies: joining in order to get benefit, misestimating the likelihood of direct medical benefit, and misconceiving the purpose of the study. The latter two are conceptually similar to measures used by Appelbaum et al. (2004), although methodologically they are distinct because coding was restricted to responses to two specific questions rather than entire interviews. The three components of our index are described below.

Reason for joining. Subjects were asked an open-ended question: “Let’s talk about your decision to join the [XX

study]. Why did you decide to join?” Their reasons for joining were coded according to whether they mentioned only benefit to science and society, only benefit to themselves, or both. This taps external influences on subjects’ perceptions about the trial, prior to joining, and their general understanding of the purpose of early phase research and what is reasonable to expect. Some said they wished to benefit society by contributing to medical science:

Oh, just they needed volunteers and you should do your part. I’d never been in a study before and I know other people have done it and they’ve taken the risks so... I should also.

Other respondents joined both to advance science and to seek benefit for themselves—benefit derived from the GTR intervention (direct benefit) or from standard treatment or monitoring received as part of the study.

Well, the first answer to that would be the obvious fact of personal gain.... The second reason which is pretty high on my list as well is that... maybe by doing this study, they could help somebody like me in the future....”

Most respondents mentioned only benefit to themselves:

Why? Because I had nothing else that they could do for me and I was in a lot of pain and I didn’t want to lose any more toes or lose my leg possibly. So at that time the pain that I was in, I would have went with anything that anybody offered me.

Expectation of benefit. Subjects were asked, “Did you expect that getting the gene transfer would improve your condition or help make you better? Would you say yes or no?” This question targets their expectation of direct benefit from the gene transfer specifically, not from standard treatments or from monitoring that may have been available on-study. After choosing an answer (yes, no, or do not know) they were asked what they expected and why. We coded transcripts to capture whether yes or no answers were qualified with language conveying uncertainty, such as “Yes, it’s possible,” or “No, but I hope.” “Yes” answers were coded as either “unqualified” or “qualified”:

Yes... because I believe in the concept behind what they’re trying to do. I think that that’s a very valid and sound approach. (“Unqualified yes”)

Yes... I never know what might hurt or help me but I’m always hoping that there is something that will help me. And I was just wishing for the best. (“Qualified Yes”)

“No” answers were similarly coded as “unqualified” or “qualified”:

No... because it was only a test study... it wasn't going to have any effect on my condition at all. (“Unqualified no”)

No... you cannot expect a definite improvement on something that is still being studied. One can hope for that to happen... but it is not a definite. (“Qualified no”)

A few said they could not choose yes or no, usually because they hoped for benefit but were uncertain. Thus, this variable had five categories: unqualified no, qualified no, do not know, qualified yes, and unqualified yes.

Study purpose. The third question was about subjects' perception of the primary intent of the study: “Thinking about the [GTR study], would you say it was *mostly* intended to help you as a patient, or *mostly* intended to gather knowledge?” This is a question about understanding the purpose of a study as research. Again respondents were asked why they chose their answers. Coding distinguished between those who viewed the study as having a dual intent (whether they emphasized one over the other or indicated that both were important), and those who appeared to view the study as having one intent without mentioning the other. The final code included five categories, illustrated below: mostly research, both but mostly research, both equally, both but mostly helping them as patients, and mostly helping them as patients.

Well, they have to have people that they can do research on and make sure that the research is validated and make sure that it's done right, and I think they have a quest to do good with this knowledge but that's what they're after right now is the knowledge whether this therapy works. (Mostly research.)

I think it's a combination of both... It's probably seventy percent to gather knowledge and probably thirty percent to take care of me as a patient... If you're taking care of me as a patient, then you would discuss radiation and hormonal treatment, pros and cons of each... [with the study doctors] it was really more the focus was on this clinical trial itself, which I would expect it to be. (Both but mostly research.)

They wanted to help heal me, and at the same time, they wanted to learn about this new drug and how it affected me and hopefully, it'd affect other people and so on. (Both equally.)

Mostly to take care of me as a patient... the other was almost secondary [because] they were both concerned

with the individual, more so than the study. But then the study is very important, too. (Both but mostly helping them as patients.)

Because there's concern about the patient... I guess all doctors and nurses are concerned for helping their patients and enlighten[ing] them on their illness. (Mostly helping them as patients.)

From the three variables just described, we created one composite additive index of TM. First, the quantitative and qualitative results were assigned numerical values ranging from 1 to 5, with higher numbers corresponding to greater TM. The three components of the measure for TM are displayed in Table 2. The median categories for each variable are: joining to gain personal benefit (a score of 5), direct benefit expectation is a qualified yes (4), and the intent of the study was both to gather knowledge and to help patients (3). The results for these three variables were added together to create an index ranging from 3 to 15. Cronbach's alpha coefficient of internal consistency for the TM index is 0.74.⁴

The distribution of the TM index across individual subjects ranged from the lowest possible score of 3 (respondent emphasized societal benefit as a reason for joining the study, had no expectation for direct medical benefit from the intervention, and thought the study was mostly intended to gather knowledge), to the highest possible score of 15 (respondent joined to get personal benefit, expected individual benefit from the intervention, and thought the study was mostly intended to help him/her as a patient). Transcript excerpts from subjects at each extreme are provided in Table 3. As expected from the high median and modal values for component variables, higher TM index scores are more common (median TM = 11, mean = 10.6).

Respondents with the same TM score combined the three components in different ways, but an individual's scores on the 3 components usually did not differ widely. Only 8 subjects scored both 1 and 5 (i.e., opposite extremes) on any two-component variables, and in each case the subject scored 1 on study purpose. For the 14 subjects whose responses fell in the modal category (10), there were four different combinations, two of which are illustrated in Table 3. The second is one of the 8 subjects who scored 1 on study purpose and 5 on one of the other

⁴Further evidence of how well the three components of TM are related comes from a confirmatory factor analysis. All three components were significantly related to the latent TM variable. The factor loadings were approximately the same size, indicating each component is an equal reflection of TM as we conceptualize it. When factor scores were used to create a weighted version of our additive index, it was correlated with the original unweighted index at 0.99. As such, weighting is unnecessary.

Table 2
Distribution of components of composite TM index

Score	Reason for joining	Frequency (%)	Expectation of direct benefit	Frequency (%)	Intent of study	Frequency (%)
1	Societal benefit	6 (8.82)	Unqualified no	3 (4.41)	Only research mentioned	20 (29.41)
2	—	—	Qualified no	10 (14.71)	Mix, mostly research	9 (13.24)
3	Both societal and personal	20 (29.41)	Do not know	4 (5.88)	Both equally	24 (35.29)
4	—	—	Qualified yes	24 (35.29)	Mix, mostly helping as patients	8 (11.76)
5	Personal benefit	42 (61.76)	Unqualified yes	27 (39.71)	Only helping as patients mentioned	7 (10.29)
Total		68 (100)		68 (100)		68 (100)

The numerical score i.e. the basis for the composite TM variable is shown in the leftmost column. The median categories for each variable are bolded.

components of the index. It illustrates that subjects could be motivated to join for personal benefit, and expect it from participation, but also understand that the intent of the study was to gather knowledge.⁵

Independent variables: subject and study/researcher characteristics

Twelve subject-level and nine study-level independent variables were hypothesized to have effects on the degree of TM, as represented by the composite index described above. Table 4 displays frequencies for the independent variables and their mean TM index scores by category.⁶

The last four study-level independent variables assess messages about benefit from the gene transfer intervention. PIs and SCs were asked to describe how they discussed the prospect of direct benefit with subjects. Their answers, and their consent forms, were coded to capture the precise language used to describe the likelihood of any benefit.⁷ We created a composite

variable to represent the combined influence of all three sources: whether all three stated that direct benefit was unlikely (3 studies comprising 5 subjects), whether one or two stated that direct benefit was unlikely (5 studies, 8 subjects), or whether none stated that direct benefit was unlikely (11 studies, 55 subjects) (Henderson et al., 2004).

Analysis

We treated our dependent variable, the index of TM, as a continuous variable. All independent variables except age were categorical, and to measure their relationships with the TM index we tested the relationship of each category against a reference category using ordinary least squares linear regression. If the effects of two categories were not statistically different, we combined the categories unless they were non-proximate ordinal values (e.g., infectious and inherited disease were combined, but high school and graduate education were not.) Our small sample size and high number of tests suggested the use of a low alpha cut-off for p -value, but the exploratory nature of the study suggested a high alpha cut-off to minimize the chance of missing factors of potential importance. We set the significance level at $p < 0.05$ for our typical one-tailed tests and $p < 0.10$ for the one two-tailed test (gender).

Because our small sample size limited the number of independent variables we could use in a multivariate regression, we screened variables in two steps. First, we dropped independent variables not significantly related to TM in bivariate regression (including independent variables significantly related in the opposite direction from that hypothesized). Second, we nested independent variables in two separate multivariate models, one for

⁵We re-ran the final regression model (Table 5) without these 8, arguably subjects whose TM scores might be questioned by researchers with a different definition of TM than ours, and the results did not change.

⁶Details about question wording, response options, and coding categories are provided in Appendix 1, available from the authors upon request.

⁷No PI, SC, or consent form conveyed a message that direct benefit was likely or probable. All used language in which the likelihood of direct benefit was difficult to determine: “you may not benefit”; benefit is “possible” or “hoped for”; there is “no guarantee”; or the chance of benefit is “unknown.” A minority of PIs, SCs, and consent forms also included more specifically negative language, saying that direct benefit was “unlikely” or not expected, or that subjects “will not benefit” from the experimental intervention.

Table 3

Excerpts from four representative subject transcripts illustrating component responses and derivation of composite therapeutic misconception index

TM value	Score	Question (code)	Subject response
3	1	<i>Why join?</i> (Societal benefit)	I refer to it as, as paying interest on borrowed time that I've been living on.... You know the medical community has kept me alive a long time and you know this is an opportunity for me to give something back.
	1	<i>Expect benefit?</i> (Unqualified no)	Absolutely no.... Because that's what they told me.
	1	<i>Intent of study?</i> (Research)	The study as a whole was intended to gather knowledge... cause that's what the study's for, you know, the study is not to help me as a patient, it's to just to find out if this works.
10	5	<i>Why join?</i> (Self)	...I thought there was minimum risk and there was potential benefits to doing it, to me.
	4	<i>Expect benefit?</i> (Qualified yes)	I would say yes... because of what the doctor said and what [SC name] said that there was no guarantee but there was a good possibility that this technique would work.
	1	<i>Intent of study?</i> (Mostly research)	I think what the contract says is that it's mostly to gather knowledge. The primary purpose... I wouldn't delusion the fact that while I thought it was going to help me that, that was not the primary purpose and I think they made that clear that there was no guarantee that, that this would help me.
10	3	<i>Why join?</i> (Both self and societal benefit)	...I'm always wanting to help, if it's something that's gonna help me and others, I was up for trying something that's going to improve my lifespan... I don't like taking the meds, and if this was another way of being able to help resist and fight the [disease name], I was up for it.
	4	<i>Expect benefit?</i> (Qualified yes)	I would say yes, that I was in hopes that it'll improve my condition...I always have hopes in any new meds that they come out, um, I always keep a open mind with things in the hopes that whatever I try is going to improve my life, versus, you know, make it worse... the way [SC name] explained the study program, it just sounded very promising....
	3	<i>Intent of study?</i> (Helping patients and research)	I was looking at it as to help me as a patient, plus for them to be able to gain knowledge... I'm always up for anything that's gonna be able to improve my health, if it's gonna be able to gain knowledge for others down, down the road... if this is gonna work, I want to be able to be a part of that.
15	5	<i>Why join?</i> (Self)	Well [SC name] explained it to me, told me it would be the best thing I could get a hold of at that time, and I accepted it. Cause [SC name] is a nice lady... She just explained that it was good and I just accepted it.... I wanted to be cured. And she thought it would do it, and that's the reason I done it.
	5	<i>Expect benefit?</i> (Unqualified yes)	I thought it would cure me...
	5	<i>Intent of study?</i> (Helping patients)	Well I felt like it would help me, and I'm hoping the next [gene transfer intervention] will cause they say they have stronger genes this time.... I just trust 'em, they're just nice people and I just feel like whatever they can do, they try and do what's right for me.

subject-level variables and one for study-level variables, and dropped variables that were non-significant in each multivariate model. We chose to consider disease as a study-level variable because it was shared by all subjects in a study. In addition, the impact of disease at the individual level was measured separately by the variables

of perceived illness severity and options. The remaining variables from each level model were tested together in a final model.

Our regression analyses corrected for the lack of independence among the subjects within a study. We analyzed our data using SUDAAN, software developed

Table 4
Distributions of independent variables and TM index scores

Independent variable	Frequency (%)		Mean TM index scores
1. <i>Subject-level</i>	<i>N</i> = 68		<i>N</i> = 68
<i>Age in years</i>			
Mean, median (range)	59.8, 62.5 (20–83)		NA
<i>Gender</i>			
Male	60 (88)		11.00
Female	8 (12)		7.37
<i>Level of education</i>			
High school or less	15 (22)		11.67
College or some college	31 (46)		9.39
Graduate degree or some graduate school	22 (32)		11.50
<i>Severity of disease or condition</i>			
Severe or moderate to severe	20 (29)		10.5
Mild, mild-moderate, moderate, or varied/DK	48 (71)		10.6
<i>Options for treatment</i>			
Options are limited or non-existent	32 (47)		11.66
Limited options not mentioned by subject	36 (53)		9.61
<i>PI considered to be personal doctor</i>			
Before enrolling in study	9 (13)		9.00
PI is not MD/ became MD during study	59 (87)		10.81
<i>Subject perception of PI</i>			
Mostly providing care	30 (45)		10.57
Both equally or DK	13 (19)		10.31
Mostly research	24 (36)		10.75
<i>N</i> = 67 because one subject did not know PI			
<i>Subject perception of SC</i>			
Mostly providing care	27 (42)		9.96
Both equally or DK	10 (15)		10.50
Mostly research	27 (42)		11.37
<i>N</i> = 64 because two subjects had no SC, 1 DK, 1 ref			
<i>Previous research involvement</i>			
Yes	18 (26)		9.11
No	50 (74)		11.10
<i>Media coverage of study</i>			
Yes	26 (38)		10.64
No	42 (62)		10.46
<i>Promotion of study by disease organizations</i>			
Yes	8 (12)		9.13
No	60 (88)		10.77
<i>How long subject took for decision-making</i>			
Decided same day	17 (25)		11.53
A few days to a week	17 (25)		10.65
Within a month	20 (29)		10.40
More than a month	14 (21)		9.57
2. <i>Study-level: study design (from CF)</i>	<i>N</i> = 19	<i>N</i> = 68	Mean TM index scores
<i>Phase of study</i>			
Phase I/II or II	10 (53)	46 (68)	9.91
Phase I	9 (47)	22 (32)	10.89

Table 4 (continued)

2. Study-level: study design (from CF)	N = 19	N = 68	Mean TM index scores
<i>Disease^a</i>			
Cancer	9 (47)	44 (65)	11.32
Vascular	3 (16)	10 (15)	12.50
Infectious or inherited	7 (37)	14 (21)	6.86
<i>Components of study</i>			
Study includes standard treatment or investigational intervention other than GT	8 (42)	41 (60)	10.93
Gene transfer alone	11 (58)	27 (40)	10.04
<i>Components of study</i>			
Study includes standard treatment	4 (21)	25 (37)	10.92
Gene transfer alone or investigational intervention other than gene transfer	15 (79)	43 (63)	10.37
<hr/>			
3. Study-level: Consent form language	N = 19	N = 68	Mean TM index scores
<i>Benefit to society</i>			
Consent form mentions benefit to society	16 (84)	48 (71)	9.90
Consent form does not	3 (16)	20 (29)	12.20
<i>Likelihood of benefit in consent form</i>			
Consent form states benefit is unlikely	5 (26)	7 (10)	4.14
Consent form does not	14 (74)	61 (90)	11.31
<hr/>			
4. Study-level: PI characteristics	N = 19	N = 68	Mean TM index scores
<i>Likelihood of benefit in PI's discussion with subjects</i>			
PI tells subjects benefit is unlikely	5 (26)	8 (12)	5.25
PI does not	14 (74)	60 (88)	11.28
<hr/>			
5. Study-level: SC characteristics	N = 18	N = 66	Mean TM index scores
<i>Likelihood of benefit in SC's discussion with subjects</i>			
SC tells subjects benefit is unlikely	5 (28)	9 (14)	7.67
SC does not	13 (72)	57 (86)	11.11
<hr/>			
6. Study-level: PI, SC, and consent form	N = 19	N = 68	Mean TM index scores
<i>Likelihood of benefit: overall message</i>			
All state clearly that benefit is unlikely	3 (16)	5 (7)	4.60
Some state clearly that benefit is unlikely	5 (26)	8 (12)	8.50
None state clearly that benefit is unlikely	11 (58)	55 (81)	11.42

^a‘Cancer’ included prostate, hematologic, and gynecologic malignancies, and other; ‘vascular’ included coronary and peripheral artery disease; ‘infectious disease’ included HIV; ‘inherited disease’ included hemophilia, cystic fibrosis and others.

specifically for data from complex survey designs.⁸ The number of subjects in each study ranged from 1 to 19 (Table 1). We tested all bivariate relationships for two samples of subjects: the full sample of 68 subjects, and a

subset of 49 subjects that excluded the 19 subjects from the largest cluster. The bivariate findings were similar for both sets and we therefore used the full set of 68 in all subsequent multivariate models. We tested for multicollinearity in all multivariate models. We found that the descriptions of likelihood of benefit for subjects by PI, SC, and consent forms were multicollinear, so we combined these into one composite variable as described above. In addition, because we found that age and

⁸All bivariate and multivariate regression was conducted using stratification with replacement. For clusters with only one subject, the square of the Taylorized deviation was used as the contribution to the variance.

disease were collinear, we included disease and excluded age in our final multivariate model because we understood age to be causally prior to disease.

Results

GTR subjects

Most of the 68 subjects were male (88%) and participated in cancer trials (Table 4). Their average age was 60 (range 20–83). Fifteen (22%) had a high school education or less; most had at least some college and 22 (32%) had completed some graduate work. Age and education were related to disease type: older subjects tended to cluster in cancer and vascular disease trials, and younger subjects in inherited and infectious disease trials. Most of those with some graduate education were in cancer trials; half of those who reported high school education or less were in vascular disease trials. Asked to rate the severity of their disease or condition on a 5-point scale from mild to severe, 20 (29%) considered it to be severe or moderate-to-severe. A qualitative assessment of full transcripts found that 32 subjects (47%) felt their treatment options were limited or non-existent.

In response to questions about gene transfer, many subjects mentioned the promise of the new genetic technology. As one subject said, “I just figured it was something new and... you [always] saw someone in the newspaper talking about gene therapy.” A frequent theme was the logic of the genetic fix. Some who had been in other types of research thought GTR sounded more promising and exciting: “I thought this was actually going to be the breakthrough, I really did.”

Bivariate regression

Five subject level variables (Table 5) were related to greater TM in bivariate regression analyses: older age, male gender, less education (high school compared to college), limited or non-existent treatment options, and having no previous research experience. The following did not affect TM in subjects: severity of disease, whether the PI was the subject’s personal physician prior to enrollment in the study, subjects’ perception of the PI and the SC as mostly providing care compared to mostly conducting research, subjects’ perception that their study was covered by the media or promoted by disease organizations, and taking a short time to decide to join the study. Surprisingly, two of these non-significant variables were significant in the opposite direction from that hypothesized: subjects with graduate education had higher TM scores than those with college education, and those who considered the PI to be their

physician prior to the study had lower TM scores than those who did not.

Three study level variables (Table 5) were significantly related to TM. Subjects with cancer and vascular disease had higher TM scores than subjects with infectious or inherited disease. Subjects whose consent forms mentioned benefit to society as a goal of the research had lower TM scores, as did subjects whose PI, SC, and consent form stated clearly that direct benefit was unlikely. Study phase, the presence of standard treatment, or the presence of standard treatment and other investigational treatment did not have an effect on subjects’ level of TM.

Multivariate regression: subject level, study level, and full model

Analysis of the multivariate subject-level model showed that two of the five subject level variables related to TM remained significant: age and education. (Table 5) Thus, when age and education are controlled for, gender, perception of one’s treatment options, and previous research involvement are not related to TM for this sample of subjects.

In the multivariate study level model, two of the three variables tested remained significant: disease and the combined message about the likelihood of benefit. (Table 5) When these were controlled for, it did not appear to matter whether the consent form mentioned societal benefit.

The combined multivariate model examined the four remaining variables: age, education, disease, and combined message about likelihood (Table 5, last column). As noted above, age was removed from the model because age and disease were collinear. Thus the final model has three variables, each of which is independently associated with TM: education, disease, and combined message about likelihood. Together, these variables explain 53% of the variation in TM scores ($R^2 = 0.53$ at $p < 0.001$). Those with a high school education or less scored 1.83 units higher on the TM index compared to those with a college education. Those with graduate education scored 1.04 units higher than those with a college education, though we had hypothesized the reverse. People in cancer trials scored almost 3 units higher on the TM scale, as did people in vascular disease trials, compared to those in trials for infectious and inherited disease. Finally, those subjects who received a consistent message from the PI, SC and consent form that they were unlikely to benefit from the gene transfer intervention scored 4 units lower on average than those who did not receive any such messages. An inconsistent message, in which one or two (but not all three) sources of information say benefit is unlikely, did not lead to TM scores that were different

Table 5
Regression of therapeutic misconception on subject-and study-level variables^a

Independent variables [Reference category]	Bivariate models	Multivariate models	
		Subject level model	Combined model
<i>(1) Subject-level variables</i>			
<i>Age in years:</i>			
Mean, median (range)	0.11***	0.10**	[Removed]
<i>Gender:</i>			
Male [Female]	3.6**	0.66	—
<i>Level of education</i>			
High school or less	2.28** *	2.07*** **	1.83** *
Graduate or some grad [College or some college]	2.11†	1.45†	1.04†
<i>Severity of disease or condition</i>			
Severe or mod-to-severe [mild, mild-moderate, moderate, or varied/DK]	−0.10	—	—
<i>Options for treatment (coded qualitatively)</i>			
Options limited nonexistent [limited options not mentioned by subject]	2.05***	0.77	—
<i>PI considered personal MD</i>			
Before enrolling in study [PI not MD/became MD in study]	−1.81†	—	—
<i>Subject perception of PI</i>			
Mostly providing care	−0.18	—	—
Both equally or DK [Mostly research]	−0.44	—	—
<i>N = 67 one subject did not know PI</i>			
<i>Subject perception of SC</i>			
Mostly providing care	−1.41	—	—
Both equally or DK [Mostly research]	−0.87	—	—
<i>N = 64 two had no SC, 1 DK, 1 ref</i>			
<i>Previous research involvement</i>			
Yes [No]	−1.99*	0.24	—
<i>Media coverage of study</i>			
Yes [No]	−0.18	—	—
<i>Promotion of study by disease organizations</i>			
Yes [No]	−1.64	—	—
<i>How long subject took for decision-making</i>			
Decided same day	1.96*	—	—
A few days to a week	1.08	—	—
Within a month [More than a month]	0.83	—	—
Intercept		2.49*	
R ²		0.48***	

Table 5 (continued)

Independent variables [Reference category]	Bivariate models	Multivariate models	
		Subject level model	Combined model
<i>(2). Study-level variables</i>			
<i>Disease</i>			
Cancer	4.46*** ***	2.57*** **	2.89*** ***
Vascular [infectious or inherited]	5.64***	3.43***	2.79***
<i>Phase of study</i>			
Phase I/II or II [Phase I]	0.98	—	—
<i>Components of study</i>			
Study includes std tx or investigational intervention other than GT [Gene transfer alone]	0.89	—	—
<i>Components of study</i>			
Study includes std tx [Gene transfer alone + /or investigational intervention other than GT]	0.55	—	—
<i>Benefit to society</i>			
Consent form mentions [Consent form does not mention]	-2.3***	-0.70	—
PI, SC, and Consent form			
<i>Likelihood of benefit:</i>			
<i>Overall message</i>			
All state clearly that benefit is unlikely	-6.82*** ***	-4.12*** **	-4.01*** **
Some state clearly benefit is unlikely [None state clearly benefit is unlikely]	-2.92**	-1.83	-1.6
Intercept		9.42***	8.03***
R^2		0.50***	0.53***

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

†N.S. in the hypothesized direction, but significant in the opposite direction at the 0.05 level or less. All tests are one-tailed, except for sex and joint tests of dummy variables.

The vertical lines represent simultaneous tests of dummy variable groups.

^aCluster sampling effects controlled through stratification with replacement using SUDAAN. For studies with only one subject, SUDAAN used the square of the Taylorized deviation as the contribution to the variance.

from those studies in which no source stated that benefit was unlikely.

Discussion

A majority of subjects in our study (74%) scored 10 or higher on our composite TM scale, out of a maximum of 15. This finding of high TM corroborates results from studies of subjects in other early phase clinical trials. Our analysis also demonstrates the importance of both subject- and study-level factors as predictors of TM. These results must be interpreted in light of our study's strengths and weaknesses. Its strengths lie in the use of qualitative and quantitative data, the ability to combine subject factors with information from the research team

and consent forms, and the assessment of their relative importance by means of multivariate analysis. Interviews were not limited to one institution or one disease type, potentially making findings more generalizable. On the other hand, our data come from retrospective interviews and, although the studies were representative of the larger population of ongoing GTR trials, our subject selection criteria excluded subjects too sick to consider an hour-long interview. Finally, although combining our small sample size with a moderate significance level for our hypothesis testing means that those factors we found to be statistically significant are likely to be meaningful, we cannot say with certainty that statistically non-significant factors are not important. Our findings need to be tested using a larger sample of subjects.

Our study sample allowed us to compare TM in different disease populations. TM has been documented extensively in subjects in early phase cancer trials, but is less studied in other groups. We hypothesized that subjects with an inherited disease might have high expectations for direct benefit from a genetic “fix,” although this had not been studied before. In fact, our results suggest that subjects in GTR trials for inherited and infectious disease tend to have low TM, while those in trials for cancer and vascular disease have higher TM. In multivariate analysis, these disease-related differences in TM are not explained by subjects’ education, previous research experience, disease severity, or perceived options outside trial participation.

Disease is a complex, multifaceted variable that encompasses not only individual- and study-level factors, but also the socially and culturally determined meanings associated with particular diseases. The category of disease may represent different perspectives on clinical research by patient or clinical trial activists or advocacy organizations (Dresser, 2001). It may tap subjects’ impressions of technical aspects of the intervention they are receiving (e.g., cancer gene “vaccines” may sound more like therapy than does the insertion of a corrective gene for inherited disease). It is also possible that people with inherited disease, who have lived with a lifetime diagnosis, expect benefit in a more long-sighted fashion than do those with cancer. Clinical researchers conducting GTR trials in different disease populations (Henderson et al., 2004) may similarly be affected in a range of ways by the culture of clinical trials for different diseases. Our data do not allow us to explain why disease is such an important determinant of TM, but our results merit further study.

In our multivariate analysis, some subject characteristics remain important, even when controlling for study-level factors. Age is a strong predictor of TM although, because age and disease are highly related in this group of subjects, it is not possible to distinguish their effects on TM. We found education to be a significant predictor of TM, confirming what has been shown in most other studies. Yet the relationship between TM and education is not linear for subjects in our sample: college education reduces TM, but graduate education does not. This may be due in part to the preponderance of cancer trial participants among those with graduate education in our sample, or it may reflect a tendency among those with advanced degrees to investigate and be persuaded by the sophisticated science of GTR. More research is needed to know if this is a meaningful finding, but it suggests a more complex relationship between TM and education than was previously recognized.

We were surprised by the lack of significant effects for factors relating to care-giving relationships, though findings of non-significance for our small sample must

be interpreted with caution. First, we found that subjects recruited by their physicians were not more likely to exhibit TM than others, contrary to the prevailing assumption that combining provider and recruitment roles will increase TM. In our sample, subjects who reported that their PIs were their physicians prior to the study tended to have inherited disease and HIV—groups with lower levels of TM. We also found that many subjects adopt PIs as their doctors during the study, and that this is related to higher levels of TM. We do not know if subjects’ TM causes them to see PIs as their doctors, or if the PIs’ becoming their doctors causes higher TM. Either way, these findings speak to the need for ongoing and transparent communication, in which PIs who are also providers or who become providers during the course of a study discuss the transition to a dual-role relationship.

Second, we found that TM was not related to subjects’ descriptions of their PI and/or SC as providing ‘mostly care’ or doing ‘mostly research’ in the study. This suggests that subjects can identify a care-giving role in research without viewing treatment as the purpose of research and without overestimating the likelihood of direct benefit. This finding underscores the complexity of clinical research, which incorporates features of both clinical and research relationships (Grunberg & Cefalu, 2003; Miller & Rosenstein, 2003; Richardson & Belsky, 2004).

Our findings suggest that TM scores are significantly lower when study teams and consent forms state clearly that benefit is unlikely. Furthermore, the three sources of those messages—consent forms and discussion by PIs and SCs—were usually in agreement. We do not know why; perhaps this agreement reflects institutional or research team norms or a general reluctance to describe benefit as unlikely in clinical research. Nevertheless, our findings point to one strategy for reducing TM in early phase GTR: clearly communicating to subjects that direct benefit is unlikely.

Finally, despite the overall high TM scores in this sample, our results reveal a wide range of expectations by subjects in early phase trials, and a level of complexity of TM not previously demonstrated. On the low end of the index scale, people with TM scores of 3 or 4 are clear about the purpose of research and realistic about what to expect. On the high end, people with scores of 14 or 15 are not clear about research intent and not realistic about what to expect. Scores in the middle resist simple categorization; some subjects score high on one component but low on another (Table 3), although the results of our multivariate model appear to be robust to the inclusion of subjects whose TM scores had components at opposite extremes. Some people in the middle categories understand that they are in a research study but still seem to have unrealistic expectations that the intervention will help them, while a

few are realistic about the likelihood of direct benefit but confused about the goals of research.

That this particular index of TM is hard to interpret in the middle range should not be surprising. This difficulty reflects a lack of clarity about TM in the literature and in clinical research in general. Most agree that it is important to distinguish research from treatment. There is considerable difference of opinion, however, about what benefit is reasonable to expect from participation in early phase clinical research. This lack of consensus is exhibited in current publications (e.g., Agrawal & Emanuel, 2003); in ambiguous communications with subjects documented in this (King et al., 2005) and other studies (e.g., Dresser, 2002); in variation in investigators' personal expectations for subjects in their studies (Henderson et al., 2004); and in the heated debate regarding how to be realistic in describing a potential benefit without demoralizing patient-subjects who see such trials as their "last hope" (e.g., Miller, 2000). The TM index presented here is a step toward constructing a measure of how people interpret the meaning of early phase clinical research. It still may not fully capture what is clearly a complex, multi-dimensional concept, but no measure of TM will be adequate until we agree about exactly what should be measured.

Acknowledgments

This work was supported by a grant from the Program on Ethical, Legal, and Social Implications of the Human Genome Project, National Human Genome Research Institute (NIH 1R01HG02087). We very much appreciate the methodological and statistical assistance provided in the early stages of this paper by Erin E. Leahey, Assistant Professor of Sociology, University of Arizona, and very helpful comments by two anonymous reviewers.

References

- ACHRE (1995). Advisory committee on human radiation experiments. *Final report of the advisory committee on human radiation experiments*. Washington, DC: US Government Publishing Office.
- Ackerman, T. F. (1995). The ethics of phase I pediatric oncology trials. *IRB: Ethics and Human Research*, 17(1), 1–5.
- Agrawal, M., & Emanuel, E. J. (2003). Ethics of phase I oncology studies: Reexamining the arguments and data. *Journal of the American Medical Association*, 290(8), 1075–1082.
- Appelbaum, P. S., Lidz, C. W., & Grisso, T. (2004). Therapeutic misconception in clinical research: Frequency and risk factors. *IRB: Ethics and Human Research*, 26(2), 1–8.
- Appelbaum, P. S., Roth, L. H., & Lidz, C. (1982). The therapeutic misconception: Informed consent in psychiatric research. *International Journal of Law and Psychiatry*, 5(3–4), 319–329.
- Cheng, J. D., Hitt, J., Koczwara, B., Schulman, K. A., Burnett, C. B., Gaskin, D. J., Rowland, J. H., & Meropol, N. J. (2000). Impact of quality of life on patient expectations regarding phase I clinical trials. *Journal of Clinical Oncology*, 18(2), 421–428.
- Churchill, L. R., Collins, M. L., King, N. M. P., Pemberton, S. G., & Wailoo, K. A. (1998). Genetic research as therapy: Implications of "gene therapy" for informed consent. *The Journal of Law, Medicine, & Ethics*, 26(1), 38–47.
- Daugherty, C. K., Banik, D. M., Janish, L., & Ratain, M. J. (2000). Quantitative analysis of ethical issues in phase I trials: A survey interview of 144 advanced cancer patients. *IRB: Ethics and Human Research*, 22(3), 6–14.
- Daugherty, C., Ratain, M. J., Grochowski, E., Stocking, C., Kodish, E., Mick, R., & Siegler, M. (1995). Perceptions of cancer patients and their physicians involved in phase I trials. *Journal of Clinical Oncology*, 13(5), 1062–1072.
- Dresser, R. (2001). *When science offers salvation: Patient advocacy and research ethics*. Oxford: Oxford University Press.
- Dresser, R. (2002). The ubiquity and utility of the therapeutic misconception. *Social Philosophy & Policy*, 19(2), 271–294.
- Friedmann, T. (1996). Human gene therapy—an immature genie, but certainly out of the bottle. *Nature Medicine*, 2(2), 144–147.
- Grunberg, S. M., & Cefalu, W. T. (2003). The integral role of clinical research in clinical care. *New England Journal of Medicine*, 348(14), 1386–1388.
- Henderson, G. E., Davis, A. M., King, N. M., Easter, M. M., Zimmer, C. R., Rothschild, B. B., Wilfond, B. S., Nelson, D. K., & Churchill, L. R. (2004). Uncertain benefit: investigators' views and communications in early phase gene transfer trials. *Molecular Therapy*, 10(2), 225–231.
- Hornig, S., & Grady, C. (2003). Misunderstanding in clinical research: Distinguishing therapeutic misconception, therapeutic misestimation, and therapeutic optimism. *IRB: Ethics and Human Research*, 25(1), 11–16.
- Horstmann, E., McCabe, M. S., Grochow, L., Yamamoto, S., Rubinstein, L., Budd, T., Shoemaker, D., Emanuel, E. J., & Grady, C. (2005). Risks and benefits of phase I oncology trials, 1991 through 2002. *New England Journal of Medicine*, 352(9), 895–904.
- Joffe, S., Cook, E. F., Cleary, P. D., Clark, J. W., & Weeks, J. C. (2001). Quality of informed consent in cancer clinical trials: A cross-sectional survey. *Lancet*, 358(9295), 1772–1777.
- King, N. M. P. (1999). Rewriting the "points to consider": The ethical impact of guidance document language. *Human Gene Therapy*, 10(1), 133–139.
- King, N. M. P., Henderson, G. E., Churchill, L. R., Davis, A. M., Hull, S. C., Nelson, D. K., Parham-Vetter, P. C., Rothschild, B. B., Easter, M. M., & Wilfond, B. S. (2005). Consent forms and the therapeutic misconception: The example of gene transfer research. *IRB: Ethics and Human Research*, 27(1).

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miller, M. (2000). Phase I cancer trials. A collusion of misunderstanding. *Hastings Center Report*, 30(4), 34–43.
- Miller, F. G., & Rosenstein, D. L. (2003). The therapeutic orientation to clinical trials. *New England Journal of Medicine*, 348(14), 1383–1386.
- Richardson, H., & Belsky, L. (2004). The ancillary-care responsibilities of medical researchers. *Hastings Center Report*, 25–33.
- Schaeffer, M. H., Krantz, D. S., Wichman, A., Masur, H., Reed, E., & Vinicky, J. K. (1996). The impact of disease severity on the informed consent process in clinical research. *The American Journal of Medicine*, 100(3), 261–268.
- Temple, R. J. (1995). A regulatory authority's opinion about surrogate endpoints. In W. S. Nimmo, & G. T. Tucker (Eds.), *Clinical measurement in drug evaluation* (pp. 3–22). New York: Wiley.